Data Management and Analyses in HIV-1 Therapeutic Studies

Sundhiya Mandalia

March 2014

Imperial College London
Department of Medicine

Chelsea and Westminster Hospital
369 Fulham Road
London SW10 9NH

A Thesis Presented For The Degree of Doctor of Philosophy

* *

*

To the memory of my dear parents

DECLARATION OF ORIGINALITY

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

Sundhiya Mandalia

ABSTRACT

The focus of this PhD is to demonstrate how routinely collected de-identified patient information can be used for basic science, clinical science and population science health research. Examples are provided on how routinely collected information from people living with HIV-1 (PLHIV) can provide strategic health information, enabling stakeholders to develop and implement strategies to improve the use, cost, outcome and impact of HIV-1 service provision. As more PLHIV are on combination anti-retroviral therapy (cART), the increasing number of PLHIV requiring long-term use of health services has long-term resource implications. De-identified personal information that are systematically collected and analysed can provide information to monitor and evaluate the use, cost, outcome and impact of HIV-1 service provision.

The basic science study identified PLHIV who were long-term non-progressors and HIV controllers managed in a HIV referral centre and determined factors that contributed to their atypical prolonged asymptomatic infection. The clinical science study assessed the impact of HIV-1 infection, cART and age on renal insufficiency in PLHIV managed at the referral centre. The population science study used information from an established national multicentre prospective monitoring system to report changing use and population cost of HIV-1 services in the UK, and estimated future population costs arising from projected increases in the number of PLHIV using NHS services.

Despite recognised problems associated with observational data, the utility of de-identified personal information for such research was demonstrated. Personal health information needs to be collected longitudinally to ensure development of good electronic medical records nationally and globally requiring appropriate information technology aswell as analytic expertise. Main lesson from this PhD is that de-identified individual level information can be used at local, regional and national levels to monitor and evaluate the effectiveness, efficiency, equity and acceptability of HIV-1 and other health services provided.

LIST OF TABLES

CHAPTER 1:
GENERAL INTRODUCTION

CHAPTER 2:
MATERIALS AND METHODS

CHAPTER 3:
LONG-TERM NON-PROGRESSOR COHORT ANALYSES: ESTIMATED TIME UNTIL DISEASE PROGRESSION

CHAPTER 4:
HIV AND AGEING ANALYSES: CO-EFFECTS OF HIV-1 INFECTION AND AGEING ON RESPONSE TO THERAPY, CO-MORBIDITY, PROGRESSION AND IMMUNE RE-CONSTITUTION ANALYSES

CHAPTER 5:
POPULATION COST FOR TREATING PEOPLE LIVING WITH HIV IN THE UK

CHAPTER 6:
GENERAL DISCUSSION

LIST OF FIGURES

CHAPTER 1:
GENERAL INTRODUCTION

CHAPTER 2:
MATERIALS AND METHODS

CHAPTER 3:
LONG-TERM NON-PROGRESSOR COHORT ANALYSES: ESTIMATED TIME UNTIL DISEASE PROGRESSION

CHAPTER 4:
HIV AND AGEING ANALYSES: CO-EFFECTS OF HIV-1 INFECTION AND AGEING ON RESPONSE TO THERAPY, CO-MORBIDITY, PROGRESSION AND IMMUNE RE-CONSTITUTION ANALYSES

CHAPTER 5:
POPULATION COST FOR TREATING PEOPLE LIVING WITH HIV IN THE UK

CHAPTER 6:
GENERAL DISCUSSION

APPENDIX

ACKNOWLEDGEMENTS

Many people have contributed in important ways to the work in this thesis and supported me during its development. First and foremost, I would like to thank my three supervisors, Nesrina Imami, Eddy Beck and Brian Gazzard whose patience and enthusiasm drove me on and whose compassion for those who stand to benefit from research findings is inspiring.

Many colleagues, from HIV/GUM, have very generously helped and supported me directly and indirectly during my PhD work and they include Ann Sullivan, Anton Pozniak, David Asboe, David Hawkins, Mark Nelson, Simon Barton and many others.

At Imperial College London, Department of Immunology, many colleagues provided valuable moral support and included Frances Gotch, Peter Kelleher, Anna Herasimtschuk, Nathali Grageda and many others.

A special thanks to Roshni Sangha and Gary Lo for their help at various stages of data collection processes of NPMS-HHC and help with a number of graphs presented in this thesis and Ali Hamidi for help with his desktop publishing skills in fine tuning some of the figures presented.

Throughout the duration of my PhD work, this work was supported by St Stephen's AIDS Trust and at various stages by NPMS-HHC, without whose support, completion of this PhD would not have been possible.

Finally, perhaps my greatest debt of gratitude is owed to my sister, brothers and my close extended family support. They all supported me, kept me cheered and ensured my spirits remained high through some difficult times which were the backdrop to my doctoral studies.

ABBREVIATIONS AND ACRONYMS

Other Symbols	
K	This is a pseudo-metric abbreviation where K stands for 'kilo' however in this thesis it is used as an abbreviation for 1000
Acronyms	
AIDS	Acquired immunodeficiency deficiency syndrome
ANOVA	Analysis of variance
ART	Anti-retroviral therapy
ARV	Anti-retroviral
ASCII	American standard code for information interchange
ASx	Asymptomatic
BHIVA	British HIV Association
BLD	Below the limit of detection of the plasma viral load assay
BT	British Telecom
CAC	Co-ordinating and analytic centre
CCG	Clinical commissioning group
CD3	Cluster of differentiation 3
CD4	Cluster of differentiation 4
CD8	Cluster of differentiation 8
CDC	Centre for disease control
CEBR	The centre for economics and business research ltd
CKD	Chronic kidney disease
CKD-EPI	Chronic kidney disease epidemiology collaboration
CSC	Computer sciences corporation
CSV	Comma-separated values
DNA	Deoxyribonucleic acid
DW	Dayward
eGFR	Estimated glomerular filtration rate
FDA	The food and drug administration
GALT	Gut associated lymphoid tissues
GISHEAL	Genetic and immunological studies of European and African HIV-1^{+} long-term non-progressor study
GRID	Gay related immune deficiency
GUS	Genomics unified schema
HAART	Highly active ART
HIC	HIV controllers
HIS	Hospital information system
HIV	Human immunodeficiency virus
HIV-1	Human immunodeficiency virus type 1
HIV-2	Human immunodeficiency virus type 2
HIVAN	HIV-associated nephropathy
HPA	Health protection agency
HTLV-III	Human T-lymphotrophic virus III
IBM	International business machines corporation
ICH	International conference on harmonisation

ABBREVIATIONS AND ACRONYMS

ICTV	International committee on taxonomy of viruses
IDC	International data corporation
IDU	Injecting drug user
IP	Inpatient
IT	Information technology
IQR	Inter-quartile range
IRB	Institutional review board
KDOQI	Kidney disease outcome quality initiative
LAV	Lymphadenopathy associated virus
LTFU	Lost to follow-up
LTNP	Long-term non-progressors
MDRD	Modification of diet in renal disease
MPPY	Mean per patient year
MRI	Medical record institute
MSM	Men who have sex with men
NHS	National health service
NIGB	National information governance board
NIGC	National information governance committee
NK	Natural killer
NNRTI	non-nucleoside reverse transcriptase inhibitor
NPfIT	National programme for IT in the NHS
NPMS-HHC	National prospective monitoring system - HIV health-economics collaboration
NRES	National research ethics service
NRTI	Nucleoside reverse transcriptase inhibitor
OI	Opportunistic infection
OP	Outpatient
PCT	Primary care trust
PGL	Persistent generalized lymphodenopathy
PI	Protease inhibitor
PIAG	Patient information advisory group
PLHIV	People living with HIV
PPY	Per patient year
RAM	Random access memory
RECORD	Reporting of studies conducted using observational routinely collected data
RNA	Ribonucleic acid
SAM	Sequential access media
SAS	Statistical analysis software
SIDA	Syndrome d'immunodéficience acquise
SOPHID	The survey of prevalent HIV infections diagnosed
START	Strategic timing of antiretroviral treatment
STROBE	Strengthening the reporting of observational studies in epidemiology
Sx non-AIDS	Symptomatic non-AIDS
T-cell	T-lymphocytes
UK	United Kingdom
UN	United Nations
US	United States
VAT	Value added tax

VL	Viral load
WHO	World Health Organisation
WBC	White blood cell

Publications Arising From This Thesis

PUBLICATIONS ARISING FROM THIS THESIS

My contributions to the work presented and published are listed in respective journal articles referenced below and comprise important intellectual content of the work including conception, design, analysis and writing.

PAPERS

Mandalia S, Mandalia R, Lo G, Chadborn T, Sharott P, Youle M, Anderson J, Baily G, Brettle R, Fisher M, Gompels M, Kinghorn G, Johnson M, McCarron B, Pozniak A, Tang A, Walsh J, White D, Williams I, Gazzard B, Beck EJ, for the NPMS-HHC Steering Group (2010).
Rising Population Cost for Treating People Living with HIV in the UK, 1997-2013.
PLOS one 5(12): e15677.

Mandalia S, Westrop SJ, Beck EJ, Nelson M, Gazzard BG, Imami N.
Are long-term non-progressors very slow progressors? Insights from the Chelsea and Westminster HIV Cohort, 1988 – 2010.
PLOS one 7(2): e29844 (2012).

Mandalia S, Herasimtschuk A, Farrington K, Westrop SJ, Nelson M, Gazzard BG, Imami N, Beck EJ
Effect of HIV-1 infection, first-line ART and Ageing on renal function
(Submitted for publication)

MEDIA COVERAGE OF PUBLISHED PAPERS

1) Rising Population Cost for Treating People Living with HIV in the UK, 1997-2013.
PLOS one 5(12): e15677. doi:10.1371/journal.pone.0015677
i) http://www.aidsmap.com/Annual-UK-HIV-treatment-and-care-costs-could-reach-750-million-by-2013/page/1618137/

ii) http://www.alere.co.uk/sexual-health/articles/rising-population-cost-for-treating-people-living-with-hiv-178/text.htm

iii) House Of Lords (2011) *No vaccine, no cure. HIV and AIDS in the United Kingdom. Report. House of Lords Select Committee on HIV and AIDS in the United Kingdom. Paper 188*. (July). http://www.publications.parliament.uk/pa/ld201012/ldselect/ldaids/188/18805.htm#a1

2) Are long-term non-progressors very slow progressors? Insights from the Chelsea and Westminster HIV Cohort, 1988 - 2010.
PLOS one 7(2): e29844 (2012).
i) http://www.aidsbeacon.com/news/2012/03/05/few-hiv-aids-positive-people-can-control-disease-progression-without-antiretrovirals/

ii) http://www.eatg.org/eatg/Global-HIV-News/Treatment/Are-HIV-non-progressors-really-very-slow-progressors

iii) http://www.hivandhepatitis.com/hiv-aids/hiv-aids-topics/hiv-disease-progression/3471-are-hiv-non-progressors-really-very-slow-progressors

COMMENT RECEIVED

Loreen Willenberg said:
Have Mandalia et al contributed to the vast body of knowledge of LTNPs/HIV Controllers, or have they merely found that LTNPs/HCs within their cohort “inevitably progressed” to AIDS? Indeed, in this paper, the investigators admit that “due to the small numbers of individuals in the LTNP and long-term stable CD4+ T-cell count groups, power to detect a statistical significance (of biological markers of progression) was low.” (Mandalia, p.7). They also state, “Furthermore, the infecting clade of HIV-1 and the geographic origin of infection are likely to impact the course of HIV-1 disease progression…Of the 13 patients identified as LTNPs, clade information was available for 5. Of these, 3 were infected with clade B, and 2 with clade C virus, while the infecting viral clade for

the remaining 8 LTNP has not yet been tested." (Mandalia, p. 7). The test results for all LTNP subjects would have been of interest in this case. In the many years I have been following (and participating in)LTNP/HC clinical studies, this is the first time an investigating team has, to my knowledge, managed to utilize the rarity of LTNPs/HCs within a small class of study subjects to serve as evidence that "ultimate progression" to AIDS is inevitable in the entire general population of LTNPs and HIV Controllers.

PUBLISHED ABSTRACTS

Mandalia S, Westrop S, Beck E, Nelson M, Gazzard B, Imami N.
Frequency and characteristics of long-term nonprogressors and HIV controllers in the Chelsea and Westminster HIV Cohort.
HIV Med 12 Supplement 1, 10-11 (2011).

Westrop SJ, **Mandalia S**, Moyle G, Bower M, Nelson M, Imami N.
Immunological manifestations of increasing age, ART duration and time since diagnosis within the ageing HIV-1+ cohort.
HIV Med 13 Supplement 1, 5-6 (2012).

Mandalia S, Westrop SJ, Beck EJ, Nelson M, Farrington K, Imami N, Gazzard B.
The Impact of HIV-1 infection, combination antiretroviral therapy and ageing on renal function.
HIV Med 13 Supplement 1, 48-49 (2012).

PRESENTATIONS

Mandalia S, Mandalia R, Lo G, Chadborn T, Sharott P, Youle M, Anderson J, Baily G, Brettle R, Fisher M, Gompels M, Kinghorn G, Johnson M, McCarron B, Pozniak A, Tang A, Walsh J, Williams I, Gazzard B and Beck EJ for the NPMS-HHC Steering Group
Rising Population Cost for Treating HIV Patients in the NHS 1997-2006.
Second Joint Conference of the BHIVA with BASHH conference, Manchester, UK, 20-23 Apr 2010.
(One of 5 Highly Commended Poster Presentation by BHIVA/BAASH judges which was selected for an oral presentation at the conference)

Mandalia S, Westrop S, Beck E, Gazzard B, Imami N.
The number of long-term nonprogressors and HIV controllers from 20 years of follow up of the Chelsea and Westminster HIV Cohort.
HIV Symposium, Barts & the London School of Medicine, London June 2010.

Mandalia S, Mandalia R, Lo G, Chadborn T, Sharott P, Youle M, Anderson J, Baily G, Brettle R, Fisher M, Gompels M, Kinghorn G, Johnson M, McCarron B, Pozniak A, Tang A, Walsh J, Williams I, Gazzard B and Beck EJ for the NPMS-HHC Steering Group
Past and Future Population Costs for Treating People living with HIV in the UK, 1997-2013.
XVIII International AIDS conference, Vienna, Austria, 18-23 July 2010.

Mandalia S, Westrop S, Beck E, Gazzard B, Imami N.
Enumeration and identification of current long-term non-progressors and HIV controllers from 20 years of clinical data of the Chelsea and Westminster HIV cohort. XVIII International AIDS. Conference, Vienna, 18-23 July 2010.

Mandalia S, Westrop S, Beck E, Gazzard B, Imami N.
Chelsea and Westminster HIV Cohort Study of 20 Years to Identify Long-Term Nonprogressors and HIV Controllers.
18th Conference on Retroviruses and Opportunistic Infections (CROI 2011), Boston February 27-March 2 2011.

Mandalia S, Westrop S, Beck E, Nelson M, Gazzard B, Imami N.
Frequency and characteristics of long-term nonprogressors and HIV controllers in the Chelsea and Westminster HIV Cohort.
17th Annual Conference of the British HIV Association (BHIVA). Bournemouth April 2011.

Mandalia S, Westrop SJ, Beck EJ, Nelson M, Imami N, Gazzard B.
Co-effect of HIV-1 infection and ageing on renal function.
2nd International Workshop on HIV and Aging, Baltimore USA October 2011.

Westrop SJ, **Mandalia S**, Moyle G, Bower M, Nelson M, Imami N.
Distorted virus-specific T-cell function and phenotype in the ageing HIV-1+ cohort.
2nd International Workshop on HIV and Aging, Baltimore USA October 2011.

Westrop, S, **Mandalia S**, Moyle G, Bower M, Nelson M, Imami N.
Compounded Influence of Patient Age, ART Duration and Time since HIV-1+ Diagnosis Delineates the Immune Profile of the Ageing Cohort.
19th Conference on Retroviruses and Opportunistic Infections (CROI 2012), Seattle March 2012.

Westrop SJ, **Mandalia S**, Moyle G, Bower M, Nelson M, Imami N.
Immunological manifestations of increasing age, ART duration and time since diagnosis within the ageing HIV-1+ cohort.
18th Annual Conference of the British HIV Association (BHIVA). Birmingham April 2012.

Mandalia S, Westrop SJ, Beck EJ, Nelson M, Farrington K, Imami N, Gazzard B.
The Impact of HIV-1 infection, combination antiretroviral therapy and ageing on renal function.
18th Annual Conference of the British HIV Association (BHIVA). Birmingham April 2012.

Mandalia S, Westrop SJ, Beck EJ, Nelson M, Farrington K, Imami N, Gazzard B.
Association between HIV-1 infection, combination antiretroviral therapy and ageing on renal function.
HIV Symposium, Barts & the London School of Medicine, London June 2012.

Mandalia S, Westrop SJ, Nelson M, Imami N, Farrington K, Gazzard B, Beck EJ.
The synergistic effects of HIV-1 infection, ART and age on renal function.
XIX International AIDS Conference, Washington DC July 2012.

Mandalia S, Westrop SJ, Farrington K, Imami N, Nelson M, Gazzard B, Beck EJ.
The Effect of HIV-1 infection, ART and Ageing on Renal Function in PLHIV. Chelsea and Westminster Hospital London, UK 1996-2011
IAS 2013, Kuala Lumpur, July 2013.

Chapter 1: General Introduction

CHAPTER 1: GENERAL INTRODUCTION

This PhD demonstrates how routinely collected de-identified patient data can be used for research studies. Three examples of: basic science, clinical science, and population science studies are presented. Furthermore, this PhD provides examples on how routinely collected de-identified information from people living with HIV-1 (PLHIV) can deliver strategic health information. This would enable stakeholders to develop and implement strategies so that they can successfully cope with an increasing HIV-1^{+} population in order to improve the use, cost, outcome and impact of HIV-1 service provision.

Introduction and success of highly active antiretroviral therapy has increased the number of people requiring long-term use of health services. This has financial implications including ongoing development of new therapeutic agents. As more PLHIV are prescribed combination antiretroviral therapy (cART), the increasing number of PLHIV requiring long-term use of health services has resource implications. Systematically collected de-identified personal information, and analysed data, can shed further light on clinical issues by providing information to monitor and evaluate the use, cost, outcome and impact of HIV-1 service provision. An established research cycle was used as a framework for the three examples of research studies undertaken.

The aim of the United Kingdom (UK) National Health Service (NHS) is to provide health care in order to improve treatment of chronic diseases such as heart disease, diabetes, HIV, and chronic kidney disease (CKD). People with chronic diseases are often in need of life-long treatment and care, which results in high costs for treatment and care, as these people need to be routinely monitored for disease severity and have their disease controlled by treatments that are efficacious. Health care systems have evolved and the two main areas of providing the healthcare are knowledge and process. Knowledge guides the care that should be delivered, and process aims to deliver it in a cost effective way.

We live in the information age[1] within which an increasing amount of data are generated and stored. Although there is no definition of information age, this is referred to the computer age where people have access to information from almost every part of the world. Organisations handle huge volumes of routinely stored data using terabytes[2] and exabytes[3]. With these volumes of data stored and collected, there is a requirement for an efficient management and access to such data volume. Databases are used to store a large amount of historical data in order to allow users at different levels to make effective decisions. Database and data management from which

Chapter 1:1 Information age: a period which began in the last quarter of the 20th century when information became easily accessible through publications and through the manipulation of information by computers and computer networks. Description from http://www.thefreedictionary.com/Information+era
Chapter 1:2 A terabyte is 10^{12} bytes and is equal to 1,000 gigabytes (1 gigabyte =1 billion bytes)
Chapter 1:3 An exabyte is 10^{18} bytes (1 billion gigabyte)
Information on bytes accessed from http://www.techterms.com/definition/

trends in data are derived and explained and the inferences are based on generating a hypothesis by putting a number of facts together necessary for both research conduct and integrity are the essential aspect of this PhD thesis.

The total global estimate of people living with HIV-1 infection continues to increase. By end of 2011 the global estimate was reported as 34 million (UNAIDS World AIDS Day Report, 2012) while this had increased to 35.3 million one year later in 2012 (UNAIDS Report on the Global AIDS Epidemic, 2013). If HIV-1 is left untreated most patients eventually progress to AIDS with reduced quality of life. As a consequence, their reduced productivity in society will impact on the economic and psychological aspects of well-being. With potent anti-retroviral (ARV) drugs now available and in use to manage the condition, HIV-1 is now manageable and regarded as chronic disease (Williams *et al.*, 2012). Data presented in this PhD thesis uses data collected from people living with HIV-1 who are managed at specialist centres in the NHS, and are here utilised as an example and a model to demonstrate the use of the NHS Hospital Information System (HIS) database that may help with decision making for health care professionals, business managers, policy makers, pharmaceutical companies and researchers so that social welfare and public services in addition to research conduct can be improved.

1.1 DATA MANAGEMENT

- ***DATA MANAGEMENT HISTORY (PRE-COMPUTERS)***

Herman Hollerith (29th February 1860 - 17th November 1929) was an American statistician who used Charles Babbage's[4] original idea of punch cards and electromagnetic relays to develop a motor powered tabulator for the 1890 United States of America Census (Table 1.1.1). This tabulator was developed to process the data and tabulate statistics from more than 62 million Americans. Although this was developed and custom built for the census, its basic function was to count, add and produce results from punched cards. Hollerith developed a mechanism using electrical connections to trigger a counter for recording information. A key idea was that data could be coded numerically. The idea behind this was that if numbers could be punched in rows and columns on a card then these cards could be counted or sorted mechanically and the data recorded. A description of this system, 'An Electric Tabulating System (1889)', was submitted by Hollerith to Columbia University as his doctoral thesis, and read before the Royal Statistical Society, 4th December 1894 (Hollerith, 1894).

Chapter 1:4 Charles Babbage (1791-1871), computer pioneer, designed the first automatic computing engines. Accessed from http://www.computerhistory.org/babbage/?gclid=CNTukdH7srUCFWbKtAodckQAOg

Having developed a new machine, Hollerith opened his own business under the name 'Tabulating Machine Company'. In 1924, the business merged and changed its name to 'International Business Machines Corporation' (IBM).

Table 1.1.1: Descendent of Herman Hollerith's 1890 Census Tabulator

1890	Hollerith Census Tabulator	Manual feed, wood cabinet, hardwired connections, counting only
1896	Hollerith Integrating Tabulator	Manual feed, true addition as well as counting
1900	Hollerith Automatic Feed Tabulator	First automatic-feed card reader, used in 1900 US Census
1906	Hollerith Type I Tabulator	(Type 090) Automatic feed; metal cabinet; first wiring panel
1921	Hollerith Type III Tabulator	(Type 091) First model with printer
1925	Hollerith Type 3-S Tabulator	First model with direct subtraction, removable plug board
1928	Hollerith Type IV Tabulator	(Type 301) First 80-column-card model
1931	Columbia Difference Tabulator	Unique machine for CU Statistical Bureau
1933	IBM Type 285 Tabulator	Numeric only
1933	IBM 401 Tabulator	Alphanumeric
1934	IBM 405 Accounting Machine	Alphanumeric
1948	IBM 402 Accounting Machine	Alphanumeric, with 403, 412, 417, 419 variations
1949	IBM 407 Accounting Machine	High-speed alphanumeric. 421, 444, 447 variations

IBM 405, 402 and 407 were electromechanical accounting machines
Source: http://www.columbia.edu/cu/computinghistory/tabulator.html

- ***DATA MANAGEMENT FROM 1950'S***

The next IBM computers (1959) were electronic digital computer for general purpose. Until the mid-1970s, most computer access was via punched cards Figure 1.1.1.

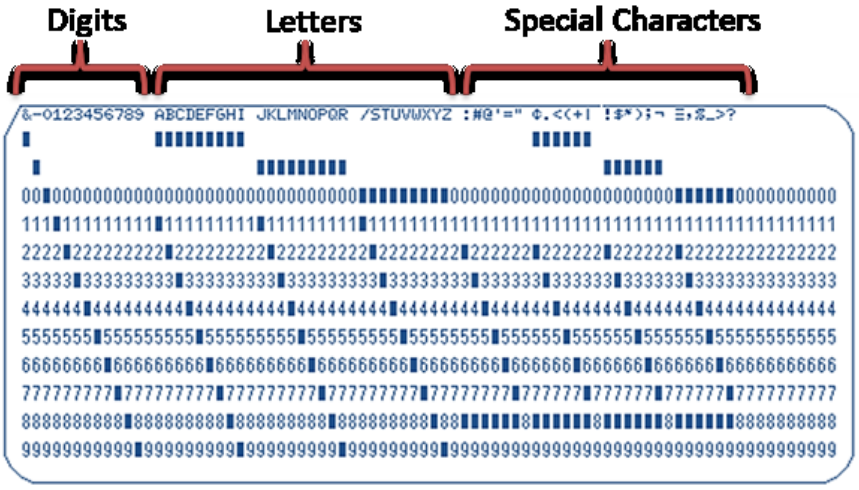


Figure 1.1.1: Punch Card Image
Adapted from Programming with punched cards 2005 Dale Fisk: http://www.columbia.edu/cu/computinghistory/fisk.pdf

In the 1980s data were processed using cards and computer programs with data punched on punch cards (Figure 1.1.1), using a key punch machine and read into a card reader. Next tapes were used for storage and retrieval of data. As technology advanced: using the punched cards, was followed by the use of tapes (sequential access media or fixed order of access), and subsequently followed by use of magnetic disks (random access media; Figure 1.1.2).

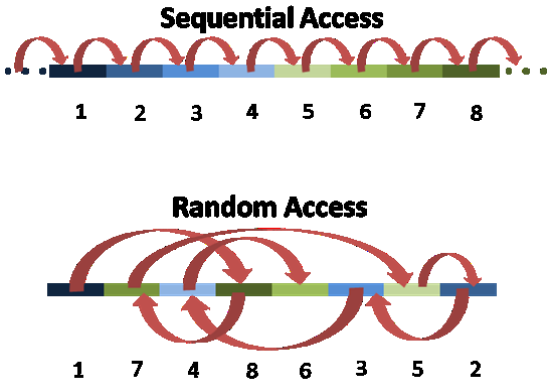


Figure 1.1.2: Tape (Sequential Access) and Disk (Random Access)
Adapted from: https://commons.wikimedia.org/wiki/File:Random_vs_sequential_access.svg

As technology began to advance it became possible to store single units of information in a single place and access that stored information or memory using random access memory (RAM). Examples of RAM are computer floppy disks and computer hard drives. As applications moved into real-time interactive applications, data management and data process management processes continued to evolve (Gray, 1996). Therefore, if the data were not defined well then there was a possibility of increased errors in data, while if the data management processes were not defined well then needs of the data users were not standardised. Research integrity in science requires good data management processes and it is important that the old data and applications continue to function as new data and applications are added (Gray, 1996).

As reported in a paper by Gray (Gray, 1996), data management has evolved over time and there are six development phases of data management advancement ard these are summarised in Figure 1.1.3. The first computational tool was developed around 4000 BC and probably involved pebbles and twigs and lines in the sand for counting. The bead frame abacus was probably invented by the Chinese sometime around the second century AD. Although abacus was not technically a computer, the operator of the abacus manipulated data mentally and with skill came speed (Dilson, 1968). The six phases of data management depicted in Figure 1.1.3 are: 1) Manually processed data; 2) Use of punched-card equipment and electro-mechanical machines to sort and tabulate records; 3) Stored data on magnetic tape and used stored program computers to perform batch processing on sequential files; 4) Introduction of the database schema and online access to data. Automated access to relational databases and added distributed and client-server processing; 5) Storage of richer data types, notably documents, images, voice, and video data. The sixth generation systems are the storage engines for the Internet, Intranets.

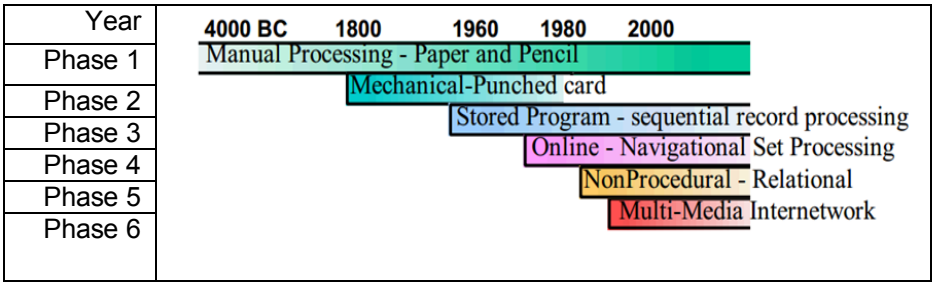


Figure 1.1.3: The Six Phases of Data Management
Source of figure (Gray, 1996). The six phases of data management advancement: from phase 1 manual methods through several phases of computerised data management.

- ***DATA MANAGEMENT***

According to the definition provided by the Data Management Association (DAMA) international, data management is '*the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets'* (Mosley, 2007). In information technology (IT) processes, data are generally represented as content in a data field which hold facts. Data become information when they are structured to provide context and meaning. Context for instance would mean: which data, how were the data collected, on whom, for what. Data management focuses on defining the data and how they are structured, stored and moved. Management of information is focused on the security, accuracy, completeness and timeliness of multiple data sources.

Data analyses are conducted on quantitative data such as age, CD4 T-cell count, HIV-1 RNA plasma load, and categorical data such as severity of disease mild, moderate, severe. The population based statistical analysis of the data was performed using various techniques such as simple descriptive statistics, linear regression analyses, survival analysis and longitudinal data analyses (Armitage *et al.*, 2001). Data analysis uses different statistical methods that help to describe data, detect patterns, provide clues and explanations, and test hypotheses (Machin & Campbell, 2005).

- ***DATABASE***

A database stores information electronically which is held over a period of time in a computer readable form. Database technology has advanced since the 1950's (see section 1.1 Data Management from 1950's). Most databases store time-varying information. For example to collect CD4 T-cell count history of a patient, the design would need a unique identifier for a person such as the hospital identifier (HOSPID) and dates describing continuous periods of time together with results.

The purpose of a database[5] is to help keep track of records, and the most commonly used type of database is the relational database. A relational database stores data in tables. Data are recorded facts and numbers. A table has rows and columns, such as those in a spreadsheet. Tables store information about a single subject and have columns that contain the different kinds of information about the subject and rows that describe all the features of the subject. A database usually has multiple tables, and each table contains data about a different type of records. A database shows the relationships among the rows of data which is an important characteristic and each row in a table is uniquely identified by a unique primary key where the values of these keys are used to create the relationships between the tables.

- ***DATABASE MANAGEMENT SYSTEM***

Managing several hundred electronic spreadsheets using folders and subfolders becomes difficult and thus access to finding patient records become a challenge. Database management systems store and handle information using relational database management systems, which manage all data in tables. Databases have evolved over time and these advances allow users to access, update, and search information based on the relationship of data stored in a database, and also allows queries to be run in multiple databases. The databases now allow, in addition to text or numeric data, storage of other data types such as sound clips, pictures, and videos. The expanding data management include data independence, data integrity, data consistency, information retrieval, multimedia retrieval, and data visualisation.

Broadly, data management involves: collection, storage, and retrieval of data. With IT advances, the volume of data collected continues to grow. The main data collected in the NHS that are used and discussed in this PhD thesis are based on patient care. Large collections of patient care data from different sources, such as laboratory, pharmacy data, activity data of outpatient and inpatient stay and many other sources of data may lead to errors such as data formats, data transfer errors, thus potentially jeopardising data integrity. To minimise these types of possible errors and thus increase confidence in predictions there is a need for appropriate processes for data collection, data storage, and data retrieval and data analysis (Figure 1.1.4).

Chapter 1:5 A database is a data structure that stores organised information electronically. http://www.techterms.com/definition/database

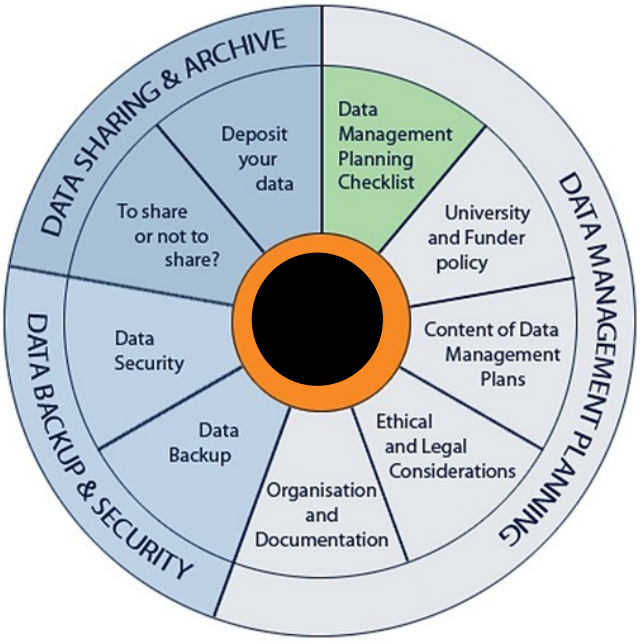


Figure 1.1.4: Data Management Process
Source: http://www.admin.ox.ac.uk/rdm/

The main ideas of database management are to ensure data independence, data integrity and data consistency where the principle aim is to ensure good quality data.

- ***TWENTY FIRST CENTURY AND THE GROWING INFORMATION AGE***

In 2008, the analyst firm, International Data Corporation (IDC) released a report '*The Diverse and Exploding Digital Universe'* (Gantz *et al.*, 2008). The report focused on the increasing size of electronic information around the world and noted that the amount of data in healthcare is large and growing. The IDC research showed that the information that is either created, captured, or replicated in computerised form was 281 exabytes in 2007. Whilst a 2010 IDC report predicted that by 2020, the information in the 'Digital Universe'[6] will grow by a factor of 44 compared to 2009; the number of 'files' to be managed in it will grow by a factor of 67, and storage capacity will grow by a factor of 30 (Figure 1.1.5) (Gantz & Reinsel, 2010).

Chapter 1:6 The Digital Universe is free online information resource, a Web application that lets users view an editorially defined subset of credible information.
http://www.digitaluniverse.net/digitaluniverse/resources/view/138616/?topic=8779#gen0

Figure 1.1.5: Worldwide Computerised Data Forecast to 2020
Source of figure: http://www.economist.com/node/18226961

As computer processor and storage capacities have grown exponentially, this has enabled the use of powerful, user-friendly software in medical practice. The UK Government has demanded an improved cost effective ways of working to produce better health outcomes for patients in order to provide an improved continuity of care. Information continuity, management continuity and relational continuity (Haggerty *et al.*, 2003) are necessary in order to improve the quality of care provided.

- ***MEDICAL RECORDS***

The two main methods of collecting patients medical data are paper based as well as electronic medical records. Data within the NHS or other health care setting in the UK are either collected in: patient's medical notes only and therefore only available in a paper format, in an electronic format captured into a specific database including local data warehouse, or are collected in a combination of both data collection methods.

Medical records contain patient care documentation where care was received and include data such as lab results, visit notes, diagnostic test results, demographics, health histories, medication information, and more. Depending on the method of data collection these are either securely stored in paper files or in a securely held local server in an electronic format. Electronic medical records mostly contain same clinical data in an electronic format so for instance electronic documentation of care history of a patient attending for care at one treatment centre. Some of the advantages of these types of data collection include: allowing patients records to be tracked over

time, administrative ease of identifying patients who are due for appointments for instance, allowing checks regarding how patients are doing following clinic visit for instance. All these allow clinicians within one treatment centre to monitor and improve overall quality of patient care and hence improving patient outcomes. While same data held only in patient notes as paper based information are often held in large folders and filed in various locations such as the doctors' offices and hospitals, and not turned into electronic records, become difficult to locate and could potentially become a time consuming task to locate individual patient folders.

Electronic patient medical records are used by health care professionals to access patient information and support decision making usually within one practice, through evidence-based health practice. Globally there is a need to ascertain the effectiveness, efficiency, equity and acceptability of health care services provided. As a result there is a commitment within the UK government for development of universal healthcare systems to be put in place (Figure 1.1.6).

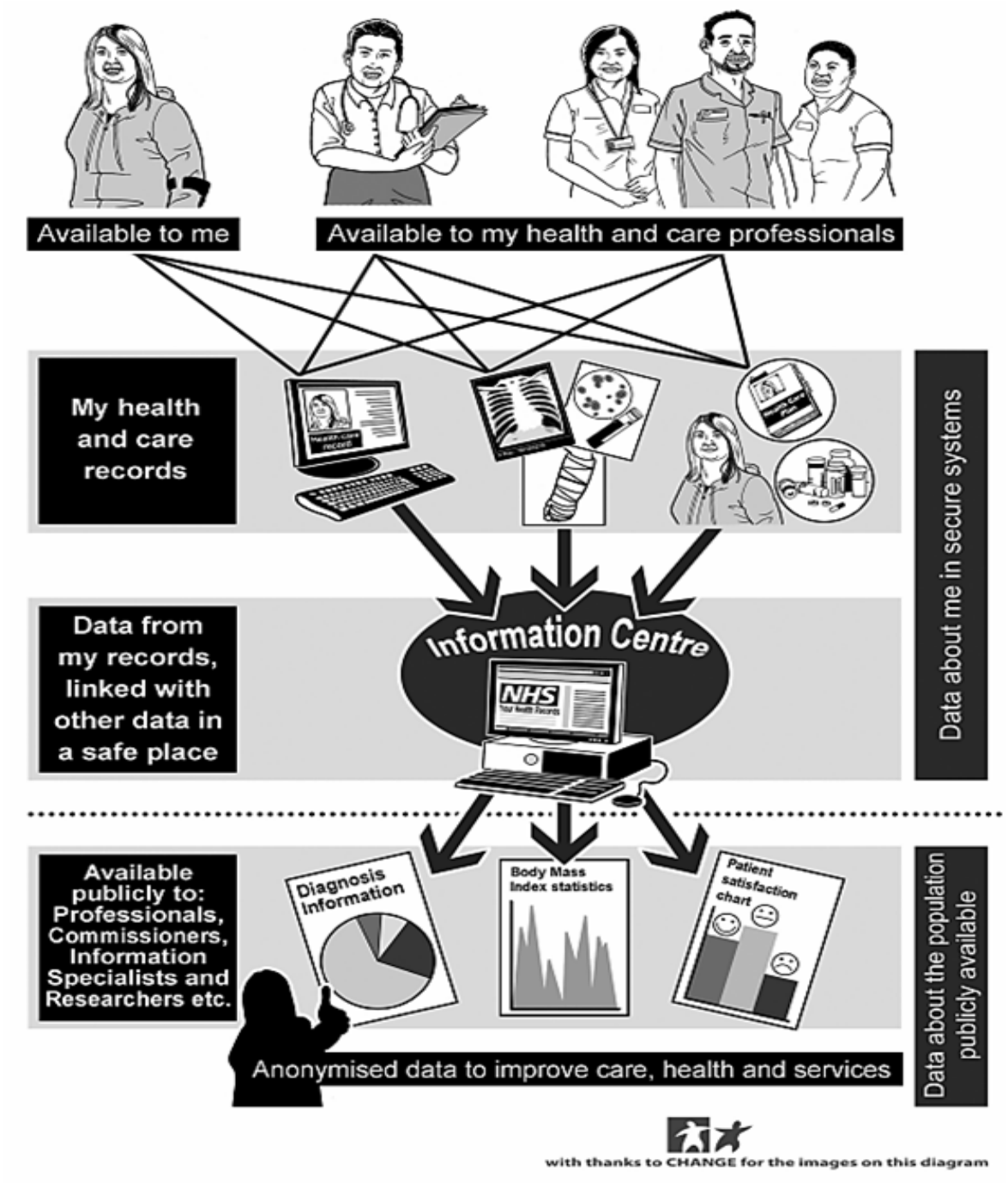


Figure 1.1.6: Access to Health and Care Information
Source: The power of information: Putting all of us in control of the health and care information we need (Department of Health, 2012)

- ***HEALTH RECORDS***

Health care is part of the information age and electronic health records focus on the total health of the patient. This is in addition to the standard electronic patient medical records collected within one practice. Health records are medical records with the capacity for greater exchange; they are the health data held on a patient at different clinical practices and when in an electronic format

allow for data exchange between different health care team workers. The disadvantages of paper based collection of data are that these types of health record data exchange has limitations as papers need to be first retrieved and exchanged via normal post or fax which is time consuming and access to these data records are with some level of delay. The electronic health records in contrast include a larger view of data about the patient's care beyond one treatment centre that originally collected the information and share information with other health care providers, such as laboratories and specialists therefore these contain information from all clinicians involved in a patient's care. They are patient-oriented, aggregated, longitudinally collected data about the health of a patient over a wide area, potentially including many geographically dispersed data sources and the information collected moves with the patient to care providers such as: to hospitals, different specialists, to nursing homes, to different hospitals in other parts of the country. An electronic health record contains each individuals aggregate, secure and confidential lifetime record of health care history recorded within the health care system that is available electronically with authorised health care providers and the individual anywhere, anytime in support of high quality care (Garets & Davis, 2006).

The databases that hold electronic health records are built so that routinely collected information can be shared with other health care providers such as laboratories, different specialist services, located locally as well as elsewhere in the country. The data from individual databases or local data warehouses are deposited into a national data repository or national data warehouses which therefore hold routine information about patients health care details from all clinicians involved with their care allowing information to be shared speedily across different care providers supporting healthcare team effort. Further, secondary uses of these routinely collected data into databases include: clinical research, service evaluation, clinical audits, surveillance and usual practice in public health (NRES, 2009) covering such areas as: public health, epidemiology, health trends, service developments, informing governments, monitoring quality of health care and informing patient choice, linkage to other data sources such as mortality data from the Office of National Statistics, national audits, cancer registries (Spencer, 2013).

The UK government's commitment to develop universal healthcare systems will use IT to improve the NHS health care provision and such improved systems will help improve decision making for health care professionals, business managers, policy makers, pharmaceutical companies and researchers so that social welfare, public services and research conduct can be improved. The routinely collected data on HIV-1 infected people who are managed long-term with HIV-1 infection was used as an example and a model to demonstrate the use of the NHS health care system database.

CHAPTER 1

GENERAL INTRODUCTION

- ***THE NATIONAL PROGRAMME FOR IT IN THE NHS: 2000-2011***

In June 2000, the UK government, responsible for NHS England, spent £11.4 billion for an update of NHS IT systems to help provide high quality services to patients through the National Programme for IT in the NHS (NPfIT). The NHS in England began deployment of electronic health records systems in 2005 together with other electronic health record initiatives in Wales, Scotland and Northern Ireland. The goal was to have the UK populations medical details on a centralised electronic health record by 2010. The UK governments aim was to replace local NHS computer systems with a modern integrated systems and make patient's health records available electronically throughout England and including other services, such as electronic prescriptions, an email and directory service for all NHS staff (NHSmail), computer accessible X-rays (Picture Archiving Communication Systems - PAC), and a facility for patients to book outpatient appointments electronically so that patient records can be shared (House Of Commons, 2007).

The implementation of planned targeted work was delayed and subsequently stopped in 2011. The failures to meet IT implementation targets were highlighted in the 2007 House of Commons report regarding the shared electronic patient clinical records. The report highlighted NPfIT were running two years behind schedule with other examples of failures described in the report. Subsequent to this report and failures to meet the set targets, the UK government announced cancellation of work contracts with NPfIT. The critical area which led to the delays in the planned work were highlighted in a publication by Deutsch and colleagues (Deutsch *et al.*, 2010) and a summarised opinion was published in computer weekly by a partner at Morrison & Forester (UK) LLP firm, with six points summary why the NHS NPfIT failed to meet the IT implementation targets (Maughan, 2010). The six points were summarised under the following broad headings: 1) Motives; 2) Buy-in; 3) More haste, less speed; 4) Poor contracting process; 5) Multi sourcing; 6) Accountability.

The Department of Health (the Department) had hired British Telecom (BT) and Computer Sciences Corporation (CSC) as two main contractors to undertake IT systems upgrade in the NHS. Although payments were made in advance to CSC, however CSC delivered very few of the systems it was contracted to supply and instead implemented a large number of interim systems as a stopgap. BT also proved unable to deliver against its original contract. One factor which contributed to failings was the Department's weak programme management. This resulted in poor accountability for project performance. The shortcomings of the National Programme included poor negotiating capability, resulting in deals which were poor value for money and weak programme management and oversight. In addition there were failures regarding understanding of complexity of the tasks, the difficulties of persuading NHS trusts to take new systems that had been procured nationally, and to get people to operate the systems effectively when they were adopted. The Department was criticised as they could have avoided some of the pitfalls and waste if they had consulted with health professionals earlier.

- ***THE CURRENT UK INFORMATION STRATEGY: FROM 2012***

In May 2012, the UK Department of Health published an information strategy policy document '*The power of information: Putting all of us in control of the health and care information we need*' (Department of Health, 2012), which set a ten year framework to transform information for health and care which would allow people control of their health and care records and to give them access to these by 2014 (see section 1.1: Figure 1.1.6). The aims are to integrate information and new technologies in order to achieve higher quality care and improve outcomes for patients and service users under the Health and Social Care Act 2012, it covers public health, healthcare and social care in adult and children's services in England (Department of Health, 2012).

The electronic capture, storage and transfer of information can provide different stakeholders with access to health care information with minimum delay (Department of Health, 2012). Improved IT will facilitate input, storage, access, analysis, and communication of data and information for clinicians and researchers thereby can help identify gaps in healthcare services. The knowledge generated from these electronic medical and health data is likely to improve resource allocations, reduce costs, and help define areas of further investigation in order to improve outcomes for patients and service users.

The NHS in the UK underwent changes which became effective from 1st April 2013, NHS England took on the functions of the primary care trusts (PCTs) with regard to the commissioning of primary care health services, as well as some nationally based functions previously undertaken by the Department and the clinical commissioning groups will be overseen by the NHS England (Figure 1.1.7). The main aim of NHS England is to improve the health outcomes for people in England.

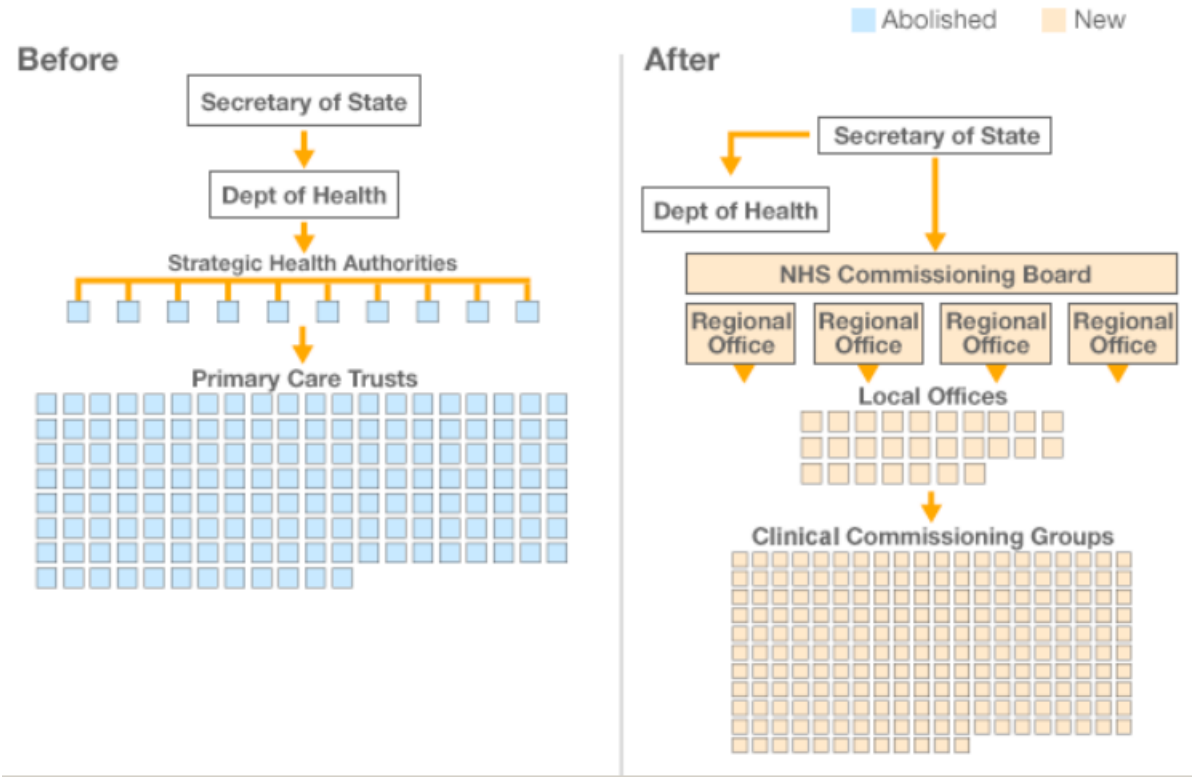


Figure 1.1.7: The Health and Care System from April 2013
Before in Figure 1.1.7 means prior to 1st April 2013, while after refers to changes that took effect from 1st April 2013
Source: http://www.bbc.co.uk/news/health-19674838

These changes are related to: which organisation will make decisions about the NHS services, how these services are commissioned, and the way money is spent. The changes that took effect are summarised in Figure 1.1.8.

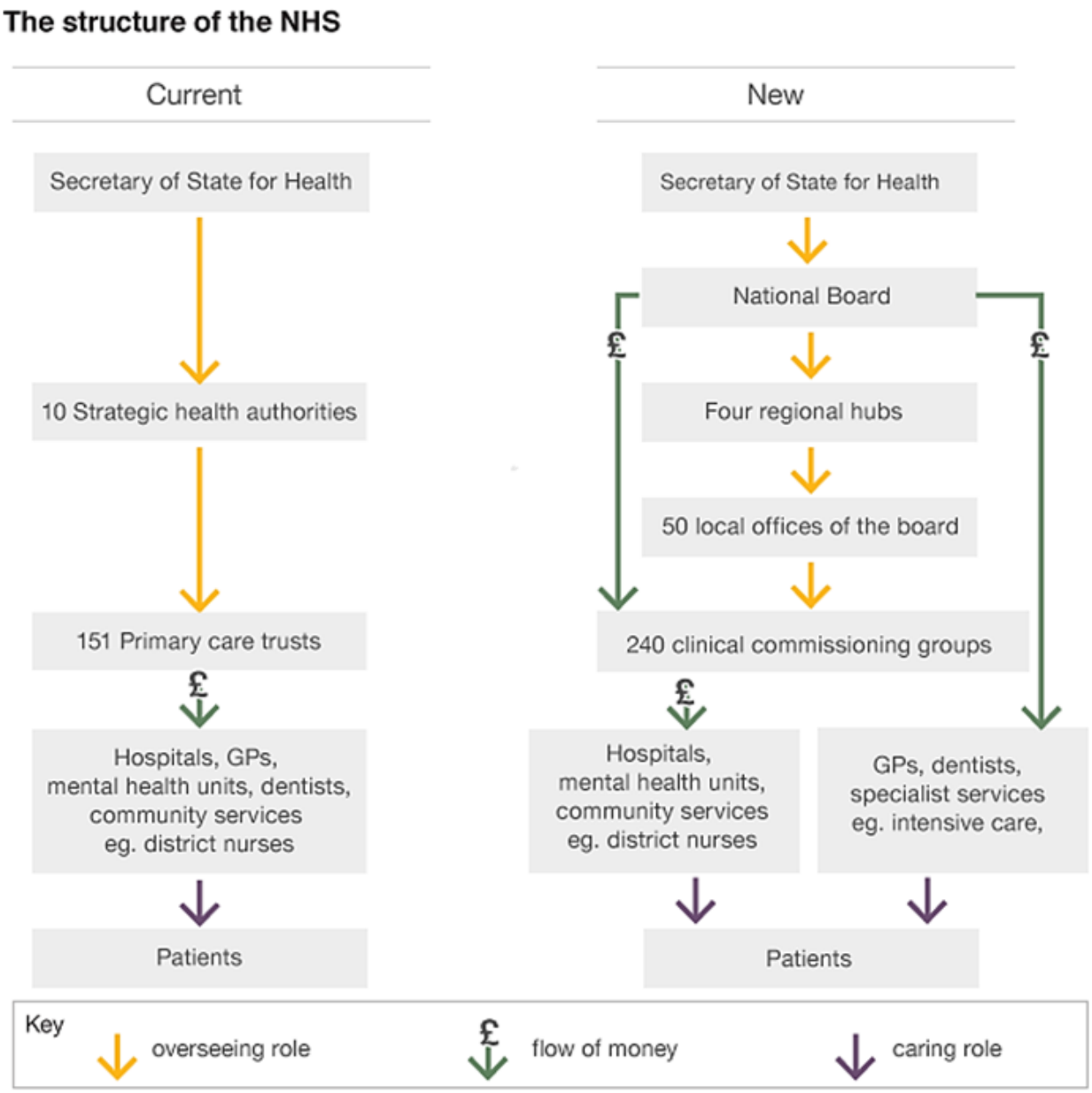


Figure 1.1.8: The Changes to NHS Structure from April 2013
Current in Figure 1.1.8 means prior to 1st April 2013, while new refers to changes that took effect from 1st April 2013
Source: http://www.bbc.co.uk/news/health-12177084

From April 2013, the clinical commissioning group (CCG) is led by GPs who are responsible for around 60% of the NHS budget. They have taken charge of overseeing the NHS from the Department. Since GPs are in a regular contact with patients, Ministers believe that CCG will be more responsive to the needs of patients and therefore will make the NHS more efficient and improve the quality of care.

The NHS Englands recently published business plan for 2013-2014 through to 2015-16 *'Putting Patients First'* (NHS England, 2013) highlighted that the NHS England will publish a technology roadmap which will set a national vision for NHS IT for commissioners investing in IT and

informatics. With regards to IT, one of the target is to have 95% of trusts using the NHS number as a unique patient identifier in clinical correspondence by January 2015. The key deliverables highlighted in this report were 100% of practices will offer the facility to order repeat prescriptions and to book appointments by March 2015 (NHS England, 2013). In addition, e-referrals service would be operational by December 2013 and 100% of referrals will be made electronically by March 2017. The plan states that health and care data are one of England's greatest public assets and 'putting it to work' is key to improving patient outcomes. In response to this the NHS England is developing the *care.data*, a programme designed to link patient data from different care sectors for the first time, which will provide linked data across the different components of the patient journey and the outcomes resulting from treatment. Once in practice, having access to such data quickly will allow early intervention which can help prevent adverse developments affecting patients' health, and therefore avoiding retrospective treatment. As a result this may have an impact on efficiency savings by reducing the cost of treatment and freeing time for hospital staff. In the current economic climate the NHS is balancing between providing high quality care and Government's intention of efficiency savings in the NHS, a saving of £20 billion a year by 2015[7].

Often used term 'big data' is not well defined but 'big data' refers to datasets whose size is beyond the ability of typical database tools to capture, store, manage and analyse (Manyika *et al.*, 2011) and is characterised by (Marr, 2013): Volume - the vast amounts of data generated every second; Velocity - the speed at which new data is generated and moves around; Variety - the increasingly different types of data (from treatment data to X-rays to voice recordings); and Veracity - the messiness of the data (e.g. text strings with hash tags, abbreviations, typos).

'*We live in the information age*' is also synonymous to '*we live in the data age'* where large amounts of data: terabytes, exabytes are being collected. Data mining (knowledge mining from large amounts of data) turns a large collection of data into knowledge. The data mining is the process of discovering patterns in data after the required information has been extracted (Figure 1.1.9). The data sources can include databases, data warehouses, the Web, other information repositories, or data that are streamed into the system dynamically, ordered/sequence data, graph or networked data, spatial data, text data, multimedia data, and the web. Data mining is currently evolving and other data sources will be included as these evolve over time.

Chapter 1:7 Department of Health, 25 March 2013 https://www.gov.uk/government/policies/making-the-nhs-more-efficient-and-less-bureaucratic

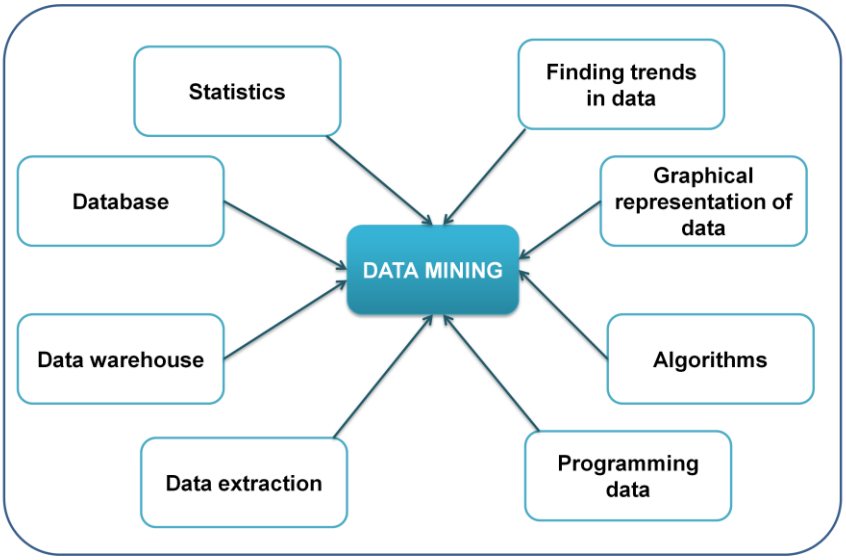


Figure 1.1.9: Data Mining Process Diagram

Definition of data mining *(Oxford dictionary: http://oxforddictionaries.com)* 'the practice of examining large pre-existing databases in order to generate new information'. The data mining process involves: Data cleaning: to remove noise and correct inconsistency in data, Data integration: merges data from multiple sources into a coherent data, Data selection: can reduce data size, for e.g., aggregating, eliminating redundant data fields, Data transformation: e.g. normalisation may be applied, where data are scaled to fall within a smaller range like log transformation, Data mining: Finding trends in data, Evaluate trends found in data, Knowledge presentation.

There are however also many other terms with a similar meaning to data mining - for example, knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, and data dredging. The ultimate goal of data mining is prediction and data mining is considered to be a blend of: statistics; database management; data base research; artificial intelligence; machine learning; pattern recognition; and data visualisation (Friedman, 1998).

Big data analytics is the process of examining large amounts of data of a variety of types (big data) to uncover hidden patterns, unknown correlations and other useful information that may have been left untapped as technology is now advanced to bring all data of a variety of types together and analyse these. The economic value of 'big data' was first termed as 'data equity'[8]. A report by The Centre for Economics and Business Research Ltd (CEBR) *'Data Equity: Unlocking the value of Big*

Chapter 1:8 The term 'data equity' was first used in an article in The Economist in May 2011 by Adrian Wooldridge entitled 'Building with big data' to mean the economic value of big data http://www.economist.com/node/18741392

Data', the use of big data analytics across the healthcare sector could deliver additional revenues of £14bn from 2012 to 2017 (CEBR, 2012).

1.2 HUMAN IMMUNODEFICIENCY VIRUS

- ***HIV HISTORY***

The earliest case of Human Immunodeficiency Virus (HIV) infection was documented in a 1959 blood sample taken from an adult male from the Democratic Republic of the Congo (Zhu *et al.*, 1998). Epidemiologists have speculated that the virus may have travelled from Africa to the United States during the 1970's where it was first transmitted via sexual intercourse among gay men in New York and San Francisco.

In 1981 a number of gay men in New York and San Francisco developed rare cancers and opportunistic infections that were resistant to treatment. On 5th June 1981 a case of AIDS was reported (MMWR, 1981; Fee & Brown, 2006). The syndrome at the time was known as gay-related immune deficiency (GRID), which on 27th July 1982, at a meeting held in Washington, was renamed to Acquired Immunodeficiency Syndrome (AIDS) (Kher, 1982). In French, Portuguese, and Spanish, it is known as syndrome d'immunodéficience acquise (SIDA).

In 1983 the cause of AIDS was described as a retrovirus when lymphadenopathy associated virus (LAV) was isolated by the Institute Pasteur, France from a French homosexual patient with generalised hyperplastic lymphodenopathy (Barré-Sinoussi *et al.*, 1983; Montagnier *et al.*, 1984). In 1984 Robert Gallo and colleagues at The National Cancer Institute, USA characterised the AIDS-associated retrovirus as human T-lymphotrophic virus III (HTLV-III) (Popovic *et al.*, 1984; Gallo & Montagnier, 2011), although both the French and US laboratories had identified the cause of AIDS, in 2008 the Nobel Prize for the discovery of HIV was awarded to Barre-Sinoussi and Montagnier. In 1987 the name human immunodeficiency virus (HIV) became internationally accepted during a World Health Organisation (WHO) Consultative Meeting and was confirmed by the International Committee on Taxonomy of Viruses (ICTV)[9].

In 1986, a second type of retrovirus (type 2; HIV-2) which belonged to the HIV group was isolated by Luc Montagnier and colleagues from West African patients with AIDS (Clavel *et al.*, 1987). While HIV-1 is found worldwide, HIV-2 is largely confined to West African regions. Although HIV-1 and HIV-2 are related, there are important differences between them which influence pathogenicity, natural history and therapy (Markovitz, 1993). Natural history studies indicate that HIV-2 is less pathogenic than HIV-1 and progression of disease among people infected with this strain alone is slower. HIV-2 infected people have shown a slower progression to AIDS with 86-

Chapter 1:9 ICTV is a committee which authorises and organises the taxonomic classification of viruses

95% of HIV-2 infected people fulfilling the definition of long-term non-progressors (LTNP) (Kanki, 1999). The work presented in this PhD thesis will be focusing on HIV-1 infection.

- ***HIV-1***

HIV-1 belongs to the lentivirus (or slow virus) family and causes slow progressive disease. Both HIV-1 and HIV-2 are ribonucleic acid (RNA) viruses and consist of two short strands of RNA along with the enzymes reverse transcriptase, protease, ribonuclease, and integrase (Tang *et al.*, 1999). This genetic material is encased in a capsid within a membrane lined by a matrix protein, the viral envelope. Projecting from this viral envelope include 72 spikes formed from the viral proteins gp120 (surface) and gp41 (transmembrane). Figure 1.2.1 shows the different components of the virus structure.

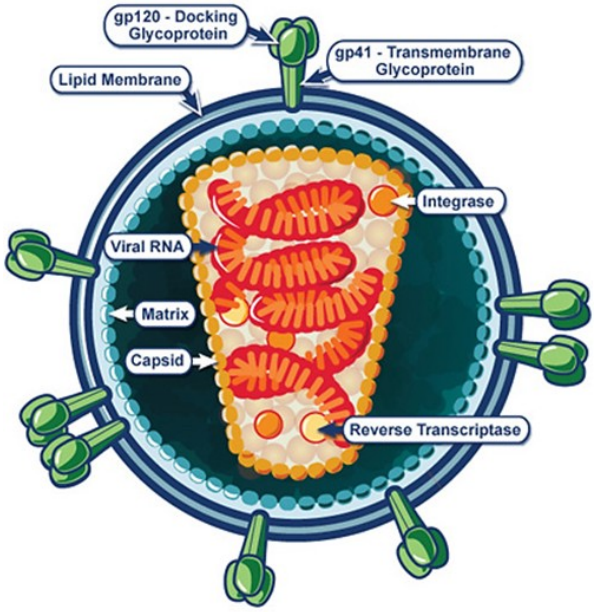


Figure 1.2.1: Structure of HIV Virus
Source: http://www.niaid.nih.gov/topics/hivaids/understanding/biology/Pages/structure.aspx

		Core Viral Enzymes	
Glycoproteins	**Core capsid**	**Reverse Transcriptase**	**Integrase**
gp120/gp41	**p24**	**p64**	**p32**
Mediate entry into host cell	Within the capsid are two identical single strands of RNA (the viral genetic material) and viral enzymes	Converts single stranded viral RNA into double stranded viral DNA	Facilitates integration / incorporation of viral DNA into host DNA

The main route of HIV-1 transmission is through sexual contact (semen, vaginal secretions), during pregnancy or vertical transmission from mother to child and through breast milk. Other possible routes of transmission include: infected blood and blood products; contaminated needle sharing among injecting drug users (IDU); contaminated blood transfusions. HIV-1 targets the host immune

system and renders it incapable of coping with opportunistic pathogens. HIV-1 infection initiates a process that leads to progressive destruction of CD4 T lymphocytes, a type of helper T white blood cell vital to the adaptive immune response. CD4 T lymphocytes are required for the normal function of the immune system, which defends the body against infection. When HIV-1 weakens the immune system due to the progressive decline in immune defence mechanisms, people with HIV-1 infection develop a number of opportunistic infections during later stages of disease that is also characterised with onset of various tumours/malignancies and becoming infected with other viruses, bacteria and parasites.

The course of the HIV-1 infection varies from person to person. A typical pattern of HIV-1 infection is divided into three stages: primary infection; asymptomatic infection; and eventually development of certain opportunistic infections, neoplasias or other conditions. The advanced stage of this is known as AIDS which may be accompanied by opportunistic infections and or decline of CD4 T-cell counts below a nationally agreed defined threshold (Figure 1.2.2). A person who tests HIV-1 positive is considered to have progressed to AIDS when the patient develops at least one of about 25 different opportunistic infections (McCarthy & Mercey, 1994), diseases that may not necessarily affect a person with a normal immune system but that take advantage of weakened immune system.

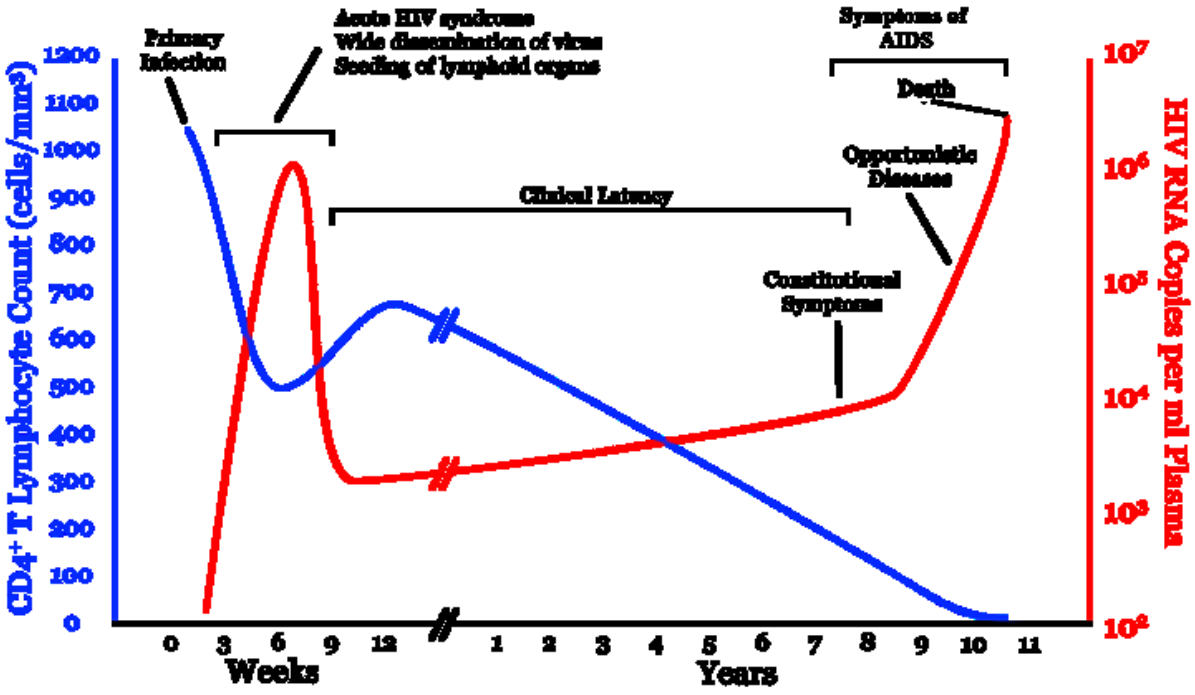


Figure 1.2.2: Typical Phases of HIV-1 Infection and the Dynamics of $CD4^+$ T-Cell Counts and HIV-1 RNA Plasma Load
Source of figure: http://upload.wikimedia.org/wikipedia/commons/a/a4/Hiv-timecourse.png
cells/µL = cells/micro litre - this is sometimes written as cells per one cubic millimetre mm^3 ($cells/mm^3$)

1. Primary infection

The primary infection, which generally lasts for 2 to 8 weeks, is typically associated with high levels of virus in the patient's plasma and rapid CD4 T-cell decline in peripheral blood. During primary infection, the number of CD4 T-cell counts decrease but return to near normal levels within three to four months after infection. Similarly, there is rapid viral replication characterised by high HIV-1 RNA plasma load during the acute phase that drops to a stable level (termed viral set-point) over a period of two to ten weeks (Figure 1.2.2).

2. The asymptomatic phase

The asymptomatic phase is characterised by a gradual decline in CD4 T-cell count, but this actual rate of decline varies substantially between individuals.

3. The symptomatic phase

The symptomatic stage occurs late in infection when opportunistic infections and malignancies occur due to poor immunity. During this stage, resurgence of HIV-1 RNA plasma load is seen and uncontrolled HIV-1 replication occurs. With the onset of AIDS, HIV-1 infected people may rapidly progress to death in the absence of appropriate treatment (Feinberg, 1996).

At present there is no cure for HIV-1 infection (Eisele & Siliciano, 2012). HIV-1 infection that was once a fatal condition is now a chronic condition managed and controlled with ARV drugs (Palella *et al.*, 1998; Deeks, 2011). These are prescribed life-long in order to sustain and control HIV-1 replication. There are a large number of approved ARV drugs prescribed for controlling replication of HIV-1: occasionally alone as mono therapy; or mostly in combinations as 3 or more ARV drugs described as Highly Active Antiretroviral Therapy (HAART). Some of the ARV drugs prescribed and are approved for use in Europe are detailed in Table 1.2.1.

Table 1.2.1: ARV Drugs Approved for Use in Europe

Cohort definition	Class of ARV	ARV drugs and year of approval for use in Europe and its availability in the UK ARV drugs are known with: their generic name, drug trade name (abbreviation)	
pre-HAART **(pre 1996)**	NRTI	Zidovudine, Retrovir (AZT)	1987
		Didanosine, Videx (DDI)	1991
		Zalcitabine, Hivid (DDC)	1992
		Stavudine, Zerit (D4T)	1995
		Lamivudine, Epivir (3TC)	1994
		Saquinavir Invirase[&], Fortovase[&&] (SQV)	1995
HAART era 1 **(1996-2000)**	NNRTI	Nevirapine, Viramune (NVP)	1996
	PI	Ritonavir, Norvir (RTV)	1996
	PI	Indinavir, Crixivan (IND)	1996
	PI	Nelfinavir, Viracept (NFV)	1997
	NRTI	Delaviradine, Rescriptor (DLV)	1997
	NNRTI	Efavirenz, Sustiva (EFV)	1998
	NRTI	Abacavir, Ziagen (ABC)	1998
	PI	Amprenavir, Agenerase (APV)	2000
HAART era 2 **(2001-2004)**		Kaletra (LOP+RTV)	2001
		Trizivir (AZT+3TC+ABC)	2001
		Combivir (AZT+3TC)	2003
	Fusion Inhibitor	Enfuvirtide, Fuzeon (T-20)	2003
HAART era 3 **(2005-2008)**		Tipranavir, Aptivus (TPV)	2005
		Truvada (TDF+FTC)	2005
		Kivexa (ABC+3TC)	2005
		Atripla (EFV+TDF+FTC)	2006
	Integrase Inhibitors	Raltegravir, Isentress (RAL)	2007
	CCR5 Entry inhibitor	Maraviroc, Celsentri^ (MVC)	2007
	NNRTI	Etravirine, Intelence (ETV)	2008
HAART era 4 **(2009-2011)**	NNRTI	Rilpivirine, Edurant (RPV)	May 2011

NRTI: Nucleoside reverse transcriptase inhibitor (NRTI), also called nucleoside analogues
NtRTI: Nucleotide reverse transcriptase inhibitor (NtRTI), also called nucleotide analogues
PI: Protease inhibitor
NNRTI: Non-nucleoside reverse transcriptase inhibitors
Fusion Inhibitor
Integrase inhibitor
CCR5 Entry inhibitor or CCR5 Fusion Inhibitor
^Celsentri is known as Selzentry in the US
[&] Hard gel capsule
[&&] Soft gel capsule

- ***GLOBAL HIV-1***

Worldwide, the total number of adults and children estimated to be living with HIV-1 continues to increase. At the end of 2011 it was 34.0 million (range: 31.4 to 35.9; Figure 1.2.3) (UNAIDS World AIDS Day Report, 2012) while by the end of 2012 it had increased to 35.3 million (range: 32.2 - 38.8) (UNAIDS Report on the Global AIDS Epidemic, 2013). The HIV-1 pandemic continues to grow and majority of new infections occur in developing countries. Of the total world population that is reported to be infected with HIV-1, 69% are estimated to be living in Sub Saharan Africa. In Western and Central Europe, 2.6% of the world's adults and children are estimated to be living with HIV-1 (UNAIDS World AIDS Day Report, 2012).

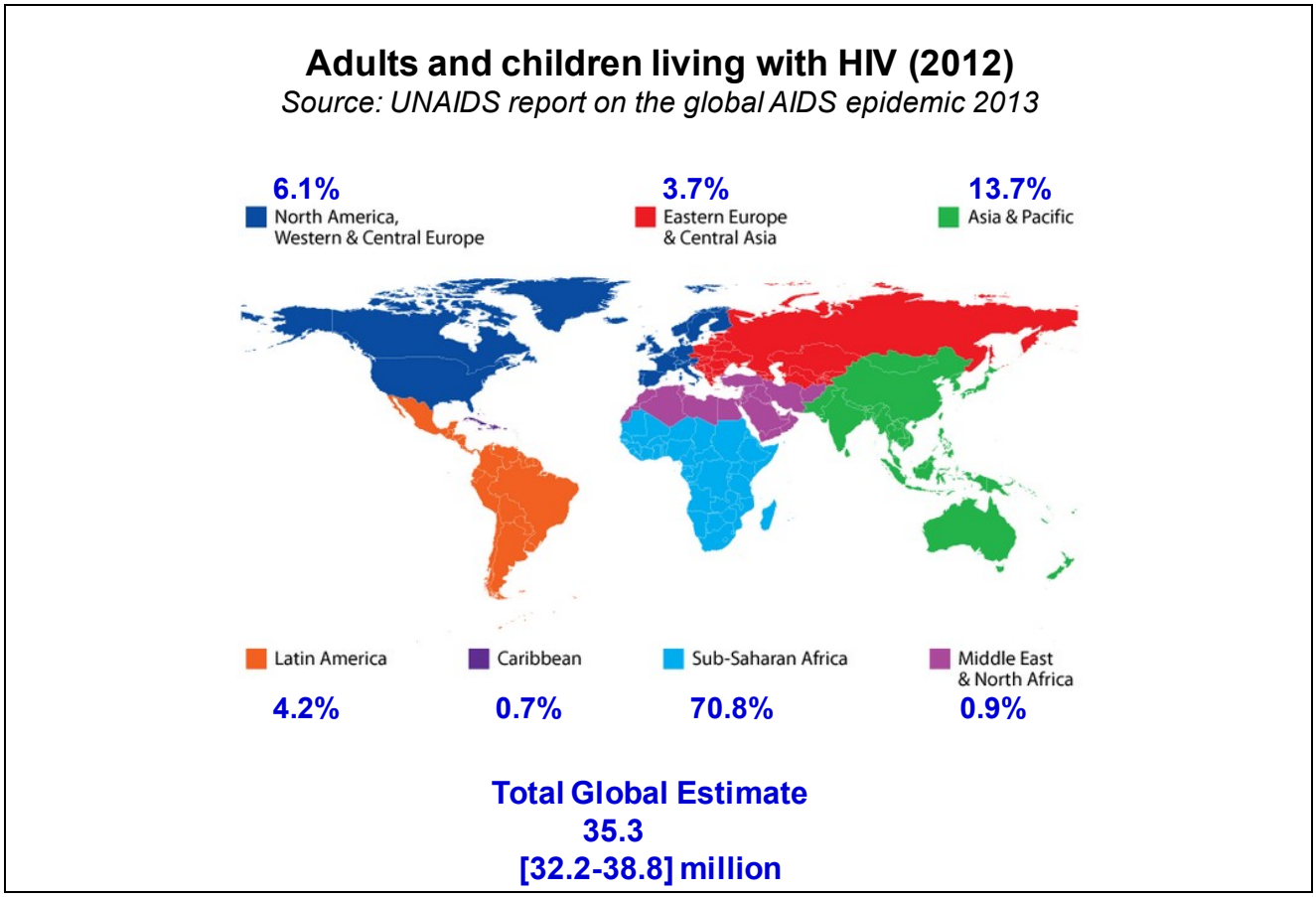


Figure 1.2.3: Proportion of Total Global 2012 Estimate of People Living with HIV-1 by Continent
Source of estimates of number of people living with HIV-1 by end of 2012: UNAIDS World AIDS Day Report, 2013
Proportion of people living with HIV-1 for each continent derived relative to the total global 2012 estimate

Many people diagnosed with HIV-1 early in the epidemic survived for a short duration. The advances in use of cART was announced in 1996 at the International AIDS Conference in Vancouver. Since entering the highly active (HA)ART era in 1996, mortality due to HIV-1 infection has declined, resulting in longer survival of HIV-1 infected people. Following the successful scale up of HIV-1 treatment access, worldwide the number of deaths attributed to AIDS related causes fell by 24% from 2.3 (2.1 to 2.6) million in 2005 to 1.7 (1.5 to 1.9) million by end of 2011 (UNAIDS World AIDS Day Report, 2012). Although there have been advances in treatment of HIV-1 infection, the search for preventive strategies on the development of prophylactic and therapeutic interventions through effective vaccines and/or immune-based therapies continues to put an end to disease, to prevent infection or delay disease progression (Blankson, 2010; Imami *et al.*, 2013; Walker & Yu, 2013).

- ***PEOPLE LIVING WITH HIV-1 IN THE UK***

The number of people living with HIV-1 in the UK increased from 96,000 in 2011 (credible interval: 90,800 to 102,500) to 98,400 in 2012 (credible interval: 93,500, 104,300), of whom an estimated

22% (credible interval: 18% to 27%) were unaware that they were infected with HIV-1. By the end of 2012 76,500 people diagnosed as HIV-1^+ had used the NHS services (credible interval: 75,000, 78,000) of whom 19,123 (25%) were older adults (≥50 years). In comparison, in 2003 4,361 (12%) HIV-1^+ people had used the NHS services in this ≥50 years age group (Aghaizu *et al.*, 2013).

- ***IMPACT OF HIV-1 INFECTION AND TREATMENT IN THE UK***

The rate of disease progression from asymptomatic HIV-1 infection to AIDS varies between patients (Pantaleo & Fauci, 1996). The basic science study presented in this PhD thesis determined factors associated with HIV-1 progression which may enable the development of prophylactic and therapeutic interventions through effective vaccines and/or immune-based therapies to put an end to disease, to prevent infection or delay disease progression (Chapter 3: long-term non-progressors - LTNP).

Many HIV-1 infected patients in the UK need long-term medical care at specialist NHS hospitals and treatment centres, and with new and potent ARV drugs now available for treating HIV-1 infection (see Table 1.2.1), people who have access to these drugs continue to live longer and therefore HIV-1 is now considered a chronic illness. Owing to the success of HIV-1 treatment and care, in combination with an increasing number of people who are older becoming infected with HIV-1, a greater number of people aged ≥50 years are living with HIV-1 (Deeks, 2011). In the UK and worldwide, the proportion of HIV-1^+ people who are aged ≥50 years is increasing rapidly. Furthermore, these older treated HIV-1^+ individuals are now facing a complex 'ageing' process involving biological, social and psychological aspects. The clinical science study presented in this PhD thesis assessed the impact of HIV-1 infection, anti-retroviral therapy (ART) and age on renal insufficiency in PLHIV managed at the HIV referral centre (Chapter 4: HIV and Ageing).

The ageing HIV-1 cohort together with increased number of people being diagnosed with HIV-1 infection has resulted in an increased number of people accessing HIV-1 services in the NHS over time (Aghaizu *et al.*, 2013). These increases in access of health care in the NHS implies that the costs associated with HIV-1 care will continue to also increase over time, if the current use of HIV-1 services persist in the NHS hospitals. This therefore directly impacts on the NHS cost burden nationwide for treating HIV-1 infected people in the NHS specialist treatment and care centres. The population science study presented in this PhD used information from an established national multicentre prospective monitoring system to report changing use and population cost of HIV-1 services in the UK, and estimated future population costs arising from projected increases in the number of PLHIV using NHS services (Chapter 5: Population Cost Study).

CHAPTER 1

GENERAL INTRODUCTION

- ***IMPACT OF HIV-1 INFECTION AND TREATMENT ON THE HEALTH CARE***

Due to success of HAART in those people who are on ART, this has increased the life-expectancy[10] of HIV-1$^+$ people. In a recent study, life-expectancy reported of an HIV-1$^+$ aged 20 years in North America and Europe ranged from 32 years for those who start ART with a CD4$^+$ T-cell count <100 cells/μL blood to 50 years for those who start ART with CD4$^+$ T-cell >200 cells/μL blood (AntiretroviralTherapyCohortCollaboration, 2008). Similarly, HIV-1$^+$ people on ART in resource limited countries can also attain near normal life-expectancy: A Ugandan study estimated that life-expectancy of an HIV-1$^+$ aged 20 years was estimated to be 27 years, for an HIV-1$^+$ aged 35 years life-expectancy was estimated to be 28 years while an HIV-1$^+$ aged ≥ 50 years life-expectancy was estimated to be 24 years. Life expectancy was observed to increase substantially with increasing baseline CD4$^+$ T-cell count with similar trends observed for older age groups (Mills *et al.*, 2011).

- ***RESEARCH DESIGN***

The design of any study from observational data is mostly implicit in the precise specification of the research question that needs to be answered. The research questions determine the general structure of the study design required together with the data requirements and the main analyses. The research design is required so that the research question can be answered systematically and scientifically. The aim is to control the variability in data as far as possible. A number of studies have examined factors associated with treatment failure and increased survival using a number of predictive analytical methods including time varying covariates. Three main areas of scientific research studies are presented in this PhD thesis and these are: ***Basic research:*** *'is the investigation of a subject to increase knowledge and understanding about it. The information gathered from basic science research is essential for 'translating' or applying new discoveries to patient care'* (Stevens *et al.*, 2002). ***Clinical research:*** *'clinical research is that component of medical and health research intended to produce knowledge valuable for understanding human disease, preventing and treating illness, and promoting health'. Clinical research involves interactions with patients, diagnostic clinical materials or data and include clinical knowledge, detection, diagnosis and natural history of disease;*

http://www.nap.edu/openbook.php?record_id=12983&page=67 Accessed 03rd Mar 2012.

Population research: Although the term population research is not precise with a widely agreed meaning (Haaga, 2001), *'A group of individuals of the same species occupying a particular geographic area. Populations may be relatively small and closed, as on an island or in a valley, or they may be more diffuse and without a clear boundary between them and a neighbouring population of the same species. For species that reproduce sexually, the members of a population interbreed either exclusively with members of their own population or, where populations intergrade,*

Chapter 1:10 Life expectancy is the statistical figure based on the average person's length of life and is usually quoted as the number of years of life remaining at a given age. http://stats.oecd.org/glossary/detail.asp?ID=1530

to a greater degree than with members of other populations.' (The American Heritage® Science Dictionary Copyright © 2010 by Houghton Mifflin Harcourt Publishing Company. Published by Houghton Mifflin Harcourt Publishing Company. All rights reserved) (Pearce, 1996; Pearce, 1999).

- ***PHD FRAMEWORK***

Research methods that are used and presented in this PhD thesis are comparable for both developed and developing countries, but the methods to capture reliable and high quality data may differ. Many developing countries do not have the advanced databases and IT infrastructure required for data collection. To produce evidence that will be accepted by policy makers and be sustainable, a variety of research methods are used for this. A recently described research cycle (Figure 1.2.4) (Beck *et al.*, 2008) is used as a framework for the three different studies that were undertaken during my PhD work and the aims of this work are to demonstrate how routinely collected de-identified HIV-1 patient information can be used for studying basic science (Chapters 3: LTNP), clinical science (Chapter 4: HIV and Ageing) and population science research (Chapter 5: Population Cost Study). Throughout the sections in this PhD I have used chapter labels: LTNP, HIV and Ageing and Population Cost Study as a reference to the three examples of research studies presented.

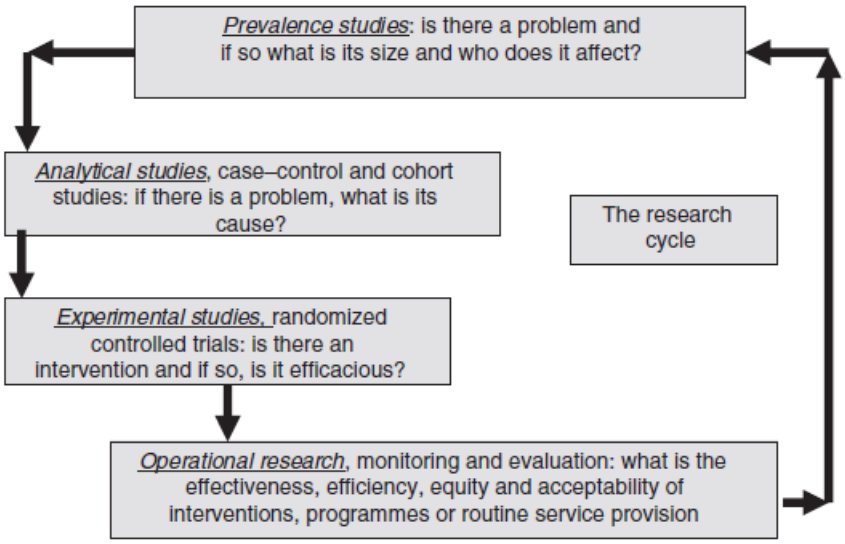


Figure 1.2.4: Framework of Research Cycle
Figure taken from (Beck *et al.*, 2008)

With reference to the research cycle (Figure 1.2.4), two analytic studies presented in this PhD are LTNP (Chapter 3) and HIV and Ageing (Chapter 4) while the Population Cost Study (Chapter 5) falls under operational research. ***Definition of analytic study:*** *is concerned with identifying or measuring the effects of risk factors or with the health effects of specific exposure(s) or interventions. Contrast descriptive study, which usually does not test hypotheses. In an analytical study, individuals in the study population may be classified according to the absence or presence (or future development) of specific disease and according to "attributes" that may influence disease*

occurrence. Attributes may include age; race; sex; other disease(s); genetic, biochemical, and physiological characteristics; social position, economic status; occupation; residence; and various aspects of the environment or personal behaviour. ***Definition of operational research:*** *The systematic study, by observation and experiment, of the working of a system (e.g., health services), with a view to improvement.* Definitions taken from Porta, 2008.

The first analytic study presented identified the prevalence of LTNP among HIV-1^+ people and then investigated which factors contributed to why disease progression did not occur as rapidly as does with most HIV-1^+ people. Once these HIV-1^+ people are identified as LTNP, they could potentially provide insights for HIV-1 preventive or therapeutic vaccine development, providing a useful model to study the mechanisms of viral control as a step towards development of prophylactic vaccines and immune therapeutic approaches. The second analytic study presented looked at the effect of HIV-1 infection on ageing HIV-1^+ people on ART and deteriorating renal function. The findings from this study also informs the current debate on the time when HIV-1 infected individuals should be started on ART. The Strategic Timing of Antiretroviral Treatment (START) trial, which is an experimental study that randomised HIV-1^+ people with $CD4^+$ T-cell count ≥500 cells/μL blood to immediate or deferred ART treatment when $CD4^+$ T-cell count dropped to below 350 cells/μL blood is due to report in 2014. The Population Costing study presented falls under operational research and estimated the population costs of HIV service provision in the UK 1997-2006 and predicted future population NHS costs 1997-2013 based on increasing numbers of HIV-1^+ people as they move from first- to second-, to third- and to forth- line ART.

1.3 THESIS AIMS AND STRUCTURE

Society now lives in information age and the volume of data that are collected are a result of advancing computerisation with availability of ever increasing powerful data collection computer systems and availability of large storage capacity. In medicine this has led to vast amounts of data being collected daily. Analysis of routinely collected data is an important need, and data mining provides tools for building on existing knowledge as well as discovery of new knowledge and increased awareness. This is achieved through collection of good quality data from which new trends can be found which may provide missing information to advance existing knowledge (see section 1.1: Figure 1.1.9).

There is an ongoing need for personal health information to be collected longitudinally to ensure development of good electronic medical and health records locally, nationally and globally requiring appropriate IT as well as analytic expertise.

CHAPTER 1

GENERAL INTRODUCTION

With the development of universal healthcare systems, the need is to ascertain the effectiveness (the outcome or impact of programmes or services), efficiency (focuses on the resources required to achieve a certain outcome or impact), equity (who benefits from programmes or services) and acceptability (acceptability of programmes or services to users and providers and to the improvements in quality of life they achieve) of services provided. In the UK there is a requirement that good individual patient data are collected so these can be used: optimally for NHS use of services by individuals; to ascertain the effectiveness, efficiency, equity and acceptability of the services provided.

In order to collect such data, databases need to be developed at local hospitals and also as part of the HIV response to HIV information systems. In 2005, UNAIDS published a report entitled *'The Three Ones in Action: Where We Are and Where We Go from Here'* (UNAIDS, 2005) that was developed within the response to HIV information systems within the context of the short-term emergency response to the country's HIV epidemic.

In response to this 'three ones principle', a good database need to be put in place that collects good quality data at a specialist hospital as well as develop databases as part of the HIV response to collect HIV information. The three ones principle was offered to countries in order that they can optimise the roles and relationships in the fight against HIV.

- ***THREE ONES PRINCIPLES***

1. One: Agreed HIV/AIDS action framework that provides the basis for co-ordinating the work of all partners.
2. One: National AIDS Coordinating Authority, with a broad based multi-sector mandate.
3. One: Agreed country level Monitoring and Evaluation System.

(UNAIDS, 2004)

Of the three ones principle, the 'Third One' was developed within the context of the short-term, emergency response to the country's HIV-1 epidemic. With the ageing of HIV-1^+ population requiring treatment and care from other specialists, further development of monitoring and evaluation systems are required and these are needed to be expanded to other clinical specialist areas.

1.4 PRESENTED STUDIES

Since HIV-1^+ patients are seeking long-term medical care there are many specialist NHS hospitals in the UK with historical HIV databases. Analysing such data is important to discover knowledge from collected data. Therefore the availability of HIV-1 specific database with historical data both locally and nationally require data mining methods to be used to establish the patterns in data using statistical analyses.

CHAPTER 1

GENERAL INTRODUCTION

Some HIV-1$^+$ individuals are able to maintain stable CD4$^+$ T-cell count within the normal healthy range between 450-1650 cells/μL blood; local laboratory reference range (Westrop *et al.*, 2009) for a prolonged length of time and do not exhibit signs of HIV related opportunistic infections such as bacterial pneumonia, cytomegalovirus and other infections (Castro *et al.*, 1992), despite not receiving ART (Chapter 3: LTNP). The patho-physiological changes observed in HIV-1$^+$ patients are similar to those seen in older people not infected with HIV-1, and these changes are characterised by ongoing inflammation, immune activation and increased coagulability (Kuller *et al.*, 2008; Tien *et al.*, 2010; Eastburn *et al.*, 2011), comparable to changes associated with 'immunosenescence' (Gress & Deeks, 2009). Furthermore, being treated long-term on ART, together with ageing, is a complex 'ageing' process involving biological, social and psychological aspects. Identifying and investigating the interacting effects of age and HIV-1 infection and ART is likely to guide improved therapeutic intervention in HIV-1$^+$ people with advancing age (Chapter 4: HIV and Ageing).

HAART, which typically consists of at least 3 or more ARV drugs, is used to keep control of the HIV-1 RNA plasma load in the blood. Tests are used to establish how much virus is circulating in HIV-1$^+$ person's plasma that allow monitoring effectiveness of the ARV drugs. Monitoring patients on treatment, both in terms of effectiveness and also efficiency, could help distribution of limited financial resources where money is needed for managing HIV-1$^+$ people, and needs to be spent in the NHS (Chapter 5: Population Cost Study).

The three results chapters presented in this PhD thesis together are themed in terms of the use of cohort databases, the data management and the use of the data stored and used both from a programming and statistical point of view. The three results chapters (Chapters 3, 4 and 5) use observational databases to evaluate how long-term follow-up of HIV-1 infected patients can be evaluated, giving further insights about their disease course, and the use and cost of HIV-1 services and use three research areas of scientific research studies: basic science, clinical science, and population science.

This chapter introduced data management, a statement of the problem using HIV-1 infection as an example. Objectives of the HIV-1 research, the analytical framework used to achieve these aims and the methods used to collect data. Chapter 2 describes the analytical methods used in the next three chapters. Work presented in: Chapter 3: LTNP; Chapter 5: Population Cost Study have appeared or will shortly appear (Chapter 4: HIV and Ageing) as papers in peer reviewed scientific journals and findings from these papers are included in this PhD thesis with slight modifications. References of these published papers can be found in the section 'publications arising from this thesis'. Because it was intended that each chapter could be read by itself, some overlap between chapters is unavoidable. The concluding Chapter 6 of this PhD thesis highlights novel results found

using statistical analysis applied in previous three chapters (Chapters 3, 4 and 5) and a general discussions about clinical database developments within the NHS and recommendations.

Main lesson from this PhD is that de-identified individual level patient information can be used at local, regional and national levels to monitor and evaluate the effectiveness, efficiency, equity and acceptability of HIV-1 and other health services provided.

Chapter 2: Materials and Methods

CHAPTER 2: MATERIALS AND METHODS

2.1 OVERVIEW (for all three studies)

- ***QUANTITATIVE DATA***

Quantitative data is any category of information or outcome/result that is expressed in a numerical form such as statistics, proportions, and/or percentages. Data collected can be used to answer specific questions raised using different types of research methods such as basic, clinical and population science research (see Chapter 1: section 1.2 for further details on these). The objective of quantitative research is to develop hypotheses that are related to some occurrence because it provides relationships between an observation and statistical summary of quantitative associations. The whole process of measuring this occurrence is central to quantitative research because it provides relationships between an observation and statistical summary of quantitative associations. Data are analysed using statistics and the ultimate aim is to derive an unbiased estimate that could be used to generalise into a defined population. In Medicine and Economics, quantitative research is used widely and quantitative methods can be used to verify or refute hypotheses. The quantitative research (hypothesising and generalising) methods are considered using data analyses where quantitative data are used to answer questions such as 'how many?' or 'how frequently', which are reported numerically and allow categorisation of set of amalgamated data, statistical analysis and mathematical modelling.

- ***STRENGTH AND WEAKNESSES OF RESEARCH DESIGN***

There are a number of different methodologies used to conduct research data analyses and in particular observational study is described and categorised with strengths and weaknesses (Table 2.1.1). The research questions determined the type of methodology that is used to carry out the research. In this PhD thesis the methods used within the studies presented in Chapters 3, 4 and 5 are described.

Table 2.1.1: Methodology Strengths and Weakness

Methodology	Weakness	Strengths
Observational research *(quantitative, epidemiological methods)* *Correlational studies as a means of looking for relationship between variables when experiments cannot be done* Includes longitudinal data collection	Cohorts can be difficult to identify due to confounding variables No randomisation therefore this could mean that imbalances in patient characteristics could exist Outcome could take time to occur Cannot draw conclusions about causality Costly and subjects may drop out or lost to follow up over time While it is possible to study how individuals change using longitudinal designs, it is more difficult to understand why they change as they do	Cheaper than a randomised controlled trial Standardisation of outcome is possible Subjects in cohorts could be matched which minimises the effect of confounding variables Can be used to examine complex relations among variables Can examine changes in variables over time Longitudinal designs control for cohort differences

- ***OBSERVATIONAL STUDY***

One of first observational longitudinal study was designed in the 1950's to identify risk factors for cardiovascular disease (Dawber *et al.*, 1951). Observational longitudinal studies have since been undertaken to detect rare or late adverse effects of treatments, and provide an indication of what is achieved in daily medical practice. Evidence regarding adverse effects of therapy is undertaken after larger, post-market, observational studies have been performed (Papanikolaou *et al.*, 2006). Observational studies are useful for demonstrating the benefits and harms of medical interventions (Black, 1996). Recommendations for the reporting of observational research was developed by a network of methodologists, researchers, and journal editors and is the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement (VonElm *et al.*, 2007) and more recently start of *'the REporting of studies Conducted using Observational Routinely-collected Data'* (RECORD[11]) initiative to develop an extension of the STROBE statement.

- ***OBSERVATIONAL STUDY LIMITATION***

Observational study is conducted by: identifying a group of subjects who are at risk of a condition being investigated; determining the exposure status of each individual; and assessing the individuals over time for the occurrence of a condition. Usually this involves a comparison of the outcomes for the two (or more) groups being investigated. However, observational studies are subject to bias. Bias is defined as *"Systematic deviation of results or inferences from truth. Processes leading to such deviation. An error in the conception and design of a study - or in the*

Chapter 2:11 http://www.equator-network.org/resource-centre/library-of-health-research-reporting/reporting-guidelines-under-development/reporting-guidelines-under-development/#10
www.record-statement.org

collection, analysis, interpretation, reporting, publication, or review of data - leading to results or conclusions that are systematically (as opposed to randomly) different from truth" (Porta, 2008). Different sources of bias have been classified into three broad and overlapping groups: the *selection bias*, *information bias*, and *confounding* between the risk factor and other possible factors (Miettinen & Cook, 1981; Delgado-Rodríguez & Llorca, 2004). Selection bias refers to systematic differences in the characteristics between those who are selected for study and those who are not. In any study, a selection process can lead to external bias if the selected study participants are unrepresentative. Information bias is related to the collection of data about the exposure or the outcome. Various forms of information bias derive from errors in measuring accurately. Confounding is an intertwining of the causal effect of the exposure with causal effects attributable to other factors. A potential confounding factor is a variable that is thought to affect the outcome independently of the treatment for instance.

The main aim of research findings is to demonstrate the presence or lack of an association (eg, between an exposure and disease) or differences in parameter estimates (eg, means, standard deviations, proportions) between populations. Associations or differences may be real, may occur by chance (sampling error), or may be related to other factors such as bias. Bias when present introduce systematic errors that decrease the validity of estimates presented. Biases are present and are difficult to completely eliminate however, attempts are made to reduce this as far as possible so that the results presented have greater generalisability. Many types of bias have been described, with overlapping terminology (Miettinen & Cook, 1981; Delgado-Rodríguez & Llorca, 2004). Some bias (eg, confounding) can be adjusted for in the statistical analysis. Uncorrected bias have an effect on validity and can lead to data misinterpretation and limit the applicability or generalisability of a study. The statistical methods described in this chapter have attempted to address sources of bias and in the analytical approach undertaken. The main limitation of an observational study is unadjusted confounding factors which may be due to unknown and therefore unidentified which may therefore distort the results. Observational data have been used in this PhD thesis to exploit rich single and multi-centre HIV-1 databases.

2.2 STUDY POPULATION (for all three studies)

The two main datasets used for data analyses in this PhD thesis are derived from routinely collected data from the NHS hospitals. These are on HIV-1^+ people attending for care within the specialist centres within the NHS. The sources of data that were used were: 1) data from the Chelsea and Westminster HIV cohort which covers routinely collected data on all patients who ever attended for specialist care since 1988 up to February 2010 Chapter 3 (LTNP) and up to May 2011 used in Chapter 4 (HIV and Ageing). The routinely collected HIV-1 cohort data presented in Chapters 3 and 4 are updated annually on average. Data presented in Chapter 3 were based on data extracted from

the local data warehouse in March 2010 while data presented in Chapter 4 are based on data extracted from the data warehouse in June 2011; 2) data from a multi-centre cohort of HIV-1 infected patients attending for routine care within the participating sites and cover data between 1st January 1996-31st December 2006 and these have been presented in Chapter 5. Data collection and work on this data commenced from middle of 2007 and data collection completed March 2008.

2.3 ETHICS STATEMENT (for all three studies)

The Singapore Statement on Research Integrity (2010) for ethical consideration comprises four principles and 14 responsibilities. The four principles aim to balance the rights, interests, and duties of participating patients, volunteers, industry representatives, and medical scientists fairly[12]. There are four broad contexts to ethics principles that are required for research and these are: 1) autonomy (respect for individual rights) - informed consent, information, understanding, consent. 2) Beneficence (do good) - working with personal data, data documentation (the research protocol), storage of data, publication, exercise of judgement, scientific misconduct. 3) non-maleficence (do no harm). 4) Justice.

Research work is required as people have the right to know about exposure of risk to their health and to make evidence-based informed choices regarding treatment and prevention. From an ethical point of view, it is essential that research must also be of good quality. Bad research may lead to wrong decisions that may have an impact on people's health and well-being.

Data analyses in Chapter 3 (LTNP) had ethics approval. Data analyses were conducted on anonymised data and this study was approved by the Riverside Research Ethics Committee, now known as London Riverside National Research Ethics Service (NRES) Committee. This work was carried out under the ethical approval of the Riverside Research Ethics Committee (RREC1108). Chapter 4 (HIV and ageing) data analyses was conducted on routinely collected clinical management data and thus all treatments that the patients were prescribed were based on physician-client interaction and based on British HIV Association (BHIVA) guidelines and no random allocation of treatment was used. Based on these national BHIVA guidelines operative in the UK there is no requirement to seek approval from an ethics board or institutional review board (IRB) provided implementation of NHS confidentiality, security and privacy guidelines are upheld. The data presented in Chapter 4 (HIV and Ageing) were thus based on anonymised patient data. Data analyses resulting from the data collected by NPMS-HHC participating sites are described as

Chapter 2:12 Singapore Statement on Research Integrity 2nd World Conference on Research Integrity, July 21–24, 2010. Available from: http://www.singaporestatement.org/statement.html. Accessed Oct 23, 2012.

a service evaluation[13] (Newton *et al.*, 2000). The NPMS-HHC cohort therefore does not require informed consent and/or ethics approval as the project collected anonymised data and has had approval by National Information Governance Board for Health and Social Care (NIGB) formally known as Patient Information Advisory Group (PIAG). One of its functions was to allow the common law duty of confidentiality to be set aside in specific circumstances. However, with the changes that took effect within the NHS from 1st April 2013 (see Chapter 1: section 1.1 NHS England from April 2013), the NIGB's functions for monitoring and improving information governance practice was transferred to the Care Quality Commission, which is currently (May 2013) establishing a National Information Governance Committee (NIGC) to oversee this work. The policy aims at maintaining the patient confidentiality and security and of HIV information which the international guidelines have reinforced the appropriateness of such procedures (UNAIDS, 2007).

2.4 STUDY SITE, SAMPLING AND COHORT DESIGN (for the two analytic studies)

- ***CONFIDENTIALITY***

All data extracted from the local database were ensured that confidentiality[14] and security of patient data were maintained. Data were kept in de-identified or pseudo-anonymised format. De-identified ensured that the data were totally anonymised, all personal identifiers (see Table 2.4.1) and other identifying information had been stripped and data could no longer be linked to the original source of the information. Pseudo-anonymised meant that data were stored in a form stripped from identifiers, but could be traced back to the original source through a key. The key for this was held at a centre where the data originated from.

Chapter 2:13 '*a set of procedures to judge merit by providing a systematic assessment of its aims, objectives, activities, outputs, outcomes, and costs'. Evaluation provides practical information to help decide whether a development or service should be continued or not. Evaluation also involves making judgments about the value of what is being evaluated.*

Chapter 2:14 Confidentiality is the principle in medical ethics that the information a patient reveals to a health care provider is private and has limits on how and when it can be disclosed to a third party.
http://medical-dictionary.thefreedictionary.com/confidentiality

Table 2.4.1: A List Through Which Persons Could be Identified in Their Own Right or in Combination

Name: first, middle and last name
Address
Full postal code
Telephone number
Fax number
Email address
Date of birth
Sex
NHS Number
Ethnicity
Social security number
National Insurance number
Occupation
Employer information
Photographs

- ***GEOGRAPHICAL LOCATION***

The Chelsea and Westminster HIV cohort comprises data based on three HIV clinics located in London within easy reach and these are: Kobler clinic (based at Chelsea and Westminster Hospital); West London Centre for Sexual Health (based at Charing Cross Hospital); 56 Dean Street (located in the Soho).

- ***DATA COLLECTION INSTRUMENT***

Data from the local database were extracted by IT trained staff from the local data warehouse annually in either ASCII[15] delimited text files or Microsoft Excel comma-separated values (CSV) files. Figure 2.4.1 demonstrates the flow of patient data to hospital data warehouse. The HIV cohort data are extracted periodically, usually annually from the Chelsea and Westminster Hospital Foundation Trust Oracle data warehouse and a broad list of data are listed in Table 2.4.2.

Chapter 2:15 ASCII stands for American Standard Code for Information Interchange codes and represent text in computers communications equipment, and other devices that use text. http://www.ascii-code.com/

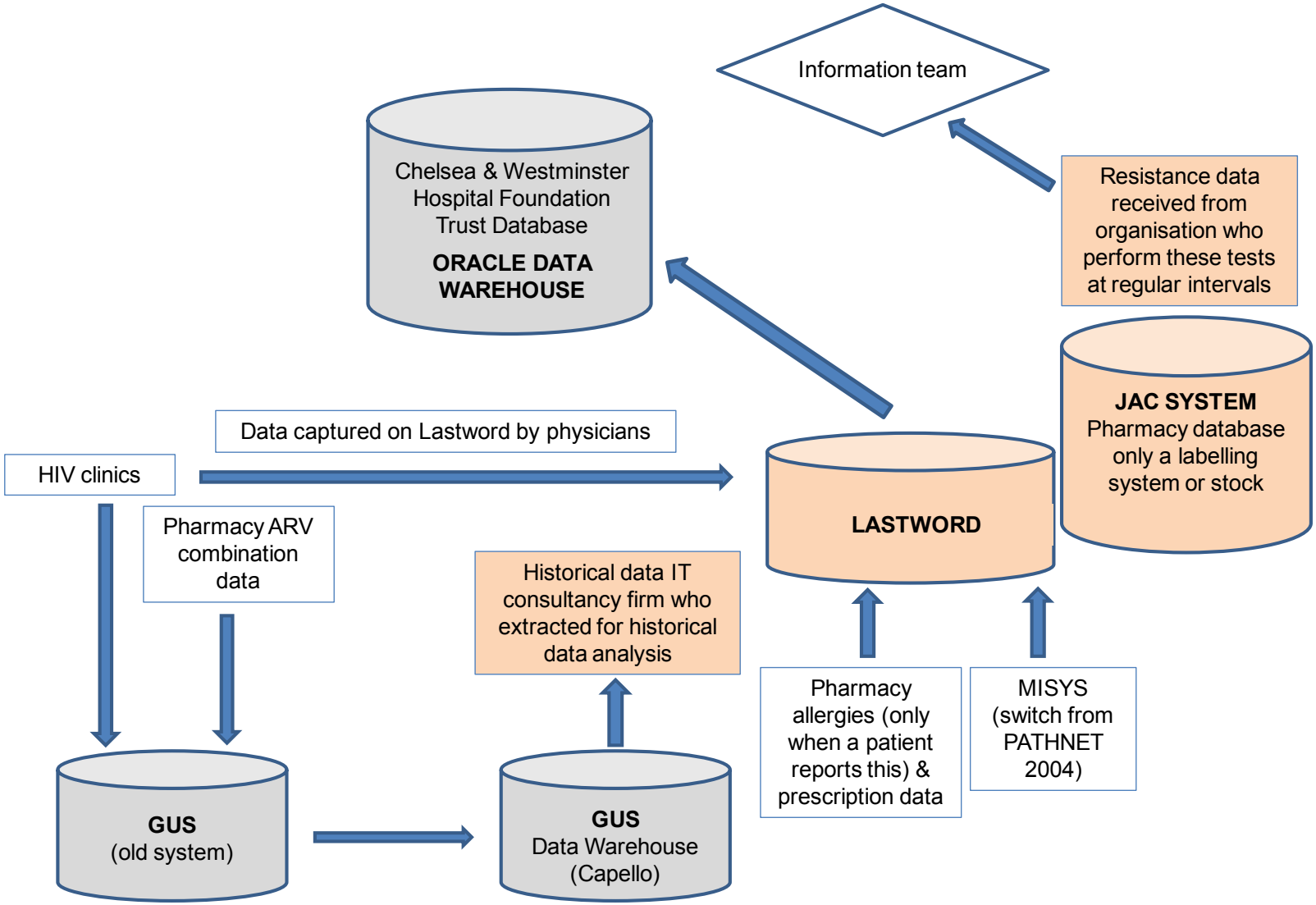


Figure 2.4.1: Flow of Data Transfer to Trust Data Warehouse[16]

Chapter 2:16 A data warehouse is storage of information collected from multiple sources, stored under a unified schema, and usually residing at a single site.

- ***DATA TRANSFER***

Once authorised, the data are transferred for statistical analyses and this occurs, on average, annually (Figure 2.4.2) with appropriate level of security, in accordance with the local NHS Trust IT policy, in place. Security measures required include data encryption and data encryption software such as the AXCrypt, a web-based freeware software http://www.axantum.com is one such encryption software that is used locally to encrypt data before they are transferred for research and observational cohort data analyses (Figure 2.4.2). All data extracted are either de-identified or pseudo-anonymised.

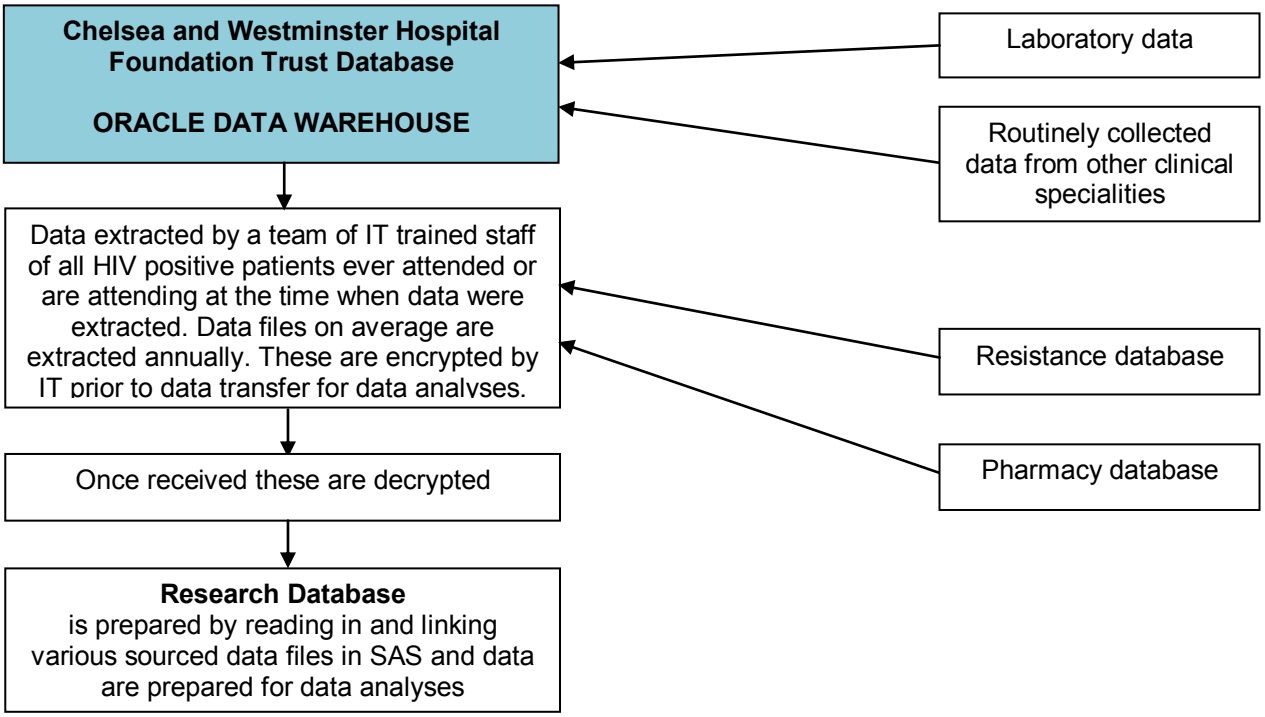


Figure 2.4.2: Flow of Data Transfer to Research Database

- ***STUDY DESIGN***

Chelsea and Westminster HIV cohort is based on routinely collected prospective data where HIV-1 diagnosed patients are followed up routinely over time and a variety of variables related to demographic characteristic, results of tests performed, diagnoses, treatments prescribed over a period of time during which patients are under follow up are collected (Table 2.4.2).

Table 2.4.2: A List Showing the Type of Data Items Available From Local Database for Research and Cohort Data Analyses

Research Database
Demographic data: date of birth, date of death, gender, ethnicity
HIV data: risk factors (sexual orientation, injecting drug use), year of HIV diagnosis, HIV diagnoses (eg opportunistic infections)
Non-HIV diagnoses
Use of hospital service data: outpatient visits (for HIV clinics), inpatient stay (including discharge dates) and day ward visits
Treatment data: Drugs prescribed for HIV and for non-HIV related conditions and drug allergies (including the nature of the reaction)
Test Results: results of all routine clinical investigations are stored including - Routine haematology - Biochemistry (liver, renal, bone profiles; lipids, glucose, CRP, urinalysis) - Hepatitis B/C serology - Toxoplasma, CMV and syphilis serology - Immunology including CD4 cell count & %, auto-antibodies - HIV-RNA levels - Genotypic and, if available, phenotypic HIV resistance testing - Co-receptor use (phenotypic and genotypic assays) - HLA typing - Other tests as required (eg haematinics, vitamin D, tumour markers) - Imaging (X-ray, CT, Ultrasound, MRI, DEXA) - Specialist investigations (PET, nerve conduction studies, endoscopy) - Microbiology (cultures, sensitivities, virus isolation) - Histology and cytology (including cervical screening) where appropriate

- ***DATA COLLECTED***

Data were requested from the IT personnel to provide all historical HIV-1^+ patient data since first attendance at clinic to the most recent time available as broadly listed in Table 2.4.2. Data collected reflected the remit of the local cohort data analyses work. The data are provided in an electronic format and transferred for data analyses in accordance with the local data protection policy where data are pseudo-anonymised.

- ***DATA LINKING AT PATIENT LEVEL***

Over time the source database from which HIV cohort data were extracted changed from Genomics Unified Schema (GUS) to Oracle Data Warehouse (Figure 2.4.1). While data on GUS were managed by a unique patient identifier (pat_ref), historically due to attendances at three HIV clinics affiliated to Chelsea and Westminster Hospital (further details are under section 2.4.1), individual patients were assigned multiple hospital identifiers (HOSPIDs') and these were linked using pat_ref. The switch over from the GUS database system to Lastword database occurred in 2008/2009 and multiple patient identifiers are no longer assigned to new HIV-1 diagnosed patients or a new attendance at the local institution, instead a unique patient identifier is assigned: the MRN number.

The HIV cohort data extracted are for all HIV-1^+ patients who ever attended for care: historical as well as newly diagnosed patients. The data linking of various extracted data files required

development of data linking process. All PAT_REF unique identifiers were extracted from the old GUS database system together with multiple hospital identifiers associated with the unique PAT_REF. In addition all HOSPIDS held under new unique identifier MRN number on Lastword database were extracted for all patient records held. All new patients who attended the HIV centre since the switchover from GUS to Lastword database in 2008 were assigned a unique MRN number on Lastword database while cohort of patients who continued to be followed up continued to use their existing hospital identifier. The ID file from GUS was used as a base to derive a unique identifier data file that amalgamated identifiers extracted from both GUS and Lastword databases.

For example:

pat_ref	hospid
0000100023	99123456
0000100023	99910111
0000100023	99141516
0000100023	99181920
0000100023	99222324
0000100023	99262728

pat_ref	hospid1	HospID2	HospID3	HospID4	HospID5	HospID6
0000100023	99123456	99910111	99141516	99181920	99222324	99262728

Note: the above identifiers are made up and do not identify a patient.

All data received were linked using a unique identifier, PAT_REF. St Stephen's AIDS Trust (SSAT) is a registered data controller with the Information Commissioners Office[17] and any personal information held and used are in accordance with the Data Protection Act 1998[18]. There are eight broad principles to this act and these are:

- Processing personal data fairly and lawfully (Principle 1);
- Processing personal data for specified purposes (Principle 2);
- Information standards
  - The amount of personal data you may hold (Principle 3);
  - Keeping personal data accurate and up to date (Principle 4);
  - Retaining personal data (Principle 5);
- The rights of individuals (Principle 6);
- Information security (Principle 7);
- Sending personal data outside the European Economic Area (Principle 8)

- ***DATA QUERIES AND RESOLUTION***

Once the data were linked these were reviewed by eye in table format in statistical package Statistical Analysis Softwwere (SAS; SAS Institute Inc., Cary, NC, USA) so that data errors such as incorrect input of decimal number, spelling and punctuations could be corrected in accordance with the International Conference on Harmonisation (ICH) of Technical Requirements for Registration of

Chapter 2:17 The Information Commissioner's Office is the Government organisation that enforces the Data Protection Act 1998 and subordinate legislation.
Chapter 2:18 Data Protection Act 1998 principles taken from http://ico.org.uk/for_organisations/data_protection/the_guide

Pharmaceuticals for Human Use Guideline for Quality Assurance and Quality Control (Manghani, 2011). Any anomalies found were reported to the team of personnel responsible for quality control of HIV cohort data.

- ***QUALITY CONTROL PROCESS***

Once all data were linked queries were raised on any data that were not linked and these were resolved. In addition patient records were checked between files, for instance number of patients in the demographics data file needed to tally with other data files and *vice versa*. Where discrepancies were found these were raised and resolved. Some of the pre-planned queries included:

- To test whether numbers of patients found in different files were in agreement. Where this was not the case queries will be raised and resolved at source clinic.
- By checking consistencies in data codes found in each data variable provided.
- Data captured on ARV were checked for consistencies in start and end date.
- All laboratory data such as $CD4^+$ T-cell counts and HIV-1 RNA plasma load data were checked for consistency regarding date of result and study time point as well as plausibility of data in terms of availability of data and also missingness.

Once queries were generated and resolved these were then documented and data changes made using SAS program.

- ***DUPLICATE PATIENT DATA***

All duplicate records were removed prior to commencing data analyses. This includes duplicate data checks within each individual file received before merging to form a large linked file. After all queries were resolved the data were then prepared using SAS programming to structure these for statistical analyses. At each stage of data restructuring consistency checks in data were performed.

2.5 LONG-TERM NON-PROGRESSOR ANALYSES

The rate of disease progression from asymptomatic HIV-1 (see Chapter 1: Figure 1.2.2) infection to AIDS varies between HIV-1^+ individuals (Pantaleo & Fauci, 1996). Establishing factors that were associated with HIV-1 disease progression aims to enable development of therapeutic strategies to delay or reverse disease progression. The gradual decline in peripheral blood $CD4^+$ T-cell count which occurs throughout the typical course of HIV-1 infection and is accompanied by excessive activation of the immune system and the sequential loss of immune responses, first to HIV-1, then to other pathogens, allogeneic and finally mitogenic stimuli (Clerici *et al.*, 1989; Douek *et al.*, 2009).

AIDS defining opportunistic infections were originally used to indicate disease progression in the clinic, however in 1993 the definition of HIV disease progression was widened to also include a

drop in $CD4^+$ T-cell count to <200 cells/μL blood or <14% of lymphocytes (Castro *et al.*, 1992; MMWR, 1992).

Historically some people living with HIV-1 remain healthy with $CD4^+$ T-cell counts more than 500 cells/μL blood for 10 years or more in the absence of ART (Klein & Miedema, 1995) and these were referred to as long-term survivors. In addition, in the absence of antiretroviral therapy they show no signs of opportunistic infection and remained asymptomatic. These groups of HIV-1^+ patients have been referred to as LTNP. However these terms and definitions used to identify these patient groups were not uniform between studies and sites.

Terms used to describe this unique and atypical group of HIV-1^+ people were not uniform across published studies in the UK and internationally. Although studies to date, have identified only a small number of factors that were consistently associated with LTNP status (Vento *et al.*, 2004). Establishment, definition and characterisation of HIV-1^+ LTNP patient cohort could be used to identify factors that were associated with non progression status, and therefore aid in determining potentially successful prophylactic vaccines and novel therapeutic approaches.

2.5.1 COHORT SELECTION CRITERIA

- ***STUDY POPULATION***

Chelsea and Westminster HIV cohort is a large HIV cohort and is one of the largest single centre HIV cohorts in Europe. Routine data were prospectively collected at routine visits when HIV-1^+ people were seen at emergency or regular intervals for assessment, follow-up and immunologic, virologic and clinical assessments. The study population used for data presented in Chapter 3 (LTNP) of this PhD thesis included all patients who had attended for HIV care since start of HIV-1 patient data collection commenced from January 1988. Data were extracted from the hospital Trust data warehouse on 28th February 2010 and therefore data presented in Chapter 3 were based on data collected until this date. Therefore last data collected were based on the data collected at the patients last recorded visit at the clinic for HIV-1 care. The entire HIV-1 cohort that was available was used to screen and identify LTNP, HIV controllers (HIC), patients with long-term stable low $CD4^+$ T-cell counts, and amongst these a subset of patients who also controlled HIV-1 RNA plasma load to below the limit of detection (BLD) as detailed in the following sections.

- ***LONG-TERM NON-PROGRESSORS***

The Chelsea and Westminster HIV Cohort LTNP selection criteria were applied to identify patients firstly with documented HIV-1 infection for greater than seven years (Pantaleo & Fauci, 1996). The duration of HIV-1 infection was derived from the date of first positive HIV-1 test result or, if this was unavailable prior to enrolment into the local HIV cohort, then the earliest clinic visit date was used. The second criterion for patient selection was never receiving ART (Imami *et al.*, 2002; Westrop *et*

al., 2009). Thirdly, these patients were screened for history of opportunistic infection (OI). OI was defined as any symptomatic manifestation of HIV-1 illness recorded in the medical record. Entries of 'acute infection', 'persistent generalized lymphadenopathy (PGL)', and 'asymptomatic infection' were not defined as OI. The fourth selection criterion used was 'patients who had a stable $CD4^+$ T-cell count'. This was defined as a $CD4^+$ T-cell count slope of ≥0 cells/μL blood from entry into cohort up to and until the most recent available $CD4^+$ T-cell count (Mandalia *et al.*, 2012). Analyses were performed by 3 monthly grouped time periods because, on average, clinically stable HIV-1^+ patients were seen at quarterly intervals at the local institution where data originated from.

- ***HIV CONTROLLERS***

Within the group of individuals classified as LTNP, a minority also control HIV-1 replication to BLD of the conventional HIV-1 RNA plasma load assay (50 copies/mL plasma or histocally, 500 copies/mL was used at the time when sensitive RNA plasma load assays were not available). These individuals were selected and have been known as HIC (Westrop *et al.*, 2009).

- ***HIV NON-PROGRESSORS USING CONTROL OF VIRAL LOAD CRITERIA***

Recent review suggested that due to the low prevalence of HIV non-progressors patients need not be infected for a long period of time. Therefore two low level HIV-1 RNA plasma load measurements used to select these atypical patients would be sufficient (Gaardbo *et al.*, 2012). The HIV cohort data were therefore further investigated using relaxed selection criteria to identify atypical patient group from the cohort. Measurements of low level HIV-1 RNA plasma load (<500 copies/mL plasma in the absence of ART at the time of HIV-1 RNA plasma load assay) were initially used to identify and select patients. Further analyses were performed to identify individuals who fulfilled the atypical patient group selection criteria. Firstly, three consecutive HIV-1 RNA plasma load results of <500 copies/mL plasma within one year were chosen to allow for: potential HIV-1 RNA plasma load assay failure; individuals who missed attending their routine follow-up HIV-1 care visit at the Chelsea and Westminster Hospital; subsequent HIV-1 RNA plasma load results which may occur at very close time points. The data were further refined to ensure that the three HIV-1 RNA plasma load results spanned for at least one year. This length of time was selected as on average, HIV-1^+ patients who were stable were seen in clinic at quarterly intervals which would equate to 4 routine clinic visits per year. A conservative estimate of this (to enable rescheduling/missed appointment) is estimated at three visits per year.

2.5.2 STATISTICAL METHODS

- ***NON-PARAMETRIC METHOD***

The Mann-Whitney U test was used for between group comparisons of non-parametric data while qualitative data are presented as numbers with percentages and compared using the chi-squared test statistic and where appropriate Yates' correction was applied.

- ***TIME TO PROGRESSION: HIV-1$^+$ LONGER THAN 7 YEARS AND NO ART N=1,204***

Survival analysis was used to estimate time from HIV-1$^+$ diagnosis until disease progression in a group of patients who had been diagnosed HIV-1$^+$ for longer than 7 years, and who had not been prescribed ART (n=1,204). HIV-1 progression in this group was defined as presentation with symptomatic or AIDS defining opportunistic illness or declining CD4$^+$ T-cell count during follow-up (see Chapter 3: Figure 3.2.1).

- ***TIME TO PROGRESSION: HIV-1$^+$ LONGER THAN 7 YEARS, NO ART AND ASYMPTOMATIC N=312***

Time from HIV-1$^+$ diagnosis until HIV-1 disease progression was estimated in 312 patients, a subset of the initial 1,204, who had been diagnosed HIV-1$^+$ for longer than 7 years, had not been prescribed ART and who remained asymptomatic. Progression in this group was assessed through evidence of unstable CD4$^+$ T-cell counts during follow up, or death.

Joint probabilities of non-progression for more than 7 years, and remaining ART-naïve, in patients who fulfilled criteria of LTNP but eventually exhibited HIV-1 disease progression, was used to estimate progression time. This was estimated from the date of HIV-1$^+$ diagnosis until each of the defined criteria indicative of HIV-1 disease progression. This was defined and included: 1) presentation with symptomatic or AIDS defining opportunistic illness or declining CD4$^+$ T-cell count during follow-up; 2) evidence of unstable CD4$^+$ T-cell counts during follow up, or death. All subjects whose CD4 count was stable since cohort entry were assumed to have either no change or an increase in CD4 count since entry to cohort indicated by a zero or positive slope whilst those patients whose CD4 count slope was negative since cohort entry were assumed to have an unstable CD4 count over time and were therefore removed from further analyses. Data were censored at the most recent visit to the clinic. The progression times were estimated and presented stratified by groups of patients whose CD4$^+$ T-cell count either remained within, or was outside on at least one occasion, the normal range.

- ***LONGITUDINAL METHOD***

Linear mixed models, using PROC MIXED in SAS are an extension of the general linear model. This method is an extension of techniques for estimating parameters (means, variances, regression coefficients, and standard errors), and are much more flexible. In repeated measures

data collection and longitudinal studies, the observations are clustered within a subject. This means the observations, and their residuals, are not independent but correlated. There are two approaches taken to deal with this correlation using either: the marginal model or the mixed model.

- ***THE MARGINAL MODEL (POPULATION AVERAGED MODEL OR GENERALISED ESTIMATION EQUATION)***

This model alters the co-variance structure of the residuals, this means that instead of assuming that all observations are independent, as is the case when using a general linear model, the marginal model assumes the residuals from a single subject are related and their co-variances are non zero. This model therefore requires an estimate of the co-variances among all the residuals from a single subject. The difference between the marginal model and a linear model is that the residuals are not assumed to be independent with constant variance. So in cases where the assumptions of equal variances and equal correlations are not met, a much better fit of models are derived by using a marginal model. A covariance matrix (covariance structure) used in statistics is symmetric. This means all correlations on the diagonal of a matrix are 1, because they are the correlation of each variable with itself:

Covariance matrix		
Covariance 1,1= **Variance 1**	Covariance 1,2	Covariance 1,3
Covariance 1,2	*Covariance 2,2*= **Variance 2**	Covariance 2,3
Covariance 1,3	Covariance 2,3	*Covariance 3,3*= **Variance 3**

Covariance is a non standardised version of correlation. A correlation is derived by dividing covariance by the standard deviation of both variables to remove units of measurement. So a covariance is just a correlation measured in the units of the original variables. Unlike the correlation value which lies between -1 and 1, covariance is not constrained to lie between -1 and 1. However, the sign remains the same, and a covariance value of 0 has the same meaning as a correlation value of 0 i.e. no linear relationship. The diagonal variables are the variances of each variable: a covariance of a variable with itself is simply the variance. The covariance structure that is specified in SAS MIXED procedure includes the following:

- Compound Symmetry: compound symmetry structure means that all the variances and all the covariances are equal. This is used when all variables in the model are measured on the same scale.

- Variance Components: this assumes each variance is different, and all covariances are 0. Used when all variables in the model are measured on the different scale.

- Unstructured: This assumes that there is no pattern in covariance structure. Each variance and each covariance is completely different and has no relation to the others.

- ***THE MIXED MODEL***

Another method used to deal with residuals that are not independent is by controlling for subject, by using it as a factor in the model. In this model the residuals are a distance between a data point and the mean of that subject and this is achieved by adding subject as a random factor in the model. In addition MIXED model also allows to control for time varying (such as $CD4^+$ T-cell counts, renal function results measured over time on the same subject) as well as variables that are not time varying (such as baseline demographics e.g. gender) by including these as fixed variables in the model. The resultant estimates from the model gives mean of for instance $CD4^+$ T-cell counts over time adjusted for baseline variables such as gender. A mixed model contains both the fixed and the random factors.

The main feature of longitudinal studies is that the patients are followed up over time and various immunological and clinical parameters together with age of a patient are measured repeatedly. Multiple measurements are obtained from each patient at different times and with changing conditions and therefore there may be a large amount of variation among patients in the number and timing of observations. One problem with data collected longitudinally is that the intended measurements are not taken for reasons such as patient lost to follow up when patients exit the cohort prematurely and unknown to the clinic or for reasons unknown, data are unavailable. The missing data process can be classified to three different categories (Rubin, 1976): 1) missing completely at random (MCAR) where the probability of the missingness does not depend on the measurements; 2) missing at random (MAR) if the probability of missingness does not depend on the unobserved measurements; and 3) when observations are neither MCAR nor MAR, they are classed as missing not at random where the probability of missingness depends on unobserved measurements and may be on the observed measurements. For example, it may happen that after a series of visits with good outcome, a patient drops out due to lack of efficacy. In this situation the analysis model based on the observed data, including relevant covariates, is likely to continue to predict a good outcome, but it is unreasonable to expect the patient to continue to derive benefit from treatment.

As longitudinal data on $CD4^+$ T-cell count were available, with multiple assessments of $CD4^+$ T-cell count on the same patient at different time points, within patient assessments at different time points were expected to be correlated. This correlation needed to be accounted for when selecting analytical methods (Diggle *et al.*, 1994).

The statistical analyses using longitudinal data needed to consider four main characteristics: i) time may be an explanatory variable; ii) repeat $CD4^+$ T-cell count measurements for a patient are likely to be correlated; iii) the co-variables may be time-dependent, showing variation through time for a patient; and iv) missing data in the successive $CD4^+$ T-cell count measurements may induce a bias. A repeated measures linear mixed model was used to derive time adjusted $CD4^+$ T-cell count slope, or rate of change per quarter in $CD4^+$ T-cell count since entry to cohort (Zeger *et al.*, 1988).

The MIXED procedure in SAS (Version 9.1.3; SAS Institute Inc., Cary, NC, USA) was used by fitting values of $CD4^+$ T-cell counts from all available results at different time points as a dependent variable grouped into 3 monthly intervals. A random intercept model was fitted using MIXED procedure in SAS. The independent variables included the fixed effects of subjects and clinic visit time points grouped into 3-month intervals. This assumed that the intercept for each patient was random and the effects of $CD4^+$ T-cell count changes (slopes) from baseline (at entry into the cohort) were also random and differed between subjects. A compound symmetry covariance matrix was used to model the within patient errors. This assumed that the variances were homogeneous. Correlation existed between two separate measurements, but it was assumed that the correlation was constant regardless of the time between successive $CD4^+$ T-cell count measurements (Kincaid, 2005). Estimates of $CD4^+$ T-cell count slopes were obtained from patient by time interaction. All patients whose $CD4^+$ T-cell counts were stable since cohort entry were assumed to show either no change or have an increase in $CD4^+$ T-cell count since entry to cohort, indicated by a slope ≥0 cells/μL blood (i.e. no decline in $CD4^+$ T-cell count). Those patients whose $CD4^+$ T-cell count slope was negative were assumed to have an unstable, declining $CD4^+$ T-cell count over time, and were therefore excluded from selection. After excluding patients who had died, a further refined selection criterion was applied to patients who had stable $CD4^+$ T-cell counts. Patients were defined as LTNP, who had a stable $CD4^+$ T-cell count which continuously remained within the normal range (450-1650 cells/μL blood) throughout clinical follow-up. These were compared with those patients whose $CD4^+$ T-cell counts were stable but had at least one recorded $CD4^+$ T-cell count below the normal range and referred to as patients with long-term stable low $CD4^+$ T-cell counts.

2.6 HIV AND AGEING ANALYSES: CO-EFFECTS OF HIV-1 INFECTION AND AGEING ON RESPONSE TO THERAPY, CO-MORBIDITY, PROGRESSION AND IMMUNE RE-CONSTITUTION ANALYSES

2.6.1 COHORT SELECTION CRITERIA

- ***STUDY POPULATION***

The Chelsea and Westminster HIV cohort data were investigated from 1988 until May 2011, tracking the age of individuals and therefore the age profile of the HIV-1 positive people. Data for this study were extracted from the database on all adults ≥18 years including patients who were

ART naïve or treatment experienced and HAART was defined as therapy consisting of at least three antiretroviral drugs in accordance with published guidelines (BHIVA Writing Group, 2012). The work presented in Chapter 4 (HIV and Ageing) of this PhD thesis is based on a cohort of HIV-1^+ people, who, since 1st January 1988 attended and were followed up until 31st May 2011, the time when data were extracted from the local Trust data warehouse.

Age was calculated to the most recent visit to HIV clinic. Whilst cohorts were grouped into pre HAART era (1988 to 31st December 1995), early HAART era (1st January 1996 to 31st December 2000), post HAART era (1st January 2001 to 31st December 2004) and the late HAART era (1st January 2004 to date of data extraction). The choice of these cohort strata were selected as these coincided with the introduction, availability and prescription of new antiretroviral drugs during the cohort strata as part of HAART therapy (see Chapter 1: Table 1.2.1). Comparisons of immune changes were performed by assessing HIV-1^+ patient groups following unbiased division according to age-defined quartiles for the entire Chelsea and Westminster HIV Cohort. Analyses were also performed by age grouped into decades due to large data size and large number of data points meant that using age as an ungrouped variable was computer intensive on modelling method used to derive estimates by age. Therefore using ungrouped age data did not allow estimates to be derived. Decade grouped age was practical from computing perspective as well as describing data by age groups commonly defined and used in literature with description of commonly used terminology (see Chapter 4: Table 4.1.1). These grouped age categories reflect 'young adults' 18-39 years, and 'middle-aged' adults age ≥40 years when, on average, decline in thymic function commences and subsequently, on average start of immune senescence from age ≥50 years or 'older' adults in this PhD thesis (Spits, 2002).

As indicated above, comparisons of immunological parameters were performed by assessing HIV-1^+ patient groups following immune considered age group strata for the entire HIV cohort relating to the estimated time of optimal thymic involution and decrease in thymic output/activity. For this, people in 3 age categorised groups were compared and described:

1) Less than 40 years of age (prior to the time of maximal thymic involution).
2) Between 40 and 49 years of age (encompassing the estimated time of maximal thymic involution, decreased thymic output and the current median age of the HIV-1 infected population).
3) Age 50 years or above (immune senescence in the 'older' HIV-1 infected population).

2.6.2 STATISTICAL METHODS

- ***CROSS SECTIONAL DATA***

Quantitative data with hyper geometric distribution have been presented as median with inter-quartile ranges, while qualitative data has been presented as numbers with proportions. Between group comparisons of quantitative data with hyper geometric distribution were assessed using Mann-Whitney U test, while qualitative data are compared by study arms using χ^2 test and where appropriate due to small sample in cells Yates' correction has been applied (Agresti, 2002).

- ***LONGITUDINAL METHOD***

Longitudinal analytic technique were applied to longitudinally collected renal function variables described in Chapter 4 (HIV and Ageing). For further details see section 2.5.2. The renal function data were routinely collected on all patients since routine collection of these data commenced. As a result of availability of these longitudinal cohort data with repeat measurements longitudinal methods used for analyses was the same as those described in detail under section 2.5.2.

eGFR and protein:creatinine ratio were analysed using longitudinal linear mixed models assuming that any data missing were missing at random. The linear mixed model was used to derive time adjusted eGFR and protein:creatinine ratio changes with age since exposure to either NRTI, PI or NNRTI drugs or the first-line regimens, 2NRTI+NNRTI or 2NRTI+boosted PI. MIXED procedure in SAS was used to generate random intercept models where all eGFR and protein:creatinine ratios from available 10 year age categories were fitted as dependent variable. Independent variables included the fixed effects of ARV drug class exposure or first-line regimen, cohort assessment of renal function results at each age group category where age-categories were based on age at the time of eGFR or protein:creatinine result and the interaction of ARV exposure with changing age. The eGFR and protein:creatinine ratio results were fitted into the model as time updated and described by age at the time of their respective results. An unstructured covariance matrix was used to model the within patient errors in order to account for dependence of observations across time within individuals.

Estimates of change in eGFR and protein:creatinine ratio from baseline were obtained from treatment by time interaction. Trends have been described for ART naïve PLHIV and those exposed to different ARV drug classes and first-line HAART regimens. Trends over time are presented as point estimates derived using the maximum likelihood method while an unstructured covariance matrix was used to allow for association of eGFR and protein:creatinine ratio results over time within individuals. Factors considered in the model were age at time of eGFR and protein:creatinine ratio results and were entered as a covariate grouped into decades. Linear and quadratic equations were explored and the equations presented were those that best explained the greatest amount of variance of longitudinal eGFR trends based on exposure to ARV drug class or

first-line regimens, along with interaction terms involving age and treatment groups. The primary interest was to look at contribution from a number of factors on renal function over time and the importance of age, HIV-1^+ infection and ARV class exposure combined or by themselves. While the majority of creatinine is excreted into the urine by glomerular filtration, the renal tubules also secrete creatinine. Creatinine clearance is thought to overestimate GFR by 10-20% (Breyer & Qi, 2010), and therefore creatinine clearance values based on these 'healthy' individuals were also used as comparators after they were reduced by 20%. An equation based on 'the healthy' individuals creatinine clearance values were obtained from the Baltimore Longitudinal Study of Ageing (Rowe *et al.*, 1976).

- ***TIME TO EVENT OF RENAL FUNCTION MARKER***

Nadir estimated glomerular filtration rate (eGFR) values ≤30 ml/min per 1.7m^2, indicative of CKD stages 4 and 5 or moderate to advanced renal failure was grouped into binary outcome. Cox's proportional hazards survival analysis was used to model time from first measurement of eGFR to either moderate to advanced renal failure and data were censored at most recent eGFR result. Univariate Cox's proportional hazards regression analysis was used to identify factors associated with the likelihood of moderate to advanced renal failure since cohort entry. Variables found to be significant in the univariate model ($p<0.2$) were used to build multivariable model which allowed the risk of a particular prognostic variable to be assessed while controlling for the others in the model. The final multivariable model presented was tested for its proportional hazards distributional assumptions using the complimentary log-log plot where data were plotted on the x-axis of the log of survival function and on the y-axis the log of the negative log of the estimated survivor function [(log(-logS(t)]. The final model presented was adjusted for age at nadir eGFR, gender, sexual orientation, injecting drug use, ethnicity, duration since HIV-1^+ diagnosis, PI exposure, max viral load, recent renal function results, and nadir $CD4^+$ T-cell count for possible confounding or residual effects.

2.7 POPULATION COST ANALYSES

The data from the NPMS-HHC study monitors prospectively the effectiveness, efficiency, equity and acceptability of treatment and care in participating HIV units since 1996. Using an agreed minimum dataset, standardised data are routinely collected in clinics and transferred to the NPMS-HHC Coordinating and Analytic Centre, ensuring both patient and clinic confidentiality. The data analyses are performed both at clinic and aggregate levels. The clinic specific analyses remain confidential, while aggregate analyses become public documents (Beck & Mandalia, 2000a, 2000b).

With growing number of HIV-1^+ people accessing care in the NHS as described in Chapter 1: section 1.2, one of the effects has been that number of people living with HIV-1 infection in the UK and elsewhere has increased. The implication of this is that increasing number of HIV-1^+ people will

require long-term use of health services, which will have financial implications. In addition, on-going development of new therapeutic drugs will be needed. Such development should include novel ART or immune therapies particularly since it is likely that currently available ARV drugs will eventually fail, requiring the sequential use of different therapeutic regimens in order for HIV-1 infected subjects to remain alive and well.

2.7.1 STUDY SITES, SAMPLING AND COHORT DESIGN

- ***CONFIDENTIALITY***

At the Co-ordinating and Analytic Centre (CAC) each HIV-1^+ individual is assigned an NPMS ID which is derived from and consists of soundex, gender and date of birth. Derivation of a unique identifier is based on these three fields. This therefore has a limitation that if a subject moves between clinics they will uphold their name and date of birth in particular. Soundex is generated from patients name and therefore ascribing a unique identifier assumes this is the same if a patient moves between clinics. Due to large scale of data checking, verification of this information on the very large numbers of patients included in the NPMS-HHC data cohort was considered to be unrealistic and not pragmatic and therefore for this work it is assumed otherwise.

All data are linked using the NPMS ID and no names or other patient identifiers are requested by the NPMS-HHC CAC (see section 2.3 for further details on ethics statement). All clinic level analyses that are summarised are confidential and are only sent to the source clinic where the actual data originated from.

- ***GEOGRAPHICAL LOCATION***

The geographical locations of the 14 sites who contributed to the data analyses from NPMS-HHC participating sites are from 7 London and 7 out of London centres with one centre located in Scotland.

- ***DATA COLLECTION INSTRUMENT***

The NPMS-HHC data collection form containing comprehensively specified list of required fields document (see Appendix A5.I) were sent to each one of NPMS-HHC participating sites. Data were requested using this standard data collection document detailing a list of required fields. This document lists the demographics and risk factor data. In addition all hospital activity data since 1996 together with the HIV-1 related diagnosis, HIV-1 RNA plasma load and $CD4^+$ T-cell count history together with a history ARV drug prescription.

- ***DATA TRANSFER TO CO-ORDINATING AND ANALYTIC CENTRE (CAC)***

In order to answer the research questions data needed to be requested from the NPMS-HHC participating sites in a standardised format. These were requested and transferred to the CAC through secure methods and with security passwords in place (see Data Collection Instrument section for further details). Safety procedures, to ensure patient and clinic confidentiality discussed and agreed between the NPMS-HHC data manager and each of the clinic's information officers, were maintained and enforced.

- ***STUDY DESIGN***

NPMS-HHC cohort include a group of HIV-1^{+} patients followed-up over time and a variety of variables are collected routinely by participating centres and relevant information as described in Appendix A5.1 are provided and that forms part of the study.

- ***DATA COLLECTED***

Data were requested from the NPMS-HHC participating centres of all HIV-1^{+} people who had ever attended for care since 1996 which is the point when HAART was first introduced into routine HIV care in the UK. Data collected reflected the remit of NPMS-HHC that was to collect data on use of hospital activity. Each participating clinic extracted data as per the NPMS-HHC list of required fields from their local database. The data were provided in an electronic format and sent to the CAC in accordance with the local data protection policy ensuring no patient identifying data were sent such as patient names. Instead data were pseudo-anonymised where use of patient's soundex, date of birth and gender allowed a generation of a study identifier that was used to link different data sets across different files within a clinic.

- ***DATA LINKING AT PATIENT LEVEL***

Each clinic provided different types of data in individual data files. For example, all demographics data were contained in one file, while all antiretroviral treatment history in another. This was the case for all data types and all follow-up data contained respective dates. All these files were linked using NPMS-HHC study derived identifier. Before each file was linked duplicate records were checked for and removed. There were variations in data received from each clinic, for instance some clinics data included use of service data whilst ARV history data were not available from some clinics. All available data were linked firstly by each patients derived unique identifier.

- ***DATA QUERIES AND RESOLUTION***

Once the data were linked these were reviewed by eye in table format in SAS so that data errors such as incorrect input of decimal number, spelling and punctuations could be corrected in

accordance with sections 5.1, 5.4, and 5.5 of the ICH Guideline for Quality Assurance and Quality Control (ICH/GCP, 1996).

- ***QUALITY CONTROL PROCESS***

Standard quality control processes were followed as described under section 2.4 Quality Control Process. Once all data were linked queries were raised on any data that were not linked and these were resolved. In addition patient records were checked between files, for instance number of patients in demographics file should tally with other files and vice versa. Where discrepancies were found these were raised and resolved. Some of the pre-planned queries include: to test whether numbers of patients found in different files were in agreement. Where this is not the case queries will be raised and resolved at source clinic; by checking consistencies in data codes found in each data variable provided. For instance gender may be coded as M and F and if any codes were found which do not match with what the clinics describe in their data extraction codes will be pointed out and resolved; to check whether all visit dates were consistent with patients clinic attendances; data captured on ARV were checked for consistencies in start and end date; all laboratory data such as $CD4^+$ T-cell counts and HIV RNA load data were checked for consistency regarding date of result and study time point as well as plausibility of data in terms of availability of data and also missingness. In the event where more than 5% of the patients found in demographics data where lab results were found to be missing, clarifications for such an occurrence was sought as a form of a data query at clinic source.

Once queries were generated and resolved these were then documented and data changes made using SAS program. In the event when updated data were provided these would be rechecked as detailed above.

- ***DATA LINKING AT DIFFERENT HIV CLINIC LEVEL***

The data files received from each clinic were first read into the statistical package, SAS. Subsequently, individual files for each clinic were linked to form a large SAS database per clinic. As file structure differed between clinics, separate SAS programs were written for each of the clinics in order to manipulate the data and perform the statistical analyses. Specific issues that included data discrepancies and cleaning of the data files were dealt with by contacting the information officer from the respective clinic. Agreement between the HIV disease classificatory systems of stage of HIV infection at the clinics and CAC was investigated during previous analyses and a high correlation between the two was found.

- ***DUPLICATE PATIENT DATA***

All duplicate records were checked and removed at individual clinic level. However it is known that a proportion of HIV-1 infected patients attend more than one HIV clinic and this is identified by using the NPMS-HHC derived ID. This is based on patient's soundex, gender and DOB. While this NPMS-HHC derived identifier allows identification of patients who move from one centre to another one of the limitation of this is that it assumes that when patients attend another centre they will provide identical name and date of birth. For a few this may not necessarily be the case as clinicians have reported (personal communication) some HIV-1^{+} patients attending different centres have used alias names. This however occurs in a minority of subjects.

2.7.2 COHORT SELECTION CRITERIA

- ***STUDY POPULATION***

At the time of data analyses for this work 24 clinics agreed to participate with the NPMS-HHC. However, not all clinics could contribute data for this work some due to problems with IT while others due to paper based medical data collection and therefore unavailability of these data in an electronic format. The prospectively collected data from the participating NPMS-HHC sites provided information on HIV-1 positive patients attending the NHS hospitals for routine HIV care. Information on the use of outpatient (OP), inpatient (IP) and day-ward (DW) services between 1st January 1997 and 31st December 2006, were obtained from computerised information systems from the participating centres together with the demographic characteristics, risk factors, data on opportunistic HIV disease, history of ARV treatment combinations and data on $CD4^{+}$ T-cell count and viral load results over time. These data were requested and transferred to the CAC through secure methods with security passwords in place. Data security procedures discussed and agreed between the NPMS-HHC Data Manager and each of the clinics' Information Managers to ensure patient and clinic confidentiality were maintained and enforced. Data were transferred from the participating NPMS-HHC centres to the CAC between November 2007 and April 2008. Once the data were received these were standardised at the CAC before commencing the data analyses.

2.7.3 STATISTICAL METHODS

- ***USE AND COST OF HOSPITAL SERVICES***

Information on the use of hospital IP, OP and DW services between 1st January 1997 and 31st December 2006, were obtained from 14 hospitals participating in this analysis. The mean number of IP days, OP and DW visits per patient-year (PPY) by stage of HIV-1 infection - asymptomatic (ASx), symptomatic non-AIDS (Sx non-AIDS) or AIDS - were calculated PPY from 1997 to 2006. The methods used for calculating the mean use of hospital services PPY were similar to those employed in previous studies (Beck *et al.*, 2004; Badri *et al.*, 2006).

The denominator consisted of the total duration of follow up for all patients during a calendar year, from when they were first seen during that year till the end of the year if still alive, or when they died, or if they were lost to follow up, which ever came first. If patients had attended prior to 1st January 1996 (start of study period), data were left censored at 1st January 1996, whilst if patients were found to have used HIV services during the second half of 2006, then the data were right censored at 31st December 2006. Numerators were calculated by summing the use of IP, OP or DW services and mean use of services PPY were calculated using the following formula:

$$M = \frac{\quad}{\quad} \times 365$$

where S_{ij} = use of service of individual i at jth time point;
t_{ij} = time since diagnosis of stage of HIV infection of individual i and remaining within the same stage
M = mean of services S per patient-year by stage of HIV-1 infection

Annual cost PPY estimates for HIV service provision for individual HIV-1^+ people by stage of HIV-1 infection were produced by linking mean number of IP days, OP visits and DW visits with their respective unit costs. Costs were in UK pounds at 2006 prices. The unit cost estimates for an average IP day was £475, average OP visit was £94 and £384 for a DW (Beck *et al.*, 2008).

- ***USE AND COST OF ANTIRETROVIRAL THERAPY***

Average costs for treatment regimens were based on London Specialised Commissioning Group prices of each licensed ARV drug. Average annual cost, including 17.5% value added tax (VAT) at the time of data analysis (2009), were calculated for different ART regimens - mono-, dual-, triple- or quadruple-or-more therapy - and stratified by stages of HIV-1 infection: asymptomatic, symptomatic non-AIDS or AIDS.

- ***OVERALL COST OF SERVICES BY STAGE OF HIV INFECTION AND TYPE OF ANTIRETROVIRAL THERAPY***

The costs for different ART by stage of HIV-1 infection were added to the cost for IP, OP and DW services for each of the clinical stages; costs of 'other' drugs and tests and procedures performed, were added to obtain the total direct costs for treatment and care for HIV-1^+ people by stage of HIV infection and type of ART (Table 2.7.1 and Table 2. 7.2).

Table 2.7.1: Table Showing Included Costs

Cost of Treatment & Care Services	Cost of Support Services All relating to hospital services
Direct costs Cost of tests Cost of procedures Cost of Drugs Staff costs Equipment Consumables Others	Catering Laundry Medical Records Pottering Storage and supplies Transport Administration Security Building & Maintenance

Table 2.7.2: Table Showing Costs

Type of service	Unit cost used
Inpatient visit	£475 per day
Outpatient visit	£94 per visit
Day ward visit	£384 per attendance
Tests/Procedures Asymptomatic Symptomatic non AIDS AIDS	 £271 per year £465 per year £1,405 per year
Other drugs Asymptomatic Symptomatic non AIDS AIDS	 £274 per year £1,582 per year £6,814 per year
Community care Asymptomatic Symptomatic non AIDS AIDS	 £3,091 per year £4,084 per year £5,990 per year
ARV drugs	Individual ARV cost prices received from the London HIV consortium

Unit costs are based on 2006 prices

The direct cost-estimates were from a public service perspective (Beck & Miners, 2001), but some cost-estimates also included indirect or community care costs (Beck *et al.*, 2008).

- ***HIV DIAGNOSED PATIENTS USING NHS SERVICES BY TYPE OF ART***

The annual number of HIV-1 diagnosed people using NHS services by stage of HIV-1 infection and by type of ART prescribed using NHS services, were obtained from the Health Protection Agency (HPA). Not all of these data sets were complete; for some HIV-1^{+} people, the stage of HIV-1 infection was known but the type of ART prescribed was unknown. These individuals with missing

ART data were proportionally distributed across the respective ART strata, ensuring that the proportion of subjects represented in each category remained unchanged. Secondly, HPA figures indicated that no subjects had been prescribed quadruple or more therapy in 1997 and 1998, which was unlikely and probably due to incorrect reporting during those calendar years. For these years linear least squares regression analysis (Aldrich, 2005) was used to estimate the number of HIV-1 diagnosed patients using NHS services likely to have been prescribed quadruple or more therapy.

- ***UK HIV POPULATIONS COST ESTIMATE***

To obtain the population cost estimates, the total annual treatment and care costs for a HIV-1^{+} people by stage of HIV-1 infection and type of ART were multiplied by the number of HIV-1 diagnosed people using NHS services within those categories for each year. Costs were then added across stages of HIV-1 infection and types of ART regimens to obtain a total population cost by year, while community care costs were also included (Beck *et al.*, 2008) .

- ***UK HIV POPULATIONS AND COST PROJECTION***

NPMS-HHC currently holds cost data between 1996 and 2006, however HPA population data were available since 1997, the population cost figures derived therefore commence from the calendar year 1997 onwards and the ARV drug costs were calculated using the 2006 prices. Data were projected to estimate future costs for the years 2007 to 2013 using linear least squares regression analyses (Aldrich, 2005). These projections were based on the trends observed over 1997-2006.

Three methods were investigated:

1) *First* method extrapolated total annual population costs 1997-2006 to 2007-2013;
2) *Second* method extrapolated the total number of PLHIV who used HIV services between 1997-2006 and 2007-2013. The average annual cost of treating a PLHIV across all stages of HIV infection in 2006 was calculated and this was multiplied by the projected numbers of PLHIV using NHS services for the years 2007-2013;
3) *Third* method extrapolated PLHIV for each of the three stages of HIV infection from 1997-2006 to 2007-2013. Average 2006 annual cost of treatment and care for stage of HIV infection was used to extrapolate cost 2007-2013 by stage of HIV infection. Total projected annual treatment costs were obtained by adding annual population costs of individual stages of HIV-1 infection across all stages;

  - **Method 1**

Population cost data for the years 2007 to 2013 were projected using these 2006 drug prices. Linear regression analysis in SAS was used to estimate costs since 2006 assuming the linear trends seen in previous years would continue. The total UK population costs were projected for

(regardless for the patients' stage of HIV-1 infection) and these were estimated excluding and including community care costs.

The cost data were projected using 1997-2006 UK population cost figures for HIV-1 infected patients accessing HIV care in the UK NHS hospitals. The 1997-2006 cost data produced were used in linear regression models after combined costs for the three stages of HIV-1 infection (asymptomatic, symptomatic non-AIDS and AIDS). Two models were derived: excluding; and including; community care costs in order to estimate costs for years 2007-2013 using regression models described below.

- **Method 2**

Method 2 calculated the average direct population cost for 2006 for all HIV patients without accounting for patient's stage of HIV infection. The direct population cost was derived by dividing the NPMS-HHC calculated population cost for 2006 by the number of people living with HIV-1 in the UK in 2006 (HPA data). In order to obtain the direct population cost the average population cost that was calculated for 2006 was multiplied by the projected number of people living with HIV for that year. This process was applied to each subsequent years from 2007-2013.

The figures were calculated from the average cost PPY regardless of the stage of HIV-1 infection.

A) Project the total UK population figures obtained from HPA of HIV patients accessing care in the UK.
B) For the total UK HIV population from 1997-2013 for both including and excluding community care (regardless of their stage of HIV infection)

Once these number of HIV-1 population data had been projected for 2007 to 2013 from A and B these were added to the costs derived using the 2006 costs as a base cost figures to forecast the cost figures between 2007 to 2013.

- **Method 3**

For this method, the average population cost for 2006 were estimated. These were estimated by dividing the NPMS-HHC calculated population cost for 2006 by the number of people living with HIV-1 in the UK in 2006 (HPA data). In order to obtain the average population cost by stage of HIV-1 infection the average population cost that was calculated for 2006 was multiplied by the projected number of people living with HIV-1 for that year by the stage of HIV-1 infection. This process was applied to each of the subsequent years from 2007-2013.

Linear least squares regression analyses was used for each stage of HIV-1 infection and the resulting linear regression models were used to estimate the population of patients living with HIV accessing care in the UK by stage of HIV infection. The UK HIV population costs for both including

and excluding community care from 1997 - 2013 were thus estimated for the three stages of HIV-1 infection (asymptomatic, symptomatic non-AIDS and AIDS).

2.8 GRAPHICAL REPRESENTATION OF DATA

The data presented in this PhD thesis follow standard display of data depending on the type of data analysed[19] (Armitage *et al.*, 2001). Tables, graphs and charts have been used to display data or display selection of data sequence using flow charts. All data were analysed using SAS statistical software while GraphPAD and Microsoft Excel were used to prepare graphs and charts and further pictorial display of information were prepared using either Microsoft Powerpoint and Microsoft Word.

Chapter 2:19 http://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/1-data-display-and-summary

Chapter 3: Long-Term Non-Progressor

CHAPTER 3: LONG-TERM NON-PROGRESSOR COHORT ANALYSES: ESTIMATED TIME UNTIL DISEASE PROGRESSION

Note: A modified version of this chapter has been published.
Mandalia S, Westrop SJ, Beck EJ, Nelson M, Gazzard BG, Imami N.
Are long-term non-progressors very slow progressors? Insights from the Chelsea and Westminster HIV cohort, 1988-2010. PLOS one. 2012;7(2):e29844.

The total global estimate of people living with HIV-1 infection continues to increase. By end of 2011 the global estimate was reported as 34 million (UNAIDS World AIDS Day Report, 2012) while this had increased 35.3 million one year later in 2012 (UNAIDS Report on the Global AIDS Epidemic, 2013) (further details are in Chapter 1: section 1.2). If HIV-1 is left untreated most patients eventually progress to AIDS with reduced quality of life. Although there have been advances in treatment of HIV-1 and opportunistic infection, the search for preventive strategies on the development of prophylactic and therapeutic interventions through effective vaccines and/or immune-based therapies continues which would to put an end to: disease; prevent infection; or delay disease progression. The lessons learned from a unique and atypical patient group would be informative in designing novel strategies for the treatment of HIV-1 (Blankson, 2010; Imami *et al.*, 2013; Walker & Yu, 2013).

Long-term non-progressors (LTNP) comprise a subgroup of HIV-1 infected people who naturally control HIV-1 replication without antiretroviral therapy (ART) and clinical progression of HIV-1 disease is slow, with a few of these patients develop into AIDS even after prolonged infection. A group identified as LTNP could potentially provide a useful model in order to study the mechanisms of viral control as a step towards development of prophylactic vaccines and immune therapeutic approaches.

3.1 INTRODUCTION

The majority of HIV-1 infected patients display a gradual decline in peripheral blood $CD4^+$ T-cell counts throughout the course of their illness. This is accompanied by excessive immune activation and the sequential loss of immune responses, first to HIV-1, then to other pathogens, allogeneic stimuli and finally mitogens (Clerici *et al.*, 1989; Imami *et al.*, 2001). The clinical course of HIV infection generally includes three stages: primary infection, clinical latency, and AIDS-defining illness (see Chapter 1: Figure 1.2.2).

The rate of disease progression from asymptomatic HIV-1 infection to AIDS varies between patients and according to the rate of progression, HIV infection may be divided into four major types: (i) typical progressors (where the median time from first HIV-1 infection to progression to AIDS is eight to ten years); (ii) long-term survivors (those who progress to AIDS within a time frame similar to typical progressors, but in whom both clinical and laboratory parameters remain stable

for an unusually long period of time once disease progression has occurred); (iii) rapid progressors (where progression to AIDS occurs within two to three years of primary infection); (iv) LTNP (where individuals do not experience progressive HIV-1 disease for many years - eight to ten years following HIV-1 infection) (Pantaleo & Fauci, 1996).

The rate of disease progression from asymptomatic HIV-1 infection to AIDS varies between patients and despite the varying disease progression rates, the majority of HIV-1 infected individuals eventually progress to AIDS. Determination of factors associated with HIV-1 progression aims to enable the development of prophylactic vaccines and immune based approaches in order to delay HIV-1 progression or prevent infection.

AIDS defining opportunistic infections were originally used to indicate disease progression in the clinic, however in 1993 the definition of HIV disease progression was widened to also include a drop in $CD4^+$ T-cell count to <200 cells/μL blood or <14% of lymphocytes (Castro *et al.*, 1992; MMWR, 1992). Some HIV-1^+ individuals are able to maintain stable $CD4^+$ T-cell count within the normal healthy range between 450-1650 cells/μL blood; local laboratory reference range (Westrop *et al.*, 2009) for a prolonged length of time and do not exhibit signs of HIV related opportunistic infections such as bacterial pneumonia, cytomegalovirus and others (Castro *et al.*, 1992), despite not receiving ART. The Chelsea and Westminster HIV-1 cohort of these atypical group of patients have been referred to as LTNP (Imami *et al.*, 2002; Westrop *et al.*, 2009).

Within this group of atypical patients a few suppress HIV-1 replication to BLD of the HIV-1 RNA plasma load assay (<50 and including <500 when historically <50 copies/mL plasma load assays were unavailable) (Deeks & Walker, 2007; Westrop *et al.*, 2009). These individuals are known as HIV Controllers (HIC) (Westrop *et al.*, 2009). LTNP and HIC are terms that have been used to describe people known to have been living with HIV-1 for a 'long time' without developing AIDS. While some studies claim that the prevalence of LTNP and HIC patients is low and is estimated at 1 per 100 (Deeks & Walker, 2007; Walker, 2007) there is still no joined nomenclature for such atypical subjects. Despite the identification and study of LTNP and HIC in many international single- and multi-centre HIV-1 cohorts, no international consensus definitions of the LTNP or the HIC exist and consistent definitions have not been uniformly applied across all studies and sites. Confusion about definitions used exists in the published literature, resulting in difficulties when comparing results from patient groups from different cohorts. The unique immunological and virological profiles demonstrated by LTNP and HIC individuals may provide further clues towards development of prophylactic and therapeutic interventions through effective vaccines and/or immune-based therapies and highlight the importance of global unified definitions and classifications of patient groups of interest, especially the LTNP and HIC. LTNP have been

described to differ in genetics, virology and immunology compared to typical HIV-1 progressors as summarised in Table 3.1.1 (Gaardbo *et al.*, 2012).

Table 3.1.1: Table Summarised from Gaardbo *et al.*, 2012 Describing Factors That Differ Between LTNP and Typical HIV-1 Progressors

	Genetics	Virology	Immunology
LTNP	The presence of CCR5-delta32 polymorphism, and HLA B57 allele, are overrepresented in non-progressors.	Infected with less virulent strains of HIV resulting in a more benign infection majority of non-progressors are infected with replicant-competent virus.	$CD4^+$ T-cell from controllers are less susceptible to HIV-1 compared to $CD4^+$ T-cell from progressors; $CD4^+$ T-cell from controllers are susceptible to HIV entry and productive infection.

Differences in definition and criteria used to define atypical patient groups will impact on estimates of duration to HIV-1 progression before considering genetics, virology and immunology, further emphasising the need for agreed standardised use of terminology and definitions in order that these atypical patient groups are studied in-depth (Imami *et al.*, 2013). In-depth study of such atypical patient group has the potential to provide insight into the control of HIV-1 infection in the absence of therapeutic intervention. Further, understanding the variables that are correlated with protective immunity to HIV-1 infection will guide in developing preventive HIV-1 vaccines as well as help determine treating HIV-1 infection by boosting immunity to HIV-1 (Imami *et al.*, 2002; Migueles & Connors, 2010; Imami *et al.*, 2013).

3.1.1 SYNONYMS OF HIV NON-PROGRESSORS

Many international single and multi centre HIV cohorts have identified LTNP and HIC however, terms used to describe these unique and atypical group of $HIV\text{-}1^+$ people are not uniform across published studies in the UK and internationally and there are no unified and standardised nomenclature or definitions being used resulting in problems with respect to comparability of data (Imami *et al.*, 2013). Although studies to date, have identified only a small number of factors that are consistently associated with LTNP status (Vento *et al.*, 2004). Establishment, definition and characterisation of $HIV\text{-}1^+$ LTNP patient cohort could be used to identify factors that are associated with non progression status, and therefore aid in determining potentially successful prophylactic vaccines and novel therapeutic approaches. The descriptions listed in Figure 3.1.1 portray some of the synonyms that have been used in published literature to define such atypical patient groups.

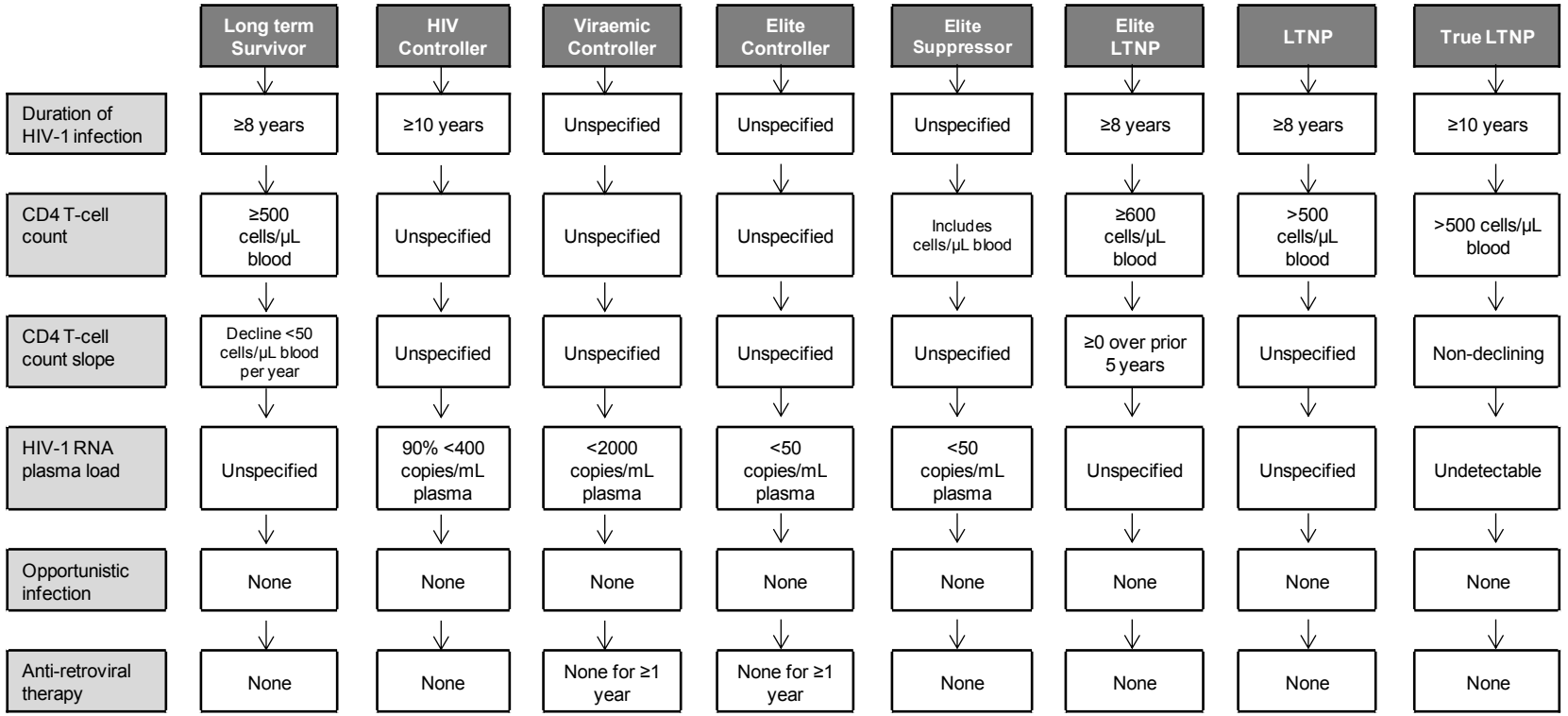


Figure 3.1.1: Synonyms of HIV Non-Progressors

Figure taken with slight modifications from (Mandalia *et al.*, 2012)

CHAPTER 3

LONG-TERM NON-PROGRESSORS

- ***LONG-TERM SURVIVOR***

During the asymptomatic period, in some individuals with normal or lower than average $CD4^+$ T-cell counts, a lack of disease progression is observed, and these individuals have been described as 'long-term survival' (Rutherford *et al.*, 1990). Historically, individuals who were HIV-1^+ remained healthy with $CD4^+$ T-cell count >500 cells/μL blood for ≥10 years without ART (Klein & Miedema, 1995) were referred to as long-term survivors. A number of synonyms were used to describe such patients and included: long-term asymptomatic; symptomless subjects; healthy long-term seropositives; long-term non-progressors; long-term survivors with high- or low- $CD4^+$ T-cell count.

These synonyms were then incorporated into a generic term, long-term survivors (Klein & Miedema, 1995) and these groups of individuals were defined as HIV-1^+ subjects who had been followed up for ≥8 years in combination with one or more of the following end points: they were AIDS free; they had normal $CD4^+$ T-cell count (≥500 cells/uL blood); minimal $CD4^+$ T-cell count loss of <50 cells/uL per year; had normal $CD4^+$ T-cell count to CD8 T-cell ratio (≥1.00); and they were not ARV exposed.

- ***HIV CONTROLLERS***

HIV controllers were defined as HIV-1^+ patients who had been followed up for ≥10 years, who had received no ARV treatment, and for whom >90% of the HIV-1 RNA plasma load measurements were <400 copies/mL plasma (Lambotte *et al.*, 2005; Grabar *et al.*, 2009).

- ***VIREMIC CONTROLLERS***

In the paper by (Walker, 2007), viremic controllers were defined as people with asymptomatic HIV-1 infection who maintain HIV-1 RNA load to below 2000 copies/mL plasma with no ARV therapy for 1 year or longer. However episodes of viremia, the presence of virus in the blood, were defined as acceptable as long as they represented the minority of all available viral load determinations.

- ***ELITE OR AVIREMIC CONTROLLERS***

Definition of elite controllers included people with asymptomatic HIV-1 infection who maintain HIV-1 RNA plasma load to below 50 copies/mL plasma with no ARV therapy for the previous 12 months or longer. Episodes of viremia were defined as being acceptable as long as there was no evidence of consecutive positive HIV-1 RNA plasma load tests (Deeks & Walker, 2007; Walker, 2007).

- ***ELITE SUPPRESSORS***

In a publication by Han and colleagues (Han *et al.*, 2008) an elite suppressor is referred to as '*an HIV-1$^+$ subject who is untreated and who maintains undetectable viral load*''. Although there is no clarity regarding repeat of CD4$^+$ T-cell count measurements over time in these subjects, the publication displayed data on the most recent CD4$^+$ T-cell count in this group of individuals. One individuals lowest CD4$^+$ T-cell count value displayed was 399 cells/μL blood.

- ***ELITE LONG-TERM NON-PROGRESSORS***

Grabar and colleagues (Grabar *et al.*, 2009) described elite long-term non-progressors with the following criteria: asymptomatic HIV-1$^+$ infected patients naïve to ART; documented CD4$^+$ T-cell count and VL; known date of HIV-1$^+$ diagnosis; age >13 years; and length of HIV-1 infection ≥8 years; at least 3 measurements of HIV-1 RNA plasma load and CD4$^+$ T-cell count in the 5 last years; a CD4$^+$ T-cell count nadir ≥ 600 cells/μL blood; with CD4$^+$ T-cell count slope ≥ 0 over the last 5 years.

- ***LONG-TERM NON-PROGRESSORS***

Historically in a publication by Aiuti and colleague, long-term non-progressors (LTNP) were defined as HIV-1$^+$ subjects who were asymptomatic for at least 8 years and had never received ARV treatment, with CD4$^+$ T-cell count levels always above 500 cells/μL blood (Aiuti & D'Offizi, 1995).

HIV-1$^+$ infected patients who immunologically did not progress to AIDS were termed LTNP and defined because of: documented HIV-1$^+$ infection for more than 7 years; absence of symptoms; and the ability to maintain stable, high CD4$^+$ T-cell count (>600 cells/μL blood) over many years; without receiving ART (Pantaleo *et al.*, 1995). In this study the criteria used to define non-progression included documented HIV-1$^+$ infection for more than 7 years, stable CD4$^+$ T-cell count greater than 600 cells/μL blood, the absence of symptoms, and no ARV therapy.

A paper published in 1997 by Lefrere and colleagues described that despite being infected with HIV-1 for more than 10 years, a number of cohort studies had shown that a small percentage positive individuals remain symptomless and maintain a high CD4$^+$ T-cell count without any therapeutic intervention of ART (Hardy, 1991; Learmont *et al.*, 1992; Cao *et al.*, 1995). These subjects who at the time maintained a CD4$^+$ T-cell count >500 cells/μL blood over a 10-year period were described as LTNP.

- ***TRUE LONG-TERM NON-PROGRESSORS***

In a recent publication (Migueles & Connors, 2010) it was discussed that LTNP were historically defined as individuals who remained healthy: $CD4^+$ T-cell count >500 cells/μL blood for ≥10 years; without ART (Klein & Miedema, 1995). With the availability of HIV-1 RNA plasma load tests, cohorts of LTNP classified using this 'early' definitions were deemed heterogeneous with respect to HIV-1 RNA plasma load levels. The authors further conferred that a small subset of patients originally classified as LTNP maintained HIV-1 RNA plasma load BLD of the HIV-1 RNA plasma load assay (Lefrère *et al.*, 1997; Migueles *et al.*, 2000; Westrop *et al.*, 2009). These LTNP with undetectable viral loads were exhibiting truly non-progressive HIV-1^+ infection, with non-declining $CD4^+$ T-cell count and were free from opportunistic infections during prolonged follow-up. These patients were therefore termed 'true' LTNP.

3.1.2 THE CHELSEA AND WESTMINSTER HIV COHORT AND DEFINITIONS OF ATYPICAL PATIENT GROUPS

A small number of HIV-1^+ patients who maintain a stable $CD4^+$ T-cell count within the normal healthy range of 450-1650 cells/μL blood, who in addition have no history of symptomatic opportunistic infection or AIDS defining illness, despite never receiving ART. A subset of this group control HIV-RNA plasma load to BLD of HIV-1 RNA plasma load assay and are known as HIC. Within Chelsea and Westminster HIV-1 cohort, these patients have been referred to as 'true' LTNP and/or LTNP. However as aforementioned these terms are not uniform between studies and sites as described in section 3.1.1, Figure 3.1.1 and there are no internationally agreed standardised synonyms or definitions of LTNP and HIC. The LTNP and HIC were identified from the entire HIV-1 cohort: duration since HIV-1 infection was longer than 7 years, the median time until disease progression; no history of symptomatic HIV-1 disease; no history of ever prescribed ART; a stable $CD4^+$ T-cell count which remained within the healthy normal range. Within the LTNP group, individuals who control HIV-1 replication to BLD of the conventional HIV-1 RNA plasma load assay (50 HIV-1 RNA copies/mL plasma) were also identified (Chapter 2: section 2.5.1). Based on a recent review (Gaardbo *et al.*, 2012), data were further interrogated to establish whether any further individuals could be identified and subsequently classified into LTNP and/or HIC using only control of viral load criteria as described in detail in Chapter 2: section 2.5.1. The aims were: to estimate time until HIV-1 disease progression in different groups of patients identified during the study period by the application of different selection criteria; to establish the frequency of HIV-1^+ patients who fulfilled the criteria of LTNP, long-term stable low $CD4^+$ T-cell count, HIC and long-term stable low $CD4^+$ T-cell counts of whom controlled HIV-1 RNA plasma load to BLD; and to report immunological and virological profiles in the identified groups of individuals.

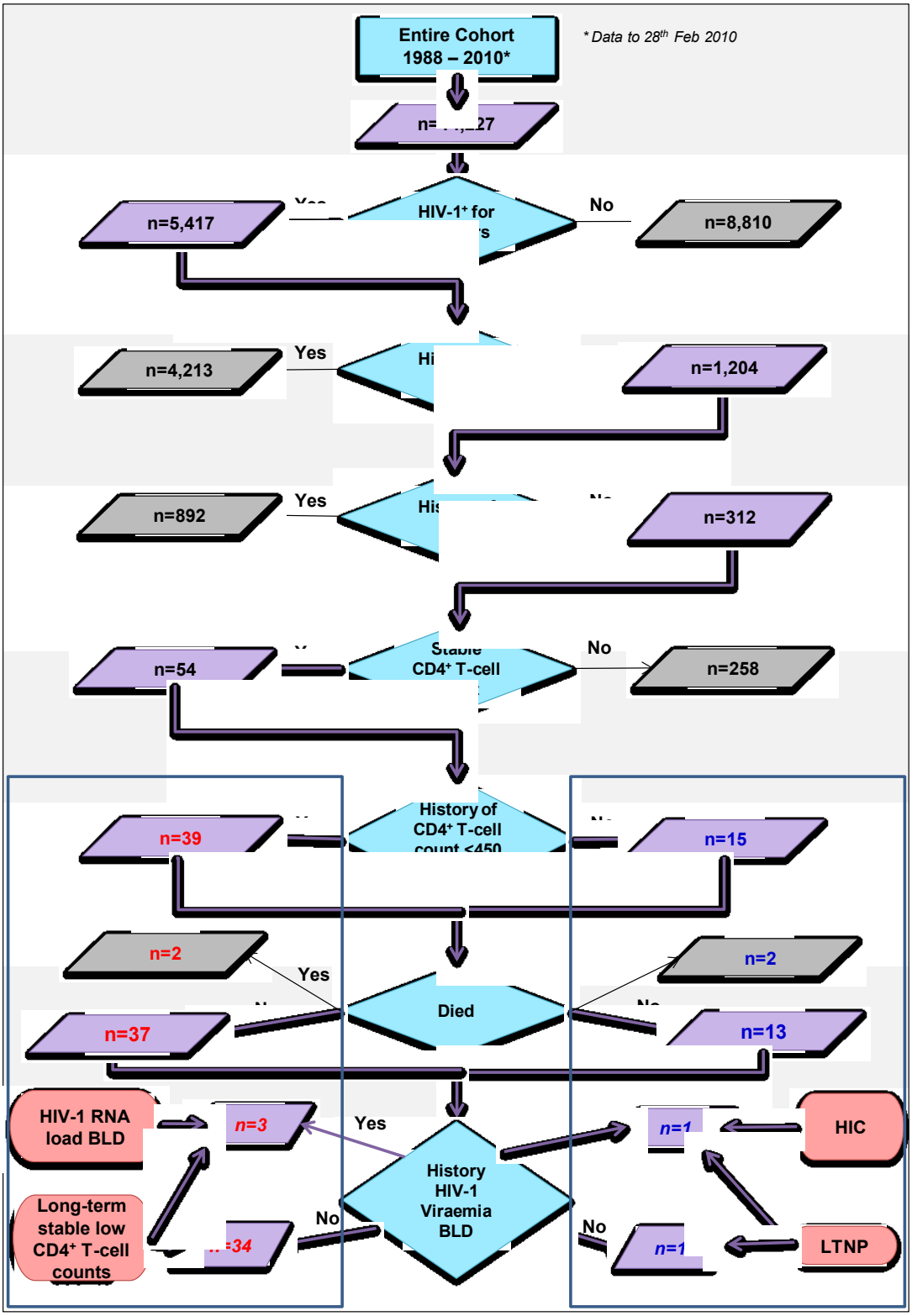


Figure 3.2.1: Flow Chart Showing Identification of LTNP and HIC from the Chelsea and Westminster HIV Cohort

Figure taken with slight modification from (Mandalia *et al.*, 2012)
ART=Anti-retroviral therapy
OI=Opportunistic infection
Stable CD4$^+$ T-cell count = rate of change in CD4$^+$ T-cell count ≥0 cells/μL blood per quarter since cohort entry
BLD=Below the limit of detection of the HIV-1 RNA plasma load assay (<500 or <50 HIV-1 copies/mL plasma since availability of this assay at our institution)
HIC=HIV controller
LTNP=Long-term non-progressor

3.2 RESULTS

- ***HIV-1 COHORT CHARACTERISTICS***

From a total of 14,227 HIV-1$^+$ patients in the study cohort, 5,417 subjects were identified to have been HIV-1$^+$ for more than 7 years. Of whom, 1,204 had no history on the database of having been prescribed ART (Figure 3.2.1). Of 1,204 individuals, 312 in addition had no recorded history of opportunistic infection on the database. Time until HIV-1 progression in the two groups 1) individuals with duration of HIV-1$^+$ greater than 7 years and no history of ART and 2) individuals with duration of HIV-1$^+$ greater than 7 years, no history of ART and no history of opportunistic infection were estimated by those whose CD4$^+$ T-cell count history did not remain within the normal range and those whose CD4$^+$ T-cell count history remained within the normal range (Figure 3.2.2: a; Figure 3.2.2: b).

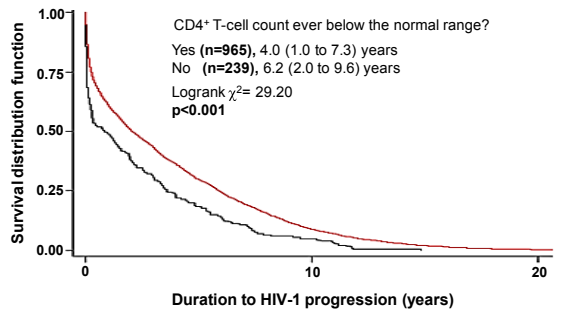
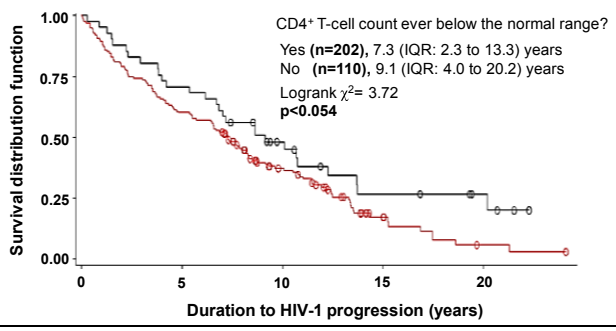
a: Individuals with Duration of HIV-1$^+$ > 7 Years and No History of ART	**b: Individuals with Duration of HIV-1$^+$ > 7 Years, No History of ART and No History of Opportunistic Infection**
N=1,204	**N=312**
	

Figure 3.2.2: Survival Plots Showing Time to HIV-1 Progression

- ***TIME UNTIL HIV-1 DISEASE PROGRESSION***

Of the 1,204 patients, 239 (19.9%) patients whose CD4$^+$ T-cell counts remained within the normal range had an estimated median time until HIV-1 disease progression since HIV-1$^+$ diagnosis of 6.2 (IQR: 2.0 to 9.6) years, compared to 4.0 (IQR: 1.0 to 7.3) years for the 965 patients who had at least one CD4$^+$ T-cell count below the normal range (Logrank χ^2 test: 21.26, p<0.001; Figure 3.2.2: a; Table 3.2.1). In addition, 312/1,204 (26%) patients had no recorded symptomatic opportunistic infection in the database and therefore were classified as remaining asymptomatic during the course of their routine HIV-1 follow up. Of whom 110 (35.3%) maintained CD4$^+$ T-cell counts within the normal range and the estimated median (IQR) time to progression in this group, defined as CD4$^+$ T-cell count decline or clinical progression showed some evidence of significantly longer 9.1 (4.0 to 20.2) years compared to 7.3 (IQR: 2.3 to 13.3) years in 202 patients who had at least one CD4$^+$ T-cell count fall below the normal range (Logrank χ^2 test: 3.72, p=0.054; Figure 3.2.2:b; Table 3.2.1).

Table 3.2.1: Time Since HIV-1$^+$ Diagnosis Until Disease Progression in Patients Who Have Been Infected With HIV-1 for More Than 7 Years and Who Remain Symptomless in the Absence of ART

Patients are stratified according to a history of at least one CD4$^+$ T-cell count below the normal range (<450 cells/μL blood)

	HIV-1$^+$ >7 years, no ART†			HIV-1$^+$ >7 years, no ART†, no OI†			HIV-1$^+$ >7 years, no ART†, no OI†, unstable CD4$^+$ T-cell count‡		
	n=1,204	**Duration until disease progression$^\#$**	***p- value***	**n=312**	**Duration until disease progression$^\#$**	***p-value***	**n=258**	**Duration until disease progression$^\#$**	***p-value***
CD4$^+$ T-cell count ever below the normal range?									
No	**239**	6.2 (2.0 to 9.6)	*<0.001*	**110**	9.1 (4.0 to 20.2)	*0.054*	**95**	5.8 (2.3 to 8.6)	*0.833*
Yes	**965**	4.0 (1.0 to 7.3)		**202**	7.3 (2.3 to 13.3)		**163**	4.6 (1.8 to 8.4)	

$^{\dagger, \ddagger,}$ See Figure 3.2.1 for definition of patient groups
$^\#$ median (IQR) years
p-values using the Logrank χ^2 test

Table taken from (Mandalia *et al.*, 2012)

- ***FREQUENCY OF UNSTABLE CD4⁺ T-CELL COUNT***

Of the 312 ART-naïve patients diagnosed as HIV-1$^+$ for longer than 7 years who remained asymptomatic during the course of their routine HIV-1 follow up, 258 (82.7%) had unstable or declining CD4$^+$ T-cell counts. Of these 95 (36.8%) had CD4$^+$ T-cell counts consistently within the normal range and the estimated median time until HIV-1 disease progression in this group was 5.8 (IQR: 2.3 to 8.6) years, this was similar to the 163 patients (63.2% of 258) who had a history of CD4$^+$ T-cell counts below the normal range with an estimated time until HIV-1 disease progression of 4.6 (IQR: 1.8 to 8.4) years (Logrank χ^2 test: 0.045, p=0.833; Figure 3.2.1, Table 3.2.1).

- ***DEMOGRAPHIC CHARACTERISTICS OF INDIVIDUALS WITH DURATION OF HIV-1⁺ GREATER THAN 7 YEARS AND NO HISTORY OF ART***

Of 1,204 (22.2%) who had no history of being prescribed ART (Figure 3.2.1) the demographic, immunology and virology data were compared between groups who had a history of CD4$^+$ T-cell count which remained ≥450 cells/μL blood and those whose CD4$^+$ T-cell count fell to <450 cells/μL blood (Table 3.2.2). There were no significant differences between patients whose CD4$^+$ T-cell count had been documented below the normal range compared with patients whose CD4$^+$ T-cell count remained within the normal range; in terms of gender (male 92.3% and 90.8% respectively; p=0.432) or median age (40.2 years and 39.9 years respectively; p=0.501). Significantly greater numbers of Caucasians were seen in the group whose CD4$^+$ T-cell count had been recorded on at least one occasion as below the normal range compared to patients whose CD4$^+$ T-cell count remained within normal range (68.5% and 58.2% respectively; p=0.003).

Amongst those whose CD4$^+$ T-cell count fell to <450 cells/μL blood, significantly greater proportion deaths were observed (36.3%), significantly lower median (IQR): nadir CD4 T-cell count 123 (IQR: 16 to 276) cells/μL blood; recent CD8 T-cell count 737 (IQR: 406 to 1,115) cells/μL blood; and nadir CD19 B-cell count 87 (IQR: 49 to 147) cells/μL blood; were observed compared to those who had a history of CD4$^+$ T-cell count which remained ≥450 cells/μL blood (Table 3.2.2). In those whose CD4$^+$ T-cell count fell to <450 cells/μL blood, a significantly higher median maximum HIV-1 RNA plasma load was observed (47,148 copies/mL, IQR: 11,300 to 145,951) compared to those who had a history of CD4$^+$ T-cell count which remained ≥450 cells/μL blood.

Table 3.2.2: Demographic, Immunological and Virological Parameters of Patients With Duration of HIV-1$^+$ >7 years and No History of ART

	Individuals with Duration of HIV-1$^+$ > 7 Years and No History of ART†		
	N(%) Total = 1204		
	CD4$^+$ T-cell count history <450 cells/µL blood Total=965	CD4$^+$ T-cell count history ≥450 cells/µL blood Total=239	^p-value
Demographics			
Died			
No	615(63.7)	225(94.1)	<0.001
Yes	350(36.3)	14(5.9)	
Gender (%)			
Female	74(7.7)	22(9.2)	0.432
Male	891(92.3)	217(90.8)	
Ethnicity (%)			
Caucasian	661(68.5)	139(58.2)	0.003
Black African	44(4.6)	9(3.8)	
Other	260(26.9)	91(38.1)	
Median age at last visit, years (IQR) [range]	40.2(35.0 to 46.7) [25.8 to 77.6]	39.9(34.4 to 45.9) [22.6 to 70.2]	0.501
HIV risk (%)			
MSM	831(86.1)	205(85.8)	0.007
Heterosexual	80(8.3)	30(12.6)	
Bisexual	54(5.6)	4(1.7)	
&Median time since HIV-1$^+$ diagnosis, years (IQR)	9.7 (8.2 to 12.6)	10.4 (8.2 to 13.7)	0.126
Immunology			
Median nadir CD4$^+$ T-cell count, cells/µL blood (IQR)	123 (16 to 276)	551 (491 to 670)	<0.001
Median most recent CD8$^+$ T-cell count, cells/µL blood (IQR)	737 (406 to 1115)	997 (764 to 1484)	<0.001
Median nadir CD19$^+$ B-cell count, cells/µL blood (IQR)	N=448 87 (49 to 147)	N=94 114 (88 to 212)	<0.001
Median nadir CD16/56 natural killer cell count, cells/µL blood (IQR)	N=370 40 (21 to 81)	N=84 54 (24 to 118)	0.068
Virology			
Median highest recorded HIV-1 RNA plasma load, copies/mL plasma (IQR)	N=265 47148 (11300 to 145951)	N=85 11867 (1607 to 43898)	<0.001

† See Figure 3.2.1 for definition of patient groups
MSM: men who have sex with men
^p-value using Mann-Whitney U test for quantitative data and χ^2 tests with Yates' correction for qualitative data

- ***DEMOGRAPHIC CHARACTERISTICS OF INDIVIDUALS WITH DURATION OF HIV-1$^+$ GREATER THAN 7 YEARS, NO HISTORY OF ART AND NO HISTORY OF OPPORTUNISTIC INFECTION***

Of 1,204, 312 (25.9%) were found to have no recorded history of opportunistic infection (Figure 3.2.1) and the demographic, immunology and virology data were compared between groups of individuals who had unstable CD4$^+$ T-cell count (n=258) and those who were classified as having long-term stable CD4$^+$ T-cell count (n=54), a group which included LTNP and HIC (Table 3.2.3). Among those with unstable CD4$^+$ T-cell count 15 (5.8%) had recorded death date while 4 (7.4%) in those with stable CD4$^+$ T-cell count had death date that was recorded. Between group comparisons of gender, ethnicity, duration since HIV-1$^+$ diagnosis, nadir CD4$^+$ T-cell count, recent CD8$^+$ T-cell count, nadir CD19$^+$ B-cell count and nadir CD16/56 natural killer (NK) cell count were significantly no different between the two groups. While significant between-group differences amid those with unstable and long-term stable CD4$^+$ T-cell count were found for age, HIV risk and highest recorded HIV-1 RNA plasma load (Table 3.2.3).

Of the 50 patients alive with long-term stable CD4$^+$ T-cell counts, 37 (74%) were identified and categorised into long-term stable low CD4$^+$ T-cell count group while 13 (26%) fulfilled LTNP criteria who had CD4$^+$ T-cell counts within the normal range and stable. Amongst whom 1 (7.7%) was identified as HIC where all but one HIV-1 RNA plasma load was 93 copies/mL plasma (Figure 3.2.1). Those who exhibited long-term stable low CD4$^+$ T-cell counts, all had at least one recorded CD4$^+$ T-cell count below the normal range, 3 of whom controlled HIV-1 RNA plasma load to BLD. No significant differences were found between long-term stable low CD4$^+$ T-cell count group and LTNP group for: gender, ethnicity, age, HIV risk, nadir CD19$^+$ B-cell count, nadir CD16/56 NK cell count, and highest recorded HIV-1 load. While significant differences in: duration since HIV-1$^+$ diagnosis, nadir CD4$^+$ T-cell count, and recent CD8$^+$ T-cell count were found between patients identified as long-term stable low CD4$^+$ T-cell count and LTNP (Table 3.2.3).

Table 3.2.3: Demographic, Immunological and Virological Parameters of Patients With Duration of HIV-1$^+$ >7 years, No History of ART and No History of Opportunistic Infection

	N=312			N=54		
	Unstable CD4$^+$ T-cell count (Total=258)	**Stable CD4$^+$ T-cell count (Total=54)**	**^p-value**	**Long-term stable low CD4$^+$ T-cell count (Total=39)**	**LTNP (Total=15)**	**^p-value**
Alive at most recent time point	n=243‡ (Died n=15)	N=50† (Died n=4)		n=37† (Died n=2)	n=13† (Died n=2)	
Demographics						
Gender (%)						
Female	26(10.7)	4(8.0)	0.751	3 (8.1)	1 (7.7)	0.584
Male	217(89.3)	46(92.0)		34 (91.9)	12 (92.3)	
Ethnicity (%)						
Caucasian	157(64.6)	34(68.0)	0.766	22 (59.5)	12 (92.3)	0.090
Black African	16(6.6)	2(4.0)		2 (5.4)	0 (0.0)	
Other	70(28.8)	14(28.0)		13 (35.1)	1 (7.7)	
Median age at last visit, years (IQR) [range]	39.9 (35.2 to 46.3) [23.5 to 69.5]	42.7 (36.8 to 49.9) [24.9 to 68.0]	0.078	41.7 (36.4 to 50.6) [24.9 to 67.1]	40.2 (39.2 to 45.4) [29.6 to 68.0]	0.782
HIV risk (%)						0.234
MSM	213 (87.7)	44 (88.0)	0.021	34 (91.9)	10 (76.9)	
Heterosexual	28 (11.5)	3 (6.0)		2 (5.4)	1 (7.7)	
Bisexual	2 (0.8)	3 (6.0)		1 (2.7)	2 (15.4)	
Median time since HIV-1$^+$ diagnosis, years (IQR) [range]	10.1 (8.3 to 13.7) [7.0 to 23.2]	10.2 (8.1 to 14.2) [7.0 to 24.1]	0.995	9.3 (7.6 to 13.0) [7.0 to 23.7]	11.9 (9.4 to 19.4) [7.4 to 22.3]	0.048
Immunology						
Median nadir CD4$^+$ T-cell count, cells/μL blood (IQR) [range]	303 (226 to 415) [18 to 1055]	337 (229 to 460) [17 to 1043]	0.404	296 (194 to 350) [17 to 400]	583 (512 to 675) [460 to 1043]	<0.001
Median most recent CD8$^+$ T-cell count, cells/μL blood (IQR) [range]	978 (714 to 1302) [168 to 3255]	1067 (755 to 1293) [328 to 3168]	0.613	946 (712 to 1219) [328 to 1803]	1222 (1185 to 1624) [721 to 3168]	0.015

	N=312			N=54		
	Unstable CD4$^+$ T-cell count (Total=258)	**Stable CD4$^+$ T-cell count (Total=54)**	**^p-value**	**Long-term stable low CD4$^+$ T-cell count (Total=39)**	**LTNP (Total=15)**	**^p-value**
Alive at most recent time point	n=243‡ (Died n=15)	N=50† (Died n=4)		n=37† (Died n=2)	n=13† (Died n=2)	
Median nadir CD19$^+$ B-cell count, cells/μL blood (IQR) [range]	90 (61 to 140) [4 to 584] n=147	100 (59 to 176) [15 to 348] N=46	0.354	112 (53 to 187) [15 to 348] n=34	94 (70 to 157) [41 to 269] n=12	0.881
Median nadir CD16/56 NK cell count, cells/μL blood (IQR) [range]	38 (18 to 80) [2 to 362] n=129	52 (24 to 89) [3 to 226] N=42	0.218	52 (25 to 90) [11 to 226] n=32	52 (24 to 61) [3 to 191] n=10	0.701
Virology						
Median highest recorded HIV-1 RNA plasma load, copies/mL plasma (IQR) [range]	39,022 (10,935 to 120,580) [<50 to 544,551] n=130	12,769 (1,547 to 64,902) [<50 to 533,774] N=44	0.007	13,022 (3,424 to 66,433) [<50 to 533,774] n=32	3,113 (11,00 to 42,238) [<50 to 481,775] n=12	0.377

‡, † See Figure 3.2.1 for definition of patient groups
^p-value using Mann-Whitney U test for quantitative data and χ^2 test with Yates' correction for qualitative data
Where data is unavailable for patients within a group, the number of patients for whom data available is detailed

Table taken from (Mandalia *et al.*, 2012)

- ***FREQUENCY OF LONG-TERM NON-PROGRESSORS AND HIV CONTROLLERS***

From a total of 14,227 HIV-1$^+$ patients, 13 were classified as LTNP and one of these was defined as HIC, whilst 37 were defined as long-term stable low CD4$^+$ T-cell counts and of these 3 were also found to control their HIV RNA load BLD (Figure 3.2.1). Of 14,227 patients 8,810 had been HIV-1$^+$ for <7 years. The 5,417 patients who had been HIV-1$^+$ seropositive for ≥7 years were then investigated for history of ART. Seventy eight percent (4,213) of patients who had been HIV-1$^+$ for ≥7 years had a history of receiving ART, either past or current.

Eight hundred and ninety two out of 1,204 individuals who had never received ART had a history of OI. The 312 patients who had no history of an OI, despite never receiving ART and being diagnosed as HIV-1$^+$ for ≥7 years, were then investigated for maintaining their CD4$^+$ T-cell count profiles.

Eighty one percent (258/312) had declining, unstable CD4$^+$ T-cell counts. Fifty four of 312 patients exhibited a stable CD4$^+$ T-cell count indicated by a CD4$^+$ T-cell count slope of ≥0 cells/μL blood per quarter, and so were investigated further. Four of the 54 patients with stable CD4$^+$ T-cell counts were noted as having died during follow-up due to undocumented reasons (i.e. not related to HIV-1$^+$ infection and who were excluded from further investigation of CD4$^+$ T-cell count profile (Figure 3.2.1).

Remaining 74% of the 50 patients in the analysis had a history of a CD4$^+$ T-cell count below the normal healthy range (<450 cells/μL blood). These 37 patients were classified as long-term stable low CD4$^+$ T-cell counts while 3 of these subjects were also found to have HIV-1 RNA plasma load consistently BLD (either <500 or <50 copies/mL plasma) and with reference to Figure 3.2.1. It is hypothesised that the CD4$^+$ T-cell counts of these subjects, although below the normal range, are likely to be perched around the subject's own normal range.

- ***FREQUENCY OF LONG-TERM STABLE LOW CD4$^+$ T-CELL COUNT***

Twenty six percent of the 50 patients in the analysis had a history of a CD4$^+$ T-cell count consistently above the normal healthy range (<450 cells/μL blood). These 13 patients fulfilled the criteria of CD4$^+$ T-cell count within the normal healthy range, at all available time points. So these fulfilled the criteria of LTNP. Viral load data from these 13 patients were investigated and found to be consistently BLD in one individual. Such data were available for this patient from 7 visit dates over a 54 month period and on all but one HIV-1 RNA plasma load assay 'blip' of 97 copies/mL plasma, the HIV-1 RNA plasma load was BLD of either 500 or 50 copies/mL plasma during the follow-up period. The variation in detection limit of the assay is due to the introduction of a new viral load assay with a lower detection limit. This patient was therefore considered an HIC. The CD4$^+$ T-cell counts are depicted for the remaining 13 LTNP Figure 3.2.3 and for those with long-

term stable low $CD4^+$ T-cell count in Figure 3.2.4. In both figures (Figure 3.2.3 and Figure 3.2.4) HIV-1^+ individuals with HIV-1 RNA plasma load BLD representing HIC are highlighted on the figure legend indicated by symbols used to plot the $CD4^+$ T-cell count data for these small number of individuals who controlled their HIV-1 RNA plasma load to BLD.

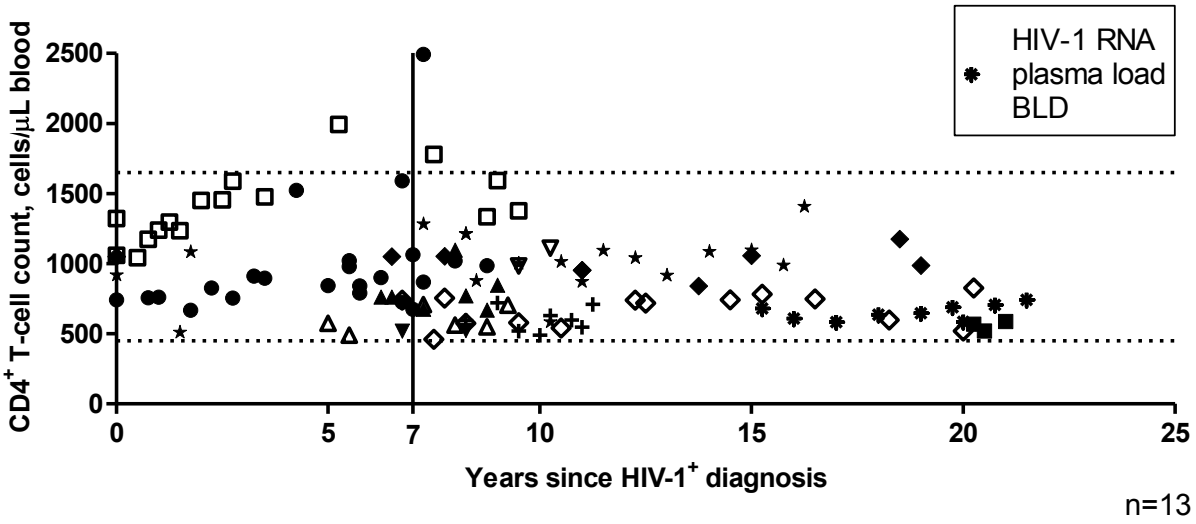


Figure 3.2.3: Distribution of $CD4^+$ T-Cell Count Since Enrolment in the Cohort in 13 LTNP
13 Long-term non-progressor including 1 HIC
(normal healthy range 450-1650 $CD4^+$ T-cell count cells/µL blood, marked with two dotted horizontal lines)
Graph taken with slight modification from (Mandalia *et al.*, 2012)

The 37 individuals with long-term stable low $CD4^+$ T-cell count profiles over time since HIV-1^+ diagnosis, count are depicted in Figure 3.2.4. All these 50 individuals displayed in Figures 3.2.3 and 3.2.4 displayed a stable slope of ≥0 cells/µL blood, per quarter.

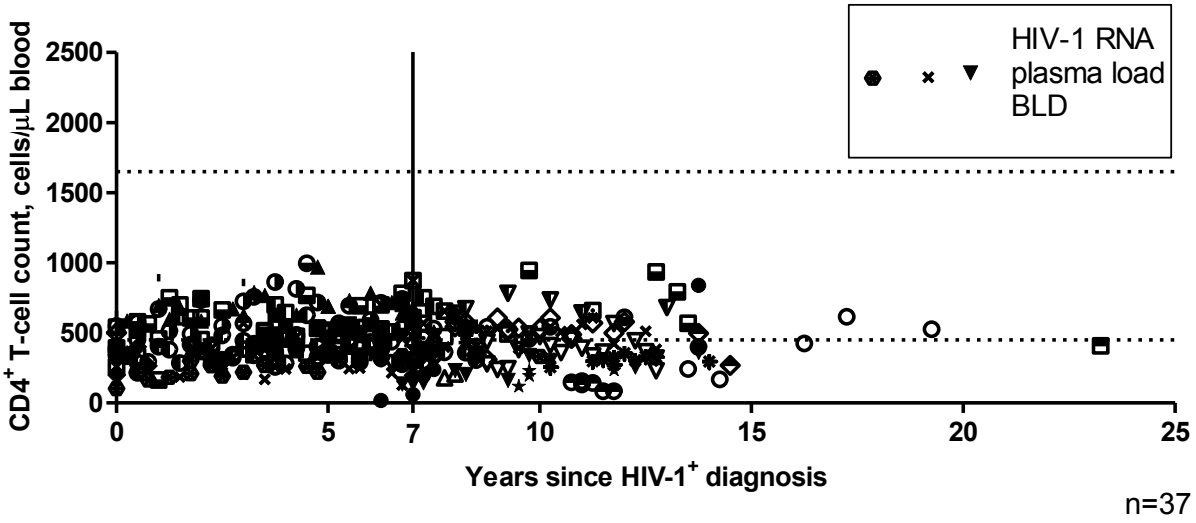


Figure 3.2.4: Distribution of $CD4^+$ T-Cell Count Since Enrolment in the Cohort in 37 Long-Term Stable Low $CD4^+$ T-cell Count
37 Long-term stable low $CD4^+$ T-cell count including 3 who controlled HIV replication to BLD
BLD = below the detection limit of the HIV-1 plasma viral load assay;
(normal healthy range 450-1650 $CD4^+$ T-cell count cells/µL blood, marked with two dotted horizontal lines)
Graph taken with slight modification from (Mandalia *et al.*, 2012)

- ***HIV NON-PROGRESSORS IDENTIFIED USING CONTROL OF VIRAL LOAD CRITERIA***

From a total HIV-1 cohort of 14,227 patients, 4,597 had 3 consecutive HIV-1 RNA plasma load <500 copies/mL plasma of whom 4,140 (90%) were on ART at the time when 3 consecutive HIV-1 RNA plasma load were <500 copies/mL plasma. Of the 4,597 patients, 457 (9.9%) were not on ART at the time when 3 consecutive HIV-1 RNA plasma load were <500 copies/mL plasma and these were further investigated to identify patients for whom these 3 consecutive HIV-1 RNA plasma load results spanned at least one year as this would discount repeat tests at close proximity in time. One hundred and two (22%) patients had 3 consecutive VL results that spanned at least one year in the absence of ART. Of these a total of 102 patients were identified and 46 (45%) of whom had no history on HIV-1 cohort database of ever receiving ART (Figure 3.2.5). Among the total of 102 patients who were identified, 7 of these were also part of a group of individuals who were identified as long-term stable low $CD4^+$ T-cell counts and LTNP groups described in Figure 3.2.1. Of these 7 individuals, 5 were classified into long-term stable low $CD4^+$ T-cell counts while the remaining 2 were classified as LTNP (Figure 3.2.5).

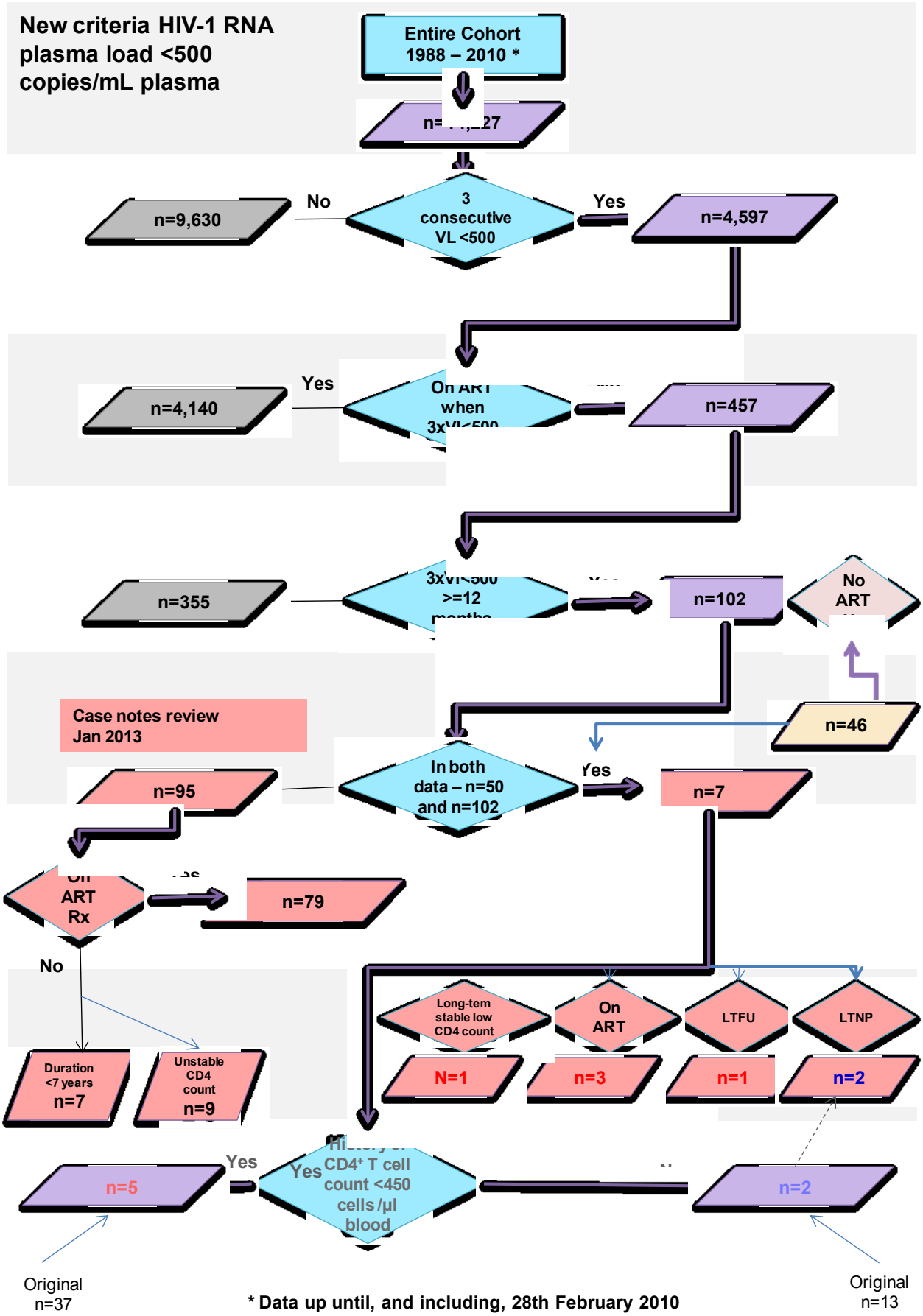


Figure 3.2.5: Flow Chart Showing Identification of HIV Non-Progressor Patient Groups Using Control of Plasma Viral Load Criteria from the Chelsea and Westminster HIV Cohort

- ***HIV CASE NOTES REVIEW***

For the 102 patients identified to have 3 consecutive HIV-1 RNA plasma load results BLD, case notes of all 102 individuals were reviewed by an independent clinician in January 2013 to establish the current (January 2013) 'atypical' status of the patients identified. Seven individuals were also part of a group of individuals who were identified as long-term stable low $CD4^+$ T-cell counts and LTNP groups. Of whom, 5 (from previous selection criteria represented in Figure 3.2.1) were also classified as long-term stable low $CD4^+$ T-cell counts while the remaining 2 were classified as LTNP using the HIV-1 RNA plasma load selection criteria (represented in Figure 3.2.5). The case note review showed that 3 of these 7 individuals had started ART by January 2013, 1 was lost to follow-up (LTFU) (using no evidence of clinic attendance within last 2 years as a definition of LTFU), while 2 of the remaining 7 continue to maintain LTNP status and the remaining 1 continues in the group of individuals with long-term stable low $CD4^+$ T-cell count (Figure 3.2.5).

Of the remaining 95 individuals, 79 (83%) were recorded on case notes to have received ART. Of the remaining 16 individuals who had no evidence of being prescribe ART, these 16 individuals were investigated further to establish reasons why these individuals were not part of the atypical individuals identified and represented in Figure 3.2.1. Of the 16 individuals who had not received ART (Figure 3.2.5): for 7 of these, duration of HIV-1^+ was <7 years at the time of analysis; while for the remaining 9 individuals, $CD4^+$ T-cell counts were defined as unstable as represented in Figure 3.2.1.

The distribution of HIV-1 plasma viral load, $CD4^+$ and $CD8^+$ T-cell counts data of these 9 individuals' are depicted in Figure 3.2.6. The descriptions of these illustrate reasons why these 9 patient were not identified in selection described in Figure 3.2.1 due to declining $CD4^+$ T-cell count slope since entry to cohort were found 'unstable' according to data represented in Figure 3.2.1 .

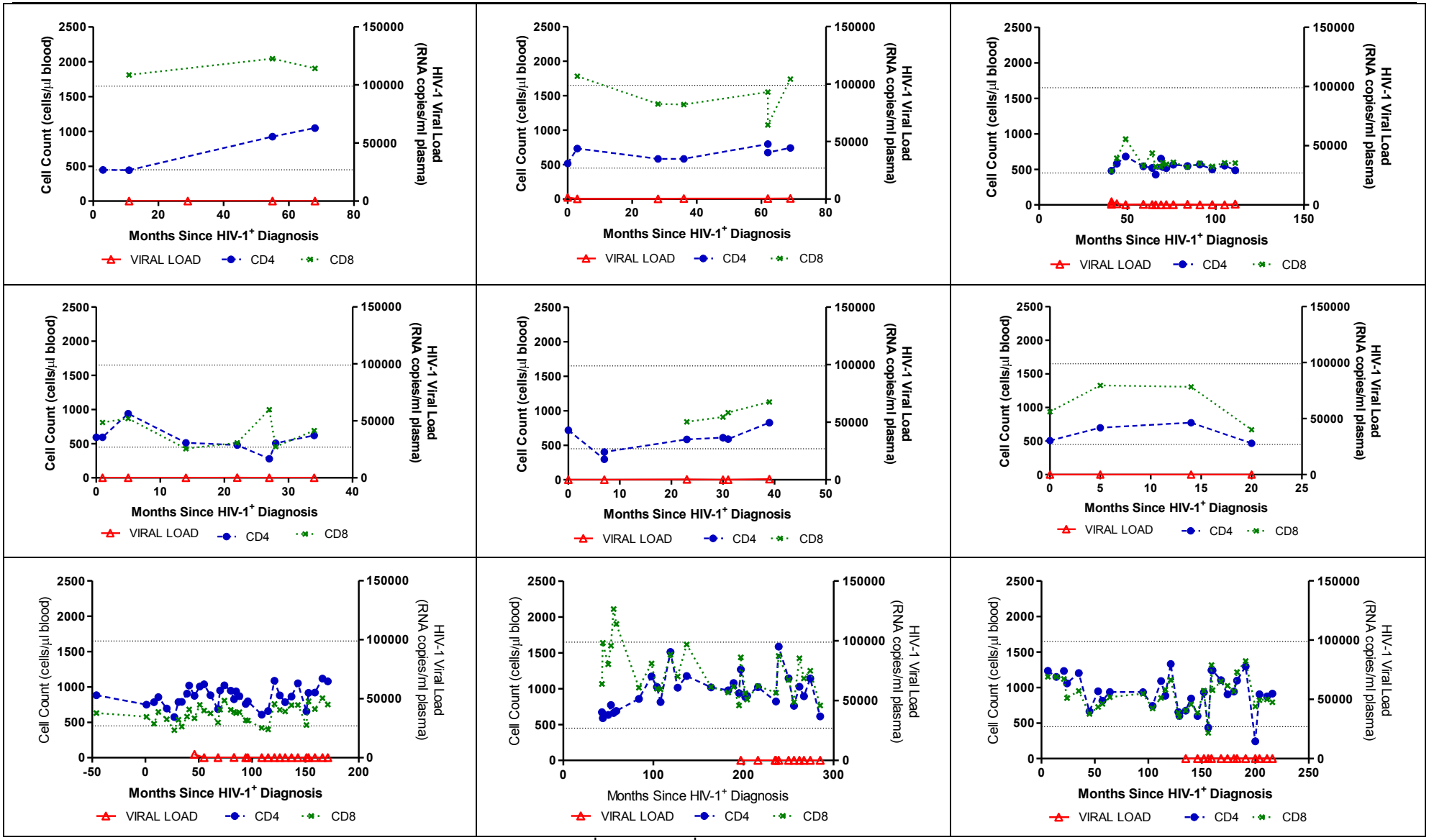


Figure 3.2.6: Distribution of HIV-1 Plasma Viral Load, $CD4^+$ and $CD8^+$ T-Cell Count Since Enrolment in the Cohort

3 consecutive HIV-1 RNA plasma load <500 copies/mL plasma within 12 month period in the absence of ART (n=9). These 9 patients were not identified in the initial selection described in Figure 3.2.1 due to unstable $CD4^+$ T-cell counts.

3.3 DISCUSSION

The caseload of HIV-1$^+$ patients attending the HIV-1 unit for treatment and care during February 2009 to February 2010 was 6,390, 13 (0.20%) were identified as LTNP, of whom one (0.02%) was an HIC. Thirty-seven (0.58%) were classified as exhibiting long-term stable low CD4$^+$ T-cell counts, of whom 3 (0.05%) were able to control their HIV-1 RNA load replication BLD. This demonstrates that the number of patients fulfilling the criteria used to define LTNP, HIC or long-term stable low CD4$^+$ T-cell counts within the Chelsea and Westminster HIV cohort is very small.

- ***NOVEL SELECTION CRITERIA APPROACH***

Studies on HIV non-progressors to date have used varying definitions of rate of CD4$^+$ T-cell count decline to identify the atypical patient groups but none have defined this with stringent criteria used as non declining since cohort entry that was used to identify LTNP and HIC. The Chelsea and Westminster HIV Cohort selection criteria included the stringent criterion of a non-declining CD4$^+$ T-cell count slope since cohort entry, combined with additional clinical criteria as reported here. Despite having used one of the largest HIV-1$^+$ cohorts in Europe to identify these atypical patients, only 50 (0.38%) exhibited maintenance of a long-term stable CD4$^+$ T-cell count, of whom only 13 could be confirmed as LTNP.

Studies on HIV-1$^+$ non-progressors to date have used varying definitions based on duration of HIV-1 infection and CD4$^+$ T-cell counts to identify atypical patient groups (Figure 3.1.1). In addition, most studies select these atypical patients who maintain their CD4$^+$ T-cell count profile within the normal range. This study suggests that by using varying selection criteria, disease progression is very likely in the majority of patients living with HIV-1. The patients who had not progressed within the study period are likely to do so, as demonstrated in the analysis of patients found to have long-term stable low CD4$^+$ T-cell counts compared to patients with unstable CD4$^+$ T-cell counts.

- ***SUMMARY OF FINDINGS***

Marked differences were observed in some of the immunological markers of HIV-1 disease progression between groups of patients. Current CD4$^+$ T-cell count, nadir CD4$^+$ T-cell count and plasma HIV-1 RNA load were shown to be statistically different between groups of patients, however no differences were observed in CD8$^+$ T-cell, CD19 B-cell or CD16/56 NK cell counts, indicating that the characteristics of HIV-1 disease, in the groups of i) LTNP, ii) patients with long-term stable low CD4$^+$ T-cell count and iii) individuals with unstable CD4$^+$ T-cell counts, may be independent of the humoral acquired and NK cell-mediated innate immune responses. However, due to the small numbers of individuals in the LTNP and long-term stable low CD4$^+$ T-cell count groups, power to detect a statistical significance was low. Small sample sizes decrease accuracy in point estimates used to describe quantitative data, including immunological and virological

markers, demonstrated by increased variability. A larger sample size would be required to reduce this variability, providing increased statistical power, sensitive hypothesis tests and decreased variability with narrower confidence intervals. In addition, studies that have investigated factors influencing $CD4^+$ T-cell counts have indicated that sex (Maini *et al.*, 1996), ethnicity (Tugume *et al.*, 1995), age (Hulstaert *et al.*, 1994) and behavioural (i.e. smoking) (Maini *et al.*, 1996) factors exert significant effects on $CD4^+$ T-cell counts.

Such potential effects on the observations reported above could not be performed due to small sample size, and relatively low numbers of women and ethnic diversity in our cohort. Furthermore, the infecting clade of HIV-1 and the geographic origin of infection are likely to impact the course of HIV-1 disease progression. However, clade data was only available for 19/50 patients with stable $CD4^+$ T-cell count, 16 of whom were infected with clade B and 3 with clade C virus. Of the 13 patients identified as LTNP, clade information was available for 5. Of these, 3 of whom were infected with clade B and 2 with clade C virus, while the clade for the remaining 8 LTNP have so far not been tested. However, the majority of HIV-1^+ individuals in the UK are infected with clade B virus (Westrop *et al.*, 2009). Infecting clade and viral fitness are important considerations for future work with the cases identified herein.

Previously published data from our group and others further indicate that disease progression in LTNP is ultimately inevitable, and virological, immunological and genetic factors have all been consistently reported to be associated with rate of disease progression (Easterbrook, 1994; Lefrère *et al.*, 1997; Westrop *et al.*, 2009). HIV-1 control is thought to be attributable to immunological, virological and genetic components.

The viral control exhibited by LTNP and HIC has been shown to be associated with HIV-1 specific $CD4^+$ and $CD8^+$ T-cell responses (Imami *et al.*, 2002; Deeks & Walker, 2007; Westrop *et al.*, 2009). Many LTNP are HLA-$B57^+$, an MHC-class I allele that demonstrates enhanced peptide presentation to cytotoxic $CD8^+$ T-cells on the surface of infected $CD4^+$ cells (Migueles *et al.*, 2000; Migueles *et al.*, 2008; Pereyra *et al.*, 2010).

In our opinion, the rate of disease progression in HIV-1^+ patients should be considered a continuous variable and not discrete. Studying patients at the extreme of this distribution may enable discovery of correlates of HIV-1^+ disease progression, leading to identification of targets to be manipulated with novel therapeutic approaches, with the ultimate goal of inducing delayed disease progression, retarding ART initiation and alleviating pill burden and toxicity. The immunological mechanisms involved need to be studied as this may also provide clues towards the development of prophylactic and therapeutic vaccines against HIV-1.

- ***LIMITATIONS***

It is possible that many more HIV-1^+ are LTNP or HIC, but due to the absence of clinical manifestations of disease these individuals have not yet attended an HIV-1 testing facility and consequently have not been diagnosed as HIV-1^+. The criteria of time from HIV-1 seropositive diagnosis should also be considered as relatively arbitrary, as the date of first HIV-1^+ test result, or date of first attendance at an HIV clinic, may not be an accurate representation of the date of infection. Instead, the date of first positive HIV-1 test should be considered a tool for healthcare professionals to identify patients exhibiting an unusual course of HIV-1 disease.

Data used for this work are based on routinely collected observation data and like for other such databases data limitations are found such as HIV-1^+ patients are lost to follow-up. Another limitation of observational data from a single centre is that HIV-1^+ patients can also be managed in other centres and may receive ARV treatment from other clinical care providers, including private practice, which may not be captured in the database of the single centre. This was one of the findings when data were further investigated using a HIV-1 RNA plasma load criteria as summarised in Figure 3.2.5 where 83% or 79 individuals who were found to have 3 consecutive HIV-1 RNA plasma load <500 spanning at least one year were recorded on patient case notes to have received ART while ART history in these individuals were not recorded on computerised HIV-1 database. Currently clinicians are reliant on patients' self reporting of ART treatment history including ART received elsewhere. The formation of a national centralised specialist observational database, highlighted by the UK governments Information Strategy (Department of Health, 2012) is a likely strategy to overcome such limitations in the future (for further details see Figure 1.1.6).

- ***REPORTS FROM OTHER STUDIES***

To date many publications have used various definitions and synonyms essentially describing a unique HIV-1 positive population who, without exposure to ARV drugs, maintain a healthy immune system and low level viremia despite being HIV-1 positive for at least 7 years. Reports from other cohorts where atypical groups of patients have been identified, even though they used different criteria from those stated here, confirm the paucity of PLHIV who could be classified as LTNP. A recent investigation of a large HIV-1 cohort in France identified only 0.4% of their sample as LTNP (Grabar *et al.*, 2009).

The data presented in this PhD thesis and other studies emphasise a need for collaboration with other cohorts when studying such atypical patients, such as work performed within the international GISHEAL consortium (Imami *et al.*, 2013), to increase the number of LTNP and HIC who can be identified and studied. In this study we demonstrate that using different criteria to define patient groups result in different estimates in time until disease progression, before even considering

functional immunology, genetic and virological factors, further emphasising the need for agreed standardised use of terminology and definitions before more in depth study of these individuals is performed. Figure 3.1.1 provides an indication of some of the different criteria used by different groups. Large cohorts of well-defined HIV-1^+ infected patients will be essential for future investigation of genetic associations with HIV-1 control and delayed disease progression (Deeks & Walker, 2007), as the unique immunological and virological responses demonstrated by LTNP and HIC may provide clues towards both the change in disease status of these patients over time, and provide insights for HIV-1 preventive or therapeutic vaccine development, sentiments reiterated by Francoise Barre-Sinoussi as part of the preparations for the June 2011 United Nations High Level Meeting in New York (Barre-Sinoussi, 2011).

- ***FUTURE WORK***

The individuals depicted in Figure 3.2.6 are currently being further investigated by a PhD student Nathali Grageda from the same department who is performing detailed immunological profiles and viral fitness in these individuals as these all show control of HIV-1 plasma viral load in the absence of ART.

Chapter 4: HIV and Ageing

CHAPTER 4: HIV AND AGEING ANALYSES: CO-EFFECTS OF HIV-1 INFECTION AND AGEING ON RESPONSE TO THERAPY, CO-MORBIDITY, PROGRESSION AND IMMUNE RE-CONSTITUTION ANALYSES

Note: A modified version of renal function data from this chapter has been submitted for publication.
Mandalia S, Herasimtschuk A, Farrington K, Westrop SJ, Nelson M, Gazzard BG, Imami N, Beck EJ
Effect of HIV-1 infection, first-line ART and Ageing on renal function

Everything that exists ages over time, and passes through three stages of development: embryo; growth and development; exhaustion and senescence, also known as the ageing process. The rate of ageing varies and there are theories of ageing that cover: genetic; biochemical; and physiological aspects. One of the problems associated with exploring senescence is distinguishing it between illnesses that are due to the ageing process or those that are caused by an underlying illness such as the HIV-1 infection.

HIV-1 diagnosed people who have been treated long-term with ARV drugs and who have consequently aged are at an increased risk of non-HIV co-morbidities that are associated with ageing. Some of these are: cardiovascular disease (e.g. myocardial infarction); cancer; haematologic disease (e.g. anaemia); neurocognitive decline (e.g. dementia); bone disease (e.g. osteoporosis); lung (e.g. hypertension); liver disease (e.g. fibrosis); kidney disease (e.g. insufficiency). Ageing in this PhD thesis chapter will be investigated as a result of disease specific or HIV-1 processes together with effect of changes to the thymus as a result of normal ageing.

4.0 EPIDEMIOLOGY OF AGEING

Epidemiological study of older adults in the general population, have been described in census studies for many years. The estimated age distribution profile described in Figure 4.0.1 is of the UK general population by gender for the years 1995, 2000, 2010 together with the expected ageing population by 2020.

The estimated ageing of the UK's population based on data from US Census Bureau 17th July 2003 illustrates the expected age distribution changes by 2020

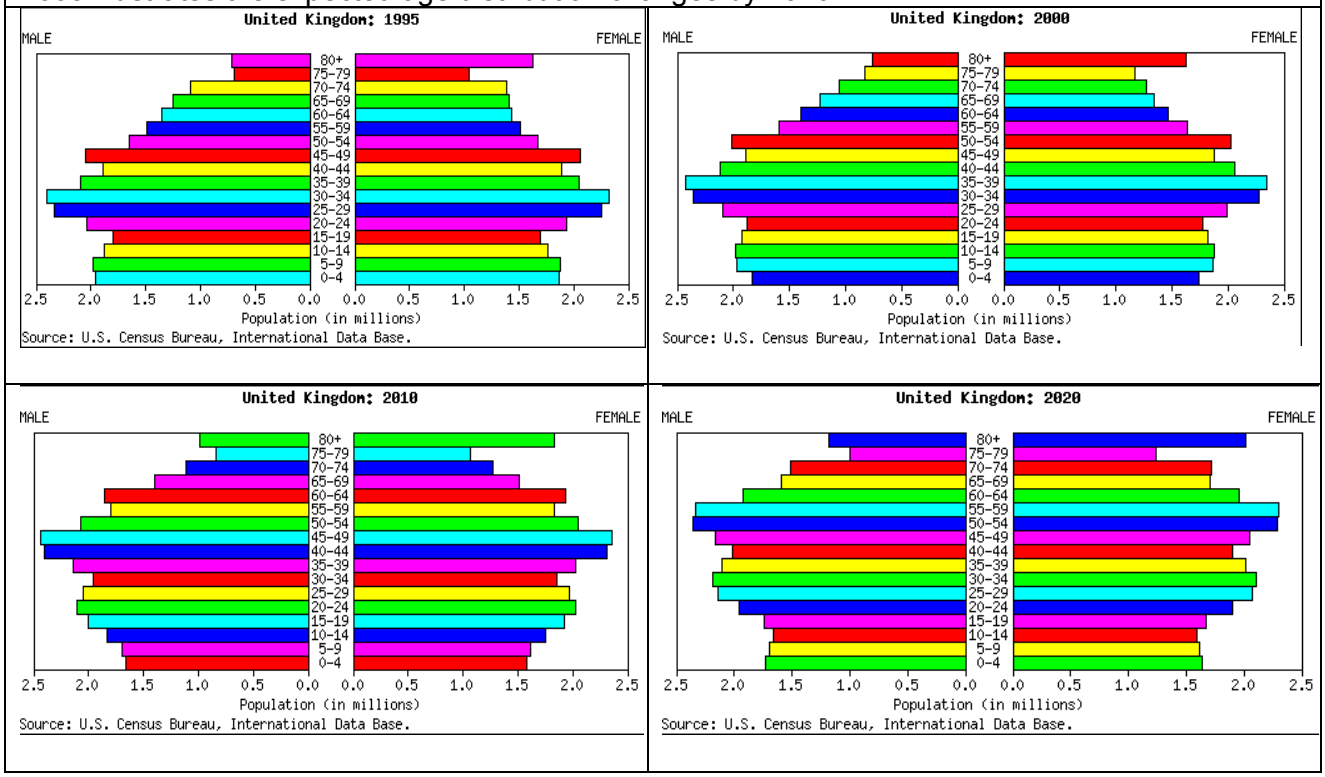


Figure 4.0.1: Population Pyramid Displaying Ageing of United Kingdom General Population by Gender
Accessed 30th March 2012: Figures taken from http://www.nationmaster.com/country/uk/Age_distribution

Based on the projected population and the increased life expectancy of the general population this has meant that there is an expected increase in older population. Due to advances in medicine, public health policies and socioeconomic development (Dorshkind *et al.*, 2009), these currently occurring and expected increased numbers of older people living in the general population has caused public heath directive for improved surveillance and control of diseases in the ageing population since susceptibility to infections and chronic diseases are often seen in advanced ages. In general there are differences in epidemiological study of 'older' compared to 'younger' and 'middle-aged' adult population. These are: differences in clinical manifestations of disease and conditions among older persons; variations of the 'natural history' trajectories and outcomes. Therefore this has led to exploration of disease occurrence, risk factors for morbidity and mortality and health outcomes.

A typical course of HIV-1 is illustrated in Figure 4.0.2 which demonstrates how HIV-1 infected patients have to make adjustments to life regarding dependence/independence. The increased susceptibility to infections and chronic diseases with ageing has societal implications as costs incurred in caring for increasing number of older HIV-1^+ people can have a significant impact on health-care system. The costs of health care services for acute and chronic conditions in general population are higher for the elderly, so the population growth in HIV-1^+ people is likely to have an

effect on the NHS health care costs. While the health care costs of caring for the elderly HIV-1$^+$ people is expected to rise, this is due to the patient's age as well as an increase in the complexity of the non-HIV co-morbidity in this group requiring tests and procedures. It is expected that as the proportion of older HIV-1$^+$ people increase, greater financial and social care support will be required for their care.

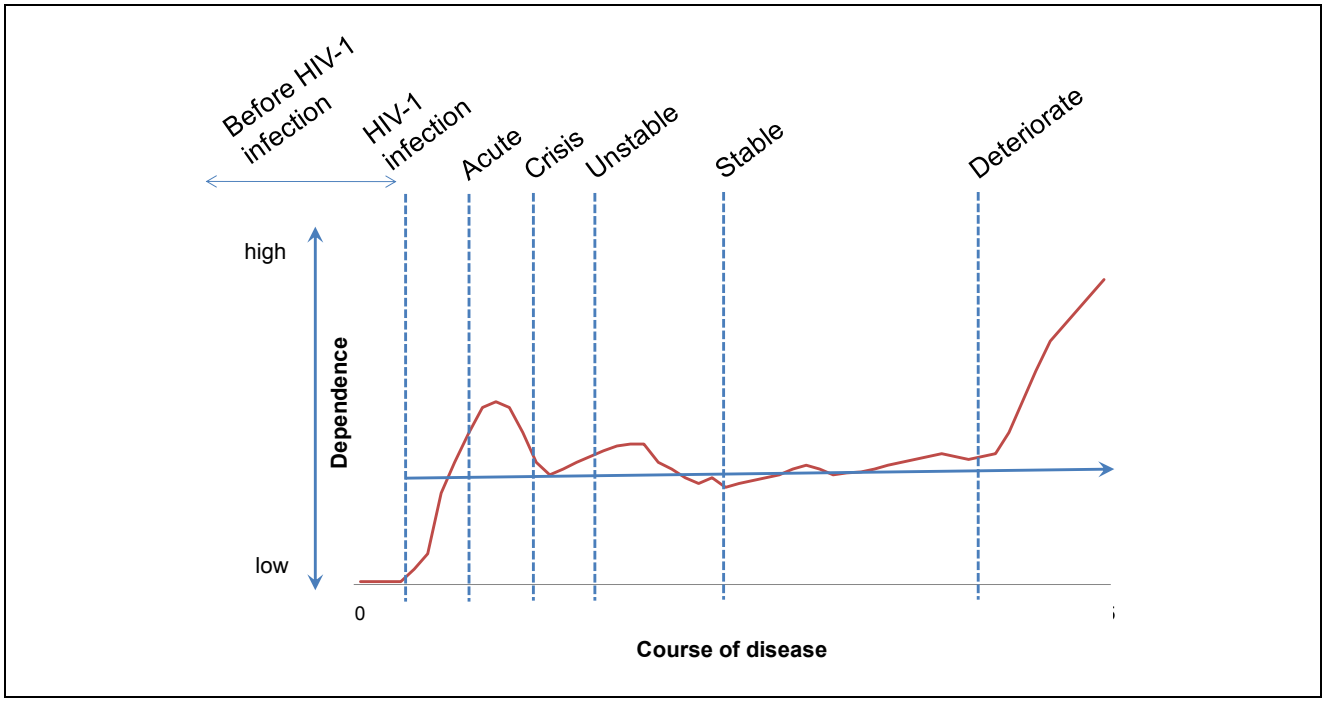


Figure 4.0.2: 'Trajectory' Illustrating the Course of a HIV-1 Infection in its Different Stages

4.1 INTRODUCTION

Globally the number of people estimated to be living with HIV-1 continues to increase. This was estimated at 34 million by end of 2011 (UNAIDS World AIDS Day Report, 2012) while at the end of 2012 it had increased to 35.3 million (range: 32.2 - 38.8) (UNAIDS Report on the Global AIDS Epidemic, 2013). By the end of 2012 the number of adults living with HIV-1 in the UK (diagnosed and undiagnosed) was 98,400 and 76,500 people diagnosed with HIV-1 had used the NHS services (credible interval: 75,000 to 78,000) (Aghaizu *et al.*, 2013). Further details of global and national HIV-1 infection are provided in Chapter 1: section 1.2.

Since the HAART era the mortality due to HIV-1 infection has declined resulting in longer survival. With the new and potent ARV drugs available and the resulting success of these together with increasing number of older people becoming infected with the virus HIV-1$^+$ patients have either survived to or become infected with the virus at age 50 years and beyond. The standard statistical reporting of HIV-1 cases in the UK by the HPA give data by 10-year intervals (<15 years, 15-24 years, 25-34 years, 35-49 years) until 49 years after which ages are grouped to ≥50 years (Aghaizu *et al.*, 2013).

CHAPTER 4

HIV AND AGEING

- ***SEXUAL ACTIVITY IN OLDER PEOPLE***

Many 'older' people are sexually active, although interest in sex and frequency of vaginal intercourse decline with age (Lindau *et al.*, 2007). Since the first FDA approval of erectile dysfunction drug in 1998 this has extended sex life of many older individuals. Many studies have shown that 'older' individuals are less likely than 'younger' individuals to practice safer sex and the use of erectile dysfunction drugs has been associated with risky sexual practices in 'older' individuals (Pantalone *et al.*, 2008).

- ***LIFE EXPECTANCY OF HIV-1$^+$ PEOPLE***

In a recent study an increased life-expectancy[20] of HIV-1$^+$ people on ART was demonstrated. Life-expectancy of HIV-1$^+$ aged 20 years in North America and Europe ranged from 32 years for those who start ART with a CD4$^+$ T-cell count less than 100 cells/μL blood to 50 years for those who start ART with CD4$^+$ T-cell greater than 200 cells/μL blood (ART Cohort Collaboration, 2008). Similarly, HIV-1 diagnosed people on ART in resource-limited countries can also attain near normal life-expectancy: Ugandan study estimated that life-expectancy of an HIV-1$^+$ individual aged 20 years, 35 years and 50 years or more was 27 years, 28 years and 24 years respectively. Life expectancy was observed to increase substantially with increasing baseline CD4$^+$ T-cell count. Similar trends were observed for individuals who were in older age groups (Mills *et al.*, 2011). The increased life-expectancy of HIV-1$^+$ people has given rise to the likelihood of developing non-HIV co-morbidities associated with ageing and the underlying process of immunosenescence (Spits, 2002).

- ***INCREASES IN 'OLDER' HIV-1$^+$ PEOPLE***

As life-expectancy increases, the number of 'older' HIV-1$^+$ people also increase, many of whom have been managed with long-term ART. Apart from increased life expectancy and increases in new HIV-1 diagnosis in people aged ≥50 years, number of people who are HIV-1$^+$ and aged ≥50 years are increasing. Some of the explanations for these increases are: with success of HAART people who are infected with HIV-1 in the 'middle aged' group are likely to move into the 'older' age category; the life expectancy in HIV-1 infected people continues to increase with 'older' age group enjoying better health which will increase the likelihood to undertake risky behaviour; age related changes in the body's immune function and protective barriers, such as the drying of vaginal mucosa in women, make older people more susceptible to the HIV-1 infection when exposed; 'older' adults may engage in risky sexual behaviours such as not using condoms, for reasons including lack of concern about birth control, lack of awareness about HIV risks, and difficulty manipulating protective devices due to conditions such as arthritis.

Chapter 4:20 Life expectancy is the statistical figure based on the average person's length of life and is usually quoted as the number of years of life remaining at a given age. http://stats.oecd.org/glossary/detail.asp?ID=1530

CHAPTER 4

HIV AND AGEING

As people who are HIV-1$^+$ grow older, they are at an increased risk of developing other non-HIV co-morbidities, in particular cardiovascular disease, non-AIDS related cancers, neurological complications, bone abnormalities and 'frailty', liver and renal problems (Goetz, 2013; Deeks, 2011). The number of HIV-1$^+$ people ≥50 years is increasing rapidly in other high income countries such as the US as it is in the UK (Vance *et al.*, 2011; Aghaizu *et al.*, 2013). As a result of these, the number of people living with HIV-1 who are ≥50 years of age continues to increase in the UK where the infection rate in this age group increased by more than three-fold between 2000 and 2007; from 299 new cases in 2000 to 710 in 2007 (Smith *et al.*, 2010).

The older HIV-1$^+$ people are now facing a complex 'ageing' process involving biological, social and psychological aspects. In addition to providing 'older' HIV-1$^+$ people with HIV-1 specific services, these need to be integrated with other specialist services: cardiovascular; renal; hepatic; neurological; endocrine; and other specialist services; as well as integrated with general geriatric services (Justice, 2010).

- ***NON-HIV CO-MORBIDITIES***

Increasing numbers of studies are reporting that HIV-1$^+$ infected people who are ageing while receiving ART treatment are at a risk of developing other chronic conditions that generally occur in 'older' age such as cardiac and metabolic complications that are associated with ageing (Broder, 2010).

Ageing is a natural process that involves a gradual progressive decline in tissue and organ function. The increasing susceptibility to environmental challenges and an increasing risk of disease and death is the biological process of ageing. The increased numbers of HIV-1$^+$ people ≥50 years seen has led to an exploration of non-HIV co-morbidity and the risk factors associated with ageing so that there is an improved surveillance and control of conditions in the older HIV-1$^+$ population. The characteristics of HIV-1 infection in different age groups will be described in this chapter. Data presented in this chapter explores one of the non-HIV co-morbidity, renal function. The co-effects of HIV-1 infection, ageing and effect of ART on renal function is explored and described.

4.1.1 BIOLOGY OF AGEING

The oldest recorded human life span is held by Jeanne Calment of France, who died in 1997 at the age of 122 years, 164 days[21]. Natural ageing is a gradual reduction in the functional capacity of an individual without the onset of severe disabilities and these long surviving people remain active into their 90's and beyond. While Jeanne Calment followed an active and a 'healthy' lifestyle many who also follow similar lifestyle do not survive to such an age.

The rate of biological ageing (based on the quality of an individual's bodily systems) varies widely and is related to a gradual decline in functional reserve and is associated with the development of morbidity and mortality in the 'elderly' and in patients with HIV-1^{+} (Deeks, 2011). There is however variability associated with the decline in functional reserve and lifestyle choices, environmental factors, genetics, and presence of co-morbidities. For example, loss of muscle strength, blood circulation, skin elasticity, joint flexibility and immune capacity occur faster in some people than in others. This is due to gene expression; the process by which information from a gene is used in the synthesis of a functional gene product (Anton *et al.*, 2005). This means our genes do not change, but their expression changes. This may be formulated as:

(life style, environment, diet) + (genotype[22]) = (phenotype[23])

(Anton *et al.*, 2005)

Ageing process involves accumulation of random molecular defects that build up within cells and tissues. These defects start to arise very early in life and over time the molecular defects increase, resulting eventually in age-related functional impairment of tissues and organs Figure 4.1.1 (Kirkwood, 2008).

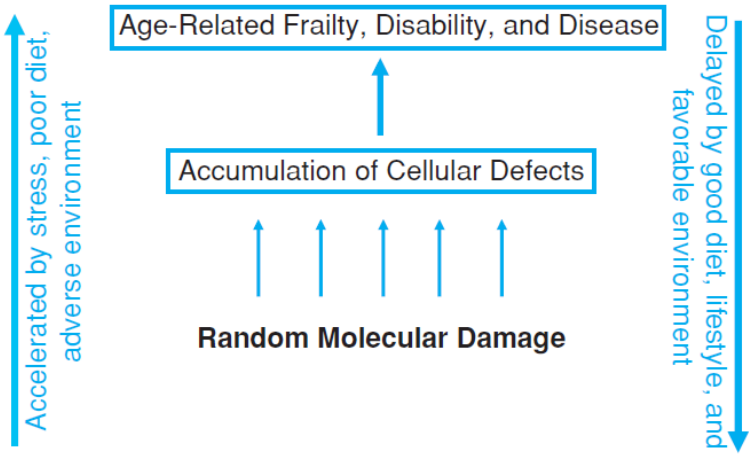


Figure 4.1.1: The Ageing Process and Accumulation of Molecular Damage
Figure taken from Kirkwood, 2008

Chapter 4:21 http://listverse.com/2010/02/07/top-10-oldest-people-ever/
Chapter 4:22 Entire set of genes in a cell in an individual
Chapter 4:23 The observable physical or biochemical characteristics of an organism, determined by genetic make-up and environmental influences

- ***VARIATIONS IN DEFINITIONS OF AGED***

Ageing is a continuous process and for example developing cataracts is considered as part of 'normal ageing' since this develops with increased age. World Health Organization report most developed countries define age ≥65 years as 'elderly' and United Nations age cut-off ≥60 years is referred to the 'older' population (World Health Organisation, 2010b). However, the Centers for Disease Control and Prevention define older HIV-1^+ people as those aged ≥50 years because originally 50 years was considered the upper age range for HIV-1 infection (Linsk, 2000).

4.1.2 IMMUNOLOGICAL AGEING

- ***AGE RELATED CHANGES TO THYMUS***

The development and maturation of T-cells occur in the thymus, which is a specialised organ of the immune system. In immunological terms, ageing is associated with a chronic inflammatory process and thymic changes that are expressed by changes in the total T-cells as well as T-cell subsets (Jamieson *et al.*, 1999), Thymic function begins to decrease after birth, a process that accelerates at puberty (Buchholz *et al.*, 2011). A steeper decline in naive $CD4^+$ and $CD8^+$ T-cells can be observed after the age of 40 years (Douek *et al.*, 1998) and thymic involution is almost complete by age of 50 years (Brunner *et al.*, 2011); after the age of 50 years, thymic epithelial space is reduced by 80% (Haynes *et al.*, 2000). While thymic involution starts in early life, the thymopoietic potential per cell remains intact till around 50 years of age (Gruver *et al.*, 2007). This immunological process is often referred to as 'immunosenesence', and it is this immunological basis that people aged 50 years or more are classified as being 'older' in the work presented in this chapter. The age group used in this thesis are summarised in Table 4.1.1.

Table 4.1.1: Defined Age Group Terms

<table>
<tr><th>Term</th><th>Age (years, inclusive)</th><th>Used term in this PhD thesis</th><th></th></tr>
<tr><td rowspan="2">Denarian</td><td rowspan="2">10 to 19</td><td></td><td rowspan="3"></td></tr>
<tr><td rowspan="3">Young adults:
18-39</td></tr>
<tr><td>Vicenarian</td><td>20 to 29</td></tr>
<tr><td>Tricenarian</td><td>30 to 39</td><td></td></tr>
<tr><td>Quadragenarian</td><td>40 to 49</td><td>Middle-aged adults:
40-49</td><td rowspan="9">Decline in thymic function commences on average from ≥40 years

+

Beginning of immune senescence on average commences ≥50 years</td></tr>
<tr><td>Quinquagenarian</td><td>50 to 59</td><td rowspan="8">Older adults:
≥50</td></tr>
<tr><td>Sexagenarian</td><td>60 to 69</td></tr>
<tr><td>Septuagenarian</td><td>70 to 79</td></tr>
<tr><td>Octogenarian</td><td>80 to 89</td></tr>
<tr><td>Nonagenarian</td><td>90 to 99</td></tr>
<tr><td>Centenarian</td><td>100 to 109</td></tr>
<tr><td>Super centenarian</td><td>110 and older</td></tr>
</table>

- ***IMMUNOSENESENCE***

Immunosenecence is the deterioration of the immune system as a result of advancing age and is associated with a decline in naïve T-cell output which is thought to contribute to the reduction in T-cells in older people which is likely to cause an increased susceptibility to infections in older people (Aw & Palmer, 2011). The patho-physiological[24] changes observed in HIV-1$^+$ people are similar to those seen in older people not infected with HIV-1 and these changes are characterised by ongoing inflammation, immune activation and increased coagulability (Kuller *et al.*, 2008; Tien *et al.*, 2010; Eastburn *et al.*, 2011) which are similar to changes associated with 'immunosenescence' (Gress & Deeks, 2009).

4.1.3 WHITE BLOOD CELL COUNT

The White Blood Cell (WBC) counts (leukocytes) are a component of the body's immune system. Indications for a WBC count include infectious and inflammatory diseases.

Composition of WBC (leukocytes)		
	Lymphoid	**Myeloid**
Mono nuclear cells	Lymphocytes Dendritic cells	Monocytes
Granulocytes		Neutrophils Eosinophils Basophils

- ***T-LYMPHOCYTES***

The thymic derived lymphocytes originate in the bone marrow, develop and differentiate within the thymus and move through the lymphatic system and bloodstream to reside in the lymph nodes and spleen. T-cells comprise 70% to 80% of the peripheral blood lymphocyte population and have an average half-life of more than 2 years. The T-cell is important in providing protection against infections by fungi (e.g. Candida), viruses (e.g. HIV-1, Cytomegalovirus), and bacteria (e.g. Chlamydia, Tuberculosis).

- ***T-HELPER AND T-SUPPRESSOR/CYTOTOXIC LYMPHOCYTES***

T-helper cells (these help the immune system by recognizing foreign substances on contact) account for approximately 40% to 60% of the peripheral blood T-lymphocyte population and they carry a protein on their surface called CD4. The T-helper cell's CD4 surface protein is the site on the cell to which the HIV-1 virus attaches. A second important subpopulation of T lymphocytes is the CD8$^+$ T-cell known as the T-suppressor/cytotoxic or killer cells that expresses the surface protein called the CD8 molecule.

Chapter 4:24 The functional changes associated with or resulting from disease

- ***LYMPHOCYTE SUBSETS***

Immunoscenescence is the result of readjustment of immune cell functions and phenotypes (Effros *et al.*, 2008). Ageing and atrophy of thymus is associated with dysfunction of stem cells and the decline in total number of T lymphocyte subsets (CD3 T-cell, CD4 T-cell, CD8 T-cell). The HIV-1 cohort data routinely measure lymphocyte subsets (including B-cell and NK-cell) to monitor HIV-1 infected patients morbidity (Burton *et al.*, 2005), and the following results section describe results of routinely collected and available data on these.

4.2 RESULTS

- ***DEMOGRAPHIC CHARACTERISTICS BY IMMUNE CONSIDERED AGE GROUPS***

The HIV-1 cohort comprised of 15,241 HIV-1^+ patients by May 2011 which was defined as the end of study period, 89% of whom were men, 70% Caucasian with a minority, 9% black African. Eighty-one percent were MSM and only 3% were past or current IDU.

The proportion of older HIV-1^+ patient was less among black African while older patients were more likely to be bisexual, less likely to have an IDU history and more likely to have been living longer with HIV-1^+ (Table 4.2.1). Relative to total women (n=1,673) and men (n=13,568) in the cohort, there were significantly fewer 'older' ≥50 women 12% (95% CI: 11% to 14%) compared to 'older' men 17% (95% CI: 16% to 18%), while significantly higher proportion of Caucasian were 'older' 19% (95% CI: 18% to 19%) compared to 'older' black Africans 13% (95% CI: 11% to 15%) and 'older' other ethnic group 11% (95% CI: 10% to 12%). Significantly higher proportion of bisexuals were 'older' 24% (95% CI: 20 to 28) compared to 'older' MSM 16% (95% CI: 16% to 17%) and 'older' heterosexual 17% (95% CI: 15 % to 18%).

Table 4.2.1: Demographic Characteristics by Recent Age of the Chelsea and Westminster HIV Cohort (1988-2011)

		Recent age (years) Grouped by immune considerations *(Age known and ≥18 year olds)*			
		18-39	**40-49**	**≥50**	
	Total cohort = 15,241	**N=7769**	**N=4961**	**N=2511**	**p-value**
Demographic characteristics:					
#Gender					
Female	1673	961 (12.4)	504 (10.2)	208 (8.3)	<0.001
Male	13568	6808 (87.6)	4457 (89.8)	2303 (91.7)	
#Ethnicity					
Caucasian	10731	5176 (66.6)	3564 (71.8)	1991 (79.3)	<0.001
Black African	1329	679 (8.7)	481 (9.7)	169 (6.7)	
Other	3181	1914 (24.6)	916 (18.5)	351(14.0)	
#Sexual orientation					
MSM	12385	6370 (82.0)	4017 (81.0)	1998 (79.6)	<0.001
Heterosexual	2355	1179 (15.2)	782 (15.8)	394 (15.7)	
Bisexual	501	220 (2.8)	162 (3.3)	119 (4.7)	
#IDU					
No	14773	7494 (96.5)	4819 (97.1)	2460 (98.0)	<0.001
Yes	468	275 (3.5)	142 (2.9)	51 (2.0)	
&Median (IQR) Duration since HIV-1$^+$ diagnosis, months	66.5 (23.6 to 131.2)	41.0 (13.1 to 81.1)	97.0 (40.3 to 158.3)	126.7 (57.6 to 188.9)	<0.001

MSM=Men who have sex with men
#p-value using χ^2 test
&Duration since HIV diagnosis is defined as HIV positive date to the most recent visit to clinic, p-value using Kruskal-Wallis test
See Table 4.1.1 for age group terms

- **Started First-line 2NRTIs+NNRTI or 2NRTIs+boosted PI**

There were 4,852 patients who started first-line 2NRTIs+NNRTI or 2NRTIs+boosted PI, 86% (n=4186) were men, 84% (n=3,506) were Caucasians and 85% (n=3,807) were MSM. Seventy-nine percent started on first-line 2NRTI+NNRTI and 21% on 2NRTI+boosted PI. There was a significant increase in NNRTI containing regimens with increasing age (χ^2 for trend, p=0.002; Table 4.2.2). The demographic characteristics of patients who started these first-line regimens were comparable with the entire HIV-1 cohort. 'Older' people (≥50 years) were less likely to be black Africans, more likely to be bisexual, less likely to be IDU and more likely to have been living with HIV-1$^+$ for longer periods (Table 4.2.2).

Table 4.2.2: Demographic Characteristics by Recent Age of Patients Who Started First-line HAART Consisting of 2NRTI+NNRTI or 2NRTI+boosted PI

		Recent age (years) Grouped by immune considerations *(Age known and ≥18 year olds)* **Patients who started first-line HAART consisting of 2NRTI + NNRTI or 2NRTI + boosted PI**			
	N= 4,852	**18-39 N=2,028**	**40-49 N=1,948**	**≥50 N=876**	**p-value**
Demographic characteristics:					
#Gender					
Female	666	335 (16.5)	238 (12.2)	93 (10.6)	<0.001
Male	4186	1693 (83.5)	1710 (87.8)	783 (89.4)	
#Ethnicity					
Caucasian	3506	1391 (68.6)	1425 (73.2)	690 (78.8)	<0.001
Black African	625	281 (13.9)	253 (13.0)	91 (10.4)	
Other	721	356 (17.6)	270 (13.9)	95 (10.8)	
#Sexual orientation					
MSM	3807	1614 (79.6)	1526 (78.3)	667 (76.1)	<0.001
Heterosexual	961	399 (19.7)	380 (19.5)	182 (20.8)	
Bisexual	84	15 (0.7)	42 (2.2)	27 (3.1)	
#IDU					
No	4756	1992 (98.2)	1902 (97.6)	862 (98.4)	<0.001
Yes	96	36 (1.8)	46 (2.4)	14 (1.6)	
&Duration since HIV-1$^+$ diagnosis, months	81.9 (42.0 to 133.2)	59.9 (30.6 to 95.4)	100.7 (56.7 to 147.3)	112.4 (63.3 to 149.5)	<0.001

MSM=Men who have sex with men
#p-value using χ^2 test
&Duration since HIV diagnosis is defined as HIV positive date to the most recent visit to clinic, p-value using Kruskal-Wallis test
See Table 4.1.1 for age group terms

- **Exposure to Different Anti-retroviral Drug Classes**

Of the total cohort who had attended for HIV-1 treatment and care (n=15,241), 6,923 (45%) were ART naive. A progressive and significant gradual increase in ART exposure was observed with increasing age (χ^2 for trend, p<0.001; Table 4.2.3) and over 70% of the 'older' (≥50 year) patients in the cohort were ART experienced. The median duration of ARV class exposure increased significantly with increasing age, from 1.2 years in 18-20 year olds to 9.6 years for those older than 70 years (Kruskal-Wallis test, p<0.001). Of the ART experienced patients (n=8,318), 4,314 (52%) had been exposed to PIs, whilst 5,982 (71%) had been exposed to NNRTIs (Table 4.2.3).

Table 4.2.3: Anti-retroviral Class Exposure by Recent Age of HIV-1 Cohort (1988-2011)

		Recent age (years) Grouped by immune considerations *(Age known and ≥18 year olds)* **N=15,241** **N(%)**			
		18-39	**40-49**	**≥50**	
	Total cohort = 15,241	**N=7,769**	**N=4,961**	**N=2,511**	**p-value**
Anti-retroviral treatment:					
#Duration of ARV exposure, years	4.4(1.6 to 9.6) Range: 0.01 to 23.5	2.4 (0.8 to 4.7) Range: 0.01 to 23.1	5.9 (2.4 to 10.8) Range: 0.01 to 23.3	8.8 (3.7 to 13.7) Range: 0.09 to 23.5	<0.001
&ART treatment ever Treatment naive Experienced	 6923 8318	 4692 (60.4) 3077 (39.6)	 1631 (32.9) 3330 (67.1)	 600 (23.9) 1911 (76.1)	 <0.001
&NRTI experienced No Yes	 40 8278	 21 (0.3) 3056 (39.3)	 10 (0.2) 3320 (66.9)	 9 (0.4) 1902(75.8)	 <0.001
&PI experienced No Yes	 4004 4314	 1725 (22.2) 1352 (17.4)	 1524 (30.7) 1806 (36.4)	 755 (30.1) 1156 (46.0)	 <0.001
&NNRTI experienced No Yes	2336 5982	1069 (13.8) 2008 (25.9)	828 (16.7) 2502 (50.4)	439 (17.5) 1472 (58.6)	<0.001
&Other ARV drug class experienced No Yes *(Includes Fusion Inhibitors, Integrase Inhibitors, CCR5 antagonists)*	 7611 707	 2922 (37.6) 155 (2.0)	 3035 (61.2) 295 (6.0)	 1654(65.9) 257(10.2)	 <0.001

#Median (inter-quartile range: IQR), p-value using non parametric method test to test for differences in averages between the three age group strata using the Kruskal-Wallis method
&p-value for qualitative data are based on χ^2 test
No ART history where no ARV data available on the database for HIV-1^+ patients in the cohort
Total duration of ARV exposure defined as cumulative duration of all combinations prescribed over time
See Table 4.1.1 for age group terms

4.2.1 CROSS SECTIONAL DATA

This cross-sectional method used included results of various parameters taken at a specific time. Below are results based on most recent results that were available in people with HIV-1 infection. Although the age changes are not measured directly, the average differences between the age groups is presented. Therefore, the effect of ageing was inferred from the between age group differences in averages. These differences are however confounded by the cohort effect using epidemiological definition as described in a paper by Keyes and colleagues (Keyes *et al.*, 2010).

- ***CHARACTERISTICS OF IMMUNE MARKERS BY IMMUNE CONSIDERED AGE GROUPS***
  - **CD4 T-Cell Count**

From the total cohort who had attended for HIV-1 treatment and care the median (IQR) nadir CD4 T-cell count by age showed a significant inverse trend with a low median nadir CD4 T-cell count among the 'older' people (≥ 50 years) 125 (IQR: 39 to 232) cells/μL blood compared to those who were between 18-39 and 40-49 respectively: 227 (IQR: 64 to 378); and 158 (IQR: 47 to 281) cells/μL blood (Table 4.2.4). Similarly, median CD4 T-cell count at cohort entry was significantly lower among the 'older' individuals 297 (IQR: 142 to 485) cells/μL blood compared to those who were 18-39 years (354, IQR: 195 to 524 cells/μL blood) or 40-49 years (333, IQR: 166 to 515 cells/μL blood). In contrast recent median CD4 T-cell among 'younger' people aged 18-39 years was lower (391, IQR: 144 to 582 cells/μL blood) compared to those who were 40-49 years or ≥50 years and these were respectively: 450 (IQR: 227 to 652); and 458 (IQR: 249 to 662) cells/μL blood

Table 4.2.4: Immunological and Virological Markers by Recent Age of HIV-1 Cohort (1988-2011)

		Recent age (years) Grouped by immune considerations *(Age known and ≥18 year olds)* **N(%)**			
		18-39	**40-49**	**≥50**	
	Total cohort = 15,241	**N=7769**	**N=4961**	**N=2511**	**p-value**
Immunological information:					
#Nadir CD4 T-cell count, cells/μL blood	179 (50 to 322)	227 (64 to 378)	158 (47 to 281)	125 (39 to 232)	<0.001
#First ever CD4 T-cell count, cells/μL blood	338 (174 to 513)	354 (195 to 524)	333 (166 to 515)	297 (142 to 485)	<0.001
#Recent CD4 T-cell count, cells/μL blood	423 (192 to 619)	391 (144 to 582)	450 (227 to 652)	458 (249 to 662)	<0.001
#First ever CD4 T-cell percent	20.3 (12.0 to 26.6)	20.3 (5.0 to 21.3)	20.0 (11.9 to 28.1)	18.0 (10.0 to 26.6)	<0.001
#Recent CD4 T-cell percent	24.2 (13.6 to 32.3)	22.9 (11.7 to 31.1)	25.7 (15.0 to 33.4)	25.3 (16.2 to 32.9)	<0.001
Cytotoxic T-cell counts:					
#Recent CD3 T-cell count, cells/μL blood	1356 (947 to 1776)	1335 (920 to 1748)	1370 (977 to 1804)	1378 (954 to 1800)	<0.001
#Recent CD3 T-cell percent	78.6 (71.6 to 84.2)	79.9 (73.0 to 85.2)	77.9 (71.0 to 83.6)	76.8 (69.9 to 82.7)	<0.001
#Recent CD8 T-cell count cells/μL blood	792 (530 to 1111)	789 (520 to 1102)	790 (537 to 1106)	804 (544 to 1141)	0.174

#Recent CD8 T-cell percent	50.3 (40.7 to 61.0)	52.0 (42.6 to 62.8)	48.7 (39.4 to 59.8)	48.2 (38.6 to 58.8)	<0.001
B cell counts:					
#Recent CD19 B-cell count, cells/μL blood	179 (114 to 271)	166 (107 to 245)	193 (125 to 291)	185 (113 to 293)	<0.001
Natural killer cells:					
#Recent CD56 NK cells count cells/μL blood	111 (62 to 190)	97 (54 to 168)	114 (64 to 191)	142 (81 to 235)	<0.001
Virological information:					
#Highest recorded HIV-1 RNA plasma load, copies/mL plasma	69,358 (10,231 to 251,204)	55,429 (10,067 to 185,859)	81,534 (11,404 to 291,856)	93,666 (9,588 to 311,919)	<0.001

#Median (inter-quartile range: IQR), p-value using non parametric method test to test for differences in averages between the three age group strata using the Kruskal-Wallis method
See Table 4.1.1 for age group terms

- **CD4 T-Cell Percentage**

First median CD4 percent for people aged: 18-39; 40-49; ≥50 years were 20.3 (IQR: 5.0 to 21.3), 20.0 (IQR: 11.9 to 28.1), and 18.0 (IQR: 10.0 to 26.6) respectively. While recent CD4 percent were: 22.9 (IQR: 11.7 to 31.1); 25.7 (IQR: 15.0 to 33.4); 25.3 (IQR: 16.2 to 32.9); for 18-39; 40-49; ≥50 years respectively.

- **CD3 T-Cell Count**

The recent median CD3 T-cell count result of people who were 18-39 years was lower which was 1,335 (IQR: 920 to 1748) cells/μL blood compared to those who were 40-49 years where this was 1,370 (IQR: 977 to 1804) cells/μL blood and it was 1,378 (IQR: 954 to 1800) cells/μL blood for those who were ≥ 50 years.

- **CD3 T-Cell Percentage**

The most recent available median CD3 T-cell percentage by immune considered categorised age groups of 18-39, 40-49 and those who were ≥ 50 years were: 79.9 (IQR: 73.0 to 85.2); 77.9 (IQR: 71.0 to 83.6); and 76.8 (IQR: 69.9 to 82.7) respectively.

- **CD8 T-Cell Count**

The recent $CD8^+$ T-cell count respectively for 18-39, 40-49 and those who were ≥ 50 years were: 789 (IQR: 520 to 1102); 790 (IQR: 537 to 1106); and 804 (IQR: 544 to 1141) cells/μL blood.

- **CD8 T-Cell Percentage**

The recent $CD8^+$ T-cell percent for 18-39, 40-49 and those who were ≥ 50 years respectively were: 52.0 (IQR: 42.6 to 62.8); 48.7 (IQR: 39.4 to 59.8); and 48.2 (IQR: 38.6 to 58.8).

- **CD19 B-Cell Count**

The recent median CD19 B-cell count in individuals aged 18-39 years was significantly lower 166 (IQR: 107 to 245) cells/μL blood compared to those who were 40-49 years this was 193 (IQR: 125 to 291) cells/μL blood and it was 185 (IQR: 113 to 293) cells/μL blood for those who were ≥ 50 years.

- **CD56: NK Cell Count**

Similarly, recent CD56 for 18-39, 40-49 and those aged ≥ 50 years were: 97 (IQR: 54 to 168) cells/μL blood; 114 (IQR: 64 to 191) cells/μL blood; and 142 (IQR: 81 to 235) cells/μL blood respectively.

- **HIV-1 RNA Plasma Load**

The highest recorded median HIV-1 RNA plasma load for 18-39, 40-49 and those aged ≥ 50 years respectively were: 55,429 (IQR: 10,067 to 185,859); 81,534 (IQR: 11,404 to 291,856); and 93,666 (IQR: 9,588 to 311,919) copies/mL plasma

- ***CHARACTERISTICS OF HEPATITIS B MARKERS BY IMMUNE CONSIDERED AGE GROUPS***

Significantly more ‘older’ people (≥50 years) had tested hepatitis B positive, hepatitis B core antibody positive compared to ‘young’ adults (18-39 years) and ‘middle-aged’ (40-49 years) adults Table 4.2.5.

Table 4.2.5: Co-Infection Parameters by Immune Considered Age of HIV-1 Cohort (1988-2011)

		Recent age (years) Grouped by immune considerations *(Age known and ≥18 year olds)* **N(%)**			
		18-39	**40-49**	**≥50**	
	Total cohort = 15,241	**N=7769**	**N=4961**	**N=2511**	**p-value**
Co-infections ever:					
#Hepatitis B surface antibody					
Never tested	13779	7425 (95.6)	4301 (86.7)	2053 (81.8)	<0.001
Negative	613	181 (2.3)	272 (5.5)	160 (6.4)	
Ever Positive	849	163 (2.1)	388 (7.8)	298 (11.9)	
#Hepatitis B core antibody					<0.001
Never tested	8939	5282 (68.0)	2524 (50.9)	1133 (45.1)	
Negative	3836	1865 (24.0)	1401 (28.2)	570 (22.7)	
Ever Positive	2466	622 (8.0)	1036 (20.9)	808 (32.2)	
&Recent Hepatitis B DNA, copies/mL plasma	N=357 500 (48 to 500) Range: 1 to 2,837,604	N=77 500 (48 to 721) Range: 1 to 602,676	N=172 500 (68 to 500) Range: 34 to 400,000	N=108 500 (34 to 500) Range: 34 to 283,7604	0.211

&Median (inter-quartile range: IQR), p-value using non parametric method test to test for differences in averages between the three age group strata using the Kruskal-Wallis method
#p-value for qualitative data are based on χ^2 test
See Table 4.1.1 for age group terms

- ***CHARACTERISTICS OF LIPID MARKERS BY IMMUNE CONSIDERED AGE GROUPS***

The recent lipid parameters appear to show statistically significant differences by age (Table 4.2.6), with patients who were middle-aged adults (40-49 years) and older adults (≥50 years) presenting with a significantly higher total cholesterol compared to younger HIV-1 infected adults (18-39 years).

Table 4.2.6: Lipid Parameters by Immune Considered Age of HIV-1 Cohort (1988-2011)

		Recent age (years) Grouped by immune considerations *(Age known and ≥18 year olds)* **N=15,241 N(%)**			
		18-39	**40-49**	**≥50**	
	Total cohort = 15,241	**N=7769**	**N=4961**	**N=2511**	**p-value**
Recent lipids:					
&Total cholesterol, mmol/l	4.7 (4.1 to 5.4)	4.5 (3.9 to 5.1)	4.9 4.2 to 5.6)	4.8 (4.2 to 5.6)	<0.001
&High density lipoprotein, mmol/l	1.13 (0.93 to 1.36)	1.12 (0.92 to 1.33)	1.13 (0.92 to 1.36)	1.15 (0.94 to 1.39)	0.004
&Low density lipoprotein, mmol/l	2.75 (2.24 to 3.36)	2.63 (2.14 to 3.21)	2.91 (2.35 to 3.52)	2.74 (2.25 to 3.35)	<0.001
&Triglycerides, mmol/l	1.51 (1.06 to 2.28)	1.35 (0.96 to 1.98)	1.60 (1.13 to 2.42)	1.73 (1.19 to 2.52)	<0.001

&Median (inter-quartile range: IQR), p-value using non parametric method test to test for differences in averages between the three age group strata using the Kruskal-Wallis method
See Table 4.1.1 for age group terms

- ***CHARACTERISTICS OF LIVER FUNCTION MARKERS BY IMMUNE CONSIDERED AGE GROUPS***

Similarly, significant differences were seen in liver function test results (Table 4.2.7) by age.

Table 4.2.7: Liver Function Parameters by Immune Considered Age of HIV-1 Cohort (1988-2011)

		Recent age (years) Grouped by immune considerations *(Age known and ≥18 year olds)* **N(%)**			
		18-39	**40-49**	**≥50**	
	Total cohort = 15,241	**N=7769**	**N=4961**	**N=2511**	**p-value**
Recent liver function:					
&Alalnine transaminase, U/l	30 (20 to 45)	28 (19 to 45)	31 (21 to 47)	30 (22 to 44)	<0.001
&Alkaline phosphotase, U/l	89 (68 to 132)	90 (68 to 146)	88 (69 to 123)	89 (70 to 121)	0.002
&Aspartate transaminase, U/l	33 (25 to 50)	34 (25 to 56)	32 (24 to 48)	31 (24 to 46)	<0.001
&Gamma-glutamyltransferase, U/l	47 (27 to 99)	38 (22 to 85)	51 (29 to 109)	51 (30 to 101)	<0.001
&Calcium, mmol/l	2.36 (2.29 to 2.43)	2.37 (2.29 to 2.43)	2.36 (2.28 to 2.42)	2.36 (2.28 to 2.43)	<0.001

&Phosphate, mmol/l	1.03 (0.90 to 1.19)	1.05 (0.90 to 1.21)	1.03 (0.90 to 1.18)	1.00 (0.87 to 1.13)	<0.001
&Albumin, g/l	40 (37 to 42)	40 (37 to 43)	40 (37 to 42)	39 (37 to 42)	<0.001
&Globulin, g/l	35 (32 to 39)	36 (32 to 40)	35 (32 to 39)	34 (31 to 38)	<0.001

&Median (inter-quartile range: IQR), p-value using non parametric method test to test for differences in averages between the three age group strata using the Kruskal-Wallis method
See Table 4.1.1 for age group terms

- ***CHARACTERISTICS OF RENAL MARKERS BY IMMUNE CONSIDERED AGE GROUPS***

Table 4.2.8 describes the renal parameters by age and significant differences were observed in the renal markers by age. Patients who were ≥50 years had estimated glomerular filtration rate (eGFR) lower than those in the 'younger' age groups. This table provides a cross-sectional description of the HIV-1 cohort population data by each age strata and presented are estimates of renal function markers available from the most recent clinic attendance. Although data presented in Table 4.2.8 were available longitudinally, per each individual, these have not been fully explored in presented data in Table 4.2.8.

Table 4.2.8: Renal Parameters by Immune Considered Age of HIV-1 Cohort (1988-2011)

		Recent age (years) Grouped by immune considerations *(Age known and ≥18 year olds)* **N(%)**			
		18-39	**40-49**	**≥50**	
	Total cohort = 15,241	**N=7769**	**N=4961**	**N=2511**	**p-value**
Recent renal:					
&Sodium, mmol/l	140 (138 to 141)	139 (136 to 141)	140 (138 to 141)	140 (138 to 141)	<0.001
&Potassium, mmol/l	4.2 (3.9 to 4.4)	4.1 (3.9 to 4.4)	4.2 (3.9 to 4.4)	4.2 (4.0 to 4.4)	<0.001
&Chloride, mmol/l	104 (102 to 106)	104 (102 to 105)	104 (103 to 106)	105 (103 to 106)	<0.001
&Urea, mmol/l	5.2 (4.2 to 6.3)	4.9 (4.0 to 6.0)	5.2 (4.3 to 6.3)	5.7 (4.6 to 6.9)	<0.001
&Estimated glomerular filtration rate (eGFR), ml/min per 1.73m^2 *The modification of diet in renal disease (MDRD) formula was used to calculate eGFR.*	82 (72 to 90)	88 (78 to 90)	80 (71 to 90)	76 (65 to 85)	<0.001

&Median (inter-quartile range: IQR), p-value using non parametric method test to test for differences in averages between the three age group strata using the Kruskal-Wallis method
See Table 4.1.1 for age group terms

HIV-1 diagnosed people on long-term with ART are at an increased risk of non-HIV co-morbidities such as cardiovascular disease (e.g. myocardial infarction), cancer, haematologic disease (e.g. anaemia), neurocognitive decline (e.g. dementia), bone disease (e.g. osteoporosis), lung (e.g. hypertension), liver disease (e.g. fibrosis), and amongst others kidney disease (e.g. insufficiency) that are associated with ageing. While all non-HIV co-morbidities are of interest, due to space and time constraints one of the commodity is investigated further and presented in the following sections. The following section 4.2.2 explores in some detail the synergistic effect of HIV-1 infection, long-term ART, ageing and risk of developing chronic kidney disease (CKD).

In order to describe how the renal function parameters changed by age the longitudinal availability of renal function data are discussed further. Firstly, in order to describe renal function changes over time by age, ages were categorised into decades (Table 4.1.1) that encompass stages of age and ageing and secondly the renal parameter data are described using longitudinal availability of results so that the ages are described at the time of renal function result rather than cross-sectionally as described in Table 4.2.1. For instance, the renal function data at successive points in time resemble each other to a greater degree than for example the renal function data of patients say from 10 years previously. Many factors change over time, making measurements close in time more similar than those across a long period of time.

4.2.2 LONGITUDINAL DATA

- ***NON-HIV CO-MORBIDITY: CHRONIC KIDNEY DISEASE***

One of the functions of kidneys is to remove all waste products of metabolism (such as urea, creatinine, and phosphorus), and to regulate the volume and composition (e.g. sodium and potassium) of body fluids. Glomerular filtration rate (GFR) measures the renal function and therefore effectiveness of the kidneys. Chronic Kidney Disease (CKD) is an indicator of level of damage to the kidneys and could occur as a result of physical injury to the kidneys or conditions such as diabetes, high blood pressure, or HIV-1 infection. Once the kidneys are damaged, they are unable to filter blood efficiently.

There are 3 tests that could be used to determine the kidney disease. These are: 1) **Blood pressure:** High blood pressure is both a risk factor for kidney disease and a complication of kidney disease; 2) **Protein in urine**: Body require protein however proteinuria[25] may be an indicator of kidney damage. Proteinuria does not increase with age; its appearance in the urine generally is due to some disease process. Proteinuria is evaluated using (Johnson *et al.*, 2004b) (i) total protein:creatinine ratio, (ii) albumin:creatinine ratio; 3) **Serum creatinine:** Creatinine is muscle waste product and when the kidneys are insufficient, creatinine from blood filtration is impaired. In

[25] protein in urine

addition since muscle mass decreases with age the creatinine in serum that is derived from muscle metabolism means the creatinine clearance may decrease substantially with age without increase in the serum creatinine concentration as muscle mass declines with age or with advancing HIV-1 disease. Serum creatinine levels also increase with increased protein intake.

Renal function deterioration is a common occurrence with advancing age. In a general population of 'healthy' men kidney function described by creatinine clearance was shown to decline by 30% between ages 30 and 80 years in male cohort Table 4.2.9 (Rowe *et al.*, 1976).

Table 4.2.9: Longitudinal Changes in Creatinine Clearance by Age in 'Healthy' Men

Age (Years)	No. Subjects	Creatinine Clearance ml/min/1.73m²	Creatinine Clearance Slope ml/min/1.73m²/yr
17 - 24	1	125.3	-1.75
25 - 34	20	140.4 ±4.6	-1.09 ±0.70
35 - 44	64	132.7 ±2.0	-0.11 ±0.36
45 - 54	95	128.1 ±1.6	-0.73 ±0.30
55 - 64	60	121.8 ±1.9	-1.64 ±0.41
65 - 74	36	110.0 ±2.6	-1.30 ±0.57
75 - 84	17	97.0 ±3.4	-1.07 ±0.77
17 - 84	293	124.7 ±1.1	-0.90 ±0.18

Note: Values indicate mean ± 1 SEM

Table taken from (Rowe *et al.*, 1976)

In 2002 the Kidney Disease Outcome Quality Initiative (KDOQI) established a classification for CKD (National Kidney Foundation, 2002) defined as GFR below 60 ml/min per 1.73 m^2 for 3 months or more.

- ***GLOMERULAR FILTRATION RATE***

GFR is the filtering of the blood also referred to as creatinine clearance which gives a measure of functional capacity of kidneys. GFR is estimated from serum creatinine concentration. Two commonly used formulas to calculate creatinine clearance (or GFR) are the Cockcroft-Gault formula and MDRD formula (Table 4.2.10).

The Cockroft-Gault formula for estimating creatinine clearance from serum creatinine concentration in older individuals has been improved upon using data from the MDRD study group (Levey *et al.*, 2009).

Table 4.2.10: Formulae to Calculate Creatinine Clearance

Formulae to calculate estimated creatinine clearance and eGFR)	
Cockcroft-Gault *ml/min per 1.73m²*	$\frac{\text{(140 - Age)} \times \text{Weight}}{72 \times \text{SCr}}$
Modification of diet in renal disease (MDRD) *ml/min per 1.73m²*	Estimated GFR = 175 x SCr$^{-1.154}$ x Age$^{-0.203}$ x (1.212 if Black) x (0.742 if Female) (Levey *et al.*, 2006) *This formula gives normal range of eGFR: 90 to 120 ml/min per 1.73m²*

SCr: Serum creatinine concentration in ml/min per 1.73m^2, age in years, weight in kilograms
eGFR: Estimated glomerular filtration rate

An eGFR is one of the pathways to assess kidney disease. Once the eGFR is impaired, progression to end-stage renal disease is likely to occur at high rates. There are 5 defined stages of CKD according to the eGFR measurement (Table 4.2.11).

Table 4.2.11: National Kidney Foundation Classification of Stages of CKD

Stage of CKD	National Kidney Foundation	Glomerular Filtration Rate *(ml/min per 1.73m²)*
1	Kidney damage (eg. protein in urine) with normal GFR	≥90
2	Kidney damage with mild decrease in GFR	60-89
3	Moderate decrease in GFR	30-59
4	Severe reduction in GFR	15-29
5	Kidney failure	<15

CKD: Chronic Kidney Disease
Table adapted from (Johnson *et al.*, 2004a)

The main aim is to understand the synergistic effects of HIV-1, ART and age on renal function. The symptoms of CKD are associated with complications on all organ systems. CKD related anemia, bone diseases, cardiovascular disease (Johnson *et al.*, 2004a). The risk of complications like heart disease, anemia, or high blood pressure increases as the GFR falls below 60 ml/min per 1.73m^2. The risk of co-morbidities increases when GFR declines to below 60 ml/min per 1.73m^2.

Table 4.2.12: Risk Factors of CKD In 'Normal' HIV-1 Uninfected People

Risk factor	Description	Example
Susceptibility factors	Increased risk of kidney damage	Ageing, diabetes, hypertension, family history of CKD, Black ethnicity,
Initiation factors	Initiation of kidney damage	Diabetes, high blood pressure, systemic infections, urinary tract infection, urinary stones, drug toxicity
Progression factors	Declining kidney function post initiation of kidney damage	Higher levels of proteinuria, high blood pressure, high blood glucose, smoking, hyperlipidemia, anaemia, cardiovascular disease
End stage	Increased likelihood of morbidity and mortality due to kidney failure	Anaemia, low serum albumin levels

Adapted from (Parmar, 2002)

CKD is associated with premature immunological ageing (Betjes *et al.*, 2011) and renal impairment is also associated with a number of conditions including diabetes, hypertension, cardiovascular disease and others (Table 4.2.12). Among HIV-1^+ people, having a low $CD4^+$ T-cell count and high HIV-1 RNA plasma load (Mocroft *et al.*, 2007; Fernando *et al.*, 2008) are both associated with CKD. If left untreated, CKD can lead to moderate or advanced renal failure (Gilg *et al.*, 2011) although the majority of patients with early CKD may not progress to these stages (Dalrymple *et al.*, 2011). Since the advent of ART, studies have reported that people diagnosed with HIV-1 receiving long-term ART are at risk of developing CKD (Mocroft *et al.*, 2007; Levey *et al.*, 2009; Mocroft *et al.*, 2010; Young *et al.*, 2012). However, HIV also directly infects renal cells, resulting in an inflammatory process or HIV-associated nephropathy (HIVAN) leading to CKD (Rao *et al.*, 1984; Ando *et al.*, 2012).

- ***DEMOGRAPHIC SHIFT SHOWING AGEING OF HIV-1 COHORT LOCALLY***

The success of HAART combined with older people being newly diagnosed as HIV-1^+ has resulted in a demographic shift in age of HIV-1^+ patients attending for care locally. A striking and progressive increase in the number of HIV-1^+ individuals was observed over time with increasing age (Figure 4.2.1).

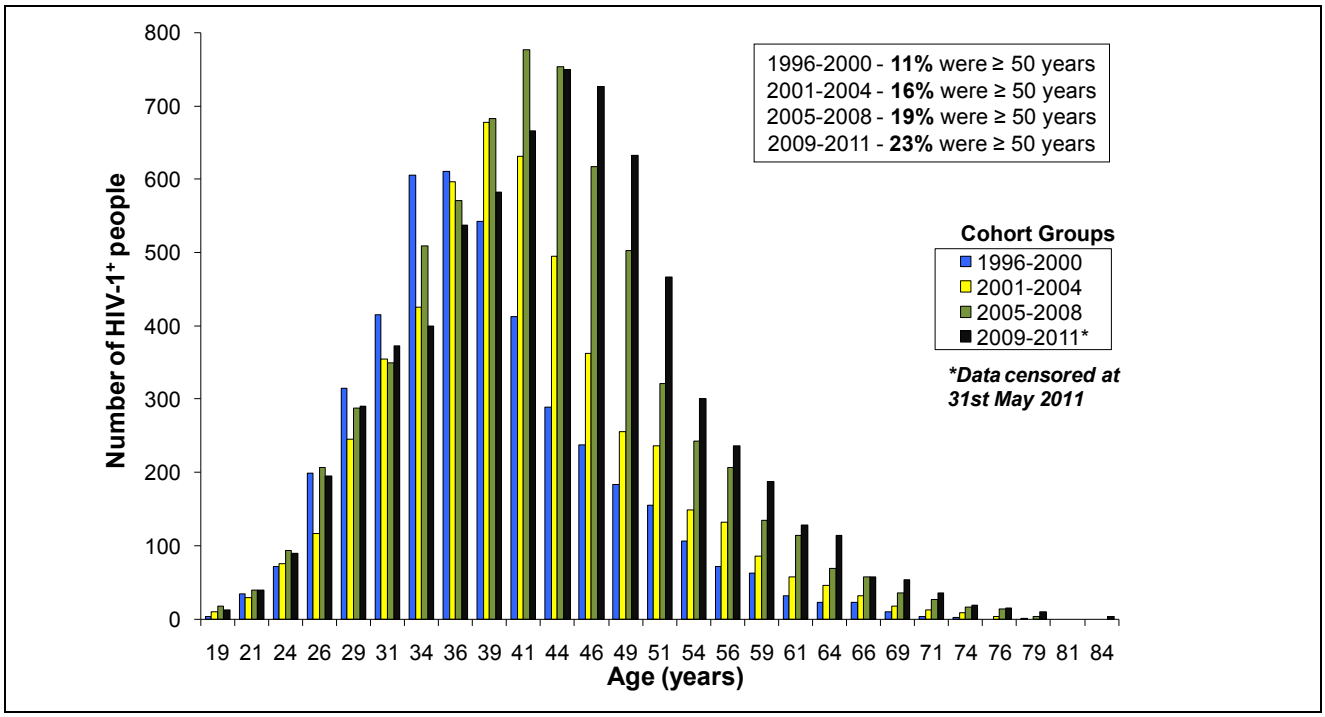


Figure 4.2.1: Age Distribution Shift by Cohort Strata Demonstrating an Increase in Older HIV-1^{+} People

Figure taken, with slight modification, from oral presentation slides: Mandalia et al., 2011
2nd International Workshop on HIV & Aging Baltimore, USA, 28th October 2011, Session 7 Epidemiology and Pathogenesis. http://regist2.virology-education.com/2011/2Aging/docs/24_Mandalia.pdf

Over time a progressive increase in the number of 'older' (≥50 years) HIV-1^{+} patients was observed (Figure 4.2.1) comparing the period of HAART era 1 (1996-2000), HAART era 2 (2001-2004), HAART era 3 (2005-2008) and HAART era 4 (2009-2011)[26]. The choice of these cohort strata were selected as these coincided with the introduction, availability and prescription of new ARV drugs during the year as part of HAART therapy (see Chapter 1: Table 1.2.1).

HIV-1^{+} people who attended during 2009-2011 accounted for the largest numbers of patients aged ≥40 years (n=4,399) compared to people who attended during 1996-2000 (n=1,609), 2001-2004 (n=2,527), 2005-2008 (n=3,893). Compared to 1996-2000 (n=488) 11% were aged ≥50 years, the respective numbers aged ≥50 years who attended during 2001-2004 (n=783) were 15.5%, 2005-2008 (n=1,241) 18.7%, 2009-2011 (n=1,621) 23.4%. These findings reflected increased numbers of HIV-1^{+} population aged ≥50 years attending for care within the cohort.

Chapter 4:26 Data were censored at 31st May 2011 this is the time when data were extracted from the local Trust Data Warehouse

- ***DEMOGRAPHIC CHARACTERISTICS OF ART NAÏVE BY AGE***

Of the 6,923 patients who were ART naive, 91% (n=6304) were men, 66% (n=4,564) Caucasians and 84% (n=5,795) were MSM. The older PLHIV was more likely to have been living longer with HIV-1^{+} (Table 4.2.13). The percentage of women 50 years or older was 8% (95% CI: 6% to 11%), similar to the percentage of older men at 9% (95% CI: 8% to 10%). Similarly, 10% of Caucasian (95% CI: 9% to 10%) were 50 years or older compared with 8% of black Africans (95%CI: 5% to 11%) and 7% of patients from other ethnic groups (95% CI: 6% to 8%). Sixteen percent of bisexuals (95% CI: 11% to 20%) were 50 years or older compared with 10% of heterosexuals (95% CI: 8% to 12%) and 8% of MSM (95%CI: 7% to 9%).

- ***DEMOGRAPHIC CHARACTERISTICS OF PATIENTS WHO START FIRST-LINE HAART (2NRTI+NNRTI OR 2NRTI+BOOSTEDPI) BY AGE***

Of the 4,852 patients on first-line ART, 79% were on 2NRTIs+NNRTI and 21% 2NRTIs+boosted-PI (Table 4.2.13). Uptake of NNRTI containing regimens increased with age; of those who started first-line ART 86% (n=4,186) were men, 84% (n=3,506) Caucasians and 85% (n=3,807) were MSM. The percentage of older PLHIV was less among black Africans and higher among those persons who lived longer with HIV (Table 4.2.13). Significantly fewer women 50 years or older were in the ART cohort: 14% (95% CI: 11% to 17%) compared with 22% men (95% CI: 20% to 23%). Similarly, 20% of Caucasians (95% CI: 18% to 21%) were older compared with 15% of black Africans (95%CI: 12% to 17%) and 13% of patients from other ethnic groups (95% CI: 11% to 16%). Thirty-two percent of bisexuals (95% CI: 22% to 43%) were 50 years or older compared with 19% of heterosexuals (95% CI: 17% to 21%) and 18% of MSM (95%CI: 16% to 19%).

Table 4.2.13: Demographic Characteristics by Age Group of ART Naïve and Patients on First-line 2NRTI+NNRTI or 2NRTI+boosted-PI

ART naïve patients		**N=6,923**	**18-19 years N=31**	**20-29 years N=1,735**	**30-39 years N=2,926**	**40-49 years N=1,631**	**50-59 years N=477**	**60-69 years N=112**	**≥70 years N=11**	**p-value**
Gender	Female	619	6 (19)	218 (13)	250 (8)	94 (6)	40 (8)	10 (9)	1 (9)	<0.001
	Male	6,304	25 (81)	1517 (87)	2676 (92)	1537 (94)	437 (92)	102 (91)	10 (91)	
Ethnicity	Caucasian	4,564	23 (74)	1052 (61)	1956(67)	1096(67)	341 (71)	87 (78)	9 (82)	<0.001
	Black African	420	2 (6)	111 (6)	187 (6)	87(5)	28 (6)	5 (5)	0 (0)	
	Other	1939	6 (19)	572 (33)	783 (27)	448(28)	108 (23)	20 (17)	2 (18)	
Sexual orientation	MSM	5,795	24 (77)	1448 (84)	2447 (83)	1407 (86)	378(79)	83 (74)	8 (73)	<0.001
	Heterosexual	838	7 (23)	228 (13)	367 (13)	150 (9)	67 (14)	17 (15)	2 (18)	
	Bisexual	290	0 (0)	59 (3)	112 (4)	74 (5)	32 (7)	12 (11)	1 (9)	
IDU	No	6,685	31 (100)	1679 (97)	2798 (96)	1588 (97)	467 (98)	111 (99)	11 (100)	0.008
	Yes	238	0 (0)	56 (3)	128 (4)	43 (3)	10 (2)	1 (1)	0 (0)	
Median (IQR) months since HIV diagnosis		31 (8 - 71)	4 (1 - 14)	15 (3 - 42)	35 (10 - 73)	45 (12 - 95)	51 (13 - 100)	66 (18 - 110)	118 (37 - 223)	<0.001
Patients on first-line ART		**N=4,852**	**N=3**	**N=414**	**N=1,611**	**N=1,948**	**N=647**	**N=181**	**N=48**	**p-value**
	2 NRTI+NNRTI	3,836	2 (67)	308 (74)	1253 (78)	1560 (80)	525 (81)	149 (82)	39 (81)	0.073
	2NRTI+boostedPI	1,016	1 (33)	106 (26)	358 (22)	388 (20)	122 (19)	32 (18)	9 (19)	
Gender	Female	666	3 (100)	91 (22)	241 (15)	238 (12)	73 (11)	15 (8)	5 (10)	<0.001
	Male	4,186	0 (0)	323 (78)	1370 (85)	1710 (88)	574 (89)	166 (92)	43 (90)	
Ethnicity	Caucasian	3,506	0 (0)	274 (66)	1117 (69)	1425 (73)	495 (76)	159 (88)	36 (75)	<0.001
	Black African	625	2 (67)	57 (14)	222 (14)	253 (13)	75 (12)	11 (6)	5 (10)	
	Other	721	1 (33)	83 (20)	272 (17)	270 (14)	77 (12)	11 (6)	7 (15)	
Sexual orientation	MSM	3,807	2 (67)	323 (78)	1289 (80)	1526 (78)	497 (77)	134 (74)	36 (75)	0.002
	Heterosexual	961	1 (33)	91 (22)	307 (19)	380 (20)	130 (20)	40 (22)	12 (25)	
	Bisexual	84	0 (0)	0 (0)	15 (1)	42 (2)	20 (3)	7 (4)	0 (0)	
IDU	No	4,756	3 (100)	411 (99)	1578 (98)	1902 (98)	633 (98)	181 (100)	48 (100)	<0.001
	Yes	96	0 (0)	3 (1)	33 (2)	46 (2)	14 (2)	0 (0)	0 (0)	
Median (IQR) months since HIV diagnosis		81.9 (42 - 133)	24.1 (20 - 220)	38.3 (20 - 64)	67.6 (35 - 106)	100.7 (57 - 147)	116.2 (65 – 152)	102.7 (49 – 147)	127.4 (79 – 183)	<0.001

CHAPTER 4

HIV AND AGEING

The eGFR is a widely reported marker of renal function in HIV-1^+ people and this is reported to detect progression of renal dysfunction. The following analyses will seek to describe longitudinal changes in renal function, as measured by eGFR (by MDRD method), by age in a cohort of HIV-1^+ people and analyse the impact of HIV-1 infection and long-term ART in HIV-1^+ people as they age. Age was grouped into decades and derived at time of renal function results in people who had been exposed to the three widely used ARV drug classes and assess whether the renal function is stable or progressive with changing age. This includes HIV-1^+ people who were ART naïve and those who had been exposed to different classes of ARV drugs, including nucleoside reverse transcriptase inhibitors (NRTI), protease inhibitor (PI) or non-nucleoside reverse transcriptase inhibitors (NNRTI). As Highly Active ART (HAART) usually comprise the use of at least two classes of ARV drugs concurrently, longitudinal changes in renal function were also analysed for those starting the most commonly prescribed first-line HAART therapy comprising either 2NRTI+NNRTI or 2NRTI+boosted PI regimens by age. Finally variables that independently predicted the occurrence of CKD stage 4 or 5 (moderate or advanced renal failure) (see Table 4.2.11: eGFR <30 ml/min per 1.73m^2) were identified.

- ***CHANGING EGFR FOR ART NAÏVE BY AGE***

In ART naïve individuals, eGFR remained relatively stable at 84 ml/min per 1.73m^2 for the 'young' adults aged 18-39 years. A progressive decline was subsequently observed at around 40 years, when thymic function declines, with a mean eGFR of 80 ml/min per 1.73m^2 for the 'middle-aged' adults 40-49 years and by the time HIV-1^+ adults were 70 years or older the mean eGFR had decreased to 59 ml/min per 1.73m^2.

- ***CHANGING EGFR BY ARV CLASS EXPOSURE AND AGE***

eGFR results were available for 7,980 patients. Similar findings were observed by age in those patients who had been exposed to ART including NRTIs, PIs or NNRTIs (Figure 4.2.2; Table 4.2.14). Polynomial models were generated that predicted mean eGFR by age for each of the four groups: ART naïve; NRTI; PI; or NNRTI exposed in addition a polynomial model based on creatinine clearance values from 'normal' HIV-1 uninfected people by age (Rowe *et al.*, 1976). For each of these five groups, the models with the 'best fit' predicted a declining eGFR with increasing age for HIV-1^+ and uninfected people (Table 4.2.15). The predicted eGFR for: ART naïve; NRTI; PI; or NNRTI exposed HIV-1^+ people at age 20 years were respectively predicted as: 84; 84; 83; and 85 ml/min per 1.73m^2. Similarly, the predicted eGFR for an HIV-1^+ person aged 60 years for: ART naïve; NRTI; PI; or NNRTI exposed patients were 78, 76, 75 and 77 ml/min per 1.73m^2 respectively. Predicted creatinine clearance for 'healthy' uninfected people was 103 ml/min per 1.73m^2 at age 20 years and 98 ml/min per 1.73m^2 at age 60 years.

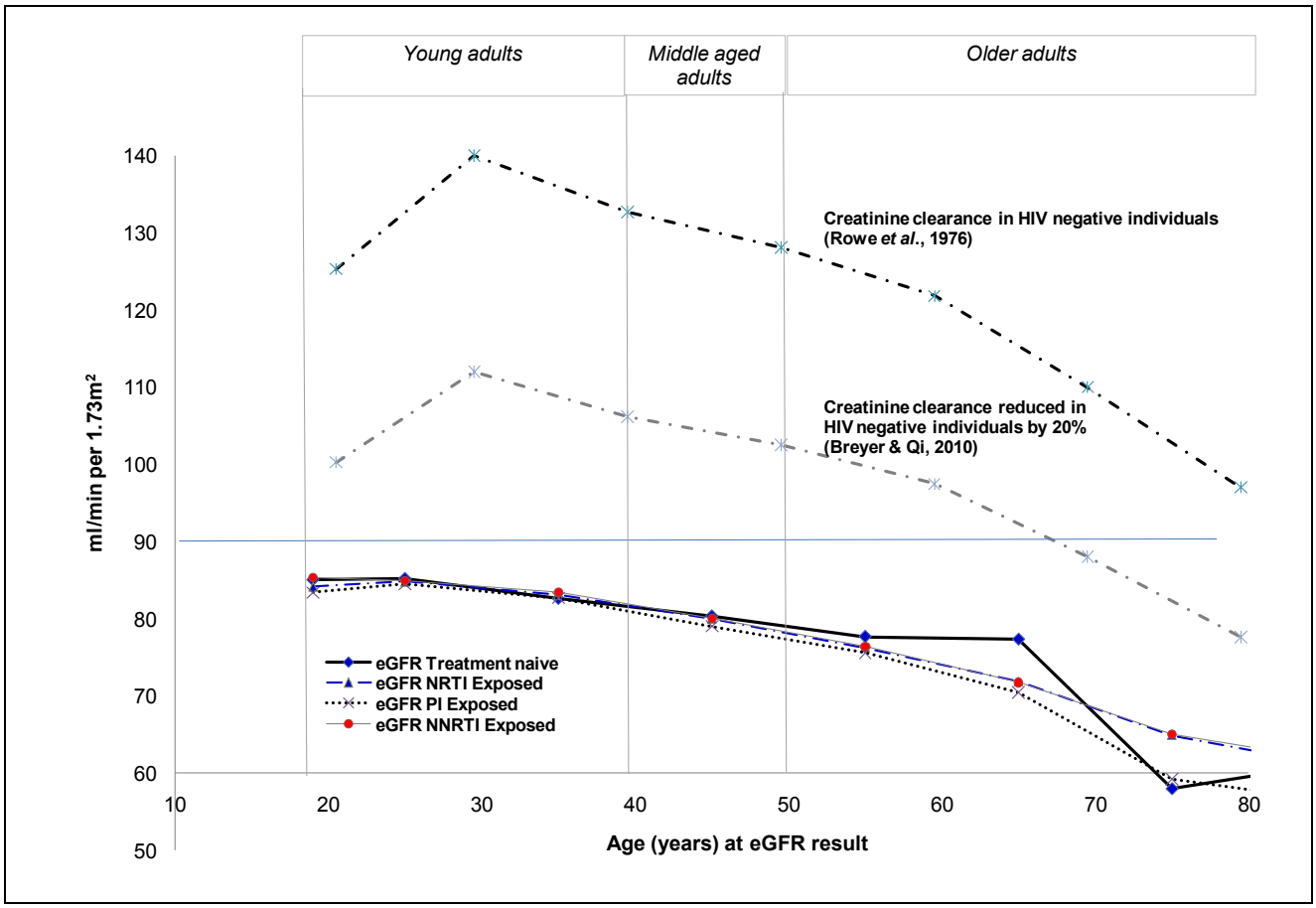


Figure 4.2.2: Longitudinal Changes in eGFR Age and ARV Class Exposure and Creatinine Clearance by Age in HIV Negative People and Reduced by 20% (Rowe *et al.*, 1976; Breyer & Qi, 2010)

eGFR of >90 ml/min per 1.73m^2 is considered normal
eGFR of 60 to 90 ml/min per 1.73m^2 indicates mildly reduced kidney function defined as stage 2 CKD
eGFR of 30 to 59 ml/min per 1.73m^2 represent CKD stage 3
eGFR of <30 ml/min per 1.73m^2 represent CKD stages 4 and 5

Table 4.2.14: Anti-retroviral Class Exposure and eGFR, by Age at the Time of Result of HIV-1 Cohort (1988-2011) and Patients Who Started First-line HAART Consisting of 2NRTI+NNRTI or 2NRTI+boosted PI

	Recent Age of HIV-1 Cohort (1988-2011) N=15,241						
	18-19 N=34	**20-29 N=2,309**	**30-39 N=5,426**	**40-49 N=4,961**	**50-59 N=1,876**	**60-69 N=529**	**>=70 N=106**
ART naive *N=6923*	85.0 (80.5 to 89.4)	85.1 (84.2 to 86.0)	82.5 (81.8 to 83.3)	80.3 (79.4 to 81.2)	77.6 (76.1 to 79.1)	77.3 (74.1 to 80.5)	58.9 (53.7 to 63.2)
	Age at time of renal result (years) Mean eGFR (95% CI) ml/min per 1.73m^2 Total number of patients=7980						
	18-19 N=33	**20-29 N=1204**	**30-39 N=3293**	**40-49 N=3721**	**50-59 N=1482**	**60-69 N=435**	**>=70 N=85**
NRTI experienced	84.1 (79.1 to 89.1)	84.8 (83.7 to 85.8)	83.0 (82.2 to 83.8)	79.9 (79.0 to 80.9)	76.2 (74.7 to 77.8)	71.8 (68.5 to 75.0)	64.8 (59.8 to 69.8)
PI experienced	83.3 (83.1 to 83.6)	84.4 (83.8 to 85.0)	82.7 (82.2 to 83.1)	79.0 (78.5 to 79.4)	75.6 (69.2 to 82.0)	70.4 (64.1 to 76.8)	59.2 (53.7 to 66.8)
NNRTI experienced	85.3 (82.9 to 87.6)	84.9 (84.4 to 85.3)	83.4 (83.1 to 83.7)	80.0 (70.1 to 89.9)	76.4 (66.5 to 86.3)	71.7 (61.8 to 81.6)	64.9 (55.0 to 73.8)
	Age at time of renal result of patients who started first-line HAART N=4,852 and ART naïve N=6923 Mean eGFR (95% CI) ml/min per 1.73m^2 longitudinally estimated using MIXED model						
	18-19 N=3	**20-29 N=414**	**30-39 N=1,611**	**40-49 N=1,948**	**50-59 N=647**	**60-69 N=181**	**>=70 N=48**
2 NRTI+NNRTI N=3,836	87.0 (81.8 to 92.3)	85.1 (84.3 to 85.8)	83.4 (83.0 to 83.9)	81.3 (80.8 to 81.8)	77.9 (77.1 to 78.7)	75.4 (73.7 to 77.0)	70.8 (67.2 to 74.4)
2NRTI+boostedPI *N*=1,016	85.1 (76.8 to 93.4)	84.2 (82.7 to 85.7)	81.9 (81.0 to 82.9)	78.9 (77.9 to 79.9)	78.1 (76.1 to 80.2)	69.6 (65.6 to 73.7)	67.0 (57.7 to 76.3)

See Table 4.1.1 for age group terms

Table 4.2.15: Relationship between Creatinine Clearance, eGFR, Age, ARV Class Exposure and First-line HAART

	Trends over time in creatinine clearance or eGFR described by linear and polynomial relationship	Proportion of variance explained	Predicted values at 20 years ml/min per 1.73m^2	Predicted values at 60 years ml/min per 1.73m^2
Normal population (Rowe *et al.*, 1976)	CrCl = -0.0157 Age^2 + 1.115Age + 87.242	94.3%*	103*	98*
Treatment naive	eGFR = -0.0096 Age^2 + 0.6131 Age + 75.535	85.0%	84	78
NRTI exposed	eGFR = -0.0063 Age^2 + 0.3054 Age + 80.781	99.8%	84	76
PI exposed	eGFR = -0.0093 Age^2 + 0.5442 Age + 76.078	98.9%	83	75
NNRTI exposed	eGFR = -0.0057 Age^2 + 0.2373 Age + 82.867	99.9%	85	77
2NRTI+NNRTI	eGFR = -0.2569 Age + 93.038	98.0%	88	78
2NRTI+boosted PI	eGFR = -0.0042 Age^2 + 0.1081Age + 84.627	96.9%	85	76

***Includes a 20% reduction to compensate for creatinine clearance overestimation** (Breyer & Qi, 2010)

- ***CHANGING PROTEIN:CREATININE RATIO BY ARV CLASS EXPOSURE***

Protein:creatinine ratios were only available for 2,912 (36%) patients who also had eGFR results available. The protein:creatinine ratio in ART naïve people increased with increasing age from a mean of 15 mg/mmol for the 'young' adults (18-19 years) to 105 mg/mmol for 60 years or above. Similar trends were observed among patients who had been exposed to NRTI with the mean protein:creatinine ratio increasing from 15 to 47 mg/mmol, while protein:creatinine ratios increased from 15 to 51 mg/mmol for HIV-1$^+$ people who had been exposed to PIs and from 16 to 160 mg/mmol for patients who had been exposed to NNRTIs (Table 4.2.16; Figure 4.2.2).

Table 4.2.16: Anti-retroviral Class Exposure and Protein:Creatinine Ratio, by Age at the Time of Result of HIV-1 Cohort (1988-2011) and Patients Who Started First-line HAART Consisting of 2NRTI+NNRTI or 2NRTI+boosted PI

	Recent Age of Total cohort N=15,241					
	18-19 N=34	**20-29 N=2,309**	**30-39 N=5,426**	**40-49 N=4,961**	**50-59 N=1,876**	**>=60 N=529**
ART naive *N=6923*	15.2 (0 to 93.5)	13.4 (0 to 44.1)	23.2 (0.1 to 46.4)	24.6 (0 to 52.9)	17.0 (0 to 77.7)	104.9 (58.5 to 151.2)
	Mean Protein:Creatinine ratio (95% CI) mg/mmol **Total number of patients=2912**					
	N=6	**N=205**	**N=855**	**N=1259**	**N=531**	**N=635**
NRTI experienced	15.2 (0 to 93.5)	54.9 (22.2 to 87.6)	26.9 (3.2 to 50.5)	40.2 (11.7 to 68.8)	43.2 (0 to 104.1)	46.9 (0 to 111.5)
PI experienced	15.4 (0 to 81.3)	67.1 (40.6 to 93.6)	35.1 (16.5 to 53.7)	49.8 (17.8 to 81.9)	51.8 (0 to 114.8)	51.4 (0 to 118.1)
NNRTI experienced	16 (0 to 82.0)	61.6 (36.1 to 87.2)	38.4 (5.6 to 71.2)	52.1 (0 to 122.0)	10.0 (0 to 48.8)	160.4 (142.1 to 178.7)
	Age at time of renal result of patients who started first-line HAART N=4,852 Mean Protein:Creatinine ratio (95% CI) mg/mmol longitudinally estimated using MIXED model					
	N=3	**N=414**	**N=1,611**	**N=1,948**	**N=647**	**N=181**
2 NRTI+NNRTI N=3,836	14.0 (0 to 102.2)	44.4 (36.7 to 52.2)	20.3 (16.0 to 24.5)	20.1 (15.4 to 24.7)	39.8 (31.9 to 47.7)	36.2 (22.3 to 50.1)
2NRTI+boostedPI *N*=1,016	17.0 (0 to 101.4)	20.5 (5.4 to 35.6)	21.0 (11.8 to 30.2)	25.6 (17.1 to 34.2)	43.2 (26.0 to 60.4)	56.1 (30.3 to 81.8)

See Table 4.1.1 for age group terms

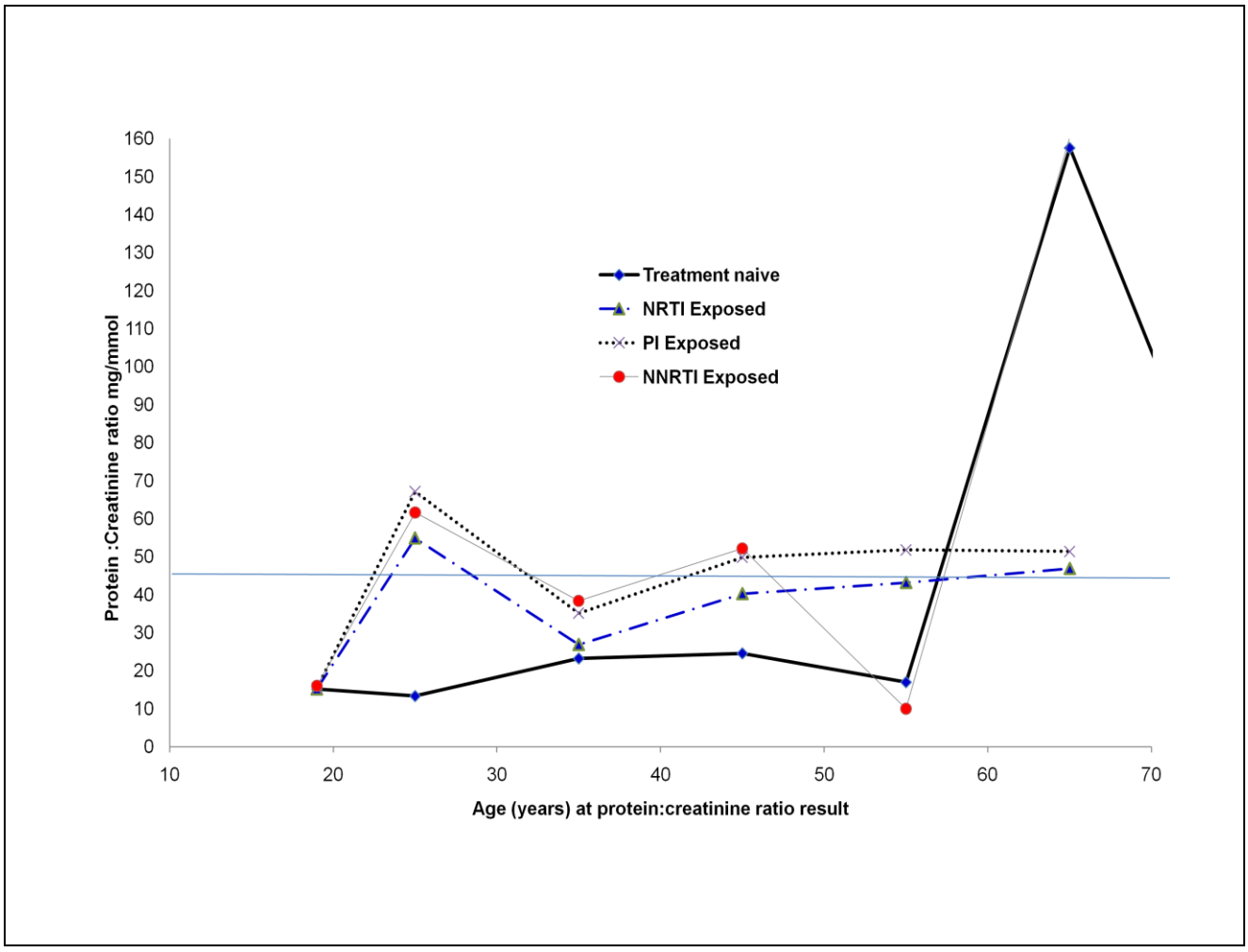


Figure 4.2.3: Longitudinal Changes in Protein:Creatinine Ratio, Age and ARV Class Exposure
Protein:creatinine ratio of ≥45 mg/mmol is indicative of proteinuria in the UK (Joint Specialty Committee on Renal Medicine, 2006; Lamb *et al.*, 2009)

- ***CHANGING EGFR FOR PATIENTS WHO START FIRST-LINE HAART (2NRTI+NNRTI OR 2NRTI+BOOSTEDPI) BY AGE***

Figure 4.2.3 describes eGFR by age, the boxes above show immune considered age group strata. The eGFR level between 60 and 89 ml/min per 1.73m^2 describe CKD stage 2 (Table 4.2.11). When comparing HIV-1$^+$ people on first-line 2NRTI+NNRTI or 2NRTI+boosted PI regimens by age (Table 4.2.14; Figure 4.2.3), people younger than 50 years did not show significant differences in mean eGFR between regimens with a mean difference in slope between the two treatment regimens of 3.7 ml/min per 1.7m^2 (95% CI: -1.2 to 8.7, p=0.136). However for people ≥50 years, a significant steeper decline in eGFR was observed between those on 2NRTI+boosted PI compared with patients on 2NRTI+NNRTI the mean difference in slope between the two treatment regimens was 5.8 ml/min per 1.73m^2 (95% CI: 3.9 to 7.7, p<0.001; using MIXED model). As can be noted the creatinine clearance values for HIV-1$^+$ people remain lower with increasing age than 'healthy' population (Rowe *et al.*, 1976) and these are observed to be well above the normal values (eGFR

≥90 ml/min per 1.73m^2) even when the healthy populations creatinine clearance were reduced by 20% (Breyer & Qi, 2010). For each of the three groups - ART naïve, PLHIV on 2NRTI+NNRTI or 2NRTI+boosted-PI - the 'best fit' models predicted a declining eGFR with increasing age for PLHIV. At age 20, the predicted mean eGFR for ART naïve individuals was 84 ml/min per 1.73m^2 compared with 88 and 85 ml/min per 1.73m^2 for those who started first-line 2NRTI+NNRTI or 2NRTI+boosted-PI respectively. The predicted eGFR for an ART naïve individual aged 60 years was 78 ml/min per 1.73m^2, while for an individual aged 60 years who started 2NRTI+NNRTI or 2NRTI+boosted-PI, the predicted eGFR was 78, and 76 ml/min per 1.73m^2 respectively.

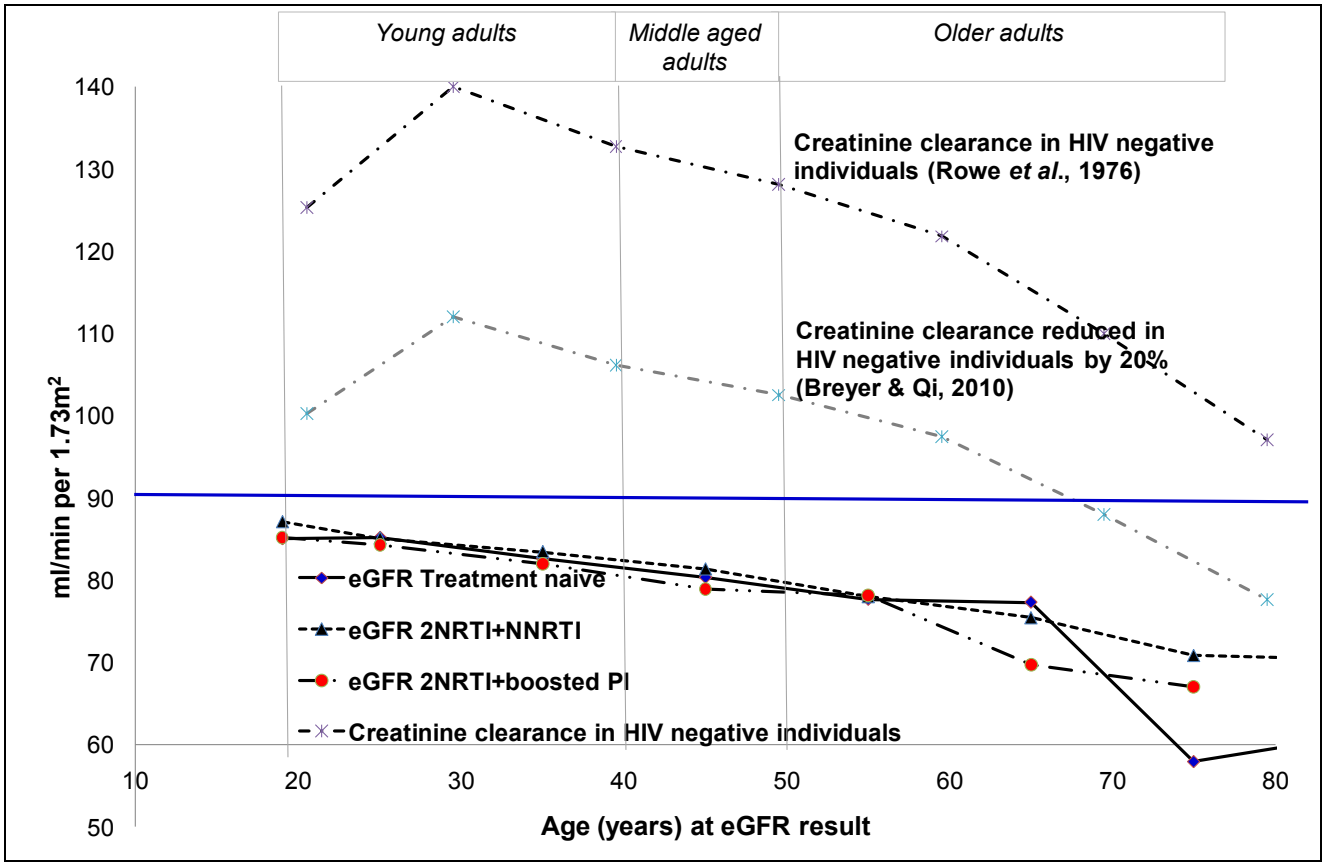


Figure 4.2.4: Longitudinal Changes in eGFR Age and Patients on First-line 2NRTI+NNRTI or 2NRTI+boosted-PI by Age HIV Negative People and Reduced by 20% (Rowe *et al.*, 1976; Breyer & Qi, 2010)
eGFR of >90 ml/min per 1.73m^2 is considered normal
eGFR of 60 to 90 ml/min per 1.73m^2 indicates mildly reduced kidney function defined as stage 2 CKD
eGFR of 30 to 59 ml/min per 1.73m^2 represent CKD stage 3
eGFR of <30 ml/min per 1.73m^2 represent CKD stages 4 and 5

Linear and polynomial 'best fit' models were derived to describe the data observed for ART naïve people and those who started first-line 2NRTT+NNRTI or 2NRTI+boosted PI (Table 4.2.15). At age 20, the predicted mean eGFR for treatment naïve people was 84 ml/min per 1.73m^2 compared with 88 and 85 ml/min per 1.73m^2 for those who started first-line 2NRTI+NNRTI or 2NRTI+boosted PI respectively. The predicted eGFR for treatment naïve people aged 60 was 78 ml/min per 1.73m^2, while for people who started 2NRTI+NNRTI or 2NRTI+boosted PI, the predicted eGFR was 78, and

76 ml/min per 1.73m^2 respectively. Even when reduced by 20%, the creatinine clearance values by age were still well above the values for HIV-1$^+$ people compared to people uninfected with HIV-1. At age 20 years, predicted 'normal' creatinine clearance values was 103 ml/min per 1.73m^2 and 98 ml/min per 1.73m^2 at age 60 (Table 4.2.15).

- ***CHANGING PROTEIN:CREATININE RATIO FOR FIRST-LINE 2NRTI+NNRTI OR 2NRTI+BOOSTED PI***

With increasing age the protein:creatinine ratio in ART naïve people increased from a mean of 15 mg/mmol for the 18-19 year old patients to 105 mg/mmol for the 60-69 year olds. Of those on first-line regimens, 836 (17%) had protein:creatinine results available on two or more occasion: 643 (77%) had been on 2NRTI+NNRTI while 193 (23%) received 2NRTI+boosted PI. People who were less than 50 years showed no significant differences in protein:creatinine ratios by age between those on 2NRTI+NNRTI or 2NRTI+boosted PI (p=0.793, using MIXED model). 'Older' adults demonstrated a significant steeper increase in the protein:creatinine ratio in patients on 2NRTI+NNRTI compared with those on 2NRTI+boosted PI (using MIXED model, p<0.001; Figure 4.2.4). However increases between ages of 60-69 years were based on a relatively small number of people and these results need to be interpreted cautiously.

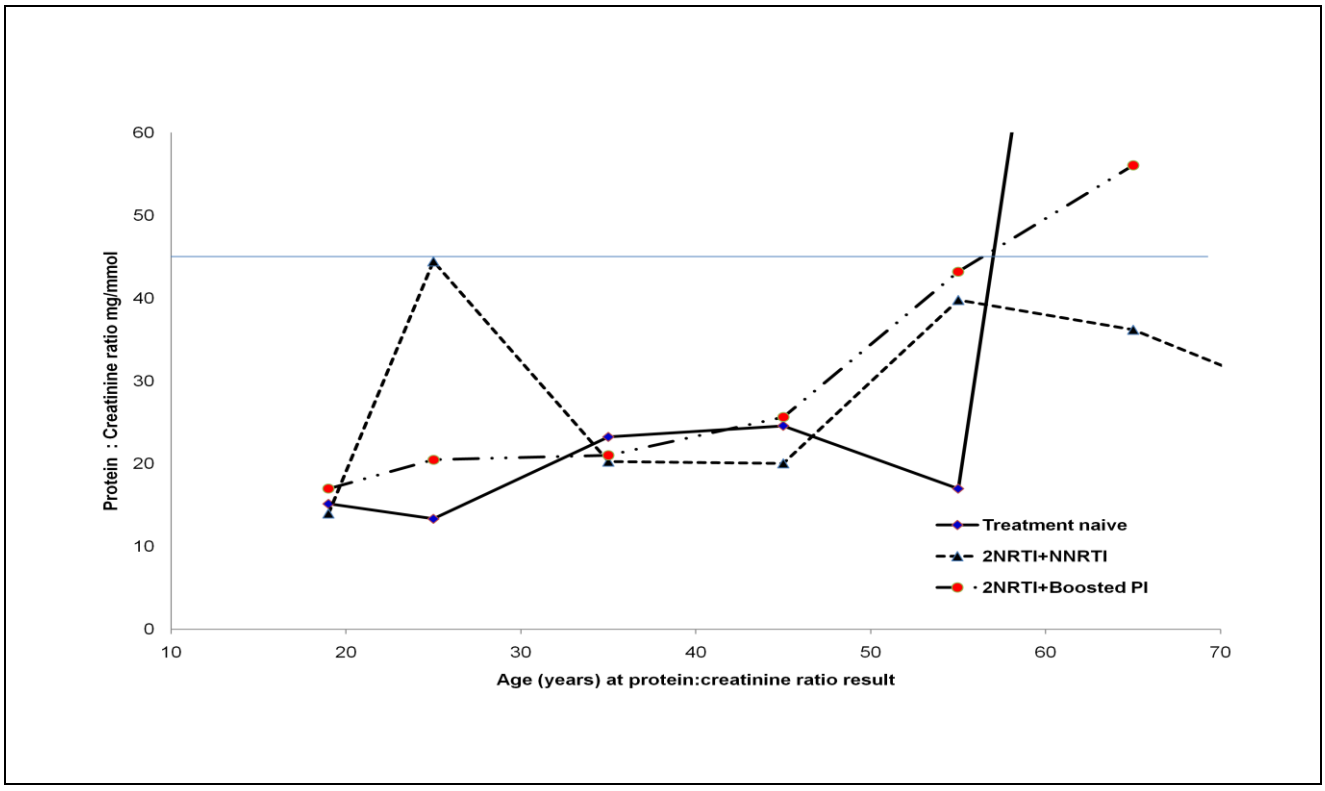


Figure 4.2.5: Longitudinal Changes in Protein:Creatinine Ratio, by Age in Patients Who Started First-line HAART

Protein:creatinine ratio of ≥45 mg/mmol is indicative of proteinuria in the UK (Joint Specialty Committee on Renal Medicine, 2006; Lamb *et al.*, 2009)

- ***UNIVARIATE AND MULTIVARIABLE COX'S PROPORTIONAL HAZARDS REGRESSION ANALYSIS***

Of the 7,980 people for whom nadir eGFR were available, 157 (2.0%) had eGFR classified as moderate to advanced renal failure. The univariate associations between demographic and other variables and a nadir eGFR<30 ml/min per 1.73m^2 were estimated (Table 4.2.17). The multivariable model indicated that HIV-1$^+$ people aged ≥50 years, IDUs and non-Caucasians and those who had PI exposure were more likely to have a nadir eGFR <30 ml/min per 1.73m^2 (Table 4.2.18: Figure 4.2.5). Similarly, those HIV-1$^+$ people with a baseline eGFR <72 ml/min per 1.73m^2 and protein:creatinine ratio >22 mg/mmol were more likely to have a nadir eGFR< 30 ml/min per 1.73m^2 as did HIV-1$^+$ people who had a nadir CD4≤ 100 cells/μL blood, while those HIV-1$^+$ people diagnosed longest with HIV-1$^+$ were less likely to develop end stage renal failure requiring dialysis or transplantation (Table 4.2.18: Figure 4.2.5).

Table 4.2.17: Univariate Cox's Proportional Hazards Regression Model showing Likelihood of Stage 4 or 5 CKD Defined Using Nadir eGFR

	Total =7,980	Stage 4 or 5 CKD Nadir eGFR <30 ml/min per 1.73m^2 n=157	Hazard ratio (HR)	95% CI	p-value
#Age at nadir eGFR (years)					
40-49	2,949	53(1.8)	1		<0.001
18-19	30	0(0.0)	0.00	(unstable)	
20-29	1,005	6(0.6)	0.68	(0.28 to 1.61)	
30-39	2,568	33(1.3)	1.12	(0.72 to 1.73)	
50-59	1,041	34(3.3)	1.88	(1.17 to 3.05)	
60-69	320	24(7.5)	4.29	(2.53 to 7.28)	
>=70	67	7(10.5)	5.73	(2.53 to 12.97)	
Gender					
Female	975	33(3.4)	1.84	(1.25 to 2.70)	0.002
Male	7,005	124(1.8)	1		
Ethnicity					
Caucasian	5,876	95(1.6)	1		<0.001
Black African	822	36(4.4)	1.47	(0.95 to 2.26)	
Other	1,182	26(2.2)	2.52	(1.72 to 3.70)	
Sexual orientation					
MSM	6,527	102(1.6)	1		
Heterosexual	1,299	50(3.9)	2.13	(1.52 to 2.99)	<0.001
Bisexual	154	5(3.3)	1.57	(0.64 to 3.85)	
IDU					
No	7,835	148(1.9)	1		0.001
Yes	145	9(6.2)	3.05	(1.56 to 5.99)	
#Duration since HIV-1$^+$ diagnosis, months					
<=35.48	1,908	30(1.6)	1		
35.49-89.98	1,904	30(1.6)	0.37	(0.22 to 0.63)	0.002
89.99-152.98	1,905	46(2.4)	0.47	(0.29 to 0.77)	
>152.98	1,907	45(2.4)	0.41	(0.25 to 0.68)	
Missing	356	6(1.7)	0.27	(0.11 to 0.69)	
ART treatment					
Naive	1,619	14(0.9)	0.95	(0.54 to 1.66)	0.857
Experienced	6,361	143(2.3)	1		

#Median Duration of ARV exposure, years (IQR)					
<=2.48	1,584	26(1.6)	1		0.249
2.49-5.82	1,560	23(1.5)	0.55	(0.31 to 0.96)	
5.83-11.09	1,577	42(2.7)	0.77	(0.47 to 1.28)	
>11.09	1,573	47(3.0)	0.82	(0.50 to 1.34)	
Treatment naive & missing	1,686	19(1.1)	0.93	(0.51 to 1.68)	
NRTI experienced					
Treatment naive	1,619	14(0.9)	0.95	(0.55 to 1.67)	0.782
No	30	1(3.3)	2.16	(0.30 to 15.40)	
Yes	6,331	142(2.2)	1		
PI experienced					
Treatment naive	1,619	14(0.9)	0.74	(0.42 to 1.31)	0.001
No	2,973	41(1.4)	0.52	(0.36 to 0.74)	
Yes	3,388	102(3.0)	1		
NNRTI experienced					
Treatment naive	1,619	14(0.9)	0.99	(0.56 to 1.74)	0.654
No	1,253	28(2.2)	1.22	(0.80 to 1.84)	
Yes	5,108	115(2.3)	1		
#Recent Pr:Cr ratio, mg/mmol					
<=10	878	10(1.1)	1		<0.001
11-14	701	5(0.7)	0.59	(0.20 to 1.74)	
15-22	608	10(1.6)	1.39	(0.58 to 3.33)	
>22	670	446.6)	5.31	(2.67 to 10.55)	
Not tested	876	88(1.7)	2.08	(1.08 to 4.00)	
#Recent eGFR, ml/min per 1.73m^2					<0.001
<72	1,867	118(6.3)	7.84	(4.65 to 13.23)	
72-82	1,991	12(0.6)	0.77	(0.37 to 1.64)	
83-90	1,611	11(0.7)	0.94	(0.44 to 2.02)	
>=91	2,511	16(0.6)	1		
#Nadir CD4, cells/µL blood					
<=108	1,989	92(4.6)	1		<0.001
109-205	1,983	38(1.9)	0.43	(0.30 to 0.63)	
206-331	1,996	19(1.0)	0.26	(0.16 to 0.42)	
>331	1,991	8(0.4)	0.17	(0.08 to 0.36)	
Unavailable	21	0(0.0)	0.00	(unstable)	
#Highest HIV RNA plasma load, copies/mL plasma					
<=9,758	1,984	36(1.8)	1		0.040
9,759-7,0415	1,983	33(1.7)	0.82	(0.51 to 1.31)	
70,416-254,845	1,984	29(1.5)	0.62	(0.38 to 1.01)	
>254845	1,983	58(2.9)	1.13	(0.75 to 1.72)	
Unavailable	46	1(2.2)	5.05	(0.69 to 37.22)	
AIDS					
No	7,893	156(2.0)	1		0.333
Yes	87	1(1.2)	0.43	(0.06 to 3.09)	

#Categories are derived using median and IQR
Pr:Cr is abbreviation used for protein:creatinine ratio
Where HR is 1, this is indicated as the reference category
Event time is defined as time since first measurement of eGFR to either moderate to advanced renal impairment or in subjects whose nadir eGFR was >=30 ml/min per 1.7m^2, data are censored at most recent eGFR result

Table 4.2.18: Multivariable Cox's Proportional Hazards Regression Model Showing Significant Independent Predictors of Likelihood of Stage 4 or 5 CKD Defined Using Nadir eGFR

	Total =7,980	Moderate to advanced renal impairment n=157	&Hazard ratio (HR)	95% CI	p-value
#Age at nadir eGFR (years)					
40-49	2,949	53(1.8)	1		
18-19	30	0(0.0)	0.00	(unstable)	0.978
20-29	1,005	6(0.6)	0.72	(0.30 to 1.75)	0.468
30-39	2,568	33(1.3)	0.85	(0.54 to 1.35)	0.499
50-59	1,041	34(3.3)	1.29	(0.76 to 2.17)	0.342
60-69	320	24(7.5)	2.07	(1.16 to 3.70)	0.014
>=70	67	7(10.5)	2.79	(1.15 to 6.78)	0.024
IDU					
No	7,835	148(1.9)	1		
Yes	145	9(6.2)	3.70	(1.79 to 7.67)	<0.001
Ethnicity					
Caucasian	5,876	95(1.6)	1		
Black African	822	36(4.4)	1.18	(0.74 to 1.87)	0.484
Other	1,182	26(2.2)	1.91	(1.12 to 3.27)	0.018
#Duration since HIV-1$^+$ diagnosis, months					
<=35.48	1,908	30(1.6)	1		
35.49-89.98	1,904	30(1.6)	0.32	(0.18 to 0.56)	<0.001
89.99-152.98	1,905	46(2.4)	0.28	(0.16 to 0.49)	<0.001
>152.98	1,907	45(2.4)	0.15	(0.08 to 0.27)	<0.001
Missing	356	6(1.7)	0.08	(0.03 to 0.20)	<0.001
Recent Pr:Cr ratio, mg/mmol					
<=10	878	10(1.1)	1		
11-14	701	5(0.7)	0.70	(0.24 to 2.06)	0.517
15-22	608	10(1.6)	1.29	(0.53 to 3.13)	0.574
>22	670	446.6)	2.86	(1.41 to 5.82)	0.004
Not tested	876	88(1.7)	2.20	(1.13 to 4.28)	0.020
PI experienced					
Treatment naive	1,619	14(0.9)	1.62	(0.80 to 3.27)	0.181
No	2,973	41(1.4)	0.65	(0.44 to 0.97)	0.033
Yes	3,388	102(3.0)	1		
#Recent eGFR, ml/min per 1.73m^2					
<72	1,867	118(6.3)	6.39	(3.66 to 11.17)	<0.001
72-82	1,991	12(0.6)	0.75	(0.35 to 1.61)	0.462
83-90	1,611	11(0.7)	0.96	(0.44 to 2.09)	0.924
>=91	2,511	16(0.6)	1		
#Nadir CD4, cells/μL blood					
<=108	1,989	92(4.6)	1		
109-205	1,983	38(1.9)	0.53	(0.36 to 0.78)	0.001
206-331	1,996	19(1.0)	0.29	(0.17 to 0.49)	<0.001
>331	1,991	8(0.4)	0.11	(0.05 to 0.26)	<0.001
Unavailable	21	0(0.0)	0.00	unstable	0.995

&Adjusted for gender, sexual orientation, PI exposure, maximum viral load and other factors in the model
#Categories are derived using median and IQR
Pr:Cr is abbreviation used for protein:creatinine ratio
Where HR is 1, this is indicated as the reference category
Event time is defined as time since first measurement of eGFR to either moderate to advanced renal impairment or in subjects whose nadir eGFR was >=30 ml/min per 1.73m^2, data are censored at most recent eGFR result

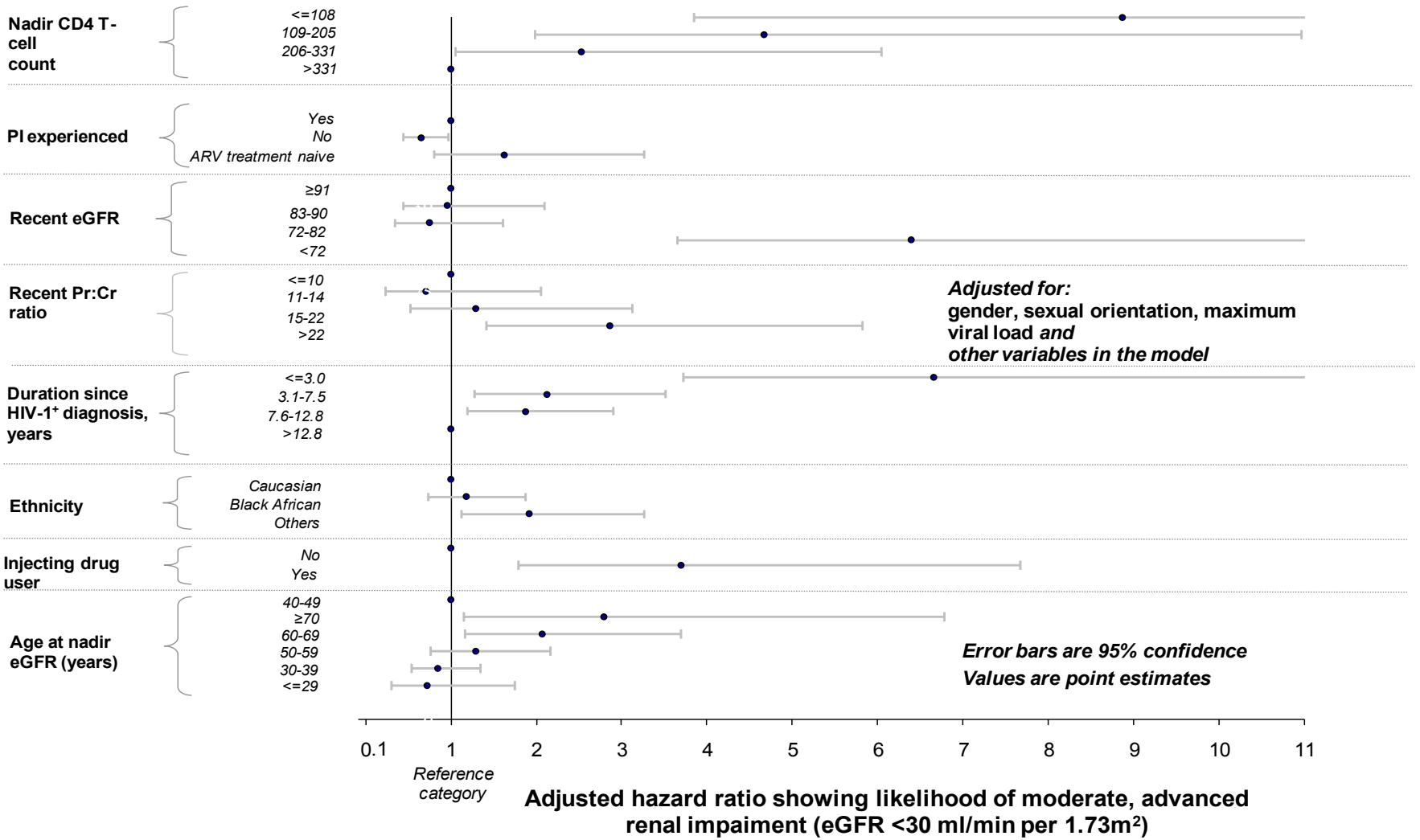


Figure 4.2.6: Multivariable Cox's Proportional Hazards Regression Model Figure Showing Significant Independent Predictors of Likelihood of Stage 4 or 5 CKD Defined Using Nadir eGFR

- ***UNIVARIATE AND MULTIVARIABLE COX'S PROPORTIONAL HAZARDS REGRESSION ANALYSIS IN PATIENTS ON FIRST-LINE ART***

Of the 4,346 PLHIV who started first-line HAART with available nadir eGFR, 87 (2.0%) had CKD 4-5 with an eGFR <30 ml/min per 1.73m^2 or moderate to advanced renal impairment. The univariate analysis investigated the association between a number of variables and nadir eGFR <30 ml/min per 1.73m^2 (Table 4.2.19). The multivariable model derived from that indicated that the following factors were significantly associated with a nadir eGFR of < 30 ml/min per 1.73m^2; PLHIV aged older than 60 years, being an injecting drug user (IDU) or non-Caucasian, having a recent protein:creatinine ratio greater than 20 mg/mmol, having had a recent eGFR below 73 ml/min per 1.73m^2 and having had nadir CD4 T-cell count below 113 cells/mm^3 (Figure 4.2.7).

Table 4.2.19: Univariate Cox's Proportional Hazards Regression Model showing Likelihood of Stage 4 or 5 CKD Defined Using Nadir eGFR In Patients on First-line ART

		Total 4,346	Moderate to advanced renal impairment n=87	HR	95% CI	p-value
Age at nadir eGFR (years)	40-49	1,645	30(1.8)	1		0.002
	18-19	13	0(0.0)	0.00	(unstable)	
	20-29	483	6(1.2)	1.03	(0.42 to 2.53)	
	30-39	1,579	23(1.5)	1.09	(0.63 to 1.88)	
	50-59	454	14(3.1)	1.71	(0.88 to 3.34)	
	60-69	142	12(8.5)	5.06	(2.52 to 10.19)	
	≥70	30	2(6.7)	3.61	(0.85 to 15.33)	
Gender	Female	575	24(4.2)	2.56	(1.60 to 4.09)	<0.001
	Male	3,771	63(1.7)	1		
Ethnicity	Caucasian	3,167	47(1.5)	1		<0.001
	Black African	522	26(5.0)	1.51	(0.83 to 2.75)	
	Other	657	14(2.1)	3.22	(1.99 to 5.20)	
Sexual orientation	MSM	3,481	50(1.4)	1		<0.001
	Heterosexual	797	36(4.5)	2.91	(1.89 to 4.46)	
	Bisexual	68	1(1.5)	0.78	(0.11 to 5.63)	
IDU	No	4,282	82(1.9)	1		0.002
	Yes	64	5(7.8)	4.16	(1.69 to 10.25)	
Duration since HIV-1$^+$ diagnosis, months	<=45.18	1,063	18(1.7)	1		0.241
	45.19-85.27	1,066	21(2.0)	0.58	(0.30 to 1.12)	
	85.28-134.85	1,063	26(2.5)	0.63	(0.34 to 1.19)	
	>134.85	1,064	21(2.0)	0.49	(0.25 to 0.96)	
	Missing	91	1(1.1)	0.25	(0.03 to 1.89)	
First-line HAART	2NRTI+NNRTI	3,422	65(1.9)	1		0.133
	2NRTI+boosted-PI	924	22(2.4)	1.45	(0.89 to 2.35)	

Duration of ARV exposure, years	<=1.89	1,080	10(0.9)	1		0.790
	1.90-4.08	1,084	21(1.9)	1.44	(0.67 to 3.06)	
	4.08-7.36	1,074	26(2.4)	1.26	(0.60 to 2.64)	
	>7.36	1,085	30(2.8)	1.30	(0.63 to 2.71)	
	missing	23	0(0.0)	0.00	(-)	
Recent Pr:Cr ratio, mg/mmol	<=10	581	8(1.4)	1		<0.001
	11-13	384	3(0.8)	0.54	(0.14 to 2.03)	
	14-20	420	4(1.0)	0.68	(0.20 to 2.25)	
	>20	417	26(6.2)	4.30	(1.95 to 9.50)	
	Not tested	2,541	46(1.8)	1.56	(0.74 to 3.31)	
Recent eGFR, ml/min per 1.73m^2	<=73	1,076	63(5.9)	6.83	(3.51 to 13.33)	<0.001
	74-82	1,057	8(0.8)	0.90	(0.36 to 2.29)	
	83-90	892	6(0.7)	0.81	(0.29 to 2.23)	
	>90	1,321	10(0.8)	1		
Nadir CD4, cells/µL blood	Unavailable	1	0(0.0)	0	(unstable)	<0.001
	<=113	1,085	46(4.2)	3.50	(1.71 to 7.17)	
	114-193	1,086	20(1.8)	1.56	(0.71 to 3.44)	
	194-279	1,084	12(1.1)	1.02	(0.43 to 2.42)	
	>279	1,090	9(0.8)	1		
Highest HIV RNA plasma load, copies/mL plasma	<=9254	1,086	22(2.0)	1		0.245
	9255-82651	1,086	26(2.4)	0.93	(0.53 to 1.64)	
	82651-273786	1,086	15 (1.4)	0.50	(0.26 to 0.97)	
	>273786	1,086	24(2.2)	0.76	(0.42 to 1.36)	
	Unavailable	2	0(0.0)	0.00	(-)	
AIDS	No	4,315	87(2.0)	1		-
	Yes	31	0(0.0)	0.0	(-)	
Year of starting first line HAART	1996	4	0(0.0)	0.0	-	0.725
	1997	29	0(0.0)	0.0	-	
	1998	81	2(2.5)	1.71	(0.15 to 19.00)	
	1999	197	8(4.1)	3.01	(0.37 to 24.27)	
	2000	160	2(1.3)	0.92	(0.08 to 10.24)	
	2001	189	4(2.1)	1.58	(0.18 to 14.21)	
	2002	195	5(2.6)	1.96	(0.23 to 16.89)	
	2003	242	5(2.1)	1.51	(0.18 to 13.00)	
	2004	332	10(3.0)	2.26	(0.29 to 17.75)	
	2005	306	6(2.0)	1.43	(0.17 to 12.00)	
	2006	322	12(3.7)	3.00	(0.39 to 23.26)	
	2007	424	9(2.1)	1.98	(0.25 to 15.71)	
	2008	555	8(1.4)	1.51	(0.19 to 12.14)	
	2009	523	11(2.1)	2.68	(0.35 to 20.79)	
	2010	574	4(0.7)	1.15	(0.13 to 10.28)	
	2011	213	1(0.5)	1		

Event time is defined as time since first measurement of eGFR to either moderate to advanced renal impairment or in subjects whose nadir eGFR was ≥30 ml/min per 1.7m^2, data are censored at most recent eGFR result
Pr:Cr is abbreviation used for protein:creatinine ratio

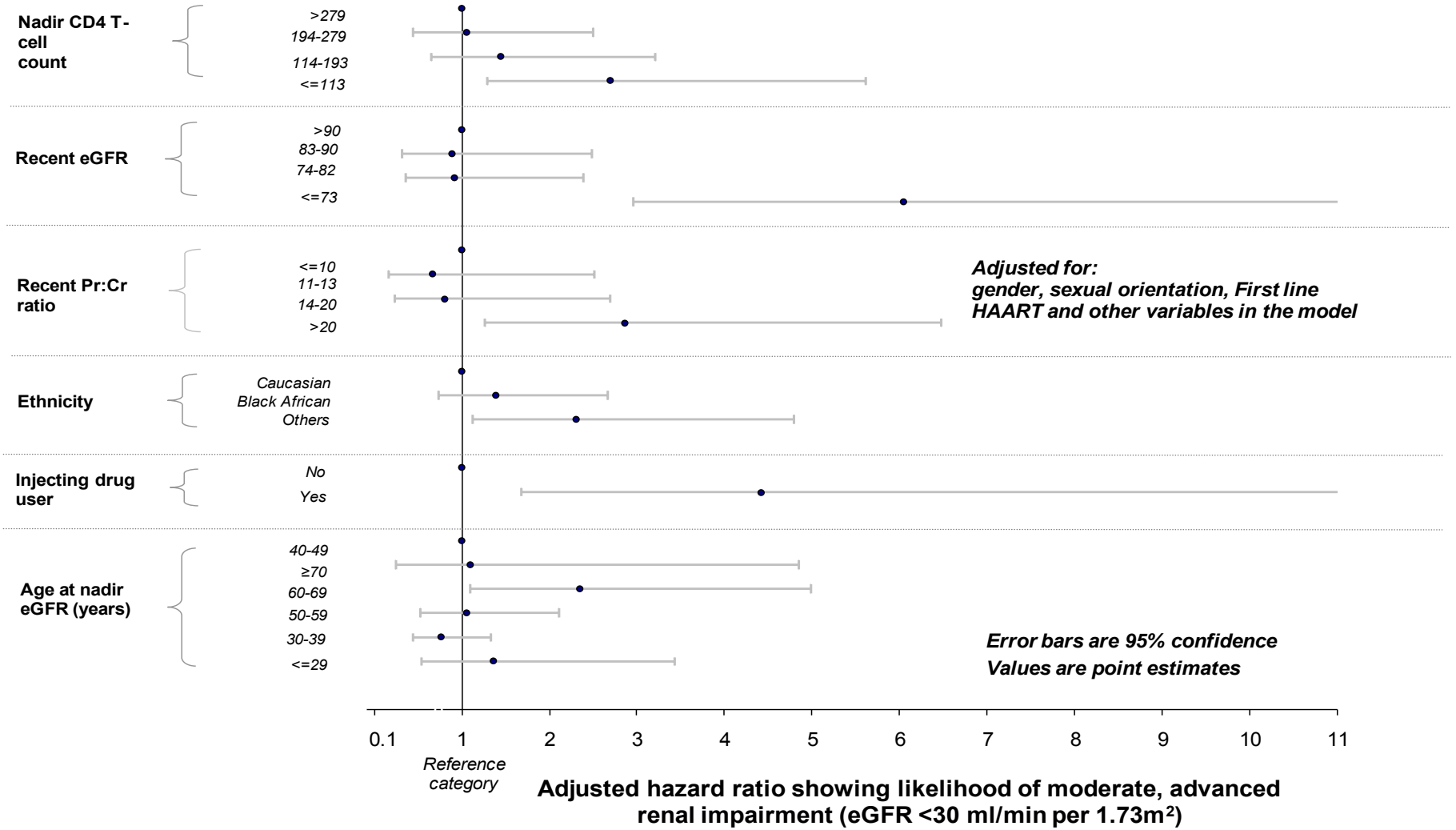


Figure 4.2.7: Multivariable Cox's Proportional Hazards Regression Model Figure Showing Significant Independent Predictors of Likelihood of Stage 4 or 5 CKD Defined Using Nadir eGFR In Patients on First-line ART

4.3 DISCUSSION

- ***NOVEL APPROACH***

Longitudinal changes in eGFR and protein:creatinine ratio associated with increasing age in adults living with HIV-1 infection have to date not been reported. The increased life-expectancy of HIV-1^+ people has given rise to the likelihood of developing non-HIV co-morbidities associated with ageing and the underlying process of immunosenescence (Spits, 2002). CKD is one of the non-HIV co-morbidities that is associated with increasing age. Using eGFR and protein:creatinine ratio as markers of renal function, longitudinal analysis indicated that renal function in HIV-1^+ patients deteriorate more with increasing age than 'healthy' older people.

- ***SUMMARY OF FINDINGS***

Cross sectional data showed that HIV-1^+ people who were 'older' aged ≥ 50 years had lower nadir and first ever CD4 T-cell count compared to 'younger adults'. While recent values of CD4, CD3, and CD8 T cells, CD56 NK-cell counts, and maximum HIV RNA plasma load were higher in the 'older' age group compared to the 'younger adults' (18-39 years), CD19 B-cell counts were lower in 'young adults' compared to 'middle aged' and 'older' adults. Patients who were 'older' aged ≥ 50 years were more likely to have ever tested hepatitis B antibody and hepatitis B core antibody positive. A number of further analyses were undertaken and summary of these results are under Appendix A4.1.

Both ART naïve and HIV-1^+ people exposed to different ARV classes and commonly prescribed ART regimens showed reductions in eGFR with increasing age well below the creatinine clearance levels observed in uninfected people. Similar but empirically weaker conclusions can be drawn from the protein:creatinine ratio analyses. ART naïve patients demonstrated increases in protein:creatinine ratio with increasing age as did patients exposed to different ARV classes and first-line regimens. However, due to the small number of patients who had their protein:creatinine ratio measured conclusions drawn from these analyses need to be viewed with caution. From the regression models, the degree of immuno-suppression, as indicated by nadir CD4 count, was predictive of likelihood of developing stage 4 or 5 CKD among HIV-1^+ people. The relationship between immunosuppression and end-stage renal disease has been demonstrated (Betjes *et al.*, 2011) and a direct relationship between severity of immuno-suppression and occurrence of other non-HIV co-morbidities also seems to exist (Sighem *et al.*, 2013). The increased probability of developing CKD is also related to the ageing process itself in PLHIV and other causes of CKD including cardiovascular disease, hypertension and diabetes mellitus. Increased duration of known HIV sero-status was associated with a reduced probability of stage 4 or 5 CKD and may be due to treatment and care received by PLHIV diagnosed earlier in their disease course (Sighem *et al.*, 2013).

- ***LIMITATIONS***

The eGFR according to the MDRD formula is standardised for body surface area, age, ethnicity and gender (Levey *et al.*, 2006). However, some remaining residual effect of age may remain and some of the changes in renal function observed may be due to a persistent age effect. Although the data show a steeper decline in renal function markers in 'middle-aged' adults 40-49 years and adults 50 years or older, the number of patients aged 70 years or older were relatively small resulting in greater variations in trends of the renal markers observed in these age groups. Furthermore, renal function measurements were taken from a diverse group of patients who contributed results at different ages. Ideally one would like to prospectively follow-up a defined group of 'young' HIV-1^+ adults as they age, and monitor changes over time, but such data would take a long time to collect. Furthermore, ART regimens are also likely to change over such long time periods, compounding longitudinal analyses.

Creatinine clearance values from HIV uninfected people should strictly speaking not be compared with eGFR from HIV-1^+ people as the creatinine clearance values were not adjusted for body surface area, age, ethnicity and gender (Rowe *et al.*, 1976). In the absence of such standardised longitudinal data from a 'healthy' populations, the current comparator including a 20% reduction to adjust for any over-estimation (Breyer & Qi, 2010) provides an insight into the degree of renal impairment associated with HIV-1 infection or ART.

While eGFRs are less reliable for assessing stage 1 and 2 CKD when eGFR is 60 to 90 ml/min per 1.73m^2, it does becomes more reliable for picking up stage 3 CKD, with eGFR rates between 30 and 59 ml/min per 1.73m^2 (Stevens *et al.*, 2007). A recent meta analyses demonstrated that the MDRD equation to calculate eGFR is not an optimal measure for all populations and eGFR ranges calculated using a single equation requires a tradeoff at higher or lower values of eGFR ranges (Earley *et al.*, 2012). The eGFR is predicted based on the MDRD formula which employs serum creatinine measurements and patients body surface area, age, ethnicity and gender (Levey *et al.*, 2006). However, some residual effect of age or other contributing factors may remain and influence the changes in renal function observed. In addition, the MDRD formula excluded patients with renal failure, though it included patients aged 18-70 years, but it has not yet been validated in an 'elderly' or 'frail' population (Levey *et al.*, 1999). Other formulae which has also appeared in the literature to estimate creatinine clearance from a single estimate of plasma creatinine without urine collection include the Cockcroft-Gault formula (Cockcroft & Gault, 1976). An alternative to the MDRD formula is the Cockcroft-Gault formula where the creatinine clearance is predicted from serum creatinine measurements and patients' age, gender and weight. While some prefer to use this formula (Cockcroft & Gault, 1976), as it is considered it to give better estimates of renal function in subjects with well preserved normal renal function (Robertshaw *et al.*, 1989), others prefer to use the MDRD

formula (Barraclough *et al.*, 2009). As routinely measured weights were unavailable for the cohort studied, the Cockcroft-Gault formula could not be used in this study.

- ***REPORTS FROM OTHER STUDIES***

Other methods to improve early diagnosis of CKD have been suggested, including the use of improved equations to calculate eGFR by the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) (Levey *et al.*, 2009), however so far these have not yet been taken up by international guidelines (Levey *et al.*, 2011) although some centres in the US are now starting to use this formula (Fine & Gallant, 2013). The CKD-EPI equation is thought to provide a more precise and accurate estimates of GFR, compared with the MDRD equation (Levey *et al.*, 2009). This was the conclusion reported in a recent review that suggested that the CKD-EPI equation could provide a more precise and less biased estimate of kidney function in HIV-infected populations (Ando *et al.*, 2012).

Clinical laboratories are routinely reporting eGFR, and computerised medical records highlight clinicians to the presence of CKD (Go *et al.*, 2004) however existing formulae to calculate creatinine clearance have to date not been validated in HIV-1^+ people (Barraclough *et al.*, 2009). Blood urea levels increase with increased protein intake. Conversely muscle mass decreases with age or advancing disease as does serum creatinine levels and these changes may mask true renal function, while variations are also observed between different ethnic groups (Stevens *et al.*, 2006). Some studies indicated that low muscle mass is common in men living with HIV and that sarcopenia affecting muscle function could be present in a substantial number of people (Buehring *et al.*, 2012).

Serum cystatin C is an alternative marker of kidney function and some argue that this better predict morbidity and mortality in “older” adults as compared with creatinine, as cystatin C seems unaffected by gender, age, ethnicity, muscle mass, and co-existing malignancy (Shlipak *et al.*, 2005; Chew *et al.*, 2008). However, some question the utility of routine cystatin C tests (Fine & Gallant, 2013) and these were not performed in our local setting and therefore this marker could not be used. Other clinics have reported a higher cost for using this assay and this may be one possible reason why it has not been introduced into routine testing in the NHS (Chew *et al.*, 2008).

ART naïve patients showed similar changes in renal function to the ART experienced patients. HIVAN is an important cause of CKD among the HIV-1^+ people. While some statistically significant differences were observed in older patients on different first-line regimens, clinically these did not

seem large. While no differences were observed between ARV classes, at a population level, specific ARVs have been associated with impaired renal function (Mocroft *et al.*, 2007; Mocroft *et al.*, 2010; Ryom *et al.*, 2013). Recently Tenofovir was shown to be associated with a greater decline in renal function than other NRTIs as indicated by a greater degree of inhibition of intra-cellular telomerase compared with other NRTIs (Leeansyah *et al.*, 2013). Renal function decline did not lead to greater rates of discontinuation of therapy and the renal function decline associated with Tenofovir and other NRTIs are considered to be a function of a longer duration of ARV treatment (Gallant *et al.*, 2005). Increased renal impairment was also observed in patients with past exposure to Indinavir, Atazanavir or Lopinavir (Mocroft *et al.*, 2007; Mocroft *et al.*, 2010; Young *et al.*, 2012; Fine & Gallant, 2013; Ryom *et al.*, 2013). Another recent study demonstrated that Tenofovir with a boosted PI led to a greater decline in eGFR than Tenofovir with a NNRTI (Young *et al.*, 2012) however the new formulations of Tenofovir may have less deleterious effect on renal function (Zolopa *et al.*, 2013).

- ***FUTURE WORK***

While specific ARV drugs are associated with impaired renal function, the findings presented in this study suggest that ART at a population level does not appear to be so nephrotoxic. This was also the conclusion of a recent Spanish study which suggested that ART may have a renal-protective effect (Masiá *et al.*, 2012) although the follow-up period of their cohort was not as long as the study presented here. One of the consequences of declining renal function with age is reduced renal excretion of drugs and metabolites, potentially increasing the likelihood of nephrotoxicity. As HIV-1$^+$ people age this poses a clinical challenge, in terms of the increasing number of patients requiring regular renal evaluation and, if progression to end-stage renal disease occurs, they may require renal dialysis. Renal function impairment may require the adjustment ARV drugs and other drug dosages in 'older' HIV-1$^+$ people.

While the time to start ART in HIV-1$^+$ people is still under discussion and awaiting the outcome of the START trial which randomised people with CD4 T-cell count ≥500 cells/mL to immediate ART versus deferred ART treatment to when CD4 T-cell count dropped to below 350 cells/mL (Babiker *et al.*, 2013) a number of clinicians and guidelines recommend that some HIV-1$^+$ people can start ART at diagnosis irrespective of CD4 T-cell count (Thompson *et al.*, 2012). The finding that HIV-1 infection and related immuno-suppression is a major driver of decreasing renal function, might add pressure to start treating HIV-1$^+$ people with ART when HIV-1 infection is first diagnosed to reduce the inflammatory and other processes associated with HIVAN especially given that ART seems to retain its immune restorative capacity on a long-term basis (Wright *et al.*, 2013).

Finally, while we can celebrate the success of improved length and quality of life that ART provides for HIV-1 infected people, as they get older they are more likely to develop other conditions,

previously rarely seen. This not only pertains to deteriorating renal function but also cardiovascular diseases, cancers, neurological problems and other morbidity patterns associated with increasing age and frailty, all of which have to be monitored and managed in HIV-1^{+} people as they live longer.

Chapter 5: Population Cost

CHAPTER 5: POPULATION COST FOR TREATING PEOPLE LIVING WITH HIV IN THE UK

Note: A modified version of this chapter has been published.

Mandalia S, Mandalia R, Lo G, Chadborn T, Sharott P, Youle M, Anderson J, Baily G, Brettle R, Fisher M, Gompels M, Kinghorn G, Johnson M, McCarron B, Pozniak A, Tang A, Walsh J, White D, Williams I, Gazzard B, Beck EJ; NPMS-HHC Steering Group.
Rising population cost for treating people living with HIV in the UK, 1997-2013.
PLOS one. 2010 Dec 30;5(12)

Population cost paper cited here.
http://www.publications.parliament.uk/pa/ld201012/ldselect/ldaids/188/18805.htm#a1

'Life is priceless'. However, decisions affecting lives are not only made by individuals but also by parliaments and public authorities on a regular basis. This implies lengthening of human life against finite resources are weighed up. The NHS lets political authorities decide on new pharmaceuticals, new therapies, and new medical devices are to be covered by the NHS. As a general rule, cost-increasing products that bring therapeutic advantages are preferred, by reducing the risk of pre-mature mortality in the at risk population. These types of new products may involve additional expenditure, for example, an installation of new dialysis equipment for patients with end-stage renal failure.

This chapter describes some of the analyses that were performed on data from the NPMS-HHC collaboration and these are based on data transferred to the NPMS-HHC Co-ordinating and Analytic Centre (CAC) in 2007/2008. At the time results presented in this chapter was carried out, data from the NPMS-HHC collaboration were available for the study period from 1st January 1996 to 31st December 2006. The use and cost of hospital services presented in this chapter covers data for this period whilst population costs are from 1997 due to unavailability of UK HIV-1^{+} population figures for 1996. The analyses of data are based on a study that routinely collect data from multi HIV centres in the UK. The data are prospectively collected HIV-1 cohort data at the participating sites and are longitudinal (Chapter 2: Figure 2.1.1). The primary aim of this multi-centre cohort is to prospectively monitor the use, cost, and outcome of HIV service provision in the UK based on data from the NPMS-HHC participating centres.

The aims of this chapter are: to estimate the use and cost of hospital services of HIV-1 diagnosed people in the UK; to estimate the population cost of treatment and care of HIV-1 population accessing HIV-1 care in the UK in the NHS by stage of HIV-1 infection based on the available population figures of number of people diagnosed and who had accessed the HIV services in the UK NHS between 1997-2006; to project population cost of treatment and care of HIV-1 for 2007-2013.

5.0 HEALTH CARE EXPENDITURE IN THE UK

Since the establishment of the Britain's National Health Service (NHS) on 5th July 1948, the UK Government spent £11.4 billion on heath during the first full year of its operation. The public spending of 2005 election briefing document showed £84.7 billion were spent on health (Emmerson & Frayne, 2005) . The figure since 1948 was over ten-fold higher £121 billion in 2010/11 (Harker, 2012). The public health care expenditure (HCE) in the UK has increased since 1948 in cost figures as well as a share of Gross Domestic Product (GDP) (Figure 5.0.1).

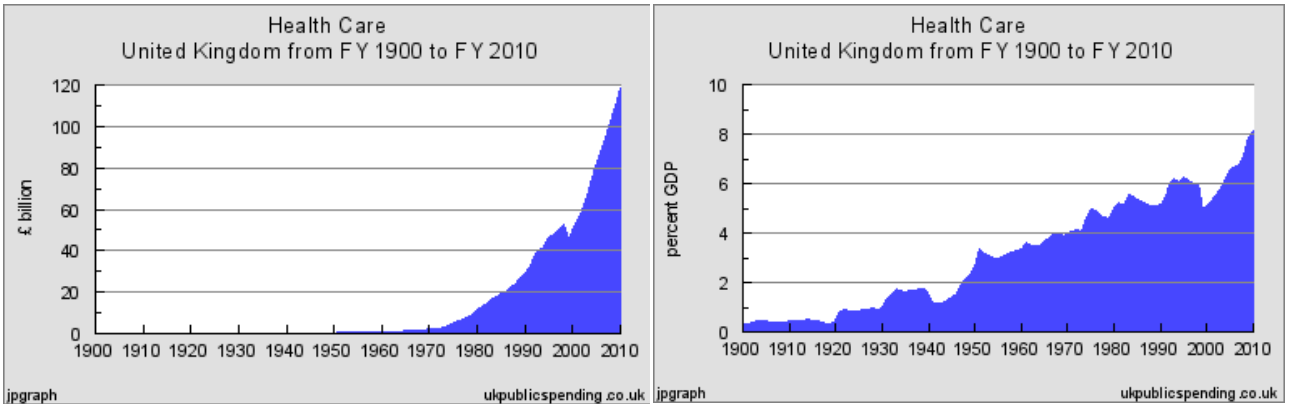


Figure 5.0.1 UK Health Care Expenditure 1900-2011
Source of graphs: http://www.ukpublicspending.co.uk/spending_brief.php
FY=Financial year

Since 1990, the UK Government increased HCE as a percentage of GDP from 5.1% in 1990, to over 8.1% in 2010 (Figure 5.0.1). To investigate whether the UK investments in health care and public health have been followed by the expected improvements in health using HIV-1 infection as one of the disease areas investigated. A recent study showed that the years of life lost (YLLs) due to premature mortality caused by HIV-1 infection in 20-54 year olds had declined between 1990 and 2010 thereby demonstrating success of HIV-1 treatment and care in the UK (Murray *et al.*, 2013).

5.1 INTRODUCTION

Further details are described in Chapter 1 of this PhD thesis. HIV-1 pandemic continues to grow and by end of 2012, 2.4% (0.86 million, range: 0.80 to 0.93) of the world's adults and children were estimated to be living with HIV-1 infection in Western and Central Europe (UNAIDS Report on the Global AIDS Epidemic, 2013). As more drugs for treating HIV-1 become available people who have access to these drugs continue to live longer and therefore HIV-1 in these group of people has become a chronic illness. By the end of 2012 76,500 people diagnosed with HIV-1 had used the NHS services (credible interval: 75,000 to 78,000) (Aghaizu *et al.*, 2013). The number of people living with HIV-1 infection continues to increase due to transmission of HIV-1 infection either in the UK or abroad (Health Protection Agency, 2009) as well as increased survival of people infected with the virus due to more effective ART regimens (Beck *et al.*, 2008).

CHAPTER 5

POPULATION COST

In 2010, 27,116 newly diagnosed cases of HIV-1 infection were reported by 28 countries of the European Union and European Economic Area (EU/EEA). Half of the HIV-1 cases reported with information on CD4 T-cell counts had low CD4 T-cell counts (<350 cells/μL blood) at time of HIV-1 diagnosis indicating late diagnosis. The UK was one of the countries to have reported highest rates of HIV-1 diagnoses and in 2010, this was reported to be 10.7 cases per 100,000 population and 61% of whom were heterosexually transmitted cases (World Health Organisation, 2011a). Therefore the increases in number of HIV-1 infected people accessing specialist care directly impacts on the NHS cost burden for treating HIV-1 infected people in UK NHS hospitals.

In the UK, HIV-1 infected people seek long-term medical help at specialist as well as non-specialist NHS hospitals and treatment centres. The number of people accessing the NHS for HIV-1 services is increasing over time and these increases in access of care in the NHS implies that the costs associated with HIV-1 care also continue to increase. With increasing number of HIV-1 infected people requiring long-term medical care in the NHS the cost burden therefore continue to increase. At the time this work was undertaken, no publically available population cost estimates were available for treating HIV-1 infected people in the UK using the NHS services. In the current economic climate which is going through periods of cutting public expenditure, there was a need for estimates of population cost for treating HIV-1 infected people in the UK so that appropriate budgets within the NHS could be allocated for treatment and care of HIV-1 infected people. This study aimed to provide an estimate of direct population cost of treatment and care of HIV-1^{+} people.

5.1.1 NPMS-HHC COHORT

The National Prospective Monitoring System on the use, cost, outcome of HIV-1 service provision in UK hospitals - HIV Health-economics Collaboration (NPMS-HHC) was set up to prospectively monitor routinely collected data on HIV-1 infected people receiving care in the participating HIV centres during the study period which for this chapter ranged from 1st January 1996 to 31st December 2006. The objective was to describe, over time, how HIV-1^{+} patients who are in different stages of HIV-1 infection and are receiving HIV-1 care in the NHS HIV-1 specialist centres in the UK the use and cost of HIV services changed over time. The particular aim of this chapter is to estimate quantitatively changing use of HIV-1 services within the NHS setting and changing costs associated with this. The study to date has generated unique descriptive information on the changing use and cost of HIV-1 services in the UK, however the primary aim of this PhD chapter is to describe and provide direct population cost estimates of those who are accessing HIV-1 care in the NHS, and address the two relevant questions: what is direct population cost of treating HIV-1^{+} people in the UK? What are the expected future HIV-1 direct population costs?

5.2 RESULTS

- ***DEMOGRAPHIC CHARACTERISTICS***

Of the 24 UK HIV centres between 1996 and 2006 participating with the NPMS-HHC data were available from 14 HIV centres 7 of these were London centres. 28,925 HIV-1^+ people were managed in these 14 centres between 1996 and 2006. 76% were men while 51% were Caucasian. Not all clinics provided details of sexual orientation of the subjects. Of those whose sexual orientation was known, 62% were MSM. A small proportion of subjects were known IDU (4.5%) (Table 5.2.1) and significantly greater proportion of men (79%) were attending London HIV centres for care (χ^2 test, $p<0.001$).

Table 5.2.1: Demographic Characteristics of Patients Seen in the 14 Participating Centres

	N(%)		
	Total number of HIV-1^+ patients attending for treatment and care	**Attendances at London clinics**	**Attendances at non-London clinics**
	N=28,925	**N=22,962**	**N=5,963**
Gender			
Male	22058(76.3)	18272(79.6)	3786(63.5)
Female	5876(20.3)	4429(19.3)	725(24.3)
Unavailable	986(3.4)	261(1.1)	863(12.2)
Ethnic origin			
Caucasian	14607(50.5)	12352(53.8)	2255(37.8)
Black African	4757(16.4)	3894(17.0)	863(14.5)
Other	5232(18.1)	4921(21.4)	311(5.2)
Unavailable	4329(15.0)	1795(7.8)	2534(42.5)
Sexual orientation			
MSM	9868(34.1)	8006(34.9)	1862(31.2)
Heterosexual	5537(19.1)	3993(17.4)	1544(24.9)
Bisexual	596(2.1)	253(1.1)	343(5.8)
Unknown	12924(44.7)	10710(46.6)	2214(37.1)
IDU	1303(4.5)	695(3.0)	608(12.2)

The ART prescription pattern changed over time from predominantly mono and dual therapy in the mid 1990's to mostly triple and quadruple HAART whilst prescription of mono and dual therapy declined over time for all stages of HIV-1 infection. Over the same time period a sharp increase in prescription of HAART consisting of triple and quadruple or more therapy was observed and this is in line with the changes that occurred in prescription of HAART containing regimens over time (Table 5.2.2, Figure 5.2.1).

Table 5.2.2: ARV Drug Prescription Trends Over Time by Stage of HIV-1 Infection

	Study period 1st Jan 1996 - 31st Dec 1996											
	Number of recent prescription of ARV drugs											
	Total	Number of patients prescribed Mono therapy (Number of ARV drugs=1)			Number of patients prescribed Dual therapy (Number of ARV drugs=2)			Number of patients prescribed Triple therapy (Number of ARV drugs=3)			Number of patients prescribed Quadruple or more therapy (Number of ARV drugs=4 or more)		
Year		ASx	Sx Non AIDS	AIDS	ASx	Sx Non AIDS	AIDS	ASx	Sx Non AIDS	AIDS	ASx	Sx Non AIDS	AIDS
1996	**3182**	1041	21	349	503	43	449	174	18	495	16	7	66
1997	**4477**	1219	25	417	590	45	334	567	67	879	49	16	269
1998	**5054**	1286	23	477	473	37	213	841	99	1047	126	23	409
1999	**5437**	1355	20	466	369	22	165	1015	133	1163	209	32	488
2000	**5627**	1343	18	457	340	17	125	1088	149	1143	285	51	611
2001	**5863**	1270	22	441	391	10	128	1181	182	1224	316	61	637
2002	**5793**	1131	17	388	317	10	116	1196	199	1294	361	70	694
2003	**5338**	875	16	354	217	10	124	1127	211	1319	300	71	714
2004	**5017**	569	15	293	170	7	131	1136	222	1226	343	77	828
2005	**5064**	368	10	212	198	8	155	1214	217	1239	475	100	868
2006	**5077**	216	10	120	166	6	154	1325	223	1229	604	107	917

A small proportion (15%) were known to have attended more than one HIV treatment and care centre within the UK during the study period. 18% of the study sample had attended more than one London HIV centre compared to 3% who had attended more than one out of London HIV centre. A small proportion (3%) had attended both London and non-London centres during the study period (1996 to 2006) and over time this has declined (Table 5.2.3).

Table 5.2.3: Proportion of HIV-1$^+$ People Who Attended More Than One HIV-1 Centres

Year	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	Overall number
Across all clinics	19%	18%	17%	15%	14%	12%	15%	14%	15%	12%	10%	15%
Attended more than one London clinic during the year	23%	22%	20%	17%	17%	15%	17%	17%	17%	16%	15%	18%
Attended more than one non-London clinic during the year	3%	3%	3%	3%	3%	4%	4%	4%	4%	3%	3%	3%
Attending a London and non-London clinic during the year	6%	5%	5%	3%	3%	3%	3%	2%	2%	2%	2%	3%

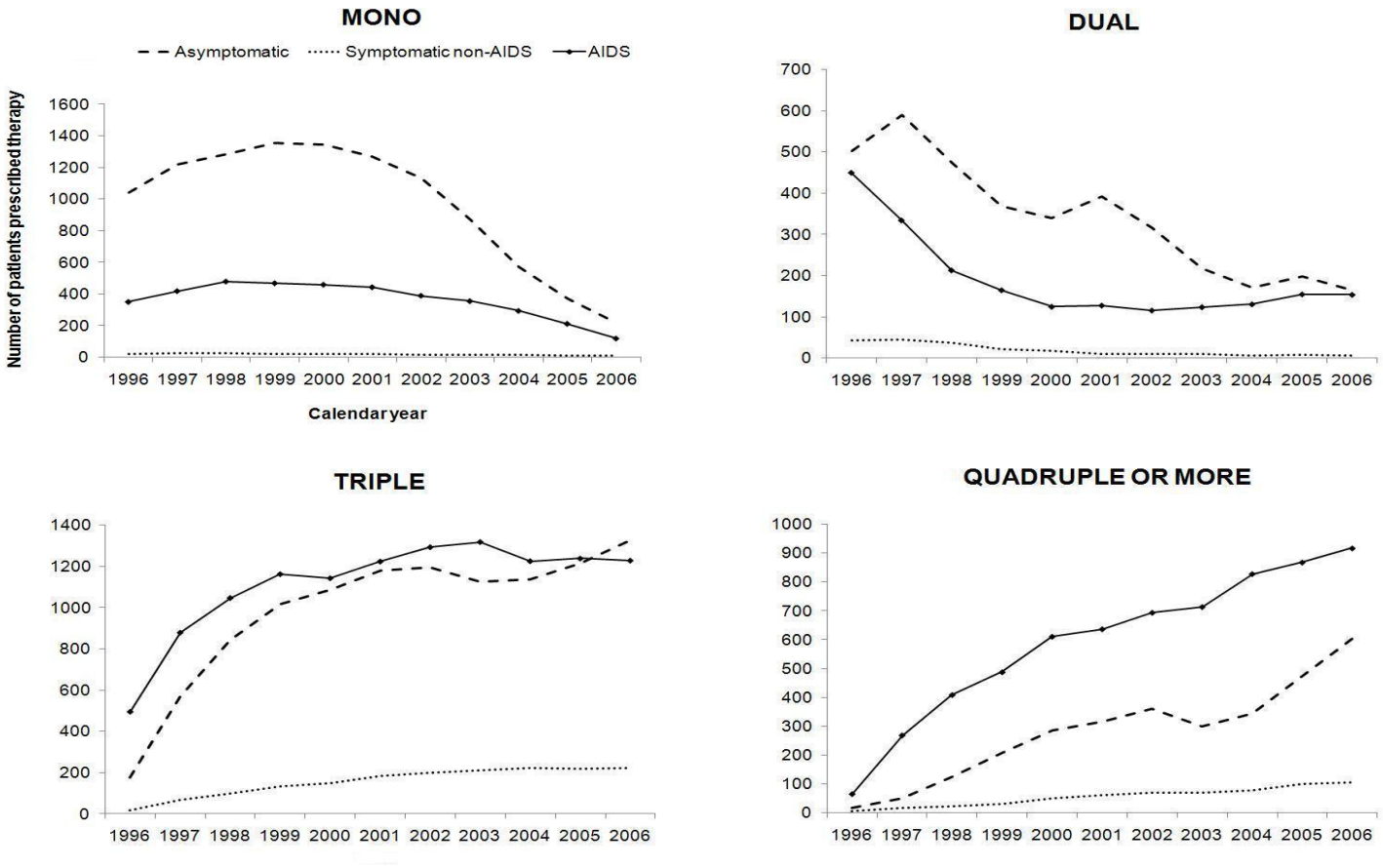


Figure 5.2.1: Changing ARV Prescription Pattern Over Time by Stage of HIV-1 Infection

5.2.1 USE OF HOSPITAL SERVICES BY STAGE OF HIV-1 INFECTION

- ***AGGREGATE***
    - **Outpatient**

Asymptomatic: The mean number of outpatient visits per patient year (PPY) increased for *asymptomatic* patients from 4.45 in 1996 to 7.95 by 2002 and thereafter a steady decline was observed and by 2006 the mean number of outpatient visits PPY was estimated at 5.83 in this group Table 5.2.4.

Symptomatic non-AIDS: Increases were observed in the mean number of outpatient visits PPY from 5.38 in 1996 to 8.98 by 2005 however a decline in outpatient visits PPY was observed since and by 2006 this was 7.62 Table 5.2.4.

AIDS: The mean number of outpatient visits per patient year increased for HIV-1^+ people with AIDS over time from 10.57 in 1996 to 11.63 by 1998 and a gradual decline was observed from 1998 and by 2006 this was 7.00 Table 5.2.4.

- **In-patient**

Asymptomatic: The mean number of inpatient days PPY decreased for asymptomatic HIV-1^+ people from 3.66 in 1996 to 2.38 by 2006 Table 5.2.4.

Symptomatic non-AIDS: The mean number of inpatient days PPY increase was seen HIV-1^+ people with symptomatic non-AIDS between 1996 to 1999 from 1.93 to 3.84 and a decline thereafter to 2.72 by 2006 Table 5.2.4.

AIDS: The mean number of inpatient days PPY amongst HIV-1^+ people AIDS stage of infection a decline was seen from 13.53 in 1996 to 7.55 by 1999 with a slight unsteady trend being observed thereafter to 2006 and by the end of the study period it was estimated to 9.26 Table 5.2.4.

- **Dayward**

Asymptomatic: The mean number of dayward visits PPY showed very little change in the asymptomatic HIV-1^+ people between 1996 when this was 0.30 and steady increases were observed thereafter and by 2006 this was estimated at 1.29 Table 5.2.4.

Symptomatic non-AIDS: The mean number of dayward visits PPY declined slightly then increased showing a slight U shaped change in HIV-1^+ people in symptomatic non-AIDS stages of HIV-1 infection between 1996 and 2006 from 0.37 to 0.84 respectively Table 5.2.4.

AIDS: The mean number of dayward visits PPY showed very little change in HIV-1^+ people with AIDS Table 5.2.4.

Table 5.2.4: Mean Number of Outpatient Visits, Inpatient Days and Dayward Attendances by Stage of HIV-1 Infection During Semester

	Aggregate PPY								
	Asymptomatic			Symptomatic Non-AIDS			AIDS		
	OP visits	***IP days***	***DW visits***	***OP visits***	***IP days***	***DW visits***	***OP visits***	***IP days***	***DW visits***
1996	4.45	3.66	0.30	5.38	1.93	0.37	10.57	13.53	1.28
1997	5.28	2.93	0.28	8.19	1.76	0.79	11.77	7.68	1.53
1998	6.40	2.60	0.35	8.51	2.48	0.73	11.63	7.09	1.16
1999	6.12	2.53	0.27	7.51	3.84	0.51	10.28	7.55	0.97
2000	5.99	2.22	0.20	7.60	2.17	0.48	9.10	7.78	0.64
2001	6.41	1.61	0.17	8.47	1.63	0.44	9.16	9.53	0.68
2002	7.95	1.52	0.41	9.24	1.94	0.54	9.90	6.73	0.78
2003	7.18	1.58	0.57	8.72	3.23	0.71	9.12	8.99	1.28
2004	7.11	2.05	0.78	8.53	3.21	0.87	8.61	7.10	1.31
2005	7.08	1.89	0.97	8.98	4.95	0.92	8.99	9.71	1.52
2006	5.83	2.38	1.29	7.62	2.72	0.84	7.00	10.90	2.04

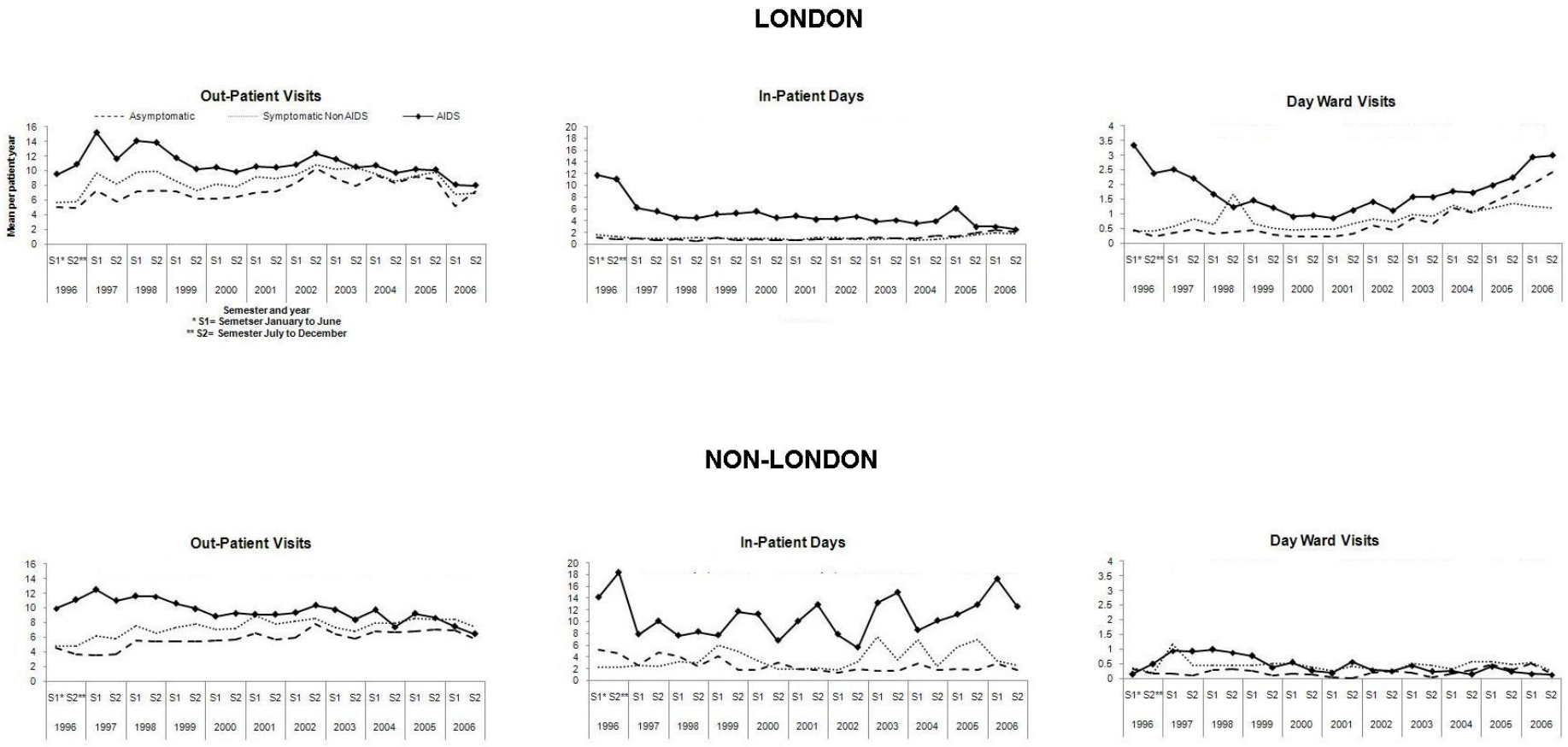


Figure 5.2.2: Use of Hospital Services by Stage of HIV-1 Infection and Location of Clinics

- ***BY SITE OF CLINIC: LONDON***
    - **Outpatient**

Asymptomatic

Similar trends between 1996 and 2002 were seen for asymptomatic PLHV in patients attending the London centres; from 5.12 to 10.32 and a decline was seen by 2006 to 7.18 Figure 5.2.2. Increasing trends in mean outpatient visits PPY between 1996 and 2002 were seen for PLHV in symptomatic non-AIDS stage of HIV-1 infection in patients attending the London centres from 5.68 to 10.85 and a decline was observed thereafter to 6.99 by 2006 Figure 5.2.2.

Symptomatic non-AIDS

Since 1996 the mean inpatient days PPY gradually declined over time since 1996 to 2002 in patients attending London HIV centres in HIV-1^{+} people in symptomatic non-AIDS stages of HIV-1 infection from 1.67 in 1996 to 0.88 and by 2006 this was slightly higher at 1.82 Figure 5.2.2.

AIDS

Similar increasing trends were seen between 1996 and 2002 for PLHV with AIDS from 9.57 to 12.37 and thereafter a decline by 2006 to 8.03 Figure 5.2.2.

- **In-patient**

Asymptomatic

Since 1996 the mean inpatient days PPY gradually declined over time since 1996 to 2006 in patients attending London HIV centres in HIV-1^{+} people in asymptomatic stages of HIV-1 infection from 1.16 in 1996 to 1.04 by 2002 and thereafter a slight increase was observed in 2006 to 2.09 Figure 5.2.2.

Symptomatic non-AIDS

AIDS

Since 1996 the mean inpatient days PPY declined substantially over time to 2006 in patients attending London HIV centres in PLHV with AIDS. The mean number of inpatient days declined from 11.8 in 1996 to 2.6 by 2006 Figure 5.2.2.

- **Dayward**

Asymptomatic

Since 1996 the mean dayward visits PPY showed no change over time between 1996 and 2002 in patients attending the London HIV centres from 0.45 to 0.45 by 2002, however a steady increases were observed thereafter and by 2006 this was estimated to be 2.44.

Symptomatic non-AIDS

The mean dayward visits PPY showed very little change over time since 1996 and 2006 in patients attending the London HIV centres with symptomatic non-AIDS stage of HIV-1 infection from 0.43 in 1996 to 0.74 by 2002 and thereafter an upward trend was observed and by 2006 this was 1.23.

AIDS

For PLHV in AIDS stage of HIV-1 infection the mean dayward visits PPY declined from 3.35 in 1996 to 1.12 by 2002 however increasing trends have been observed thereafter and by 2006 the mean dayward visits PPY were 2.6.

- ***BY SITE OF CLINIC: NON-LONDON***
    - **Outpatient**

Asymptomatic

HIV-1^+ people in asymptomatic stage and who were attending non-London HIV centres, in 1996 and 2002 mean OP visits PPY were 4.57 and 7.83 respectively and declined thereafter to 5.90 in 2006 Figure 5.2.2. On average, in non-London clinics, the mean number of OP visits PPY were 20% lower than those in London clinics.

Symptomatic non-AIDS

Similarly, in patients attending non-London HIV centres between 1996 and 2002 in HIV-1^+ people in symptomatic non-AIDS stage of HIV-1 infection, mean OP PPY increased from 4.89 to 8.58. A decline was observed thereafter and by 2006 the mean OP visits PPY was 7.43 Figure 5.2.2. On average the non-London centres had OP visits PPY that were 15% lower compared to the London centres.

AIDS

Similarly in patients attending non-London HIV centres between 1996 and 2002, the mean OP PPY was 10.81 and 10.01 respectively and a decline was observed by 2006 to 5.72 OP visits PPY Figure 5.2.2. On average the non-London centres had OP visits PPY that were 20% lower compared to the London centres.

- **In-patient**

Asymptomatic

For non-London centres the mean IP days PPY for patients who were classified into asymptomatic stage of HIV-1 infection showed a decline from 5.31 in 1996 to 1.84 by 2006 Figure 5.2.2. On average the non-London centres had mean inpatient days PPY that were two fold higher compared to the London centres.

Symptomatic non-AIDS

The trends over time showed in HIV-1^+ people in symptomatic non-AIDS stage of HIV-1 infection an increase in mean IP days PPY between 1996 and 1999 from 2.35 to 5.01 respectively and an unsteady decline thereafter and by 2006 this was observed to be 2.67 Figure 5.2.2. On average the non-London centres had inpatient days PPY that was over three fold higher compared to the London centres.

AIDS

The trends over time showed in HIV-1$^+$ people in the AIDS the mean inpatient days PPY declined from 14.17 in 1996 to 5.67 by 2002 with an unstable variation between 2003 to 2006 and by the end of the study period this was observed to be 12.59 Figure 5.2.2. On average the non-London centres had IP days PPY that were over 2 fold higher compared to the London centres.

- **Dayward**

Asymptomatic

Very little change was observed in the use of dayward visits PPY in non-London HIV centres in patients classified into asymptomatic between 1996 and 2006 and these were estimated to be 0.35 and 0.17 respectively Figure 5.2.2. On average the mean number of DW visits PPY in non-London clinics were 50% lower than those in London clinics.

Symptomatic non-AIDS

There were very little variations in the use of dayward visits PPY in non-London HIV centres in patients classified into symptomatic non-AIDS stage of HIV-1 infection between 1996 when it was 0.37 and 2006 when it was 0.28 Figure 5.2.2. On average the DW visits PPY were 40% lower in non-London centres compared to the London centres.

AIDS

The use of dayward visits PPY remained unchanged over time from 1996 when this was 0.17 and by 2006 the mean dayward visits PPY was 0.14 Figure 5.2.2. Overall the use of dayward services was observed to have been slightly higher for HIV-1$^+$ people with AIDS compared to those who were classed as asymptomatic or symptomatic non-AIDS stages of HIV-1 infection. On average the non-London centres had DW visits PPY that were 70% lower compared to the London centres.

5.2.2 COST OF HOSPITAL SERVICES BY STAGE OF HIV-1 INFECTION

- ***AGGREGATE***

The mean total hospital services costs PPY showed a steady rise over time for all three stages of HIV-1 infection.

Asymptomatic For HIV-1$^+$ people with asymptomatic stage of HIV-1, the mean total cost of hospital service was estimated at £58K PPY in 1996 and by 2006 it had increased to £72K PPY Table 5.2.5.

Symptomatic non-AIDS For HIV-1$^+$ people who were classified as symptomatic non-AIDS stage of HIV-1, this was £71K PPY in 1996 and £89K PPY by 2005 however a slight decline in cost was observed by 2006 at £84K PPY.

AIDS For HIV-1$^+$ people who were classified in AIDS stage of HIV-1 infection, the total direct costs of hospital service PPY remained relatively stable over time. In 1996 it was estimated to be £147K PPY and by 2005 a slight decline was seen £143K PPY followed by an increase of £147K PPY in 2006 Table 5.2.5.

Table 5.2.5: Total Cost of HIV Service Use in HIV-1^{+} People by Time and Stage of HIV-1 Infection

	Total Cost PPY					
	Excluding Community Care Cost			Including Community Care Cost		
Year	Asymptomatic	Symptomatic non-AIDS	AIDS	Asymptomatic	Symptomatic non-AIDS	AIDS
1996	£58,097	£70,827	£147,143	£89,007	£111,667	£199,419
1997	£59,673	£77,827	£133,573	£90,583	£118,667	£185,849
1998	£61,512	£83,361	£133,017	£92,422	£124,201	£185,293
1999	£61,619	£87,097	£137,666	£92,529	£127,937	£189,942
2000	£58,318	£75,571	£129,228	£89,228	£116,411	£181,504
2001	£59,560	£76,980	£139,464	£90,470	£117,820	£191,740
2002	£61,851	£80,945	£130,885	£92,761	£121,785	£183,161
2003	£62,337	£86,506	£143,635	£93,247	£127,346	£195,911
2004	£65,902	£84,768	£134,780	£96,812	£125,608	£187,056
2005	£65,640	£89,210	£142,820	£96,550	£130,050	£195,096
2006	£71,943	£83,665	£147,129	£102,853	£124,505	£199,405

Costs include total use of services costs (OP,IP,DW) + Other drugs + Tests and procedure+ ART (Non London ART prices were estimated at 9.5% higher than London ART prices)

- ***BY SITE OF CLINIC: LONDON AND NON-LONDON***

The overall trends of the mean total cost of hospital services PPY stratified by stage of HIV-1 infection and clinic location showed shifts in changes in costs over time.

Asymptomatic Compared to London HIV specialist centres, the HIV-1$^+$ people treated in the non-London centres with asymptomatic stage of HIV-1 infection, the total use of hospital service costs were 45% higher in 1996. However, over time the differences in hospital costs between London and non-London centres almost diminished and by 2006 the total hospital services costs for non-London centres were 5% higher compared to London centres Table 5.2.6.

Symptomatic non-AIDS For HIV-1$^+$ people with symptomatic non-AIDS stage of HIV-1 infection, in 1996 the non-London total hospital services costs PPY were 9% higher which rose to 45% higher by 1999 however decline in the differences in costs between clinic locations were observed and by 2006 the total hospital services costs PPY for non-London centres were 10% higher compared to London centres by 2006.

AIDS Variations in costs between the London and non-London centres were also found in patients with AIDS. In 1996 these costs were 10% higher in Non-London Centres and these differences in cost estimates continued to expand over time and by 2006 the total hospital service costs PPY in non-London centres were estimated to be 29% higher compared to the total cost of hospital services in the London centres Table 5.2.6.

5.2.3 COST OF HOSPITAL SERVICES BY STAGE OF HIV-1 INFECTION INCLUDING COMMUNITY CARE COST

The overall mean total cost of hospital services PPY by stage of HIV-1 infection and clinic location showed similar trends when community care costs were added to the total hospital services costs. The variations that were observed were the differences in costs between the London and the non-London Centres.

Asymptomatic After community care costs were added, in HIV-1$^+$ people in asymptomatic stage of HIV-1 infection, the mean use of hospital services costs amongst those attending non London centres these were 27% higher in 1996 compared to the London centres whilst the differences in costs by clinic location had dropped to 4% by 2006 Table 5.2.6.

Symptomatic non-AIDS Whilst for HIV-1$^+$ people in symptomatic stage of HIV-1 infection the non-London clinics' costs were 5% higher than the London clinic costs in 1996 and remained similar by 2006 at 7% higher costs for Non-London centres compared to the London centres.

AIDS Similarly, for HIV-1$^+$ people with AIDS the non London centre total hospital services costs were 7% higher in 1996 however these differences increased over time and by 2006 total hospital services costs including the community care costs in non London centres were 21% higher than the London Centres Table 5.2.6.

Table 5.2.6: Total Cost of HIV-1 Service Use in HIV-1^{+} People by Year of Attendance, Stage of HIV-1 Infection and Clinic Location

	Total Cost											
	Excluding Community Care Cost						Including Community Care Cost					
	London			Non-London			London			Non-London		
Year	Asymptomatic	Symptomatic non-AIDS	AIDS	Asymptomatic	Symptomatic non-AIDS	AIDS	Asymptomatic	Symptomatic non-AIDS	AIDS	Asymptomatic	Symptomatic non-AIDS	AIDS
1996	£23,727	£33,937	£70,188	£34,370	£36,890	£76,955	£39,182	£54,357	£96,326	£49,825	£57,310	£103,093
1997	£26,482	£36,489	£64,865	£33,192	£41,339	£68,708	£41,937	£56,909	£91,003	£48,647	£61,759	£94,846
1998	£27,090	£38,764	£63,746	£34,421	£44,598	£69,271	£42,545	£59,184	£89,884	£49,876	£65,018	£95,409
1999	£27,710	£36,197	£62,737	£33,909	£50,900	£74,929	£43,165	£56,617	£88,875	£49,364	£71,320	£101,067
2000	£26,223	£35,188	£61,031	£32,095	£40,383	£68,197	£41,678	£55,608	£87,169	£47,550	£60,803	£94,335
2001	£27,867	£36,504	£61,400	£31,693	£40,476	£78,065	£43,322	£56,924	£87,538	£47,148	£60,896	£104,203
2002	£29,774	£38,857	£62,663	£32,077	£42,087	£68,222	£45,229	£59,277	£88,801	£47,532	£62,507	£94,360
2003	£30,387	£38,825	£62,696	£31,950	£47,682	£80,939	£45,842	£59,245	£88,834	£47,405	£68,102	£107,077
2004	£31,246	£37,341	£61,800	£34,656	£47,427	£72,981	£46,701	£57,761	£87,938	£50,111	£67,847	£99,119
2005	£31,244	£38,477	£65,093	£34,395	£50,733	£77,727	£46,699	£58,897	£91,231	£49,850	£71,153	£103,865
2006	£35,023	£39,844	£64,225	£36,920	£43,821	£82,904	£50,478	£60,264	£90,363	£52,375	£64,241	£109,042

Costs include total use of services costs (OP,IP,DW) + Other drugs + Tests and procedure+ ART (Non London ART prices were estimated at 9.5% higher than London ART prices)

5.2.4 COST OF DIFFERENT TYPES OF ANTI-RETROVIRAL THERAPY BY STAGE OF HIV-1 INFECTION

Although slight variations were observed between ARV costs in London and non London clinics (Table 5.2.7) the proportion spent on ART between London and non London clinics for any of three stages of HIV-1 infection nor for the four ART treatment categories were similar (Table 5.2.7).

5.2.5 COST OF HOSPITAL SERVICES BY STAGE OF HIV-1 INFECTION AND TYPES OF ARV THERAPY

Treatment and care costs PPY for PLHIV at different stage of HIV infection and different treatment regimens increased over time (Table 5.2.8) with similar trends being observed in London sites.

- ***MONO THERAPY***

For asymptomatic individuals in 1996 the annual cost of care was £8,503, which increased to £10,020 in 2006. For those with symptomatic non-AIDS, the 1996 cost of £10,591 increased to £11,482, while for AIDS patients on mono-therapy, the annual costs decreased from £36,396 in 1996 to £30,717 in 1997 which then increased to £32,760 by 2006 (Table 5.2.8).

- ***DUAL THERAPY***

For asymptomatic PLHIV on dual therapy, the 1996 cost of £11,525 increased to £13,351 in 2006. For patients with symptomatic non-AIDS, the annual cost increased from £13,671 in 1996 to £15,692 in 2006, while for AIDS patients on dual therapy, a decreased from £39,409 in 1996 to £33,726 in 1997 was observed which increased to £36,580 in 2006 per patient-years in 2006 (Table 5.2.8).

- ***TRIPLETHERAPY***

For asymptomatic PLHIV on triple therapy, annual costs increased from £16,346 in 1996 to £18,280 in 2006. For patients with symptomatic non-AIDS annual costs between 1996 and 2006 increased from £18,374 to £21,597, while for AIDS patients annual costs varied from £44,350 in 1996 to £41,747 in 2006 (Table 5.2.8).

- ***QUADRUPLE OR MORE THERAPY***

Annual cost for those asymptomatic PLHIV on quadruple-or-more therapy ranged from £16,722 in 1996 to £23,775 in 2006. Similarly annual costs increased between 1996 and 2006 from £21,139 to £25,135 respectively for PLHIV with symptomatic non-AIDS, while for AIDS patients annual costs first decreased from £46,041 in 1996 to £41,383 in 1997 and increased to £48,055 in 2006(Table 5.2.8).

Table 5.2.7: Annual Cost Due to Different Types of Anti-Retroviral Therapy by Stage of HIV Infection, Aggregate and London Site, 2006 UK Prices

ARV costs PPY												
	Non London and London cost											
	Mono therapy			Dual therapy			Triple therapy			Quadruple or more therapy		
Year	ASx	Sx Non AIDS	AIDS	ASx	Sx Non AIDS	AIDS	ASx	Sx Non AIDS	AIDS	ASx	Sx Non AIDS	AIDS
1996	£3,504	£3,539	£3,735	£6,525	£6,619	£6,748	£11,347	£11,322	£11,689	£11,722	£14,087	£13,380
1997	£3,839	£4,337	£3,976	£6,906	£6,946	£6,985	£11,877	£12,067	£12,146	£14,369	£14,985	£14,642
1998	£4,163	£4,476	£4,339	£7,181	£7,375	£7,196	£12,062	£12,474	£12,355	£15,049	£15,438	£15,350
1999	£4,243	£3,968	£4,212	£7,309	£7,040	£7,323	£12,196	£12,466	£12,263	£15,585	£15,626	£15,786
2000	£4,144	£3,761	£4,085	£7,043	£6,814	£7,234	£12,003	£12,144	£12,052	£14,796	£14,937	£15,190
2001	£4,091	£3,975	£4,107	£7,138	£6,796	£7,372	£12,060	£12,008	£11,860	£16,846	£16,195	£16,789
2002	£4,018	£4,173	£4,077	£7,230	£6,586	£7,182	£12,130	£12,226	£11,996	£17,322	£18,686	£17,253
2003	£4,047	£4,540	£4,158	£7,189	£6,715	£7,385	£12,292	£12,579	£12,152	£17,617	£17,194	£17,763
2004	£4,233	£3,452	£4,209	£7,236	£6,663	£7,427	£12,713	£12,801	£12,532	£18,286	£17,839	£18,656
2005	£4,357	£2,116	£4,139	£7,536	£6,971	£7,437	£12,601	£12,636	£12,618	£18,651	£18,526	£19,234
2006	£4,346	£2,592	£3,660	£7,677	£6,803	£7,479	£12,607	£12,707	£12,647	£18,101	£18,061	£18,955
	London cost											
1996	£1,668	£1,685	£1,779	£3,107	£3,152	£3,213	£5,403	£5,391	£5,566	£5,582	£6,708	£6,371
1997	£1,828	£2,065	£1,894	£3,289	£3,308	£3,326	£5,656	£5,746	£5,784	£6,842	£7,136	£6,972
1998	£1,982	£2,132	£2,066	£3,419	£3,512	£3,427	£5,744	£5,940	£5,884	£7,166	£7,351	£7,310
1999	£2,020	£1,889	£2,006	£3,481	£3,352	£3,487	£5,808	£5,936	£5,840	£7,421	£7,441	£7,517
2000	£1,974	£1,791	£1,945	£3,354	£3,245	£3,445	£5,715	£5,783	£5,739	£7,046	£7,113	£7,233
2001	£1,948	£1,893	£1,956	£3,399	£3,236	£3,510	£5,743	£5,718	£5,647	£8,022	£7,712	£7,995
2002	£1,913	£1,987	£1,941	£3,443	£3,136	£3,420	£5,776	£5,822	£5,712	£8,249	£8,898	£8,216
2003	£1,927	£2,162	£1,980	£3,424	£3,198	£3,517	£5,853	£5,990	£5,787	£8,389	£8,188	£8,458
2004	£2,016	£1,644	£2,004	£3,446	£3,173	£3,537	£6,054	£6,096	£5,967	£8,707	£8,495	£8,884
2005	£2,075	£1,007	£1,971	£3,589	£3,320	£3,541	£6,001	£6,017	£6,009	£8,881	£8,822	£9,159
2006	£2,070	£1,235	£1,743	£3,656	£3,239	£3,562	£6,003	£6,051	£6,022	£8,620	£8,601	£9,026

POPULATION COST

	Non London cost											
1996	£1,835	£1,854	£1,957	£3,418	£3,467	£3,535	£5,943	£5,931	£6,123	£6,140	£7,379	£7,008
1997	£2,011	£2,272	£2,083	£3,618	£3,638	£3,659	£6,221	£6,321	£6,362	£7,527	£7,849	£7,669
1998	£2,180	£2,345	£2,273	£3,761	£3,863	£3,769	£6,318	£6,534	£6,472	£7,883	£8,086	£8,041
1999	£2,222	£2,078	£2,206	£3,829	£3,688	£3,836	£6,389	£6,530	£6,424	£8,163	£8,185	£8,269
2000	£2,171	£1,970	£2,140	£3,689	£3,569	£3,789	£6,287	£6,361	£6,313	£7,750	£7,824	£7,957
2001	£2,143	£2,082	£2,152	£3,739	£3,560	£3,861	£6,317	£6,290	£6,212	£8,824	£8,483	£8,794
2002	£2,105	£2,186	£2,135	£3,787	£3,450	£3,762	£6,354	£6,404	£6,284	£9,073	£9,788	£9,037
2003	£2,120	£2,378	£2,178	£3,766	£3,518	£3,868	£6,439	£6,589	£6,365	£9,228	£9,006	£9,304
2004	£2,217	£1,808	£2,205	£3,790	£3,490	£3,890	£6,659	£6,705	£6,564	£9,578	£9,344	£9,772
2005	£2,282	£1,108	£2,168	£3,947	£3,651	£3,896	£6,601	£6,619	£6,610	£9,770	£9,704	£10,075
2006	£2,277	£1,358	£1,917	£4,021	£3,563	£3,918	£6,603	£6,656	£6,625	£9,482	£9,461	£9,929

Table 5.2.8: Annual Cost of Treatment and Care by Stage of HIV Infection and Different Types of Anti-Retroviral Therapy, 2006 UK prices

Total Cost PPY: OP + IP + DW + other drugs + tests & procedures + ARV costs												
	Non London and London											
	Mono therapy			Dual therapy			Triple therapy			Quadruple or more therapy		
Year	ASx	Sx Non AIDS	AIDS	ASx	Sx Non AIDS	AIDS	ASx	Sx Non AIDS	AIDS	ASx	Sx Non AIDS	AIDS
1996	£8,503	£10,591	£36,396	£11,525	£13,671	£39,409	£16,346	£18,374	£44,350	£16,722	£21,139	£46,041
1997	£8,308	£12,076	£30,717	£11,376	£14,685	£33,726	£16,347	£19,806	£38,887	£18,839	£22,724	£41,383
1998	£8,642	£13,008	£29,609	£11,660	£15,907	£32,467	£16,542	£21,006	£37,626	£19,528	£23,970	£40,621
1999	£8,552	£13,481	£30,431	£11,619	£16,553	£33,542	£16,506	£21,980	£38,483	£19,894	£25,140	£42,005
2000	£8,083	£11,299	£29,274	£10,981	£14,352	£32,423	£15,941	£19,683	£37,242	£18,734	£22,476	£40,380
2001	£7,859	£11,489	£32,229	£10,906	£14,310	£35,493	£15,827	£19,521	£39,981	£20,613	£23,709	£44,911
2002	£8,145	£11,901	£29,059	£11,358	£14,314	£32,164	£16,257	£19,954	£36,978	£21,449	£26,414	£42,235
2003	£8,176	£13,435	£32,227	£11,319	£15,611	£35,454	£16,422	£21,474	£40,221	£21,747	£26,089	£45,831
2004	£8,774	£12,272	£28,988	£11,777	£15,483	£32,206	£17,254	£21,621	£37,311	£22,827	£26,659	£43,436
2005	£8,685	£12,192	£30,566	£11,864	£17,048	£33,864	£16,930	£22,713	£39,045	£22,979	£28,603	£45,661
2006	£10,020	£11,482	£32,760	£13,351	£15,692	£36,580	£18,280	£21,597	£41,747	£23,775	£25,135	£48,055
	London											
1996	£3,262	£5,085	£17,344	£4,700	£6,552	£18,778	£6,996	£8,792	£21,131	£7,175	£10,108	£21,936
1997	£3,569	£5,636	£15,123	£5,030	£6,878	£16,556	£7,397	£9,317	£19,013	£8,584	£10,707	£20,202
1998	£3,675	£6,008	£14,359	£5,112	£7,388	£15,719	£7,437	£9,816	£18,176	£8,859	£11,228	£19,602
1999	£3,746	£5,364	£14,255	£5,206	£6,827	£15,737	£7,533	£9,411	£18,089	£9,147	£10,916	£19,766
2000	£3,539	£5,221	£13,878	£4,920	£6,675	£15,377	£7,281	£9,213	£17,672	£8,612	£10,543	£19,166
2001	£3,643	£5,440	£13,697	£5,094	£6,784	£15,251	£7,438	£9,265	£17,388	£9,717	£11,259	£19,736
2002	£3,943	£5,729	£13,908	£5,473	£6,878	£15,387	£7,806	£9,564	£17,679	£10,278	£12,640	£20,182
2003	£4,034	£5,924	£13,761	£5,531	£6,960	£15,298	£7,960	£9,752	£17,568	£10,496	£11,950	£20,240
2004	£4,151	£5,239	£13,413	£5,581	£6,768	£14,946	£8,189	£9,691	£17,377	£10,843	£12,090	£20,293
2005	£4,133	£5,005	£14,117	£5,647	£7,317	£15,687	£8,059	£10,015	£18,155	£10,940	£12,820	£21,305
2006	£4,924	£5,468	£13,654	£6,511	£7,473	£15,473	£8,858	£10,285	£17,934	£11,474	£12,834	£20,938

5.2.6 NUMBER OF HIV-1$^+$ PEOPLE USING THE NHS SERVICES

- ***DURING STUDY PERIOD: 1997-2006***

The total number of HIV-1$^+$ population who were reported to the HPA to have used the NHS services in the UK for treatment and care increased during the study period. In 1997, 16,075 HIV-1$^+$ people were reported to have accessed the NHS services, which increased to 52,083 patients by 2006. These increases in the numbers of HIV-1$^+$ patients accessing the NHS for care were seen across all stages of HIV-1 infection in the UK population figures, however compared to 1997 in 2006 the proportion of HIV-1$^+$ people classified as asymptomatic increased from 38.1% to 48.7%, while the proportion of HIV-1$^+$ people classified as symptomatic non-AIDS patients decreased from 33.5% to 28.3% similarly the number of AIDS patients declined from 28.4% in 1997 to 22.9% in 2006 (Table 5.2.9).

Table 5.2.9: Number of People Living With HIV-1 in the UK Using NHS Services by Stage of HIV-1 Infection and Year 1997-2006

	UK HIV-1 Population (%)			
Year	**Asymptomatic**	**Symptomatic non-AIDS**	**AIDS**	**Total**
1997	6124(38.1)	5384(33.5)	4567(28.4)	**16,075**
1998	5835(31.8)	6975(38.0)	5525(30.1)	**18,335**
1999	6833(32.4)	7994(37.9)	6288(29.8)	**21,114**
2000	8055(35.4)	8347(36.7)	6338(27.9)	**22,740**
2001	10015(38.0)	9088(34.5)	7253(27.5)	**26,356**
2002	12787(40.5)	10635(33.7)	8115(25.7)	**31,537**
2003	15932(43.4)	11544(31.5)	9203(25.1)	**36,679**
2004	18794(45.1)	12834(30.8)	10009(24.0)	**41,637**
2005	21656(46.1)	14590(31.0)	10779(22.9)	**47,025**
2006	25385(48.7)	14750(28.3)	11947(22.9)	**52,083**

Source of data from HPA
1996 UK HIV-1 population figures unavailable from HPA (from 1st April HPA became part of NHS England)
Table from (Mandalia *et al.*, 2010)

- ***DURING PROJECTED PERIOD: 2007-2013***

In 2006 the total number of HIV-1 infected people who accessed the NHS for care was 52,083. The extrapolation showed that by 2013, the expected total numbers of HIV-1 positive subjects who are likely to access the NHS for HIV-1 care is estimated to be 78,370. An estimated 38,951 are expected to be living with asymptomatic stage of HIV-1 infection in the UK by 2013 whilst 22,267 are projected to be living with symptomatic non-AIDS stage HIV-1 infection and for AIDS patients its is estimated that by end of 2013, numbers accessing the NHS HIV services who are classified into AIDS stage of HIV-1 infection is expected to be 17,151 (Table 5.2.10, Figure 5.2.3).

Table 5.2.10: Projected and HPA Estimates of People Living with HIV-1 Using NHS Services by Stage of HIV-1 Infection and Year 2007-2013

Year	Number of people living with HIV-1 in the UK using the NHS services				
	Projected numbers n(%)				HPA estimates n
	Asymptomatic	Symptomatic non-AIDS	AIDS	Total	Total
2007	25,485 (47%)	15,979 (30%)	12,378 (23%)	53,842	56,211
2008	27,729 (48%)	17,027 (29%)	13,173 (23%)	57,930	61,019
2009	29,974 (48%)	18,075 (29%)	13,969 (23%)	62,018	65,213
2010	32,218 (49%)	19,123 (29%)	14,764 (22%)	66,106	69,298
2011	34,462 (49%)	20,171 (29%)	15,560 (22%)	70,194	73,645
2012	36,707 (49%)	21,219 (29%)	16,355 (22%)	74,282	76,500
2013	38,951 (50%)	22,267 (28%)	17,151 (22%)	78,370	N/A

Table adapted from (Mandalia *et al.*, 2010)
HPA estimates of people living with HIV-1 in the UK using the NHS services reported to the Public Health England (Aghaizu *et al.*, 2013) (from 1st April HPA became part of NHS England)
N/A – Currently (November 2013) figure for this year was not available

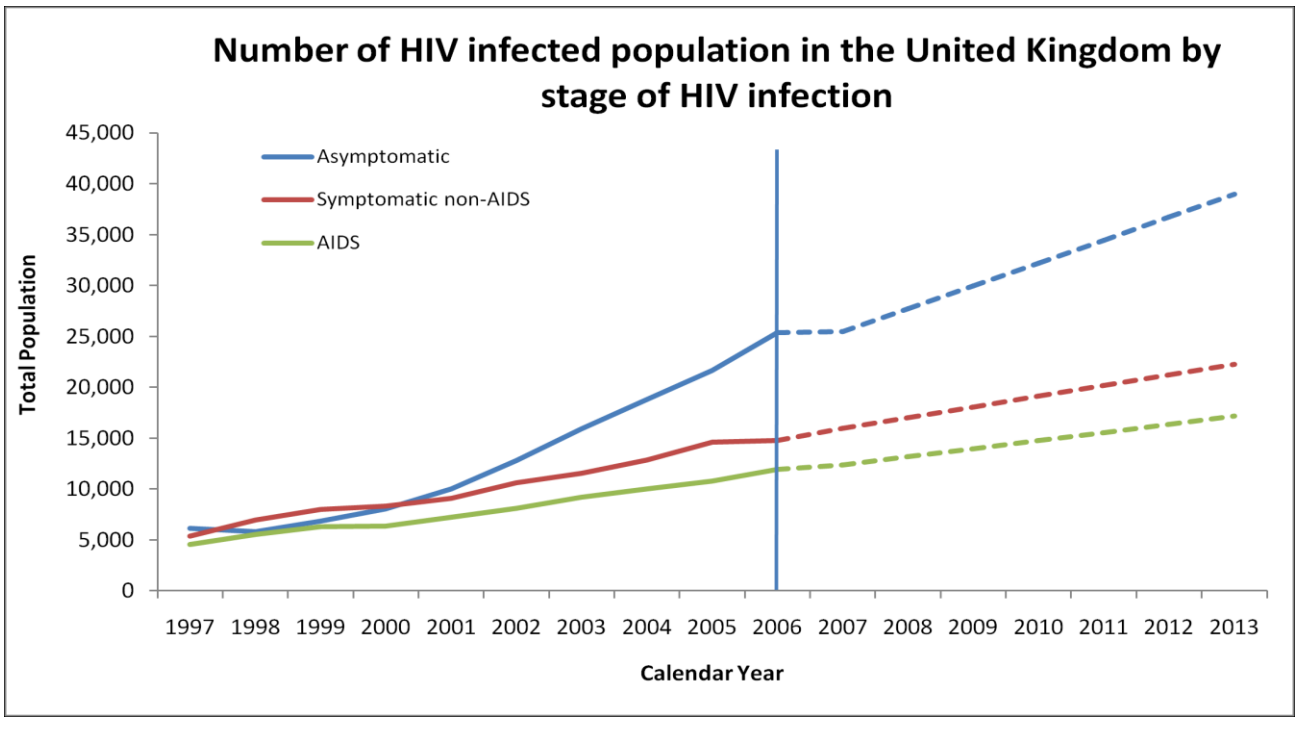


Figure 5.2.3: The Number HIV-1 Infected Population in the UK by Stage of HIV-1 Infection, Projected From 2007-2013

The accuracy of the estimated projected figures were compared to the HPA estimates published by the Public Health England (from 1st April HPA became part of NHS England) and these estimated projected population to access the NHS for HIV services by end of 2012 was 74,282 whist the observed numbers were 76,500 (see Chapter 1 for further details) thereby 2012 projected

population presented in Table 5.2.10 were underestimated by 3% impacting the population cost estimate by this magnitude.

5.2.7 ANNUAL POPULATION COST OF HIV-1^{+} PEOPLE

- ***DURING STUDY PERIOD: 1997-2006***

The direct population cost increased over time for all stages of HIV-1 infection since 1997 to 2006 reflecting in increasing numbers of subjects in the UK who are living with HIV-1 infection and accessing the NHS hospital services.

In 1997 the direct population cost of treating HIV-1^{+} in patients classified in asymptomatic stage of HIV-1 infection was estimated at £22 million whilst for the same year this was estimated to be £39 million for patients in symptomatic non-AIDS stage of HIV-1 infection and was £43 million for AIDS patients. In contrast these had increased by over 4.5 fold collectively to £483 million by 2006 and patients in asymptomatic, symptomatic non-AIDS and AIDS stage of HIV-1 infection the total population costs were £162, £146 and £176 million Figure 5.2.4.

As expected, when community care costs were accounted for, the total cost increased from £164 million in 1997 to £683 million in 2006 which was estimated to be 41% higher than £483 million, estimated when community care costs were not accounted for.

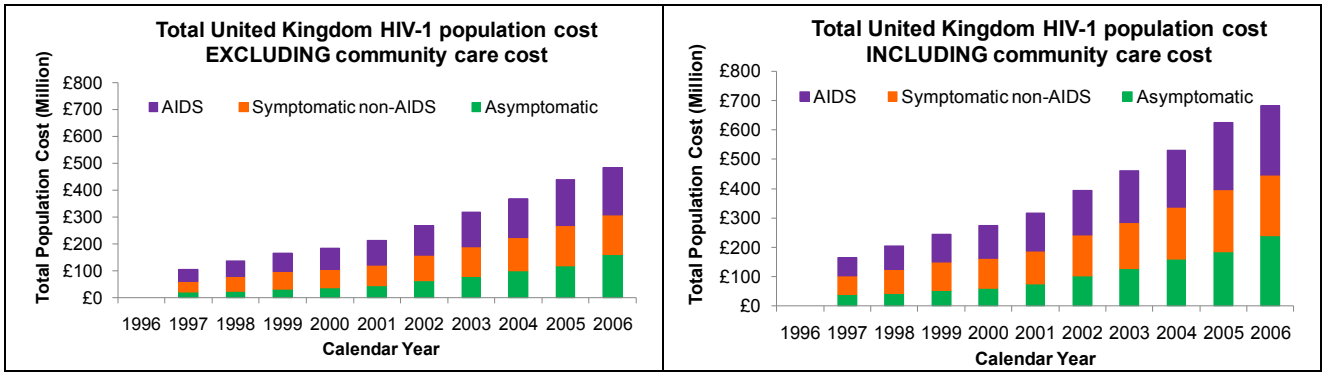


Figure 5.2.4: Total UK Population Cost by Stage of HIV-1 Infection Excluding and Including Community Care Cost

- ***DURING PROJECTED PERIOD: 2007-2013***

Three methods were investigated and the findings from these were:

- **Method 1**

The *first* method projected the annual UK population costs per year from 1997-2006 and estimated population cost from 2007-2013 (including and excluding community care) (regardless of stage of HIV-1 infection) using the population cost figures derived for 1997-2006 using regression models. In Table 5.2.11, Figure 5.2.5 the original costs values were divided by 1,000,000.

Method 1: Regression Models Used to Project Direct Total UK Population Cost

	UK direct population cost regression models (TC) Per £1,000,000 $y=\alpha+\beta x$	
	Regression model Excluding Community Care	Regression model Including Community Care
	$y=\alpha+\beta x$	$y=\alpha+\beta x$
Total	TC =-4839+(42.52*year)	TC=-116919+(58.61*year)

Where y=UK direct population cost per £1,000,000 (TC)
α=intercept
β=slope or gradient of a line
x=year

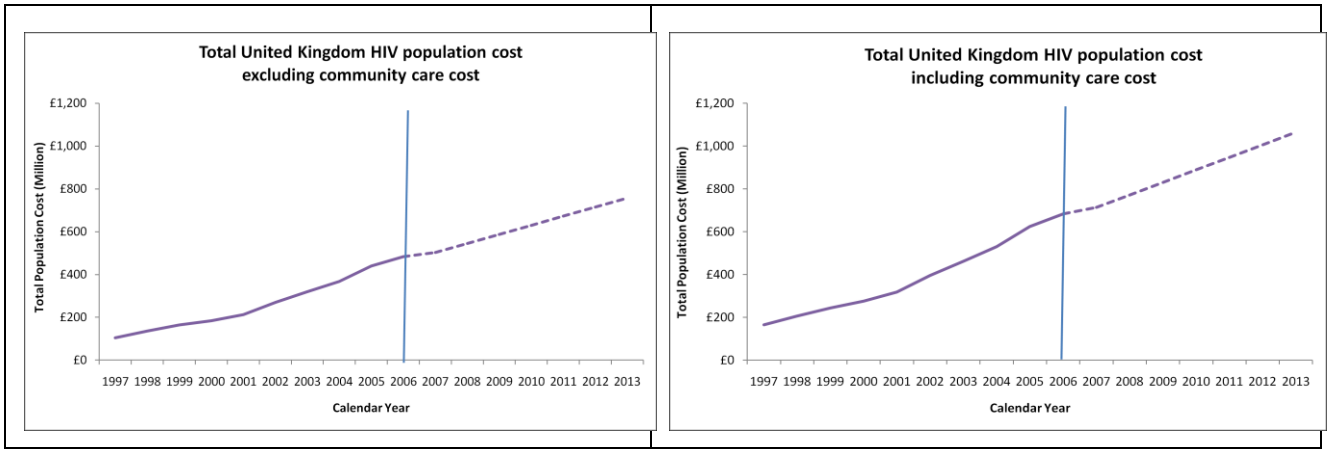


Figure 5.2.5: Method 1 - Direct UK HIV-1 Population Cost Excluding and Including Community Care Cost From 2007-2013

Table 5.2.11: Method 1 - Total UK Population Cost From NPMS-HHC and Projected

Calendar year	Excluding community care costs combined for the three HIV-1 stages *(asymptomatic, symptomatic non-AIDS and AIDS)*	Including community care costs combined for the three HIV-1 stages *(asymptomatic, symptomatic non-AIDS and AIDS)*
1996	-	-
1997	£104	£164
1998	£136	£206
1999	£165	£244
2000	£184	£276
2001	£213	£318
2002	£269	£395
2003	£319	£462
2004	£368	£531
2005	£440	£625
2006	£483	£683
2007	**£502**	**£713**
2008	**£545**	**£772**
2009	**£588**	**£830**
2010	**£630**	**£889**
2011	**£673**	**£947**
2012	**£715**	**£1,006**
2013	**£758**	**£1,065**

Figures in blue (from 2007-2013) are projected

The figures in blue are projected cost figures and the figures in black are the original cost figures and both the figures in black and blue are costs per £1,000,000. The annual direct population costs were estimated to increase to £758 million by 2013 if community care cost were not accounted for while the cost were estimated to rise to £1,065 million by 2013 if community care costs were included (Table 5.2.11, Figure 5.2.5).

- **Method 2**

The *second* method projected the population total of number of HIV-1$^+$ people by 2013 (Table 5.2.10) using the regression models.

Method 2: Regression Models Used to Project Direct Total UK Population Cost

	UK direct population cost regression models (TC) $y=\alpha+\beta x$	
	Excluding community care costs	**Including community care costs**
Total Population cost from 2006	£483,362,397	£682,824,388
Total UK population based on HPA 2006 figures	52,083	52,083
Average UK cost per patient year for 2006 (Mean)	£483,362,397 ÷ 52,083 = £9,281	£682,824,388 ÷ 52,083 = £13,110
95% confidence Interval	mean +/-1.96* 238.95	mean +/-1.96* 238.95

Where y=UK direct population cost
α=intercept
β=slope or gradient of a line
x=year

Multiplied the average 2006 PPY costs to the projected number of HIV-1$^+$ people using the NHS services by 2013, the annual population cost was estimated to increase to £727 million by 2013 when community care cost were not accounted for and £1,027 million when including community care costs (Table 5.2.12).

Table 5.2.12: Method 2 - Total UK Population Cost From NPMS-HHC and Projected

Calendar year	Excluding community care costs combined for the three HIV-1 stages *(asymptomatic, symptomatic non-AIDS and AIDS)*	Including community care costs combined for the three HIV-1 stages *(asymptomatic, symptomatic non-AIDS and AIDS)*
1996	-	-
1997	£104	£164
1998	£136	£206
1999	£165	£244
2000	£184	£276
2001	£213	£318
2002	£269	£395
2003	£319	£462
2004	£368	£531
2005	£440	£625
2006	£483	£683
2007	**£500**	**£706**
2008	**£538**	**£759**
2009	**£576**	**£813**
2010	**£614**	**£867**
2011	**£651**	**£920**
2012	**£689**	**£974**
2013	**£727**	**£1,027**

Figures in blue (from 2007-2013) are projected

- **Method 3**

The *third* method looked at the number of HIV-1^{+} people by stage of HIV-1 infection using the regression models.

Method 3: Regression Models Used to Project Direct Total UK Population Cost

	UK direct population cost regression models (TC) $y=\alpha+\beta x$					
	Excluding community care costs			Including community care costs		
	ASx	**SNA**	**AIDS**	**ASx**	**SNA**	**AIDS**
Population cost from 2006	£161,558,459	£145,622,691	£176,181,247	£240,024,772	£205,862,874	£236,936,742
UK population based on HPA 2006 figures	25,385	14,750	11,947	25,385	14,750	11,947
Average UK cost per patient year for 2006 (Mean)	£161,558,459 ÷ 25,385 = **£6,364**	£145,622,691 ÷ 14,750 = **£9,873**	£176,181,247 ÷ 11,947 = **£14,747**	£240,024,772 ÷ 25,385 = **9,455**	£205,862,874 ÷ 14,750 = **13,957**	£236,936,742 ÷ 11,947 = **19,832**
Confidence Interval calculated for each	Mean +/- 1.96*193.88	Mean +/- 1.96*46.21	Mean +/- 1.96*33.73	Mean +/- 1.96*22.94	Mean +/- 1.96*46.21	Mean +/- 1.96*33.73

Formula for projecting ASx (asymptomatic), SNA (symptomatic non-AIDS) and AIDS and total population in the UK:
$y= \alpha +\beta x$
Where α =intercept
β =slope or gradient of a line
x=year

Formula for calculating the lower bound of the confidence interval for ASx, SNA and AIDS and total population in the UK: *$y= x-1.96*\beta$, Where β =slope or gradient of a line, x=HIV population value for that year for that stage of HIV infection*

Formula for calculating the upper bound of the confidence interval for ASx, SNA and AIDS and total population in the UK: *$y= x+1.96*\beta$ Where β =slope or gradient of a line, x= HIV population value for that year for that stage of HIV infection*

These were projected and multiplied by the average annual 2006 cost for each stage of HIV-1 estimated and respectively for people with asymptomatic, symptomatic non-AIDS and AIDS were: £248; £220; and £253 million which collectively gave a total population cost of £721 million for 2013 when community care cost were not included. When community care costs were accounted for, the costs for asymptomatic, symptomatic non-AIDS and AIDS were: £368, £311, and £340 million or a total of £1,019 million (Table 5.2.13).

Table 5.2.13: Method 3 - Total UK Population Cost From NPMS-HHC and Projected

Calendar year	Excluding community care costs				Including community care costs			
	ASx	Sx Non-AIDS	AIDS	Combined for the three HIV-1 stages *(asymptomatic, symptomatic non-AIDS and AIDS)*	ASx	Sx Non-AIDS	AIDS	Combined for the three HIV-1 stages *(asymptomatic, symptomatic non-AIDS and AIDS)*
1997	£22	£39	£43	£104	£41	£61	£63	£165
1998	£24	£56	£56	£136	£42	£83	£80	£205
1999	£32	£67	£67	£166	£53	£98	£93	£244
2000	£37	£68	£79	£184	£62	£102	£112	£276
2001	£45	£75	£93	£213	£76	£113	£129	£318
2002	£63	£95	£111	£269	£103	£139	£154	£396
2003	£79	£110	£130	£319	£128	£157	£177	£462
2004	£101	£124	£143	£368	£159	£177	£195	£531
2005	£118	£151	£171	£440	£185	£210	£229	£624
2006	£162	£146	£176	£484	£240	£206	£237	£683
2007	**£162**	**£158**	**£183**	**£503**	**£241**	**£223**	**£245**	**£709**
2008	**£176**	**£168**	**£194**	**£538**	**£262**	**£238**	**£261**	**£761**
2009	**£191**	**£178**	**£206**	**£575**	**£283**	**£252**	**£277**	**£812**
2010	**£205**	**£189**	**£218**	**£612**	**£305**	**£267**	**£293**	**£865**
2011	**£219**	**£199**	**£229**	**£647**	**£326**	**£282**	**£309**	**£917**
2012	**£234**	**£209**	**£241**	**£684**	**£347**	**£296**	**£324**	**£967**
2013	**£248**	**£220**	**£253**	**£721**	**£368**	**£311**	**£340**	**£1,019**

Figures in blue (from 2007-2013) are projected

5.3 DISCUSSION

- ***NOVEL APPROACH***

This study estimated that the direct population cost for treatment and care of HIV-1$^+$ people in the UK rose by 4.6 fold between 1997 and 2006, from £104 million to £483 million respectively. The number of HIV-1$^+$ people using NHS services during this period tripled from 16,075 to 52,083 HIV-1$^+$ people: the greatest increase was seen among asymptomatic HIV-1$^+$ people and less so among HIV-1$^+$ people with symptomatic non-AIDS or AIDS.

When the annual population costs were projected using three different methods, the estimated population costs increased to between £721 million and £758 million by 2013, a 1.5 fold increase from the 2006 baseline. The three methods produced similar estimates, with only a 5% difference between the lowest and highest estimates. The accuracy of these estimated projected figures of HIV-1$^+$ people using NHS services were compared with the actual figures published by HPA for 2007. In 2007 the projected number of HIV-1$^+$ people was 53,842 compared with 56,211 according to the HPA (Aghaizu *et al.*, 2013), an under-estimate of 4%; these were up to maximum of 5% underestimated for each of the calendar years 2008, 2009, 2010, 2011 and 2012 reported by HPA when compared to the projected estimates for these years as described in Table 5.2.10.

- ***SUMMARY OF FINDINGS***

An increase in mean annual IP days and DW visits among AIDS patients were observed, while number of OP visits declined. Increased use of IP services was also observed among HIV-1$^+$ people with symptomatic non-AIDS while asymptomatic HIV-1$^+$ people used more DW services over time. These increases in annual cost of treatment and care over time were most pronounced for AIDS patients on quadruple-or-more ART. This may be due to these ART-experienced patients having to use more IP and DW services, as well as their increased use of new and more expensive ARVs in the second half of the study period.

- ***LIMITATIONS***

The majority of patients were seen in London clinics and increases in annual cost of treatment and care over time were most pronounced for AIDS patients on quadruple-or-more ART may have skewed our findings, though overall costs did not seem to differ substantially between London and non-London centres during the study period (Beck *et al.*, 2008). This analysis is furthermore contingent on the fact that most if not all HIV-1$^+$ people using NHS services during the study, were diagnosed and reported to the HPA. Examples were described in the methods section of how some of the HPA data had to be adjusted because of missing data.

Another potential source of error for estimating the population cost was the extent that HIV-services also used generic, non-HIV services to operate. While the unit costs include facility-level

overheads, other general costs covered by the NHS, such as its drug procurement system, general staff training and other more general non-HIV indirect support, were not included in these estimates. However, these points suggest that the estimates produced in this study are likely to underestimate the true population cost of delivering HIV services in the NHS. Furthermore in the UK HIV-1^+ people are entitled to standard care under the NHS and findings from this study are applicable to the UK only as the infrastructure of care of HIV-1 patients in the UK differs to those from other countries.

- ***REPORTS FROM OTHER STUDIES***

At the end of 2012 the number of people living with HIV-1 infection in the UK was estimated to be 98,400, of whom an estimated 22% (credible interval: 18% to 27%) were unaware of being infected with HIV-1 (Aghaizu *et al.*, 2013); 6,364 people were newly diagnosed with HIV-1 in 2012, or 5% of the estimated 98,400 HIV-1^+ people in the UK and 6% of the 76,500 HIV-1^+ people reportedly using NHS services during 2012. A recent study in the UK, confirmed the benefit for starting ART early with CD4 T-cell count <350 cells/μL blood (Beck *et al.*, 2010). The annual cost of treatment and care, for those who started ART with a CD4 T-cell count >200 cells/μL blood, is 30-35% less than for those who start ART with a CD4 T-cell count ≤ 200 cells/μL blood. However, between 1996 and 2006, of 5,541 HIV-1^+ people who started first-line therapy, 55% were diagnosed with a CD4 T-cell count ≤ 200 cells/μL blood, many of whom were Black Africans (Beck *et al.*, 2010). It is likely that many of the HIV-1^+ people, who are currently unaware of being infected, may well present late in their disease course with a CD4 T-cell count ≤ 200 cells/μL blood. Starting more HIV-1^+ peoples with a CD4 T-cell count <350 cells/μL blood will increase the number of people receiving HAART, which will initially add to the financial burden of the NHS. However, starting HIV-1^+ people on cost-effective regimens earlier, will maintain them in better health, resulting in the use of fewer health or social services and thereby generating fewer treatment and care costs, while enabling them to remain socially and economically active members of society.

- ***FUTURE WORK***

Increasing population costs for HIV treatment and care are raising serious concerns in high-income countries and the equally trivial answer as to where it will end is clearly the obvious limit of a finite GDP (Appleby, 2012), especially as many are going through periods of cutting public expenditure as a consequence of the global economic downturn. This issue is even more pertinent for middle- and lower-income countries. As part of universal access, many of these middle- and lower-income countries have been increasing treatment and care services and the number of HIV-1^+ people on ART reported by WHO in June 2011 had reached 6.6 million HIV-1^+ people (World Health Organisation, 2011b). While this constitutes a great success, if countries want to continue to increase their ART coverage the issue of providing sustainable quality services in the context of

limited country resources raises serious questions. Lessons from high-income countries teach us that increased coverage is going to result in increased population costs.

The recent change in WHO criteria to start ART when CD4 T-cell count is <350 cells/μL blood (World Health Organisation, 2010a) increases the number of HIV-1^+ people in need of starting on ART, and also raises the ethical issue whether HIV-1^+ people with more severe HIV disease should receive ART first, or should HIV-1^+ people with higher CD4 T-cell counts have preference because their treatment and care costs are less than those with higher CD4 T-cell counts? Starting ART early is effective in terms of decreasing rates of mortality observed (Severe *et al.*, 2010) and cost-effectiveness (Badri *et al.*, 2006) in resource limited settings and if ART is focused on these HIV-1^+ people, one could argue that there will be more resources available that will enable more HIV-1^+ people to be receiving ART, while allowing them to remain socially and economically active. These are some of the issues, which countries and resource-limited ones in particular, will have to face. Furthermore, many low-income countries will need to make these choices with the knowledge that many of them, at least for the foreseeable future, will continue to be reliant on donor agencies or countries to sustain their treatment and care services (Hecht *et al.*, 2009).

Trying to curtail the costs of service provision is one measure by which one could try and curtail the population cost. Measures which are being used range from sending laboratory test results by email, home delivery of the drugs to the patient, which is exempted from VAT in the UK, to using the most cost-effective regimens (Beck *et al.*, 2010). Even if the costs could be brought down, without reducing the quality of services provided, the fact that the number of new people being infected with HIV continue to outpace those being put on ART, will continue to drive up population cost for HIV services.

Only greater prevention efforts will reduce the number of people becoming infected with HIV. A study from the US suggests that even if incidence is reduced, HIV prevalence is likely increase creating additional demands for health care services (Hall *et al.*, 2010). While putting HIV-1^+ people on ART will reduce their infectivity and contribute to reducing the incidence of people newly infected with HIV-1 (Montaner *et al.*, 2010), this in itself will not be sufficient to reduce incidence in a large number of settings. It is now recognized that in most instances a combination of prevention interventions are going to be required to achieve significant reductions in HIV-incidence (Hankins & de Zalduondo, 2010).

While recent biomedical interventions are likely to be a cost-effective HIV prevention strategy, such as male-circumcision (Kahn *et al.*, 2006), the development of a vaccine (Rerks-Ngarm *et al.*, 2009) and microbicides (Baleta, 2010) - constitute important advances, they are most likely to be

successful when juxtaposed with relevant behavioural and structural changes (Hankins & de Zalduondo, 2010). One of the findings of the Second Independent Evaluation of the UN Joint Programme on HIV/AIDS highlighted the relative lack of success of global prevention programmes, including in high-income countries, and recommended a greater emphasis on making prevention more effective and efficient (Poate *et al.*, 2009). Policy makers and other relevant stakeholders need to use evidence-informed HIV prevention, treatment and care strategies to ensure that HIV-1^{+} people have full access to the treatments that they need based on their clinical condition, which will prolong life, reduce morbidity, reduce transmission and ultimately deliver the best for both the individual and public health agendas.

Chapter 6: General Discussion

CHAPTER 6: GENERAL DISCUSSION

The thesis concludes with a general discussion of the novel results highlighted by the use of statistical analysis applied in Chapters 3, 4 and 5.

6.1 DISCUSSION

This chapter reviews some of the empirical findings and limitations from basic, clinical and population science research presented respectively in Chapter 3 (LTNP), Chapter 4 (HIV and Ageing) and Chapter 5 (Population Cost Study) concerning the effects of HIV-1 infected people using the NHS demonstrated by the use of patient level data in these chapters. This chapter summarises what has been presented in the preceding five chapters and draws lessons from the theoretical and empirical information derived using data management processes which have broader relevance for the HIV-1 research and policy implications.

The aims of this PhD thesis were: to describe uses of data management to process data from HIV-1 specific clinical databases that arise from single and multiple data sources; to identify groups of patients who exhibited atypical characteristics of HIV-1 infection described as long-term non-progressors (LTNP) identified from a large cohort of HIV-1^+ people and identify factors that were associated with LTNP so that these patients would provide a basis to study in detail immunologically in order to help determine the most appropriate prophylactic and therapeutic approaches such as immune based therapies; being treated long-term on ART, together with ageing, is a complex 'ageing' process involving biological, social and psychological aspects. Identification of the interacting effects of age, HIV-1 infection and ART is likely to guide improved therapeutic intervention in HIV-1^+ people with advancing age; and to provide an estimate of direct population cost of treatment and care of HIV-1^+ in the UK to help determine budget allocation for treatment and care of HIV-1 infected people within the NHS.

Chapter 1 described the rationale for having and setting up databases, the use of data management for collecting routine individual patient data, and provided a general introduction to HIV-1 infection with reference to its impact both globally and nationally. Chapter 2 described data management and statistical methods used to carry out the work in the three successive chapters. Chapter 3 (LTNP) looked at local experience of using a database and data management to identify HIV-1 patients with HIV-1 atypical characteristics. Chapter 4 (HIV and Ageing) looked the synergistic effects of HIV-1 infection, ART and age on one of the non-HIV co-morbidity examined (CKD) using local database in order to investigate the synergistic effect of HIV-1 infection, ART and ageing. Chapter 5 (Population Cost Study) used data from multi-centre HIV-1 cohort and used data management and data analyses to report changing use and cost of HIV services over time since the HAART era and estimate trends in direct population cost during the study period when data were available and use these trends to estimate direct population costs in the future when

data at the time were unavailable. Evaluation of HIV-1 patient data from the NPMS-HHC cohort was used to estimate direct HIV-1 population costs during the study period as well as project these for subsequent years. When compared to the HPA estimates of people living with HIV-1 using the NHS services used to estimate population cost the projected numbers of estimated HIV-1 diagnosed people who would use the NHS services during 2007, 2008, 2009 and 2010 were underestimated by between four and five percent (see Chapter 5: Table 5.2.10).

6.1.1 STATEMENT OF PRINCIPAL FINDINGS

- ***BASIC SCIENCE RESEARCH: DATA ANALYSIS PRESENTED IN LONG-TERM NON-PROGRESSORS CHAPTER***

Chapter 3 used data from the local database to identify HIV-1^+ people with atypical characteristics of being infected with HIV-1 for long duration without developing symptoms of HIV-1 progression described as LTNP. The study presented demonstrated that the number of HIV-1^+ people fulfilling the criteria used to define LTNP, HIC or long-term stable low $CD4^+$ T-cell counts within the Chelsea and Westminster HIV cohort was very small. The caseload of HIV-1^+ patients who had attended the centre for treatment and care during February 2009 to February 2010 was 6,390, only 13 (0.20%) subjects fulfilled the criteria which defined them as LTNP. Further only 0.58% of the HIV-1^+ people in the cohort were classified as exhibiting long-term stable low $CD4^+$ T-cell counts and 0.05% were in addition also able to control their HIV-1 RNA load replication BLD. Further, the work from the LTNP study demonstrated that using different criteria to define atypical patient groups resulted in different estimates in time until disease progression, before considering functional immunology, genetic and virological factors, emphasising the need for standardised use of terminology and definitions before more in depth study of these individuals is performed. Detailed immunological work is currently under way on LTNP patients who were identified from the work carried out during this PhD including those with continued LTNP status and these identified patients are being used for detailed immunological study as part of another PhD work.

- ***CLINICAL SCIENCE RESEARCH: DATA ANALYSIS PRESENTED IN HIV AND AGEING CHAPTER***

Recent studies demonstrated an increased life-expectancy of HIV-1^+ people on ART. As life-expectancy increased, the number of people aged ≥ 50 years with HIV-1 has also increased and many were managed with long-term ART. As HIV-1^+ people grow older, they are at an increased risk of developing non-HIV co-morbidities, some of these are cardiovascular disease, non-AIDS related cancers, neurological complications, bone abnormalities, 'frailty', liver and renal problems (Goetz, 2010; Deeks, 2011). Renal function, one of non-HIV co-morbidities, was investigated and was presented in Chapter 4. Studies have reported that people diagnosed with HIV-1 infection receiving long-term ART are at an increased risk of developing CKD (Levey *et al.*, 2009). Data presented in Chapter 4 (HIV and Ageing) described the changing age distribution of an HIV-1^+

cohort of patients attending a single centre for HIV care since HAART era commenced. A description of renal function changes by age was investigated to establish the synergistic effect of HIV-1 infection, increasing age and exposure to most commonly prescribed ARV classes of drugs to HIV-1^+ people: NRTI; PI; and NNRTI; aswell as those on first-line HAART comprising 2NRTI+NNRTI; or 2NRTI+boosted PI. The longitudinal analyses presented in Chapter 4, comprised both ART naïve and ART experienced patients, and the findings demonstrated that renal function based on eGFR in HIV-1^+ people deteriorated more with increasing age than 'healthy' older people. ART naïve and HIV-1^+ people on first-line ART showed similar reductions in eGFR with increasing age well below the creatinine clearance levels observed in uninfected people. In addition it reported that although eGFR are frequently reported in publications, equations to estimate eGFR have not yet been validated in HIV-1^+ people. The similarity in eGFR reductions among ART naïve and those on first-line ART suggested that HIV-1 itself impairs renal function which may suggest that ART should be started soon after diagnosis to reduce end organ damage. As the HIV-1 infected population ages, these conditions may warrant novel therapeutic approaches, requiring improved linkage and integration of HIV services with renal services as well as services for other co-morbidities.

- ***POPULATION SCIENCE RESEARCH: DATA ANALYSIS PRESENTED IN POPULATION COST ANALYSIS CHAPTER***

Findings described in Chapter 5 of this PhD thesis showed that the number of people accessing NHS HIV services increased substantially between 1996-2006 and more people who are HIV-1^+ are now on triple or quadruple or more ART surviving longer and the relative lack of success of HIV preventing programmes, resulting in a rise of UK population costs during the study period for providing HIV services. The results presented in Chapter 5 estimated that by end of 2013 over 78,000 people diagnosed with HIV-1 in the UK will be using the NHS services and the annual cost for treatment and care was shown to vary by stage of HIV infection and type of ARV therapy, which need to be taken into consideration when resources in the NHS are being allocated and spent. The estimated direct annual population cost to treat HIV-1^+ people by end of 2013 was estimated to be £758 million. The HPA estimates of people living with HIV-1 using the NHS services used to estimate population cost during 2007, 2008, 2009, 2010, 2011 and 2012 the projected numbers were underestimated by up to maximum of five percent (see Chapter 5: Table 5.2.10) while the variations in population cost estimated using three different methods during 2007, 2008, 2009, 2010, 2011 and 2012 was shown to differ by up to five percent. In the current economic downturn, containing and reducing the direct annual costs of HIV-1 treatment and care, without reducing the quality of services, remains a challenge for healthcare professionals and commissioners working in the field of HIV. Greater efforts are needed in order to reduce the number of people becoming newly infected with HIV-1.

CHAPTER 6

GENERAL DISCUSSION

- ***SUMMARY OF METHODS USED***

In order to either enhance current knowledge or to generate new knowledge, trends that emerged from the data were used to draw inferences. While databases and data warehouses are important as they are data repositories used to store large amount of data, data management (*'the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets'* (Mosley, 2007)) and analyses were used in all three studies presented in Chapters 3, 4 and 5 to help deliver knowledge in order to help make informed decisions based on data stored in large databases or data warehouse. So all three studies presented in this PhD thesis used data management (see Chapter 1: Figure 1.1.4) and data mining processes (see Chapter 1: Figure 1.1.9).

- ***GENERAL LIMITATIONS OF METHODS USED***

Epidemiological research in HIV-1 is complex as it is a multi-factorial and progressive disease influenced by biological, psychological, social, environmental, and genetic factors that can present significant challenges to patients. Successful management of the disease requires constant consultation with specialist HIV healthcare providers. In addition, with the increasing ageing HIV-1^{+} population, consultations with other healthcare specialists are also required for management of non-HIV co-morbidities such as the renal disease for instance. When data are linked together that come from different sources, corrupted bits of data have often crept into the datasets and after these were checked the rule was to treat these as problems and corrupted bits of data were removed. With relatively small data, reducing errors and ensuring high quality of data was essential since only a relatively small data are collected it was therefore important that these were recorded as accurately as possible. In addition, monitoring patients over a long period of follow-up meant there was a higher chance that patients being followed up in the cohort would leave the cohort either because of death or because they choose to leave the cohort for other reasons. Further, it was difficult to establish and measure causal relationships among several risk factors for HIV-1. In the UK the HIV-1 infected population is diverse and mobile, making individuals more difficult to track especially when they attend other health care centres for care. Caring physician often rely on the patient to accurately give details of their HIV-1 care history prior to receiving care at their institution. While successful treatment for HIV-1 requires a specialist HIV physician to guide the patient with their disease management referring patients to other specialists which may be in a specialist centre located elsewhere from their usual HIV-1 care centre. Clinical questions of relevance in HIV-1 are mostly always based on treatment strategies which often involve more than one treatment change and the development of ARV drug toxicity being the key issue. Patients on a long-term follow-up receive ARV therapy according to the protocols and national treatment guidelines in place at their HIV-1 care centre at the time of being treated for HIV-1. The cohort of patients are followed-up in their own treatment and care centre and the follow-up data are obtained from local clinic database or patient case notes therefore data such as renal function markers

(eGFR) for instance are measured frequently but may also be measured irregularly as information will be recorded as and when a patient attends the clinic. Although data collected on patients from one clinic are usually representative of the clinic population from which the data are drawn from. In addition, patients within a clinic have a level of adherance to treatment that is consistent within the indivudual care centre and therefore the information that is derived reflects what happens in practice therefore making it possible to make comparisons by treatments such as the renal function outcome by many ARV treatments. The comparisons between treatments is unbiased if the treatment groups have the same entry criteria at baseline and are treated similarly during follow-up apart from the differences of starting different ARV treatments which may not necessarily be the case in routinely collected observational cohort. Patient prognosis may well differ between the treatment groups and this may well be the reason why some HIV-1^{+} patients were treated with one ARV combination while others with another ARV combination. For example, patients receiving PI's as part of their ARV combinations may have been at an advanced stage of HIV as this group may well have received early PI drug formulations, were at an advanced stage of HIV with poor prognosis while those receiving newer NNRTI's as part of ARV combination could have been at a less advanced stage of HIV-1 infection and have therefore have improved prognosis. The statistical approach taken to minimise these types of biases was by using longitudinal MIXED model described in Chapter 2, adjusting for known factors at baseline, however, it is not possible to adjust for all residual factors due to unknown and/or unavailable factors. With some of the limitations highlighted herein routinely collected cohort data used in this PhD thesis are however valuable resource for monitoring the longitudinal changes in incidence of an outcome in a real life clinical setting as well as monitoring the use and cost of HIV service provision.

6.1.2 LONGITUDINAL

- ***USES OF INDIVIDUAL LEVEL LONGITUDINAL DATA***

In the same individual the longitudinal data collection were repeatedly measured at several different time points. The observations over time of each individual are not independent of each other, and therefore statistical techniques taking account of longitudinal nature of data described in Chapter 2 were applied that took account of correlation of the repeated measurements taken from each individual. One of the advantages of longitudinally collected data is that the development of an 'outcome' over time can be studied and described. For instance, data presented in Chapter 4 on renal function by age describe how renal function changed with increasing age.

- **Why We Need Longitudinal Data**

One of the uses of collecting longitudinal individual health data is for the purpose of improving medical records. Managing individual medical needs such as managing chronic conditions like HIV-1 infection as described in Chapters 3, 4 and 5. Use of longitudinal data were described for

more detailed work such as that described in Chapter 4 (HIV and Ageing with reference to renal function changes over time) and Chapter 5 (Population Cost Study with reference to estimating the use of HIV services and costs per patient year). With this type of data collected lifestyle modification, medication management, patient education, lab results monitoring and adverse events can be co-ordinated across different clinical care teams. These type of longitudinally collected data provide access to evidence-based information at all points of care. Therefore this allows different clinical care teams to support a patient's care, encourage treatment adherence and allow different clinical team members from nurses, specialist physician to pharmacy to co-ordinate patient care.

With the current economic climate with tightened financial resources, efforts within the NHS are to curtail costs. Data presented in Chapter 5 showed that the projected population cost of treating HIV-1^{+} individuals in the UK was estimated to increase by 1.5 fold since 2006 to £758 million by 2013. The types of clinical databases used for such work such as NPMS-HHC cohort presented in Chapter 5 (Population Cost Study) are healthcare efforts which have shown that such use, cost and outcome data collected longitudinally can help drive clinical improvement to improve and save lives and help identify sources where costs can be saved. However, the data used for this work included data from HIV specialist individual participating centres who had data collected electronically. Those centres who were unable to provide data in an electronic format for various reasons, including not having an IT infrastructure to capture electronic medical data to difficulty with extracting data from existing databases which held electronic medical records, for these types of reasons not all clinics who agreed to participate with the NPMS-HHC could provide data and therefore not all clinics who were participating with the NPMS-HHC could be included in the work presented. This was one of the limitations of work presented in Chapter 5. Small clinics where electronic data capture was unavailable, database to collect specific fields for NPMS-HHC study participation was provided and this proved difficult even to a small clinic as capturing all required data retrospectively from patient notes was staff time and resource intensive.

6.1.3 DATA MANAGEMENT

Routinely collected health data of individuals captured electronically can eliminate rework by capturing data once at the source and presenting it for reuse as needed downstream. However, this approach assumes error free electronic medical record, which may not be achieved in an individual health care centre as described in a review by Hogan and Wagner, 1997. This review summarised studies evaluating both correctness and completeness and various sources at which errors could be introduced (Figure 6.1.1). Using an electronic medical record capture as an example, Figure 6.1.1 demonstrates some of the sources where data errors could be introduced. A review by Komaroff (Komaroff, 1979) provides a detailed review processes by which different types

of medical data are recorded and where different sources of errors may be introduced at each data collection stages.

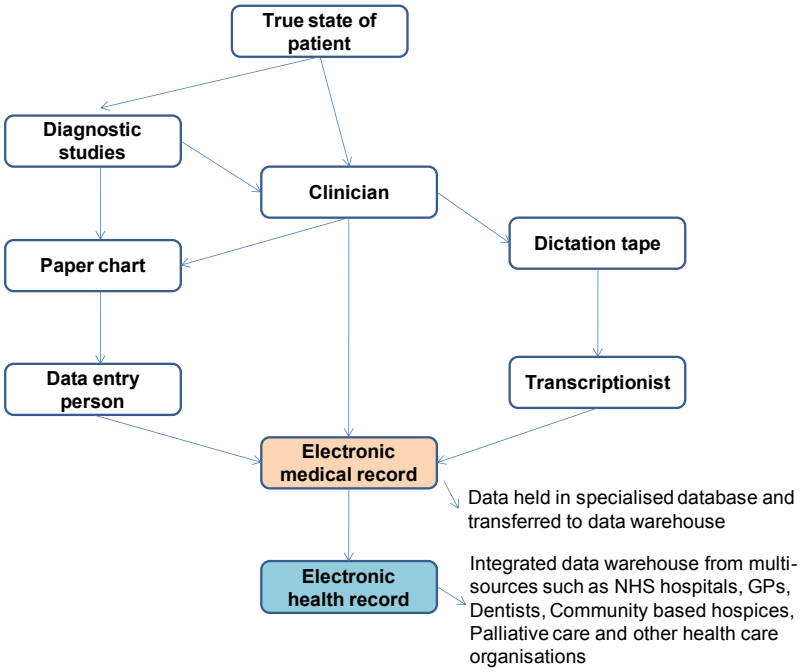


Figure 6.1.1: Electronic Medical Record – Some Of The Sources Where Errors Could Be Introduced
Figure adapted from (Hogan & Wagner, 1997)

The electronic health record can only be put in place if the electronic medical records of the various different health care stakeholders (both locally and nationally within a country), which include care centres such as the UK NHS hospitals, GP's, dentists, community based hospices, palliative care and others can create and support exchange of electronic information between different health care stakeholders.

Problems could also occur in terms of computer hardware and software support. The increased resources required in terms of employee time and costs required to train on using the new software for routine data entry including how to avoid data loss. In some instances, newly established electronic medical records system may generate errors rather than prevent them, especially early in the adoption process where such errors could harm the care of a patient and create data gaps. Sources of errors include: fragmentation of data; failure to integrate all hospital systems; and human to computer interface difficulties in the machine rules.

- ***DATA PRIVACY, CONFIDENTIALITY AND SECURITY***

Privacy protections provide the overall framework within which both confidentiality and security are implemented (UNAIDS, 2007). While privacy is related to the way a patient could be identified and contacted, confidentiality deals with the identifiable data and who has access to the patient data. With increased access to data through electronic medical records at different locations via links to databases and data warehouses, there is a potential increased risks to the security of the information possibly violating patients confidentiality, security and privacy of health records (House Of Commons Health Committee, 2007). However, there needs to be a balance between protection of confidentiality and access to information to improve public health care. Connecting for Health is an organisation who has the responsibility for NHS operational security, including strong access controls and audit systems where data can be exchanged via the N3[27] network. While the NHS has guidance for encryption policy and provides guidance to safeguard confidential information (NHS Information Governance, 2013) and despite the fact that the IT specialists continue to work hard to prevent security risks, the internet is known to transmit malicious codes too. Malware authors are continuously able to learn how to manipulate the loopholes of ever advancing technologies and are then able to perform malicious attacks and intrusions. The losses from such intrusions may include loss of data confidentiality (unauthorised disclosure), loss of data integrity (unauthorised modification), or decreased system functionality. As a result costs incurred due to needing to repair the damage which may also include loss of individual productivity. For instance what happens if a doctor loses a server connection or is exposed to a computer virus while inputting notes into the computer for electronic record keeping? Currently it is unclear whether data held in an electronic format are any less safe than the paper based systems they replace. Electronic medical and health data confidentiality is important because some diseases such as HIV-1 are stigmatised. At different times in history, some people considered HIV-1^+ status to be contagious just by touching an individual infected and that people who were HIV-1^+ were to be avoided. Patients with HIV-1 infection for instance have to provide information to their physician in order to receive treatment. To encourage open and honest exchanges, patients are assured that information about their medical records would be kept confidential. The assumption is that society benefits when patients are treated because the potential spread of the disease such as HIV-1 and cost of its treatment have been kept in check. When breaches in data security happen and become high profile with media coverage, this may affect patient-clinician relationship.

Chapter 6:27 N3 network has a variety of end user NHS organisations from GP practices to large hospitals with dedicated IT staff. It has gateways to the Internet. A number of approved NHS suppliers are connected to N3. All involved, including those who connect to it, have security responsibilities: Description taken from: http://n3.nhs.uk/technicalinformation/N3NetworkSecurity.cfm

6.1.4 DATBASES AND DATA WAREHOUSES IN THE UK

- ***DATABASES***

With ever advancing technology, more and more medical data are now captured in databases. Clinical databases hold comprehensive clinical data with restricted access given to registered users and individual databases are held at care centre. Some of the types of data that may be held on clinical database are described in Chapter 2: Table 2.4.2. One of the main limitations of holding individual clinical database it is a standalone database with no link to other sources of patient care. For instance if a single database is created to collect specific patient information which excluded their inpatient stay, a physician looking at the database would not have electronic access to inpatient stay information from that stand alone database and would therefore need to acquire this information through other sources such as patient notes.

- ***DATA WAREHOUSES***

In contrast data warehouses (see Chapter 2: Figure 2.4.1) are repository for data storage and extraction from different services. Once stored the data can be extracted using data warehousing software such as relational database.

- ***SINGLE CENTRE DATABASES AND MULTIPLE CENTRE DATA WAREHOUSES - HIV-1 AND OTHER SPECIALTIES***

The Healthcare Quality Improvement Partnership (HQIP) was established in April 2008 to promote quality in health and social care services in the UK and is led by a consortium of the Academy of Medical Royal Colleges, the Royal College of Nursing and National Voices and is contracted by the Department of Health in England to deliver outcome focused quality improvement programmes structured around collection of clinical data, including clinical audits and registers. From 1st April HQIP is reporting to the NHS Commissioning Board rather than the DH. A list of HQIP Directory of Clinical Databases and Registers lists projects registered with HQIP in the UK (by February 2012) which collect systematic data on procedures, conditions and services in healthcare and contain a list of clinical registers, databases and surveillance registers (Table 6.1.1). These are listed under the following broad clinical areas and include audits, research and service evaluation databases which were mapped (HQIP, 2012).

Table 6.1.1: Register of Clinical Databases In The UK

	Total number of Clinical databases (local, regional and national)	Number of local databases	International
Acute Care	14	3	0
Blood Transfusion	4	0	0
Cancer	20	2	0
Cardiovascular	6	1	0
Gastro-Intestinal	1	0	0
Haematology	3	0	0
Long-term	29[&]	2	1
Ophthalmology	1	0	0
Orthopaedic	3	0	0
Peri-Neonatal & Children	20	0	0
Trauma	3	0	0
Other	9	0	1
Current Audits	6	0	0

Source of information in Table 6.1.1 is derived from HQIP, 2012

[&] There are 6 registered HIV-1 infected patients clinical databases in the UK and these are:

1. *HIV and AIDS New Diagnoses Database (HPA)* (from 1st April HPA became part of NHS England)
2. *Survey of Prevalent HIV Infections Diagnosed (SOPHID) (HPA)* (from 1st April HPA became part of NHS England)
3. *UK Collaborative HIV Cohort Study*
4. *UK HIV Drug Resistance Database*
5. *UK Register of HIV Seroconverters*
6. *National Haemophilia Database (surveillance for HCV, HIV and CJD in this group.)*

- ***HEALTH AND SOCIAL CARE BODIES DATABASES***

The Health and Social Care Information Centre (HSCIC) is a delivery organisation with statutory duties which was established on 1 April 2013 as an executive non-departmental public body (ENDPB) under the Health and Social Care Act 2012. As an ENDPB, the organisation is accountable to the Secretary of State for Health for discharging its functions, duties and powers effectively, efficiently and economically. The HSCIC is a data, information and technology resource for the health and care system and plays a role in driving better care, better services and better outcomes for patients. It is the source of data and information relating to health and care (Table 6.1.2). It supports the delivery of IT infrastructure, information systems and standards to ensure information flows efficiently and securely across the health and social care system to improve patient outcomes. National data sets are generated from care records, from any organisation or system that captures the base data. They are structured lists of individual data items, which can be used to monitor and improve services (http://www.hscic.gov.uk/datasets). The data list is managed by the HSCIC (HSCIC, 2013).

Table 6.1.2: List of National Data From Health and Social Care Bodies in England

Subject	Data Collection and data sets	Comment
Community care	Community Information Data Set	There is currently no national data collection for community care, but the community information data set aims to address this by providing definitions for a national data set. In common with other secondary use data sets, this describes the data which NHS funded providers of community services will be required to submit to the centre for analysis and reporting. This is a subset of data needed for care delivery.
	Maternity and Childrens' Data Set	
Estates and facilities	Hospital Estates and Facilities Statistics	This is a data warehouse containing data on various different returns each NHS Trust has to complete, some dating back to 1999/2000, and covers a great variety of information ranging from Estates Information Returns, Fires, Defects and Failures to the spend on such things as cleanliness and food in Hospitals. The system also allows the NHS to undertake benchmarking within the returns to assist in performance management.
Hospital care	KO41a - Hospital and Community Health Services Complaints	There were fifteen million inpatient admissions to NHS hospitals in England in 2011. We have a variety of information covering all aspects of secondary care which can help you plan, monitor and improve services.
	KP90 - Admissions, Changes in Status and Detentions under the Mental Health Act	
Mental health	Mental Health Minimum Data Set (MHMDS)	In 2010/11 over 1.25 million adults accessed NHS services for severe or enduring mental health problems.
	Improving Access to Psychological Therapies (IAPT) Key Performance Indicators Data Collection	
	Improving Access to Psychological Therapies (IAPT) Data Set	
	Deprivation of Liberty Safeguards (DoLS) Data Collection	
	Community Mental Health Activities Data Collection	
	Child and Adolescent Mental Health Services Data Set	
	Guardianship Collection	
Patient experience	KA34 - Ambulance services	Improving patient experience is a key aim for the NHS. By asking, monitoring, and acting upon patient feedback, organisations are able to make improvements in the areas that patients say matter most to them.The Health and Social Care Information Centre collects and publishes statistics related to many stages of the patient journey, including ambulance services, hospital facilities and follow-up services.
	KA41a - Hospital and Community Health Services Complaints	
Prescribing	Primary Care Trust prescribing data (Opens in a new window)	Prescribing is the most common patient-level intervention in the NHS, and covers all sectors of care: primary, hospital, public and community health. It is the second highest area of spending in the NHS, after staffing costs.
	GP Practice prescribing data	
	Practice level prescribing data - released under the Transparency Agenda	
Primary care	SBE515 General Ophthalmic Services (GOS): Activities	As many people's first point of contact with the NHS, around 90 per cent of patient interaction is with primary care services. In addition to GP practices, primary care covers dental practices, community pharmacies and high street optometrists. The information gives insight into how services are provided.
	KO41b - General Practice (including Dental) Complaints	
Public health	NHS Stop Smoking Services	The choices people make can have a lasting impact on their health. Improving patient care isn't just about treating people's individual diseases, it also means changing and preventing harmful behaviours which can improve and even save people's lives.
	KT31 Community Contraceptive Services	
	KC50 Immunisation Programme	
	Sexual and Reproductive Health Activity Data set (SRHAD)	
	Smoking at Time of Delivery	

Social care	Social Care Collections	We collect and publish a wide range of information on adult social care, which you can use to plan, deliver and monitor services.
Workforce	KO41b - General Practice (including Dental) Complaints	1.3 million people make up the NHS workforce, all contributing towards the effectiveness of the NHS.
	KO41a - Hospital and Community Health Services Complaints	
	National Workforce Data Set	
	Health Visitors Minimum Data Set	
	Workforce Minimum Data Set (wMDS)	

Source of information contained in table: http://www.hscic.gov.uk/datasets (Accessed 21st May 2013)

- ***ADVANTAGES AND LIMITATIONS OF SINGLE CENTRE DATABASES AND MULTIPLE CENTRE DATA WAREHOUSES***

  - **How Are Data Collected**

Data in a data warehouse come from many different sources (Chapter 2: Figure 2.4.1). The relationships between different data added in the data warehouse may not necessarily be identified however the data field names and data types are usually reconciled. The data warehouses are now widely used as they provide a timely information for decision-making. However getting timely information in raw data format is not always beneficial as data from a large number of people usually in analysed form are beneficial and informative. Once data are integrated from multiple sources, this reduces costs to access historical data and therefore improves the turnaround time for analysis and reporting; allowing others to easily access and share data. Data held in data warehouse have the potential to withstand instances such as fire/flooding, where paper copies of patient notes could be ruined however copies of data stored in a data warehouse will remain intact.

  - **Limitations of Data Collected**

Data represents the lowest raw format of information or knowledge. Large volumes of data available in varying complexity, (For example, in a graph, M and F can depict gender or can depict Monday and Friday) cannot be processed using standard data processing methods and algorithms such data include those that are machine generated data for instance from X-rays and scanning devices. One of the limitations of handling large volume of data from data warehouse is the speed of data processing. Speed of data processing requires a combination of: hardware, software, networking, and storage. Each of these has limitations and, in a combination, these limitations affect the speed of processing large data. Some of the limitations observed using routinely collected data in specialist HIV clinics are described using basic science, clinical science and population cost studies described in the previous three chapters.

It is possible that many more HIV-1^{+} are LTNP or HIC, but due to the absence of clinical manifestations of disease these individuals have not yet attended an HIV-1 testing facility and consequently have not been diagnosed as HIV-1^{+}. The criteria of time from HIV-1 seropositive diagnosis should also be considered as relatively arbitrary, as the date of first HIV-1^{+} test result, or date of first attendance at an HIV clinic, may not be an accurate representation of the date of

infection. Instead, the date of first positive HIV-1 test should be considered a tool for healthcare professionals to identify patients exhibiting an unusual course of HIV-1 disease. Data used for this work are based on routinely collected observation data and like for other such databases data limitations are found such as HIV-1^+ patients are lost to follow-up. Another limitation of observational data from a single centre is that HIV-1^+ patients can also be managed in other centres and may receive ARV treatment from other clinical care providers, including private practice, which may not be captured in the database of the single centre. This was one of the findings when data were further investigated using a HIV-1 RNA plasma load criteria where 83% or 79 individuals who were found to have 3 consecutive HIV-1 RNA plasma load <500 spanning at least one year were recorded on patient case notes to have received ART while ART history in these individuals were not recorded on computerised HIV-1 database. Currently clinicians are reliant on patients' self reporting of ART treatment history including ART received elsewhere. The formation of a national centralised specialist observational database, highlighted by the UK governments Information Strategy (Department of Health, 2012) is a likely strategy to overcome such limitations in the future.

The eGFR according to the MDRD formula is which employs serum creatinine measurements and patients body surface area, age, ethnicity and gender (Levey *et al.*, 2006). However, some remaining residual effect of age may remain and some of the changes in renal function observed may be due to a persistent age effect. Although the data show a steeper decline in renal function markers in 'middle-aged' adults 40-49 years and adults 50 years or older, the number of patients aged 70 years or older were relatively small resulting in greater variations in trends of the renal markers observed in these age groups. Furthermore, renal function measurements were taken from a diverse group of patients who contributed results at different ages. Ideally one would like to prospectively follow-up a defined group of 'young' HIV-1^+ adults as they age, and monitor changes over time, but such data would take a long time to collect. Furthermore, ART regimens are also likely to change over such long time periods, compounding longitudinal analyses. Creatinine clearance values from HIV uninfected people should strictly speaking not be compared with eGFR from HIV-1^+ people as the creatinine clearance values were not adjusted for body surface area, age, ethnicity and gender (Rowe *et al.*, 1976). In the absence of such standardised longitudinal data from a 'healthy' populations, the current comparator including a 20% reduction to adjust for any over-estimation (Breyer & Qi, 2010) and provides an insight into the degree of renal impairment associated with HIV-1 infection or ART. While eGFRs are less reliable for assessing stage 1 and 2 CKD when eGFR is 60 to 90 ml/min per 1.73m^2, it does becomes more reliable for picking up stage 3 CKD, with eGFR rates between 30 and 59 ml/min per 1.73m^2 (Stevens *et al.*, 2007). A recent meta analyses demonstrated that the MDRD equation to calculate eGFR is not an optimal measure for all populations and eGFR ranges calculated using a single equation requires a tradeoff at higher or lower values of eGFR ranges (Earley *et al.*, 2012). The eGFR is predicted

based on the MDRD formula which employs serum creatinine measurements and patients body surface area, age, ethnicity and gender (Levey *et al.*, 2006). However, some residual effect of age or other contributing factors may remain and influence the changes in renal function observed. In addition, the MDRD formula excluded patients with renal failure, though it included patients aged 18-70 years, but it has not yet been validated in an 'elderly' or 'frail' population (Levey *et al.*, 1999). Other formulae which has also appeared in the literature to estimate creatinine clearance from a single estimate of plasma creatinine without urine collection include the Cockcroft-Gault formula (Cockcroft & Gault, 1976). An alternative to the MDRD formula is the Cockcroft-Gault formula where the creatinine clearance is predicted from serum creatinine measurements and patients' age, gender and weight. While some prefer to use this formula (Cockcroft & Gault, 1976), as it is considered it to give better estimates of renal function in subjects with well preserved normal renal function (Robertshaw *et al.*, 1989), others prefer to use the MDRD formula (Barraclough *et al.*, 2009). As routinely measured weights were unavailable for the cohort studied, the Cockcroft-Gault formula could not be used in this study.

The majority of patients were seen in London clinics and increases in annual cost of treatment and care over time were most pronounced for AIDS patients on quadruple-or-more ART may have skewed our findings, though overall costs did not seem to differ substantially between London and non-London centres during the study period (Beck *et al.*, 2008). This analysis is furthermore contingent on the fact that most if not all HIV-1$^+$ people using NHS services during the study, were diagnosed and reported to the HPA. Examples were described in the methods section of how some of the HPA data had to be adjusted because of missing data. Another potential source of error for estimating the population cost was the extent that HIV- services also used generic, non-HIV services to operate. While the unit costs include facility-level overheads, other general costs covered by the NHS, such as its drug procurement system, general staff training and other more general non-HIV indirect support, may not have been included in these estimates. However, these points suggest that the estimates produced in this study are likely to underestimate the true population cost of delivering HIV services in the NHS. Furthermore in the UK HIV-1$^+$ people are entitled to standard care under the NHS and findings from this study are applicable to the UK only as the infrastructure of care of HIV-1 patients in the UK differs to those from other countries.

6.1.5 DEVELOPMENT OF DATABASES AND DATA WAREHOUSES

- ***WORLDWIDE***

With the growing information age, many countries worldwide are in the process of implementing electronic medical records and electronic health records which is regarded as an opportunity for improvement in the public health sector to improve patient care, reduce health care costs and change the way medicine is practiced. Some of the countries, where electronic health record implementations underway are: Australia, Canada, France, Germany, Japan, The Nordics, Spain and the United Kingdom (UK). Internationally, the benefits of improvement in the public health sector are driving adoption of electronic health records worldwide despite variations in health care systems, market structures and regulatory requirements.

- ***DENMARK***

Denmark is one of the first countries to have successfully implemented electronic health records and is using it for healthcare. eHealth[28] is now commonly used in all areas of the Danish health service, and today IT supports many work processes (Kierkegaard, 2013).. The country comprises 5.5 million citizens and by 2006 98% of the Danish 3,500 GPs, the majority of specialists, all 73 hospitals, all 331 pharmacies and half of the 271 local authorities shared data over the network, about 80% of the totally exchanged healthcare information was sent electronically (Edwards, 2006; Deutsch *et al.*, 2010). In addition to financial benefits, evidence suggests with the implementation of electronic health records, clinical benefits was seen such as improved adherence to care guidelines, faster exchange of test results, fewer duplicate procedures and more time for clinicians to spend with patients (Edwards, 2006). Some of the lessons learnt with the successful implementation are summarised in a monograph by Edwards, 2006 and include: start with basic needs then add other things; establish continual monitoring and evaluation process and this must include the measurement of improvements in the quality of care; align the incentives of providers; payer organisations and vendors; develop an approach to privacy and security that satisfies the demands of clinicians and patients, and then implement it consistently; keep an appropriate balance between central co-ordination and local leadership and devote plenty of resources to local implementation and training to ensure clinician adoption.

- ***UNITED STATES OF AMERICA***

Unlike relatively small population in Denmark of 5.5 million, the US population is over 57 fold higher at 314 million. Like in the UK, in 2008 the US Congress embarked on a $30 billion program to encourage doctors throughout the country to adopt electronic health records by 2014. The purpose was to create an inter-connected system of electronic health records to improve safety

Chapter 6:28 eHealth means using digital tools and services for health. eHealth covers the interaction between patients and health-service providers, institution-to-institution transmission of data, or peer-to-peer communication between patients and/or health professionals. Description taken from http://europa.eu/rapid/press-release_MEMO-12-959_en.htm

and reduce medical costs. In the US dozens of IT companies competed for contracts most with commercial software products developed in secret competed for billions in funds. In the US, adaption of electronic health records caused problems. With so many different software products used, ensuring different software products were compatible became a challenge, therefore not allowing all clinicians to view the same patient's health information. Clinicians and hospitals reported many medical injuries and deaths and preventable heart attacks caused by problems related to computerised health records such as software errors and unreadable computer screens. Some errors resulted in drug doses that were 10 times higher than intended (Soumerai & Avery, 2010). A recent report (May 2013) in a VentureBeat, which reports tech, money and innovation news in the US, reported that more than 50 percent of doctors and 80 percent of hospitals in the US have received funding for achieving 'meaningful use' of digital records (Farr, 2013). The US healthcare therefore has a mammoth task and is currently in the process of computerising their populations medical records.

6.1.6 LESSONS FOR THE UK

- ***UK HEALTH DATA: TECHNICAL IMPLICATIONS FOR DEVELOPMENT OF LOCAL DATABASES AND REGIONAL AND NATIONAL DATA WAREHOUSES***

The three main results chapters presented used data stored in databases at individual HIV specialist centres both locally and nationally. The need for change requiring access to peoples health records is required. One of the reasons for this need is due to chronic nature of HIV-1 aswell as the growing ageing HIV-1^{+} population requiring other specialist service care both within the NHS and elsewhere such as use of social care. With more integrated services there is a requirement that different specialists have easy access to individual patient health care records quickly and efficiently with minimum time delay. As noted in previous sections, aswell as many countries worldwide, the UK government has committed to update the NHS IT systems by 2014/15. Once the identifiable electronic medical records are linked together into electronic health records these potentially could be held securely in an information centre in a national data warehouse repository. From there the UK governments commitment to publicly make available anonymised population health data from the national data warehouse repository to the individuals aswell as various professionals, commissioners, information specialists, researchers and others (see Chapter 1: Figure 1.1.5).

- ***DATA CONFIDENTIALITY AND SECURITY RISK MANAGEMENT - INFORMATION GOVERNANCE***

With the move towards electronic health records, to cover issues of access to these records regarding patient confidentiality a number of legislation, standards and guidance has been produced in the UK, in order to cover Information Governance[29]. The Health and Social Care Act

Chapter 6:29 Information governance is the term used to describe he structures, policies and practice of the DH, the NHS and its suppliers to ensure the confidentiality and security of all records, and especially patient records, and to enable the

2008 established the National Information Governance Board for Health and Social Care (NIGB) as the body with statutory duty to oversee information governance. One of its functions was to allow the common law duty of confidentiality to be set aside in specific circumstances. However, with the changes that took effect within the NHS from 1st April 2013 (see Chapter 1: section 1.1 NHS England from April 2013), the NIGB's functions for monitoring and improving information governance practice was transferred to the Care Quality Commission[30], which is currently (June 2013) establishing a National Information Governance Committee to oversee this information governance work.

- ***CHALLENGES OF ANALYSING 'BIG DATA'***

Research to date has been driven by hypotheses which are validated by collecting and analysing data on a 'small' scale. In the future the availability of 'big data' (for further details see Chapter 1: section 1.1) with large sample sizes will include structured as well as unstructured data. Some of the challenges of 'big data' will include: (i) unstructured data will include accumulation or errors and spurious correlations; (ii) unstructured data plus large sample size will require heavy use of computing and will likely incur higher costs as a result. In addition, processing very large datasets may create instability of algorithms used to process the data; (iii) a very large sample size will be as a result of amalgamated data from multiple sources at different time points using different technologies. This will create issues of heterogeneity, experimental variations, and statistical biases, and thus will require developing statistical methods that are robust to complex data. Some of the problems that 'big data' will introduce are: 1) by having so much unstructured and structured data spurious correlations will be found between outcomes and unrelated covariates, therefore caution will be required since as the number of data points and spurious variables increase, spurious correlations are also more likely to be observed which may lead to wrong statistical inference and false scientific conclusions (Fan *et al.*, 2012). In particular, Fan and colleagues (Fan *et al.*, 2012)) showed that when there are many spurious variables, σ^2 (variance) is underestimated, which leads further to wrong statistical inferences including model selection or significance tests, and false scientific conclusions; 2) unstructured data gives rise to endogeneity. The term 'endogeneity' (Engle *et al.*, 1983) means that some predictors in linear regression will be correlated with the residuals. Endogeneity happens as a result of selection biases, measurement errors, and omitted variables. In linear regression analysis, assumption is that the residuals are uncorrelated with all the predictors which is important for validity of the model. As the true model is unknown, many variables are collected that are potentially associated to the outcome, and include all variables that are collected in linear regression model in the data collection stage. Some of those collected variables may be correlated with the residuals which would violate one of the linear

ethical use of them for the benefit of individual patients and the public good.
http://www.connectingforhealth.nhs.uk/crdb/igreview/igreview.pdf
[30] www.cqc.org.uk

regression model assumption. Greater the number of covariates that are collected, the validity of the linear regression assumption is violated (that the residuals are uncorrelated with all the predictors). The endogeneity therefore creates statistical biases and causes model selection inconsistency that lead to wrong scientific conclusions; 3) 'big data' are created via aggregating many datasets from multiple and different sub-populations. Each sub-population may exhibit some features not shared by others. A very large sample size will allow hidden patterns associated with small sub-populations to be found. At the moment in the setting of 'small data', data categorised as outliers are hard to model due to small number of observations where outliers are observed. In the 'big data' era, the large sample size will allow to better understand these types of heterogeneous data, potentially shedding light such as exploring the association between certain covariates and rare outcomes (e.g., rare diseases or diseases in small populations) and will allow better understanding why certain treatments benefit a sub-population and harm another sub-population.

6.2 SUMMARY

The electronic data generated for each patient by doctors, nurses, lab technicians, emergency personnel, medical devices within a hospital, or care given at different institutions such as GP, social care, dentists through the use of data integration techniques and algorithms, will be collected and processed electronically to gain mathematical or statistical insights. These insights will provide useful information that will help improve quality of care for a given set of diseases.

In the UK there are many clinical databases that exist and a handful of these are currently registered including multiple clinical database registries for one specialty (Table 6.1.1). With each cohort working independently, it is probable that there is repetition of similar work between cohorts. This is an area of likely waste of resources and clinical data base set up would be better if there were individual specialist databases in the UK that were co-ordinated by one organisation and regulated so that such repetition of work and wasted resources could be minimised. A single database of cancer patients is currently being set up by Public Health England which will deliver 'real-time' cancer data containing detailed clinical information of 11 million historical cancer records going back 30 years (12th June 2013[31]). These types of specialist databases are also required for other clinical specialties that manage chronic conditions.

The availability of large volume of data from databases and data warehouses when linked with other data via data integration as 'big data'[32] increases the value of data (Manyika *et al.*, 2011). However, there is likely to be data management and data analysis challenges in terms of the size of the 'big data'. The volume, the velocity, the variety, and the veracity (Marr, 2013). The

Chapter 6:31 Accessed 15th June 2013 https://www.gov.uk/government/news/worlds-largest-cancer-database-launched-by-phe

Chapter 6:32 'Big data' refers to datasets whose size is beyond the ability of typical database tools to capture, store, manage and analyse.

information extracted from databases and data warehouses may not necessarily be in a format ready for analysis. For example, electronic health records, comprising data from different data sources, image data such as x-rays, multi-media data such as video. Data in these formats cannot be readily analysed and therefore these types of data will need data extraction processes developed that pull out the required information from the data sources and transform these in a format that could then be used for statistical analysis.

Using all types of available data is becoming feasible as 'big data'. Allowing for imprecision in such data may be a positive feature and used as a trade-off although in many situations in medicine accuracy will continue to matter even with 'big data'. Where imprecision trade off with 'big data' is used for instance, in return for relaxing the standards of allowable errors, it would be possible to get hold of much more data. By allowing for some imprecision with 'big data', it is likely to open a window into previously untapped data which may provide greater insights and rigorous accuracy is likely to become less important than getting a grasp of what the data shows over time using previously unused data.

Data analysis, data retrieval, and modeling are other challenges. With 'big data' available future IT requirements will need to take this into consideration as powerful IT resources will be required for storage and processing of 'big data' for data analysis and longitudinal statistical modeling which are often computer intensive. Furthermore, there are many possible sources of error: computer systems and programs to integrate data may have bugs, and the results derived could be based on erroneous data. Results derived will need to be verified by developing algorithms to check the validity of data which is likely to be a challenge due to volume and complexity of data and processes will need to be developed and worked on by data management team of personnel.

In recent years, security has become important and the goal of database security is the protection of data from accidental or intentional threats to its integrity and access. Confidential and sensitive data (such as HIV-1 infected patient data) should be stored in a separate data warehouse where only authorised users are able to get access. The conventional method for securing computer systems is by use of firewalls, authentication mechanisms, and Virtual Private Networks (VPN) that create a protective 'shield' around them. However, such security will have inevitable vulnerabilities and they do not always provide full security as new attacks are continually being adapted to exploit computer system weaknesses. So systems that hold clinical data as 'big data' will need to be continuously monitored on an ongoing basis from such unauthorised attacks.

References

REFERENCES

Aghaizu, A., Brown, A., Nardone, A., Gill, O., et al. (2013) *HIV in the United Kingdom 2013 report: data to end 2012. November 2013 Public Health England*. [Online]. (November). Available from: http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317140300680.

Agresti, A. (2002) *Categorical data analysis*. Second. DJ Balding, P Bloomfield, NAC Cressie, NI Fisher, et al. (eds.). New Jersey, John Wiley & Sons.

Aiuti, F. & D'Offizi, G. (1995) Immunologic and virologic studies in long-term non-progressor HIV infected individuals. NOPHROCO study group. Non progressors HIV + Roman cohort. *Journal of biological regulators and homeostatic agents*. 9 (3), 82–87.

Aldrich, J. (2005) Fisher and regression. *Statistical science*. [Online] 20 (4), 401–417. Available from: doi:10.1214/088342305000000331.

Ando, M., Tsuchiya, K. & Nitta, K. (2012) How to manage HIV-infected patients with chronic kidney disease in the HAART era. *Clinical and experimental nephrology*. [Online] 16 (3), 363–372. Available from: doi:10.1007/s10157-012-0585-7.

AntiretroviralTherapyCohortCollaboration (2008) Life expectancy of individuals on combination antiretroviral therapy in high-income countries: a collaborative analysis of 14 cohort studies. *Lancet*. [Online] 372 (9635), 293–299. Available from: doi:10.1016/S0140-6736(08)61113-7.

Anton, B., Vitetta, L., Cortizo, F. & Sali, A. (2005) Can we delay aging? The biology and science of aging. *Annals of the New York academy of sciences*. [Online] 1057:525–535. Available from: doi:10.1196/annals.1356.040.

Appleby, J. (2012) Rises in healthcare spending: where will it end? *British medical journal*. [Online] 345:1–4. Available from: doi:10.1136/bmj.e7127.

Armitage, P., Berry, G. & Matthews, J. (2001) *Statistical methods in medical research*. 4th edition. New York, Blackwell Scientific Publications.

Aw, D. & Palmer, D. (2011) The origin and implication of thymic involution. *Aging and disease*. 2 (5), 437–443.

Babiker, A., Emery, S., Fätkenheuer, G., Gordin, F., et al. (2013) Considerations in the rationale, design and methods of the Strategic Timing of Antiretroviral Treatment (START) study. *Clinical trials*. 10 (Suppl 1), S5–S36.

Badri, M., Maartens, G., Mandalia, S., Bekker, L., et al. (2006) Cost-effectiveness of highly active antiretroviral therapy in South Africa. *PLOS medicine*. [Online] 3 (1), e4. Available from: doi:10.1371/journal.pmed.0030004.

Baleta, A. (2010) Antiretroviral vaginal gel shows promise against HIV. *Lancet*. [Online] 376 (9738), 320. Available from: doi:10.1016/S0140-6736(10)61123-3.

Barraclough, K., Er, L., Ng, F., Harris, M., et al. (2009) A comparison of the predictive performance of different methods of kidney function estimation in a well-characterized HIV-infected population. *Nephron*. [Online] 111 (1), C39–C48. Available from: doi:10.1159/000178978.

Barre-Sinoussi, F. (2011) *Toward an HIV cure in The New York Times*. [Online] Available from: http://www.nytimes.com/2011/06/04/opinion/04iht-edsinoussi04.html?_r=0.

Barré-Sinoussi, F., Chermann, J., Rey, F., Nugeyre, M., et al. (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*. 220 (4599), 868–871.

Beck, E. & Mandalia, S. (2000a) The cost of HIV treatment and care in England since HAART – part 1. *British journal of sexual medicine*. 27 (1), 19–23.

Beck, E. & Mandalia, S. (2000b) The cost of HIV treatment and care in England since HAART – part 2. *British journal of sexual medicine*. 27 (2), 27–29.

Beck, E., Mandalia, S., Gaudreault, M., Brewer, C., et al. (2004) The cost-effectiveness of highly active antiretroviral therapy, Canada 1991-2001. *AIDS*. 18 (18), 2411–2418.

Beck, E., Mandalia, S., Lo, G., Sharott, P., et al. (2010) for the NPMS-HHC Steering Group. PI boosted or NNRTI as first-line HAART regimens? Lessons from the UK. In: *XVIII International AIDS conference*. 2010 Vienna, Austria.

Beck, E., Mandalia, S., Lo, G., Youle, M., et al. (2008) *for the NPMS-HHC steering group. Use and cost of HIV service provision in UK NPMS-HHC Sites: aggregate analyses January 1996-2006. NPMS-HHC Coordinating and Analytic Centre, St. Stephen's Centre, Chelsea and Westminster Hospital Trust*. (January 1996).

Beck, E., Mandalia, S., Youle, M., Brettle, R., et al. (2008) Treatment outcome and cost-effectiveness of different highly active antiretroviral therapy regimens in the UK (1996-2002). *International journal of STD & AIDS*. [Online] 19 (May), 297–304. Available from: doi:10.1258/ijsa.2007.007236.

Beck, E. & Miners, A. (2001) Effectiveness and efficiency in the delivery of HIV services: economic and related considerations in the effective management of HIV disease. In: B Gazzard, M Johnson, & A Miles (eds.). *Aesculapius medical press*. London, Aesculapius Medical Press. pp. 1–36.

Beck, E., Santas, X. & Delay, P. (2008) Why and how to monitor the cost and evaluate the cost-effectiveness of HIV services in countries. *AIDS*. [Online] 22 (Suppl 1), S75–S85. Available from: doi:10.1097/01.aids.0000327626.77597.fa.

Betjes, M., Langerak, A., van der Spek, A., de Wit, E., et al. (2011) Premature aging of circulating T cells in patients with end-stage renal disease. *Kidney international*. [Online] 80 (2), 208–217. Available from: doi:10.1038/ki.2011.110.

BHIVA Writing Group (2012) British HIV Association guidelines for the treatment of HIV-1-positive adults with antiretroviral therapy 2012. *HIV medicine*. [Online]. 13 Suppl 2 (April). Available from: doi:10.1111/j.1468-1293.2012.01029.x.

Black, N. (1996) Why we need observational studies to evaluate the effectiveness of health care. *British medical journal*. 312 (7040), 1215–1218.

Blankson, J. (2010) Effector mechanisms in HIV-1 infected elite controllers: highly active immune responses? *Antiviral research*. [Online] 85 (1), 295–302. Available from: doi:10.1016/j.antiviral.2009.08.007.

Breyer, M. & Qi, Z. (2010) Better nephrology for mice-and man. *Kidney international*. [Online] 77 (6), 487–489. Available from: doi:10.1038/ki.2009.544.

Broder, S. (2010) The development of antiretroviral therapy and its impact on the HIV-1/AIDS pandemic. *Antiviral research*. [Online] 85 (1), 1–18. Available from: doi:10.1016/j.antiviral.2009.10.002.

Brunner, S., Herndler-Brandstetter, D., Weinberger, B. & Grubeck-Loebenstein, B. (2011) Persistent viral infections and immune aging. *Ageing research reviews*. [Online] 10 (3), 362–369. Available from: doi:10.1016/j.arr.2010.08.003.

Buchholz, V., Neuenhahn, M. & Busch, D. (2011) CD8+ T cell differentiation in the aging immune system: until the last clone standing. *Current opinion in immunology*. [Online] 23 (4), 549–554. Available from: doi:10.1016/j.coi.2011.05.002.

Buehring, B., Kirchner, E., Sun, Z. & Calabrese, L. (2012) The frequency of low muscle mass and its overlap with low bone mineral density and lipodystrophy in individuals with HIV - a pilot study using DXA total body composition analysis. *Journal of clinical densitometry*. 15 (2), 224–232.

Burton, C., Nelson, M., Hay, P., Gazzard, B., et al. (2005) Immunological and virological consequences of patient-directed antiretroviral therapy interruption during chronic HIV-1 infection. *Clinical and experimental immunology*. [Online] 142 (2), 354–361. Available from: doi:10.1111/j.1365-2249.2005.02918.x.

Cao, Y., Qin, L., Zhang, L., Safrit, J., et al. (1995) Virologic and immunologic characterization of long-term survivors of human immune deficiency virus type I infection. *The New England journal of medicine*. 332 (4), 201–207.

Castro, K., Ward, J., Slutsker, L., Buehler, J., et al. (1992) 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *MMWR*. 41 (3), 1–18.

CEBR (2012) *Data equity. Unlocking the value of big data*. [Online]. (April). Available from: http://www.sas.com/offices/europe/uk/downloads/data-equity-cebr.pdf.

Chew, J., Saleem, M., Florkowski, C. & George, P. (2008) Cystatin C-a paradigm of evidence based laboratory medicine. *Clinical biochemist reviews*. 29 (2), 47–62.

Clavel, F., Mansinho, K., Chamaret, S., Guetard, D., et al. (1987) Human immunodeficiency virus type 2 infection associated with AIDS in West Africa. *The New England journal of medicine*. 316 (19), 1180–1185.

Clerici, M., Stocks, N., Zajac, R., Boswell, R., et al. (1989) Detection of three distinct patterns of T helper cell dysfunction in asymptomatic , human immunodeficiency virus-seropositive patients. *The journal of clinical investigation*. 84 (6), 1892–1899.

Cockcroft, D. & Gault, M. (1976) Prediction of creatinine clearance from serum creatinine. *Nephron*. 16 (1), 31–41.

Dalrymple, L., Katz, R., Kestenbaum, B., Shlipak, M., et al. (2011) Chronic kidney disease and the risk of end-stage renal disease versus death. *Journal of general internal medicine*. [Online] 26 (4), 379–385. Available from: doi:10.1007/s11606-010-1511-x.

Dawber, T., Meadors, G. & Moore, F. (1951) Epidemiological approaches to heart disease: the Framingham study. *American journal of public health*. 41:279–286.

Deeks, S. (2011) HIV infection, inflammation, immunosenescence, and aging. *Annual review of medicine*. [Online] 62:141–155. Available from: doi:10.1146/annurev-med-042909-093756.

Deeks, S. & Walker, B. (2007) Human immunodeficiency virus controllers: mechanisms of durable virus control in the absence of antiretroviral therapy. *Immunity*. [Online] 27 (3), 406–416. Available from: doi:10.1016/j.immuni.2007.08.010.

Delgado-Rodríguez, M. & Llorca, J. (2004) Bias. *Journal of epidemiology and community health*. [Online] 58 (8), 635–641. Available from: doi:10.1136/jech.2003.008466.

Department of Health (2012) *The power of information: putting all of us in control of the health and care information we need*. [Online]. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213689/dh_134205.pdf.

Deutsch, E., Duftschmid, G. & Dorda, W. (2010) Critical areas of national electronic health record programs - is our focus correct? *International journal of medical informatics*. [Online] 79 (3), 211–222. Available from: doi:10.1016/j.ijmedinf.2009.12.002.

Diggle, P., Liang, K. & Zeger, S. (1994) *Analysis of longitudinal data*. Oxford, Oxford University Press.

Dilson, J. (1968) *The abascus: a pocket computer*. [Online]. New York, St. Martin's Press. Available from: http://ia600805.us.archive.org/23/items/abacuspocketco00dils/abacuspocketco00dils.pdf.

Dorshkind, K., Montecino-Rodriguez, E. & Signer, R. (2009) The ageing immune system: is it ever too old to become young again? *Nature reviews immunology*. 9 (January), 57–62.

Douek, D., McFarland, R., Keiser, P., Gage, E., et al. (1998) Changes in thymic function with age and during the treatment of HIV infection. *Nature*. [Online] 396 (6712), 690–695. Available from: doi:10.1038/25374.

Douek, D., Roederer, M. & Koup, R. (2009) Emerging concepts in the immunopathogenesis of AIDS. *Annual review of medicine*. [Online] 60:471–484. Available from: doi:10.1146/annurev.med.60.041807.123549.Emerging.

Earley, A., Miskulin, D., Lamb, E., Levey, A., et al. (2012) Estimating equations for glomerular filtration rate in the era of creatinine standardization: a systematic review. *Annals of internal medicine*. 156 (11), 785–795.

Eastburn, A., Scherzer, R., Zolopa, A., Benson, C., et al. (2011) Association of low level viremia with inflammation and mortality in HIV-infected adults. *PLOS one*. [Online] 6 (11), e26320. Available from: doi:10.1371/journal.pone.0026320.

Easterbrook, P. (1994) Non-progression in HIV infection. *AIDS*. 8 (10), 1179–1182.

Edwards, J. (2006) *Case Study: Denmark's achievements with healthcare information exchange*. [Online]. 2006. Gartner industry research. Available from: http://www-03.ibm.com/industries/ca/en/healthcare/files/gartner-case_study-denmarks_achievementswHIE.pdf.

Effros, R., Fletcher, C., Gebo, K., Halter, J., et al. (2008) Aging and infectious diseases: workshop on HIV infection and aging: what is known and future research directions. *Clinical infectious diseases*. [Online] 47 (4), 542–553. Available from: doi:10.1086/590150.

Eisele, E. & Siliciano, R. (2012) Redefining the viral reservoirs that prevent HIV-1 eradication. *Immunity*. [Online] 37 (3), 377–388. Available from: doi:10.1016/j.immuni.2012.08.010.

Emmerson, C. & Frayne, C. (2005) *Public spending. IFS election briefing note no 2*. [Online]. (2). Available from: http://zippy.ifs.org.uk/bns/05ebn2.pdf.

Engle, R., Hendry, D. & Richard, J. (1983) Exogeneity. *Econometrica*. 51 (2), 277–304.

Fan, J., Guo, S. & Hao, N. (2012) Variance estimation using refitted cross-validation in ultrahigh dimensional regression. *Journal of the royal statistical society series B, statistical methodology*. [Online] 74 (1), 37–65. Available from: doi:10.1111/j.1467-9868.2011.01005.x.

Farr, C. (2013) Now a majority of doctors use electronic health records - what does this mean for you? *Venturebeat/health*. [Online] 28 May. Available from: http://venturebeat.com/2013/05/28/now-a-majority-of-doctors-use-electronic-health-records-what-does-this-mean-for-you/.

Fee, E. & Brown, T. (2006) Michael S. Gottlieb and the identification of AIDS. *American journal of public health*. [Online] 96 (6), 982–983. Available from: doi:10.2105/AJPH.2006.088435.

Feinberg, M. (1996) Changing the natural history of HIV disease. *Lancet*. 348239–246.

Fernando, S., Finkelstein, F., Moore, B. & Weissman, S. (2008) Prevalence of chronic kidney disease in an urban HIV infected population. *The American journal of the medical sciences*. [Online] 335 (2), 89–94. Available from: doi:10.1097/MAJ.0b013e31812e6b34.

Fine, D. & Gallant, J. (2013) Nephrotoxicity of antiretroviral agents: is the list getting longer? *Journal of infectious diseases*. [Online] 207 (9), 1349–1351. Available from: doi:10.1093/infdis/jit044.

Friedman, J. (1998) Data mining and statistics: what's the connection? *Computing science and statistics*. [Online] 29 (1), 1–7. Available from: http://docs.salford-systems.com/dm-stat.pdf.

Gaardbo, J., Hartling, H., Gerstoft, J. & Nielsen, S. (2012) Thirty years with HIV infection-nonprogression is still puzzling: lessons to be learned from controllers and long-term nonprogressors. *AIDS research and treatment*. [Online] 2012:1–14. Available from: doi:10.1155/2012/161584.

Gallant, J., Parish, M., Keruly, J. & Moore, R. (2005) Changes in renal function associated with tenofovir disoproxil fumarate treatment, compared with nucleoside reverse-transcriptase inhibitor treatment. *Clinical infectious diseases*. [Online] 40 (8), 1194–1198. Available from: doi:10.1086/428840.

Gallo, R. & Montagnier, L. (2011) HIV's leading men. *IAVI report: newsletter on international AIDS vaccine research*. 15 (3), 13–14.

Gantz, J., Chute, C., Manfrediz, A., Minton, S., et al. (2008) *The diverse and exploding digital universe. An updated forecast of worldwide*. [Online]. Available from: http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf.

Gantz, J. & Reinsel, D. (2010) *The digital universe decade - are you ready?* [Online]. 2009 (May) pp.1–16. Available from: http://uk.emc.com/collateral/analyst-reports/idc-digital-universe-are-you-ready.pdf.

Garets, D. & Davis, M. (2006) Electronic medical records vs. electronic health records: yes, there is a difference: a HIMSS analytics white paper. *Healthcare Information and Management Systems Soc*. [Online]. pp.1–14. Available from: http://www.emrsurvival.com/pdf/IEEPrivacysecurity.pdf.

Gilg, J., Castledine, C., Fogarty, D. & Feest, T. (2011) *UK renal registry 13th annual report (December 2010): chapter 1: UK RRT incidence in 2009: national and centre-specific analyses. Nephron clinical practice*. p.119 Suppl 2:c1–25.

Go, A., Chertow, G., Fan, D., McCulloch, C., et al. (2004) Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *The New England journal of medicine*. [Online] 351 (13), 1296–1305. Available from: doi:10.1056/NEJMoa041031.

Goetz, M. (2013) *Special considerations related to aging in HIV-infected patients*. [Online]. 2013. InPractice. Available from: https://www.inpractice.com/Textbooks/HIV/Management_of_Specific_Populations/ch25_Aging/Chapter-Pages/page-1.aspx [Accessed: 13 March 1BC].

Grabar, S., Selinger-Leneman, H., Abgrall, S., Pialoux, G., et al. (2009) Prevalence and comparative characteristics of long-term nonprogressors and HIV controller patients in the French hospital database on HIV. *AIDS*. [Online] 23 (9), 1163–1169. Available from: doi:10.1097/QAD.0b013e32832b44c8.

Gray, J. (1996) Data management: past, present, and future. *IEEE data engineering bulletin*. 29 (10), 38–46.

Gress, R. & Deeks, S. (2009) Reduced thymus activity and infection prematurely age the immune system. *The journal of clinical investigation*. [Online] 119 (10), 2884–2887. Available from: doi:10.1172/JCI40855.2884.

Gruver, A., Hudson, L. & Sempowski, G. (2007) Immunosenescence of ageing. *Journal of pathology*. [Online] 211:144–156. Available from: doi:10.1002/path.

Haaga, J.G. (2001) *Frontiers in population sciences*. [Online]. (December). Available from: http://www.prb.org/pdf/FrontiersPopulationSciences.pdf.

Haggerty, J., Reid, R., Freeman, G., Starfield, B., et al. (2003) Continuity of care: a multidisciplinary review. *British medical journal*. [Online] 327:1219–1221. Available from: doi:10.1136/bmj.327.7425.1219.

Hall, H., Green, T., Wolitski, R., Holtgrave, D., et al. (2010) Estimated future HIV prevalence, incidence, and potential infections averted in the United States: a multiple scenario analysis. *Journal of acquired immune deficiency syndromes*. [Online] 55 (2), 271–276. Available from: doi:10.1097/QAI.0b013e3181e8f90c.

Han, Y., Lai, J., Barditch-Crovo, P., Gallant, J., et al. (2008) A long-term durable remission with high-dose therapy and autologous stem cell transplant for stage IVB HIV-associated Hodgkins disease. *AIDS*. 22 (4), 541–544.

Hankins, C. & de Zalduondo, B. (2010) Combination prevention: a deeper understanding of effective HIV prevention. *AIDS*. [Online] 24 (Suppl 4), S70–S80. Available from: doi:10.1097/01.aids.0000390709.04255.fd.

Hardy, A. (1991) Characterization of long-term survivors of acquired immunodeficiency syndrome. The long-term survivor collaborative study group. *Journal of acquired immune deficiency syndromes*. 4 (4), 386–391.

Harker, R. (2012) NHS funding and expenditure. *Social and general statistics*. [Online]. Available from: www.parliament.uk/briefing-papers/sn00724.pdf.

Haynes, B., Markert, M., Sempowski, G., Patel, D., et al. (2000) The role of the thymus in immune reconstitution in aging, bone marrow transplantation, and HIV-1 infection. *Annual review of immunology*. 18 (3), 529–560.

Health Protection Agency (2009) *HIV in the United Kingdom: 2009 Report*. [Online]. Available from: http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1259151891830.

Health Protection Agency (2011) *Survey of prevalent HIV infections diagnosed (SOPHID). Data tables 2011*.

Hecht, R., Bollinger, L., Stover, J., McGreevey, W., et al. (2009) Critical choices in financing the response to the global HIV/AIDS pandemic. *Health affairs*. [Online] 28 (6), 1591–1605. Available from: doi:10.1377/hlthaff.28.6.1591.

Hogan, W. & Wagner, M. (1997) Accuracy of data in computer-based patient records. *Journal of the American medical informatics association*. 4 (5), 342–355.

Hollerith, H. (1894) The electrical tabulating machine. *Journal of the royal statistical society*. 57 (4), 678–689.

House Of Commons (2007) *House of commons committee of public accounts. Department of Health: the National Programme for IT in the NHS twentieth report of session 2006-07*. [Online]. (April). Available from: http://www.publications.parliament.uk/pa/cm200607/cmselect/cmpubacc/390/390.pdf.

House Of Commons Health Committee (2007) *Sixth report of session 2006-07: the electronic patient record*. [Online]. (September). Available from: http://www.publications.parliament.uk/pa/cm200607/cmselect/cmhealth/422/422.pdf.

HQIP (2012) *Directory of clinical databases and registers*. [Online]. Available from: http://www.hqip.org.uk/assets/Core-Team/Directory-of-Clinical-Databases-and-Registers-February-2012.pdf.

HSCIC (2013) *Informing better care: our plan for 2013/14*. [Online]. Available from: http://www.hscic.gov.uk/media/11860/HSCIC-business-plan-2013 14/pdf/80305_HSCIC_Business_plan_V1.0.pdf.

Hulstaert, F., Hannet, I., Deneys, V., Munhyeshuli, V., et al. (1994) Age-related changes in human blood lymphocyte subpopulations. II. Varying kinetics of percentage and absolute count measurements. *Clinical immunology and immunopathology*. 70 (2), 152–158.

ICH/GCP (1996) *Guideline for good clinical practice E6(R1)*. [Online]. 1996 (June). Available from: http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/good-clinical-practice.html.

Imami, N., Hardy, G., Burton, C., Pires, A., et al. (2001) Immune responses and reconstitution in HIV-1 infected individuals: impact of anti-retroviral therapy, cytokines and therapeutic vaccination. *Immunology letters*. 79 (1-2), 63–76.

Imami, N., Pires, A., Hardy, G., Wilson, J., et al. (2002) A balanced type 1 / type 2 response is associated with long-term nonprogressive human immunodeficiency virus type 1 infection. *Society*. [Online] 76 (18), 9011–9023. Available from: doi:10.1128/JVI.76.18.9011.

Imami, N., Westrop, S., Grageda, N. & Herasimtschuk, A. (2013) Long-term non-progression and broad HIV-1-specific proliferative T-cell responses. *Frontiers in immunology*. [Online] 4 (1), 1–16. Available from: doi:10.3389/fimmu.2013.00058.

Jamieson, B., Douek, D., Killian, S., Hultin, L., et al. (1999) Generation of functional thymocytes in the human adult. *Immunity*. 10 (5), 569–575.

Johnson, C., Levey, A., Coresh, J., Levin, A., et al. (2004a) Clinical practice guidelines for chronic kidney disease in adults: part I definition, disease stages, evaluation, treatment, and risk factors. *American family physician*. 70 (5), 869–876.

Johnson, C., Levey, A., Coresh, J., Levin, A., et al. (2004b) Clinical practice guidelines for chronic kidney disease in adults: Part II. glomerular filtration rate, proteinuria, and other markers. *American family physician*. 70 (6), 1091–1097.

Joint Specialty Committee on Renal Medicine (2006) Chronic kidney disease in adults: UK guidelines for identification, management and referral. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association*. [Online]. 21 (7). Available from: doi:10.1093/ndt/gfl351.

Justice, A. (2010) HIV and aging: time for a new paradigm. *Current HIV/AIDS reports*. [Online] 7 (2), 69–76. Available from: doi:10.1007/s11904-010-0041-9.

Kahn, J., Marseille, E. & Auvert, B. (2006) Cost-effectiveness of male circumcision for HIV prevention in a South African setting. *PLOS medicine*. [Online] 3 (12), e517. Available from: doi:10.1371/journal.pmed.0030517.

Kanki, P.J. (1999) Human immunodeficiency virus type 2 (HIV-2). *AIDS review*. 1:101–108.

Keyes, K., Utz, R., Robinson, W. & Li, G. (2010) What is a cohort effect? Comparison of three statistical methods for modeling cohort effects in obesity prevalence in the United States, 1971-2006. *Social science & medicine*. [Online] 70 (7), 1100–1108. Available from: doi:10.1016/j.socscimed.2009.12.018.

Kher, U. (1982) A name for the plague. *TIME*. [Online]. Available from: http://web.archive.org/web/20080307015307/http://www.time.com/time/80days/820727.html.

Kierkegaard, P. (2013) eHealth in Denmark: a case study. *Journal of medical systems*. [Online] 37 (6), 1–10. Available from: doi:10.1007/s10916-013-9991-y.

Kincaid, C. (2005) Guidelines for selecting the covariance structure in mixed model analysis. *SUGI*. 30:198–230.

Kirkwood, T. (2008) Understanding ageing from an evolutionary perspective. *Journal of internal medicine*. [Online] 263 (2), 117–127. Available from: doi:10.1111/j.1365-2796.2007.01901.x.

Klein, M. & Miedema, F. (1995) Long-term survivors of HIV-1 infection. *Trends in microbiology*. 3 (10), 386–391.

Komaroff, A. (1979) The variability and inaccuracy of medical data. *Proceedings of the IEEE*. [Online] 67 (9), 1196–1207. Available from: doi:10.1109/PROC.1979.11435.

Kuller, L., Tracy, R., Belloso, W., De Wit, S., et al. (2008) Inflammatory and coagulation biomarkers and mortality in patients with HIV infection. *PLOS medicine*. [Online] 5 (10), e203. Available from: doi:10.1371/journal.pmed.0050203.

Lamb, E., MacKenzie, F. & Stevens, P. (2009) How should proteinuria be detected and measured? *Annals of clinical biochemistry*. [Online] 46:205–217. Available from: doi:10.1258/acb.2009.009007.

Lambotte, O., Boufassa, F., Madec, Y., Nguyen, A., et al. (2005) HIV controllers: A homogeneous group of HIV-1 – infected patients with spontaneous control of viral replication. *Clinical infectious diseases*. 41:1053–1056.

Learmont, J., Tindall, B., Evans, L., Cunningham, A., et al. (1992) Long-term symptomless HIV-1 infection in recipients of blood products from a single donor. *Lancet*. 340 (8824), 863–867.

Leeansyah, E., Cameron, P., Solomon, A., Tennakoon, S., et al. (2013) Inhibition of telomerase activity by human immunodeficiency virus (HIV) nucleos(t)ide reverse transcriptase inhibitors: a potential factor contributing to HIV-associated accelerated aging. *Journal of infectious diseases*. [Online] 207 (7), 1157–1165. Available from: doi:10.1093/infdis/jit006 [Accessed: 21 November 2013].

Lefrère, J.J., Morand-Joubert, L., Mariotti, M., Bludau, H., et al. (1997) Even individuals considered as long-term nonprogressors show biological signs of progression after 10 years of human immunodeficiency virus infection. *Blood*. 90 (3), 1133–1140.

Levey, A., Bosch, J., Lewis, J. & Greene, T. (1999) A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. *Annals of internal medicine*. 130 (6), 461–470.

Levey, A., Coresh, J., Greene, T., Stevens, L., et al. (2006) Using standardized serum creatinine values in the Modification of Diet in Renal Disease study equation for estimating glomerular filtration rate. *Annals of internal medicine*. 145 (4), 247–254.

Levey, A., de Jong, P., Coresh, J., El Nahas, M., et al. (2011) The definition, classification, and prognosis of chronic kidney disease: a KDIGO controversies conference report. *Kidney international*. [Online] 80 (1), 17–28. Available from: doi:10.1038/ki.2010.483.

Levey, A., Stevens, L. & Coresh, J. (2009) Conceptual model of CKD: applications and implications. *American journal of kidney diseases*. [Online] 53 (3, Suppl 3), S4–S16. Available from: doi:10.1053/j.ajkd.2008.07.048.

Levey, A., Stevens, L., Schmid, C., Zhang, Y., et al. (2009) A new equation to estimate glomerular filtration rate. *Annals of internal medicine*. 150 (9), 604–612.

Lindau, S., Schumm, L., Laumann, E., Levinson, W., et al. (2007) A study of sexuality and health among older adults in the United States. *The New England journal of medicine*. [Online] 357 (8), 762–774. Available from: doi:10.1056/NEJMoa067423.

Linsk, N. (2000) HIV among older adults: age-specific issues in prevention and treatment. *The AIDS reader*. 10 (7), 430–440.

Machin, D. & Campbell, M. (2005) *Design of studies for medical research*. Chichester, John Wiley & Sons Ltd.

Maini, M., Gilson, R., Chavda, N., Gill, S., et al. (1996) Reference ranges and sources of variability of CD4 counts in HIV-seronegative women and men. *Genitourinary medicine*. 72:27–31.

Mandalia, S., Mandalia, R., Lo, G., Chadborn, T., et al. (2010) for the NPMS-HHC steering group. Rising population cost for treating people living with HIV in the UK, 1997-2013. *PLOS one*. [Online] 5 (12), e15677. Available from: doi:10.1371/journal.pone.0015677.

Mandalia, S., Westrop, S., Beck, E., Nelson, M., et al. (2012) Are long-term non-progressors very slow progressors? Insights from the Chelsea and Westminster HIV cohort, 1988-2010. *PLOS one*. [Online] 7 (2), e29844. Available from: doi:10.1371/journal.pone.0029844.

Manghani, K. (2011) Quality assurance: importance of systems and standard operating procedures. *Perspectives in clinical research*. [Online] 2 (1), 34–37. Available from: doi:10.4103/2229-3485.76288.

Manyika, J., Chui, M., Brown, B., Bughin, J., et al. (2011) *McKinsey Global Institute. Big data: the next frontier for innovation, competition, and productivity*. (May).

Markovitz, D. (1993) Infection with the human immunodeficiency virus type 2. *Annals of internal medicine*. 118 (3), 211–218.

Marr, B. (2013) *Is this the ultimate definition of "big data"?* [Online]. 2013. SmartData Collective. Available from: http://smartdatacollective.com/bernardmarr/128486/big-data-ultimate-definition.

Masiá, M., Padilla, S., Alvarez, D., López, J.C., et al. (2012) Risk, predictors, and mortality associated with non-AIDS events in newly diagnosed HIV-infected patients: role of antiretroviral therapy. *AIDS*. [Online] 27 (2), 181–189. Available from: doi:10.1097/QAD.0b013e32835a1156.

Maughan, A. (2010) Six reasons why the NHS national programme for IT failed. Opinion. *ComputerWeekly.com*. [Online]. Available from: http://www.computerweekly.com/opinion/Six-reasons-why-the-NHS-National-Programme-for-IT-failed.

McCarthy, G. & Mercey, D. (1994) The changing clinical features of HIV-1 infection in the United Kingdom. Communicable disease report. *CDR review*. 4 (5), R53–R58.

Miettinen, O.S. & Cook, E.F. (1981) Confounding: essence and detection. *American journal of epidemiology*. 114 (4), 593–603.

Migueles, S. & Connors, M. (2010) Long-term nonprogressive disease among untreated HIV-infected individuals: clinical implications of understanding immune control of HIV. *JAMA*. [Online] 304 (2), 194–201. Available from: doi:10.1001/jama.2010.925.

Migueles, S., Osborne, C., Royce, C., Compton, A., et al. (2008) Lytic granule loading of CD8+ T cells is required for HIV-infected cell elimination associated with immune control. *Immunity*. [Online] 29 (6), 1009–1021. Available from: doi:10.1016/j.immuni.2008.10.010.

Migueles, S., Sabbaghian, M., Shupert, W., Bettinotti, M., et al. (2000) HLA B*5701 is highly associated with restriction of virus replication in a subgroup of HIV-infected long term nonprogressors. *Proceedings of the national academy of sciences of the United States of America*. [Online] 97 (6), 2709–2714. Available from: doi:10.1073/pnas.050567397.

Mills, E.J., Bakanda, C., Birungi, J., Chan, K., et al. (2011) Life expectancy of persons receiving combination antiretroviral therapy in low-income countries: a cohort analysis from Uganda. *Annals of internal medicine*. 155 (4), 209–216.

MMWR (1992) *1993 Revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults*. p.41(RR–17).

MMWR (1981) CDC pneumocystis pneumonia - Los Angeles. *MMWR*. 30 pp.250–252.

Mocroft, A., Kirk, O., Gatell, J., Reiss, P., et al. (2007) Chronic renal failure among HIV-1 infected patients. *AIDS*. [Online] 21 (9), 1119–1127. Available from: doi:10.1097/QAD.0b013e3280f774ee.

Mocroft, A., Kirk, O., Reiss, P., De Wit, S., et al. (2010) Estimated glomerular filtration rate, chronic kidney disease and antiretroviral drug use in HIV-positive patients. *AIDS*. [Online] 24 (11), 1667–1678. Available from: doi:10.1097/QAD.0b013e328339fe53.

Montagnier, L., Chermann, J., Barre-Sinoussi, F., Klatzmann, D., et al. (1984) Lymphadenopathy associated virus and its etiological role in AIDS. *Retroviruses in human lymphoma/leukemia: Proceedings of the 15th international symposium of the Princess Takamatsu cancer research fund, Tokyo*. 15:319–331.

Montaner, J., Lima, V., Barrios, R., Yip, B., et al. (2010) Association of highly active antiretroviral therapy coverage, population viral load, and yearly new HIV diagnoses in British Columbia, Canada: a population-based study. *Lancet*. [Online] 376 (9740), 532–539. Available from: doi:10.1016/S0140-6736(10)60936-1.

Mosley, M. (2007) *DAMA-DMBOK guide TM (data management body of knowledge)*. [Online]. (November). Available from: http://www.dama.org/files/public/DI_DAMA_DMBOK_Guide_Presentation_2007.pdf.

Murray, C.J., Richards, M. a, Newton, J.N., Fenton, K. a, et al. (2013) UK health performance: findings of the global burden of disease study 2010. *Lancet*. [Online] 381 (9871), 997–1020. Available from: doi:10.1016/S0140-6736(13)60355-4.

National Kidney Foundation (2002) K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *American journal of kidney diseases*. 39 (Suppl 1), 266.

Newton, J., Graham, J., McLoughlin, K., Moore, A., et al. (2000) Change in primary health care: new wine in old bottles? *Journal of management in medicine*. 14 (1), 37–47.

NHS England (2013) *Putting patients first*. [Online]. Available from: http://www.england.nhs.uk/wp-content/uploads/2013/04/ppf-1314-1516.pdf.

NHS Information Governance (2013) *Guidelines on use of encryption to protect person identifiable and sensitive information*. [Online]. (1). Available from: http://www.connectingforhealth.nhs.uk/systemsandservices/infogov/security/encryptionguide.pdf.

NRES (2009) *NHS defining research*. [Online]. Available from: http://www.nres.nhs.uk/applications/guidance/research-guidance/?entryid62=66985.

Palella, F., Delaney, K., Moorman, A., Loveless, M., et al. (1998) Human immunodifficiency virus. *The New England journal of medicine*. 338 (13), 853–860.

Pantaleo, G. & Fauci, A.S. (1996) Immunopathogenesis of HIV infection. *Annual review of microbiology*. 50:825–854.

Pantaleo, G., Menzo, S., Vaccarezza, M., Graziosi, C., et al. (1995) Studies in subjects with long-term nonprogressive human immunodeficiency virus infection. *The New England journal of medicine*. [Online] 332 (4), 209–216. Available from: doi:10.1056/NEJM199501263320402.

Pantalone, D., Bimbi, D. & Parsons, J. (2008) Motivations for the recreational use of erectile enhancing medications in urban gay and bisexual men. *Sexually transmitted infections*. [Online] 84 (6), 458–462. Available from: doi:10.1136/sti.2008.031476.Motivations.

Papanikolaou, P.N., Christidi, G.D. & Ioannidis, J.P.A. (2006) Comparisons of evidence on harms of medical interventions in randomized and nonrandomized studies. *CMAJ*. 174 (5), 635–641.

Parmar, M.S. (2002) Chronic renal disease. *British medical journal*. 325:85–90.

Pearce, N. (1999) Epidemiology as a population science. *International journal of epidemiology*. 28:S1015–S1018.

Pearce, N. (1996) Traditional epidemiology, modern epidemiology and public health. *American journal of public health*. [Online] 86 (5), 676–683. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9378189.

Pereyra, F., Jia, X., McLaren, P., Telenti, A., et al. (2010) The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science*. [Online] 330 (6010), 1551–1557. Available from: doi:10.1126/science.1195271 [Accessed: 18 July 2011].

Poate, D., Balogun, P. & Attawell, K. (2009) *Evaluation team. 25th Meeting of the UNAIDS programme coordinating Board UNAIDS second independent evaluation 2002-2008 final report*. [Online]. (December). Available from: http://www.government.nl/documents-and-publications/reports/2009/10/02/unaids-second-independent-evaluation-2002-2008.html.

Popovic, M., Sarngadharan, M., Read, E. & Gallo, R. (1984) Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. *Science*. 224 (4648), 497–500.

Porta, M. (2008) *A dictionary of epidemiology*. Oxford University Press.

Rao, S., Filippone, E., Nicastri, A., Landesman, S., et al. (1984) Associated focal and segmental glomerulosclerosis in the acquired immunodeficiency syndrome. *The New England journal of medicine*. 310 (11), 669–673.

Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., Kaewkungwal, J., et al. (2009) Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *The New England journal of medicine*. 361 (23), 2209–2220.

Robertshaw, M., Lai, K.N. & Swaminathan, R. (1989) Prediction of creatinine clearance from plasma creatinine: comparison of five formulae. *British journal of clinical pharmacology*. 28 (3), 275–280.

Rowe, J.W., Andres, R., Tobin, J.D., Norris, a H., et al. (1976) The effect of age on creatinine clearance in men: a cross-sectional and longitudinal study. *Journal of gerontology*. 31 (2), 155–163.

Rubin, D.B. (1976) Inference and missing data. *Biometrika*. [Online] 63 (3), 581. Available from: doi:10.2307/2335739.

Rutherford, G., Lifson, A., Hessol, N., Darrow, W., et al. (1990) Course of HIV-I infection in a cohort of homosexual and bisexual men: an 11 year follow up study. *British medical journal*. 301:1183–1188.

Ryom, L., Mocroft, A., Kirk, O., Worm, S.W., et al. (2013) Association between antiretroviral exposure and renal impairment among HIV-positive persons with normal baseline renal function: the D:A:D study. *Journal of infectious diseases*. [Online] 207 (9), 1359–1369. Available from: doi:10.1093/infdis/jit043.

Severe, P., Juste, M.A.J., Ambroise, A., Eliacin, L., et al. (2010) Early versus standard antiretroviral therapy for HIV-infected adults in Haiti. *The New England journal of medicine*. [Online] 363 (3), 257–265. Available from: doi:10.1056/NEJMoa0910370.

Shlipak, M., Sarnak, M., Katz, R., Fried, L., et al. (2005) Cystatin C and the risk of death and cardiovascular events among elderly persons. *The New England journal of medicine*. 352 (20), 2049–2060.

Sighem, A., Kesselring, A., Gras, L., Prins, J., et al. (2013) Risk of non-AIDS defining events amongst HIV+ patients not yet on ART. In: *20th conference on retroviruses and opportunistic infections*. 2013 Atlanta.

Smith, R.D., Delpech, V.C., Brown, A.E. & Rice, B.D. (2010) HIV transmission and high rates of late diagnoses among adults aged 50 years and over. *AIDS*. [Online] 24 (13), 2109–2115. Available from: doi:10.1097/QAD.0b013e32833c7b9c.

Soumerai, S. & Avery, T. (2010) Don't repeat the UK's electronic health records failure. *The Huffington Post*. [Online] 12 January. Available from: http://www.huffingtonpost.com/stephen-soumerai/dont-repeat-the-uks-elect_b_790470.html.

Spencer, A. (2013) *Hospital episode statistics (HES): improving the quality and value of hospital data*. [Online]. Available from: http://www.hscic.gov.uk/media/1594/Executive-Summary---Hospital-Episode-Statistics-HES-improving-the-quality-and-value-of-hospital-data/pdf/HES_discussion_EXEC_SUMMARY_4pp_0411.pdf.

Spits, H. (2002) Development of alphabeta T cells in the human thymus. *Nature reviews immunology*. [Online] 2 (10), 760–772. Available from: doi:10.1038/nri913.

Stevens, L., Coresh, J., Feldman, H., Greene, T., et al. (2007) Evaluation of the modification of diet in renal disease study equation in a large diverse population. *Journal of the American society of nephrology*. [Online] 18 (10), 2749–2757. Available from: doi:10.1681/ASN.2007020199.

Stevens, L., Coresh, J., Greene, T. & Levey, A. (2006) Assessing kidney function-measured and estimated glomerular filtration rate. *The New England journal of medicine*. [Online] 354 (23), 2473–2483. Available from: doi:10.1056/NEJMra054415.

Stevens, L., Lynm, C. & Glass, R. (2002) Basic science research. *JAMA*. [Online] 287 (13), 1754–1754. Available from: doi:10.1001/jama.287.13.1754.

Tang, H., Kuhen, K. & Wong Staal, F. (1999) Lentivirus replication and regulation. *Annual review of genetics*. [Online] 33:133–170. Available from: doi:10.1146/annurev.genet.33.1.133.

Thompson, M., Aberg, J., Hoy, J., Benson, C., et al. (2012) Antiretroviral treatment of adult HIV infection 2012 recommendations of the international antiviral society - USA panel. *JAMA*. 308 (4), 387–402.

Tien, P.C., Choi, A.I., Zolopa, A.R., Benson, C., et al. (2010) Inflammation and mortality in HIV-infected adults: analysis of the FRAM study cohort. *Journal of acquired immune deficiency syndromes*. [Online] 55 (3), 316–322. Available from: doi:10.1097/QAI.0b013e3181e66216.

Tugume, S., Piwowar, E., Lutalo, T., Mugyenyi, P., et al. (1995) Hematological reference ranges among healthy Ugandans. *Microbiology*. 2 (2), 233–235.

UNAIDS (2007) *Guidelines on protecting the confidentiality and security of HIV information: proceedings from a workshop*. [Online]. (May 2006) pp.1–61. Available from: http://www.hivlawandpolicy.org/sites/www.hivlawandpolicy.org/files/confidentiality_security_interim_guidelines_15may2007_en.pdf.

UNAIDS (2005) *The "Three Ones" in action: where we are and where we go from here*. [Online]. Available from:

https://www.unfpa.org/webdav/site/global/shared/documents/publications/2005/3onesinaction_en.pdf.

UNAIDS (2004) *"Three Ones" key principles*. [Online]. pp.1–4. Available from: http://www.unaids.org/en/media/unaids/contentassets/dataimport/una-docs/three-ones_keyprinciples_en.pdf.

UNAIDS Report on the Global AIDS Epidemic (2013) *Global report: UNAIDS report on the global AIDS epidemic 2013*. [Online]. Available from: http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/UNAIDS_Global_Report_2013_en.pdf.

UNAIDS World AIDS Day Report (2012) *Global report: UNAIDS report on the global AIDS epidemic 2012*. [Online]. Available from: http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2012/gr2012/JC2417_GR 2012_Annexes_en.pdf.

Vance, D.E., McGuinness, T., Musgrove, K., Orel, N.A., et al. (2011) Successful aging and the epidemiology of HIV. *Clinical interventions in aging*. [Online] 6:181–192. Available from: doi:10.2147/CIA.S14726.

Vento, S., Lanzafame, M., Malena, M., Tositti, G., et al. (2004) Can we really identify HIV-1 long-term nonprogressors? *Journal of acquired immune deficiency syndromes*. 37 (1), 1218.

VonElm, E., Altman, D., Egger, M., Pocock, S., et al. (2007) The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Preventive medicine*. [Online] 45 (4), 247–251. Available from: doi:10.1016/j.ypmed.2007.08.012.

Walker, B. (2007) Elite control of HIV infection: implications for vaccines and treatments. *Topics in HIV medicine*. 15 (4), 134–136.

Walker, B.D. & Yu, X.G. (2013) Unravelling the mechanisms of durable control of HIV-1. *Nature reviews immunology*. [Online] 13:487–498. Available from: doi:10.1038/nri3478.

Westrop, S.J., Qazi, N. a, Pido-Lopez, J., Nelson, M.R., et al. (2009) Transient nature of long-term nonprogression and broad virus-specific proliferative T-cell responses with sustained thymic output in HIV-1 controllers. *PLOS one*. [Online] 4 (5), e5474. Available from: doi:10.1371/journal.pone.0005474.

Williams, I., Churchill, D., Anderson, J., Boffito, M., et al. (2012) British HIV association guidelines for the treatment of HIV-1-positive adults with antiretroviral therapy 2012. *HIV medicine*. [Online] 13 (Suppl 2), 1–85. Available from: doi:10.1111/j.1468-1293.2012.01029.

World Health Organisation (2010a) *Antiretroviral therapy for HIV infection in adults and adolescents, Recommendations for a public health approach, 2010 revision*. [Online]. Available from: http://whqlibdoc.who.int/publications/2010/9789241599764_eng.pdf.

World Health Organisation (2011a) *European centre for disease prevention and control/WHO regional office for europe. HIV/AIDS surveillance in Europe 2010*. [Online]. Available from: http://www.euro.who.int/__data/assets/pdf_file/0006/154554/e96000.pdf.

World Health Organisation (2010b) *Health statistics and health information systems: definition of an older or elderly person*. [Online]. 2010. Available from: http://www.who.int/healthinfo/survey/ageingdefnolder/en/.

World Health Organisation (2011b) *HIV treatment reaching 6.6 million people, but majority still in need. WHO embarks on a new HIV strategy to boost further progress in 2011-2015*. [Online]. 2011. Available from: http://www.who.int/mediacentre/news/releases/2011/hivtreatement_20110603/en/ [Accessed: 23 April 2012].

Wright, S., Petoumenos, K., Boyd, M., Carr, A., et al. (2013) Ageing and long-term CD4 cell count trends in HIV-positive patients with 5 years or more combination antiretroviral therapy experience. *HIV medicine*. [Online] 14 (4), 208–216. Available from: doi:10.1111/j.1468-1293.2012.01053.x.

Young, J., Schäfer, J., Fux, C. a, Furrer, H., et al. (2012) Renal function in patients with HIV starting therapy with tenofovir and either efavirenz, lopinavir or atazanavir. *AIDS*. [Online] 26 (5), 567–575. Available from: doi:10.1097/QAD.0b013e32834f337c.

Zeger, S., Liang, K. & Albert, P. (1988) Models for longitudinal data: a generalized estimating equation approach. *Biometrics*. 44 (4), 1049–1060.

Zhu, T., Korber, B.T., Nahmias, a J., Hooper, E., et al. (1998) An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature*. [Online] 391 (6667), 594–597. Available from: doi:10.1038/35400.

Zolopa, A., Ortiz, R., Sax, P., Brar, I., et al. (2013) Comparative study of tenofovir alafenamide vs tenofovir disoproxil fumarate, each with elvitegravir, cobicistat, and emtricitabine, for HIV treatment. In: *20th conference on retroviruses and opportunistic infections*. 2013 Atlanta.

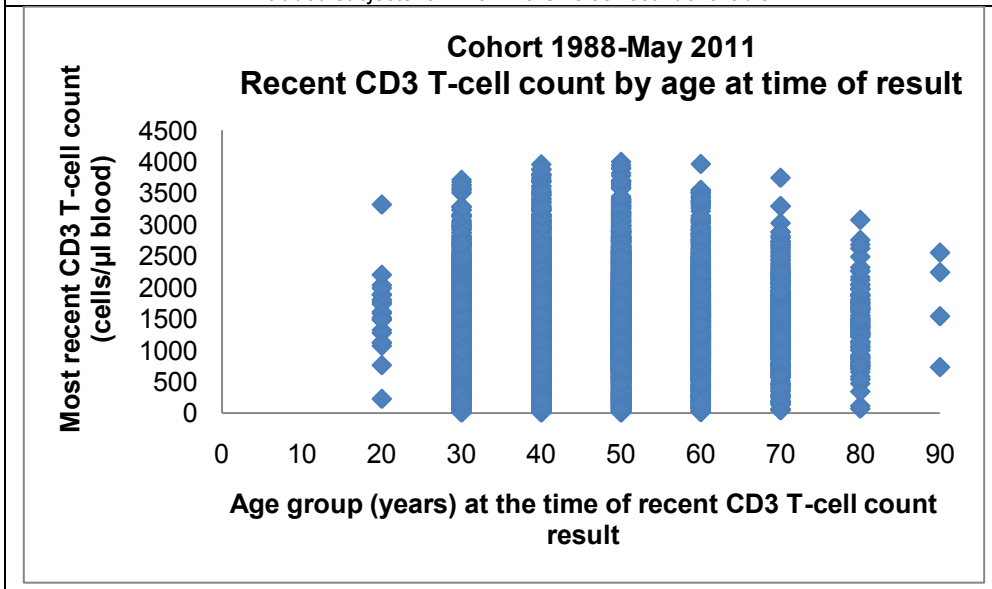
Appendix

APPENDIX

A 4.1 FIGURES SHOWING DISTRIBUTION OF IMMUNOLOGICAL PARAMETERS BY AGE AND COHORT STRATA

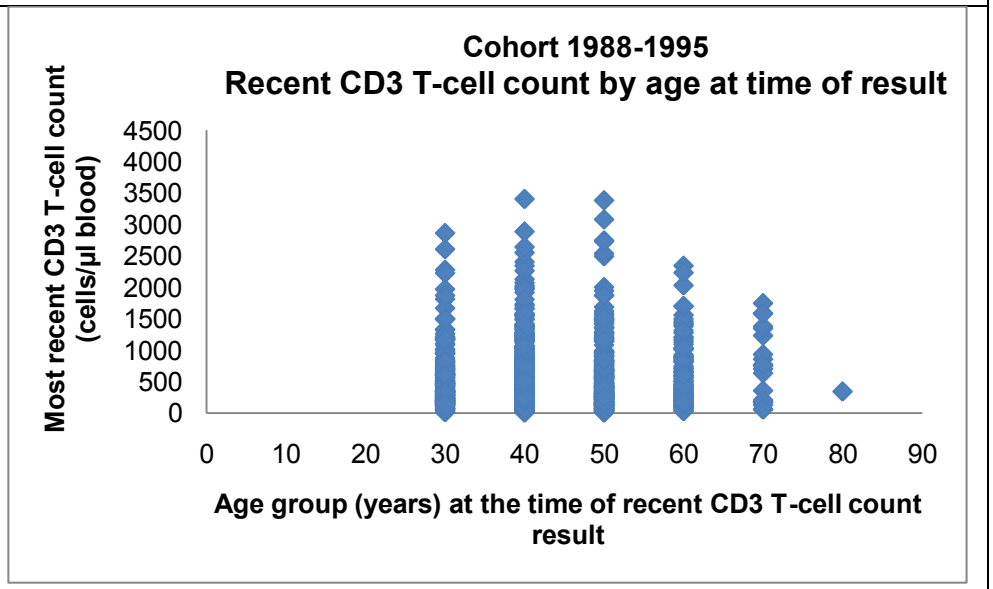
Figure A4.1.1: Distribution CD3 T-Cell Count by Age at Time of Result by Cohort Strata

Patients who attended for HIV care
1988 – May 2011
N=11,699
Excluded subjects for whom no CD3 cell count available



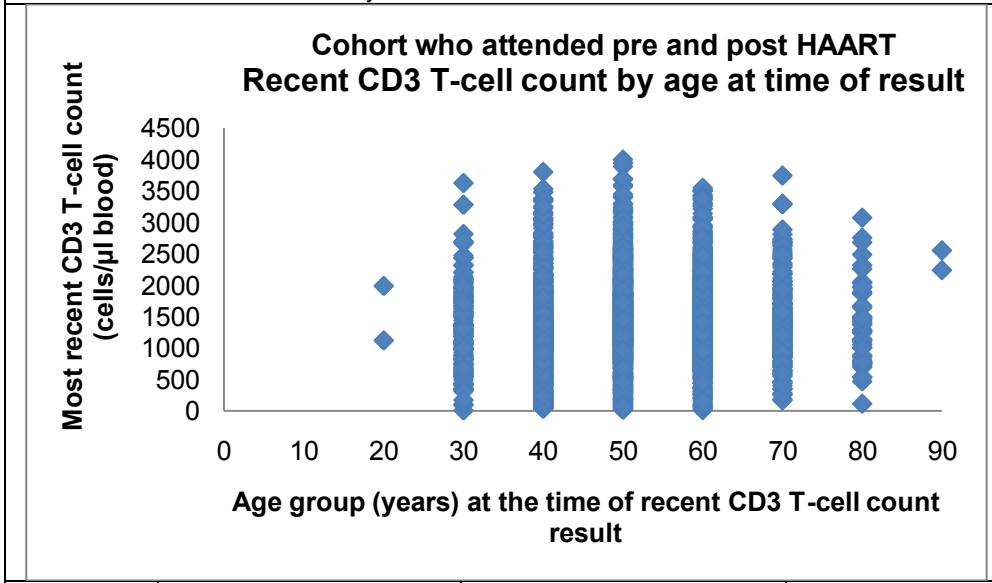
	Total=11,699 n	Median (IQR)	p-value
Age			
18-20	21	1586(1274 to 1880)	<0.001
21-30	1427	1372(961 to 1791)	
31-40	4058	1310(892 to 1721)	
41-50	4021	1370(978 to 1798)	
51-60	1615	1392(958 to 1818)	
61-70	457	1334(980 to 1742)	
71-80	96	1373(961 to 1743)	
81-90	4	1890(1135 to 2397)	

Patients who attended for HIV care
1988 – 1995 only
N=1,228
Excluded subjects for whom no CD3 cell count available



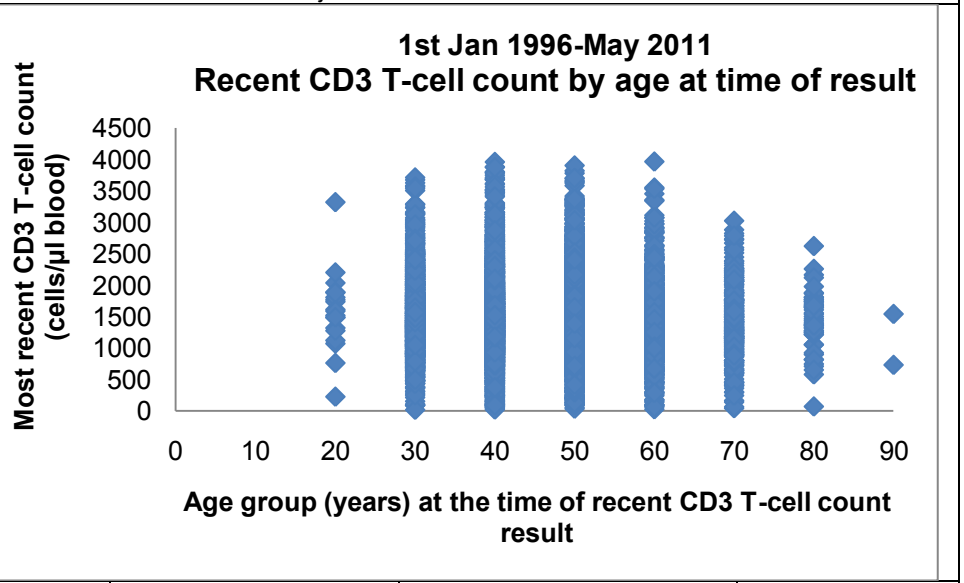
	Total=1,228 n	Median (IQR)	p-value
Age			
18-20	-	-	0.342
21-30	169	426(216 to 728)	
31-40	511	385(181 to 729)	
41-50	318	393(215 to 737)	
51-60	108	444(229 to 869)	
61-70	21	750(181 to 1230)	
71-80	1	340	
81-90	-	-	

Patients who attended for HIV care
pre HAART and post HAART
N=3,092
Excluded subjects for whom no CD3 cell count available



	Total=3,092 n	Median (IQR)	p-value
Age			
18-20	2	1120,1988	<0.001
21-30	148	1313(831 to 1820)	
31-40	734	1236(781 to 1666)	
41-50	1231	1432(1036 to 1882)	
51-60	719	1460(1028 to 1932)	
61-70	210	1369(1030 to 1703)	
71-80	46	1273(852 to 1864)	
81-90	2	2240,2254	

Patients who attended for HIV care
post HAART only
N=7,479
Excluded subjects for whom no CD3 cell count available

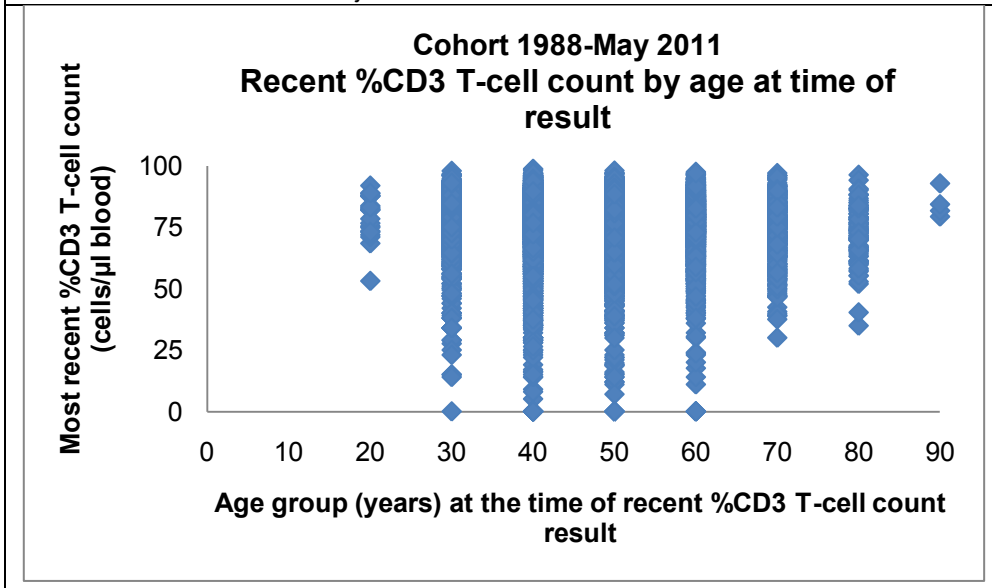


	Total=7,479 n	Median (IQR)	p-value
Age			
18-20	19	1586(1274 to 1880)	0.033
21-30	1110	1448(1124 to 1842)	
31-40	2813	1426(1089 to 1804)	
41-50	2472	1425(1082 to 1806)	
51-60	788	1405(1018 to 1765)	
61-70	226	1339(990 to 1776)	
71-80	49	1440(1216 to 1742)	
81-90	2	729,1540	

Figure A4.1.2: Distribution %CD3 T-Cell Count by Age at Time of Result by Cohort Strata

Patients who attended for HIV care
1988 – May 2011
N=11,766

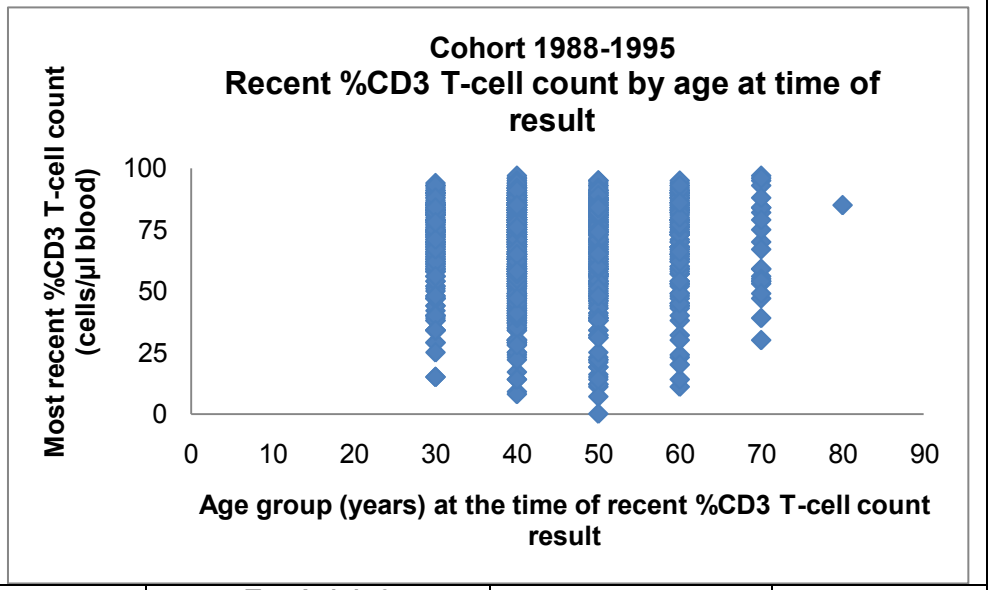
Excluded subjects for whom no %CD3 cell count available



	Total=11,766 n	Median (IQR)	p-value
Age			
18-20	21	82(73 to 88)	<0.001
21-30	1438	81(74 to 86)	
31-40	4075	80(73 to 85)	
41-50	4043	78(71 to 84)	
51-60	1625	77(71 to 83)	
61-70	463	76(68 to 83)	
71-80	97	74(66 to 81)	
81-90	4	83(81 to 89)	

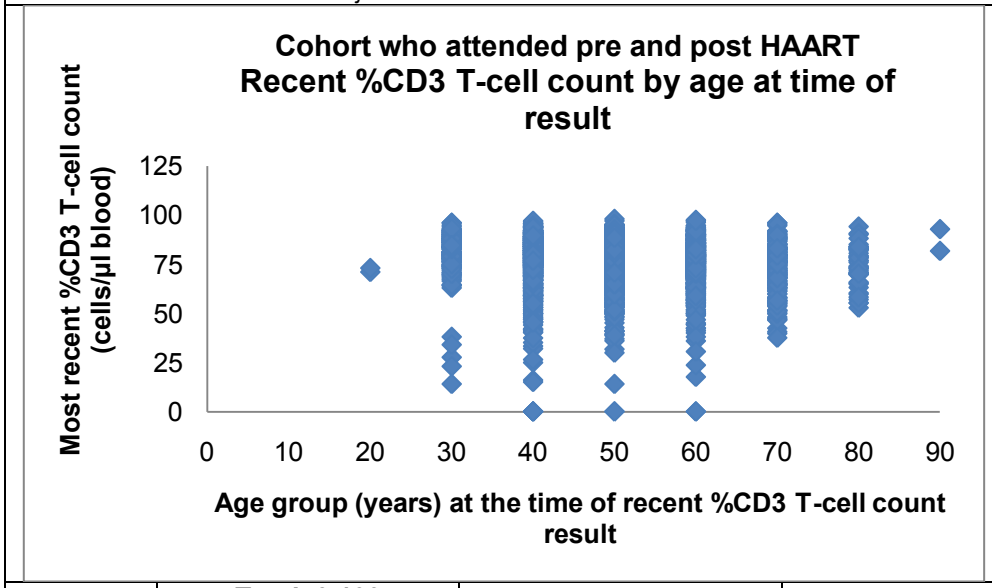
Patients who attended for HIV care
1988 – 1995 only
N=1,156

Excluded subjects for whom no %CD3 T-cell count available



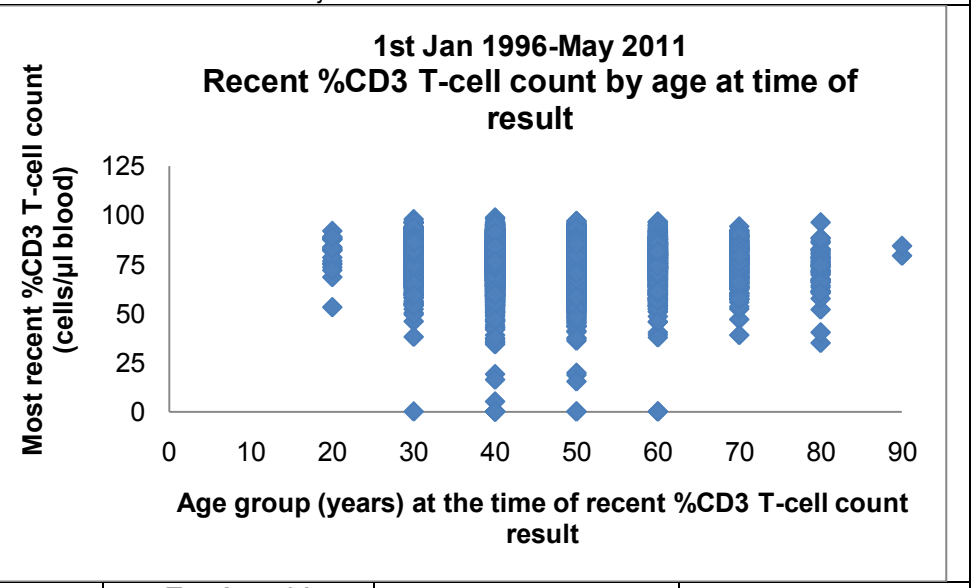
	Total=1,156 n	Median (IQR)	p-value
Age			
18-20	-	-	0.855
21-30	178	73(60 to 82)	
31-40	521	72(58 to 83)	
41-50	321	72(57 to 82)	
51-60	112	73(59 to 84)	
61-70	23	67(54 to 84)	
71-80	1	85	
81-90	-	-	

Patients who attended for HIV care
pre HAART and post HAART
N=3,109
Excluded subjects for whom no %CD3 cell count available



	Total=3,109 n	Median (IQR)	p-value
Age			
18-20	2	72(71 to 73)	<0.001
21-30	148	83(77 to 88)	
31-40	736	80(73 to 86)	
41-50	1240	79(73 to 84)	
51-60	723	77(71 to 83)	
61-70	212	76(67 to 82)	
71-80	46	74(70 to 83)	
81-90	2	82,93	

Patients who attended for HIV care
post HAART only
N=7,501
Excluded subjects for whom no %CD3 cell count available

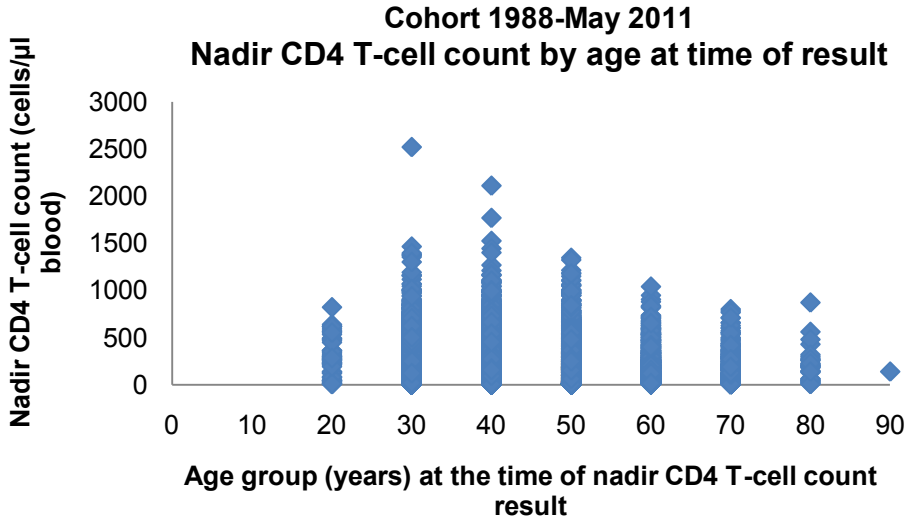


	Total=7,501 n	Median (IQR)	p-value
Age			
18-20	19	82(75 to 88)	<0.001
21-30	1112	81(76 to 86)	
31-40	2818	80(74 to 85)	
41-50	2482	78(71 to 83)	
51-60	790	77(72 to 83)	
61-70	228	77(69 to 83)	
71-80	50	74(66 to 79)	
81-90	2	79,84	

Figure A4.1.3: Distribution Nadir CD4 T-Cell Count by Age at Time of Result by Cohort Strata

Patients who attended for HIV care
1988 – May 2011
N=13,813

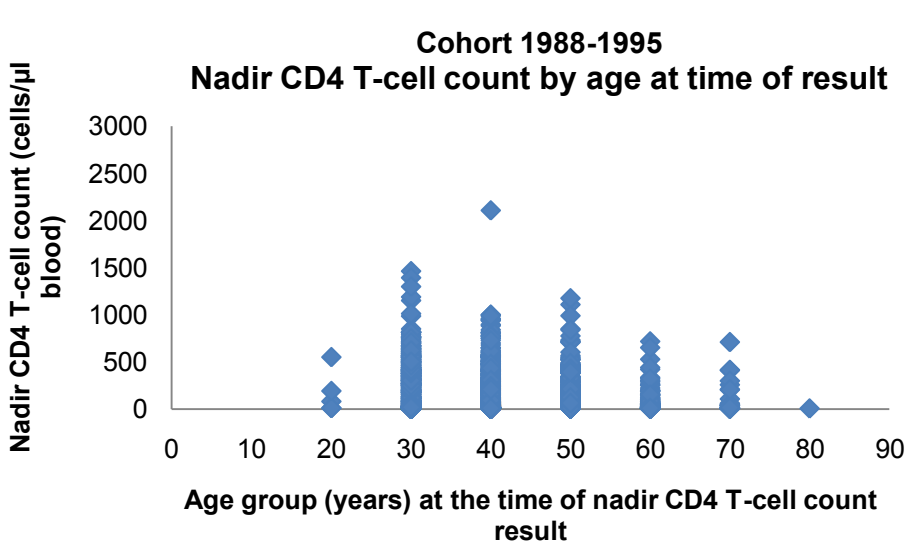
Excluded subjects for whom no nadir CD4 T- cell count available



	Total=13,813 n	Median (IQR)	p-value
Age			
18-20	55	266(120 to 467)	<0.001
21-30	2753	251(104 to 406)	
31-40	6115	175(47 to 311)	
41-50	3622	155(40 to 291)	
51-60	992	155(40 to 291)	
61-70	244	154(49 to 277)	
71-80	31	192(60 to 267)	
81-90	1	137	

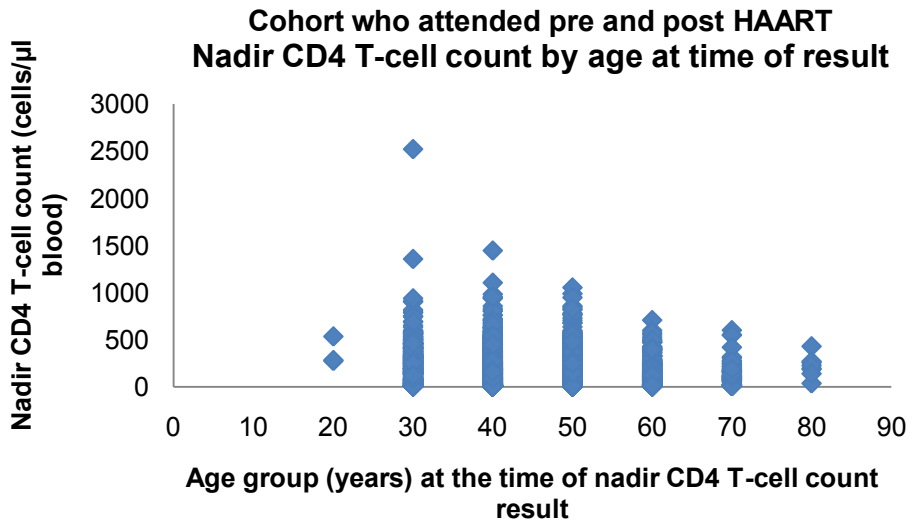
Patients who attended for HIV care
1988 – 1995 only
N=2,750

Excluded subjects for whom no nadir CD4 T-cell count available



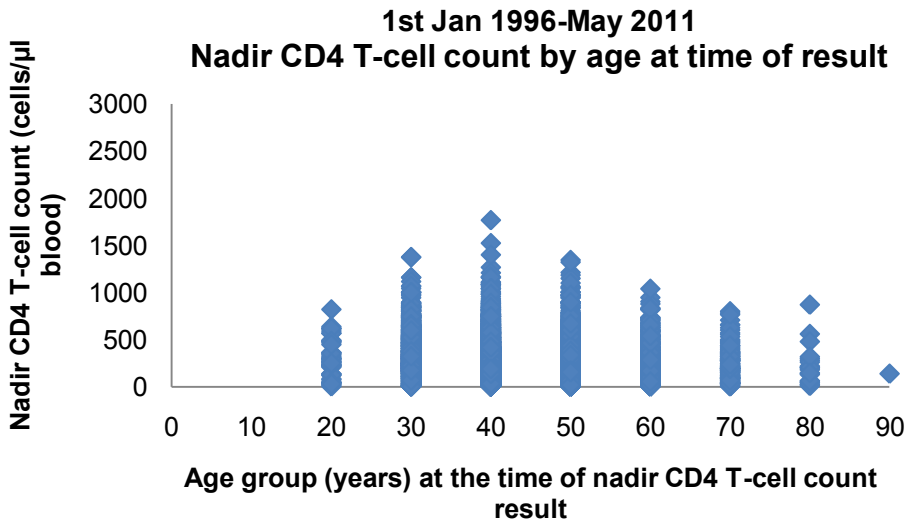
	Total=2,750 n	Median (IQR)	p-value
Age			
18-20	5	79(9 to 190)	<0.001
21-30	682	150(13 to 378)	
31-40	1170	25(6 to 154)	
41-50	673	18(6 to 88)	
51-60	183	20(8 to 60)	
61-70	36	15(6 to 54)	
71-80	1	4	
81-90	-	-	

Patients who attended for HIV care
pre HAART and post HAART
N=3,445
Excluded subjects for whom no nadir CD4 T- cell count available



	Total=3,445 n	Median (IQR)	p-value
Age			
18-20	3	281(272 to 532)	<0.001
21-30	447	186(48 to 330)	
31-40	1570	126(32 to 238)	
41-50	1030	120(35 to 206)	
51-60	322	111(35 to 198)	
61-70	65	107(58 to 175)	
71-80	8	208(161 to 261)	
81-90	-	-	

Patients who attended for HIV care
post HAART only
N=7,618
Excluded subjects for whom no nadir CD4 T- cell count available

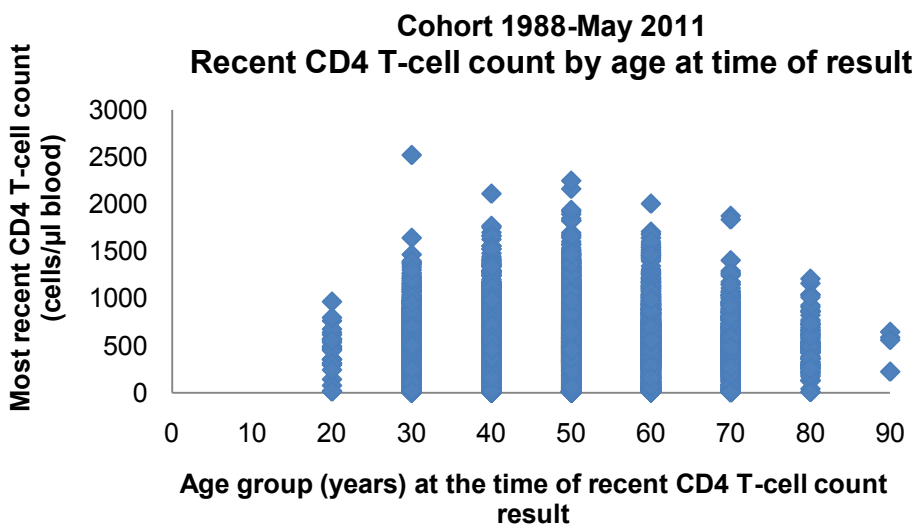


	Total=7,618 n	Median (IQR)	p-value
Age			
18-20	47	266(123 to 467)	<0.001
21-30	1624	287(176 to 433)	
31-40	3375	235(128 to 370)	
41-50	1919	220(120 to 369)	
51-60	487	204(92 to 340)	
61-70	143	194(97 to 325)	
71-80	22	188(60 to 280)	
81-90	1	137	

Figure A4.1.4: Distribution CD4 T-Cell Count by Age at Time of Result by Cohort Strata

Patients who attended for HIV care
1988 – May 2011
N=13,813

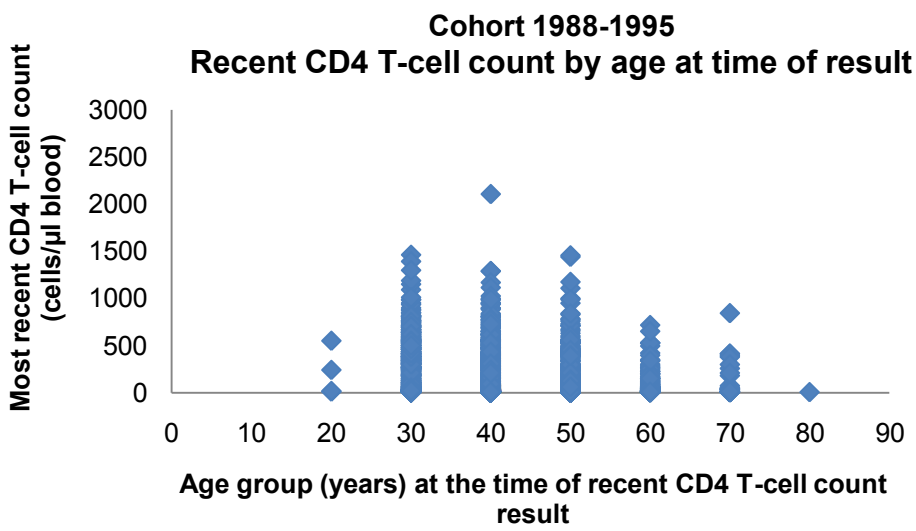
Excluded subjects for whom no CD4 T- cell count available



	Total=13,813 n	Median (IQR)	p-value
Age			
18-20	27	465(305 to 568)	<0.001
21-30	1997	406(214 to 574)	
31-40	4976	380(110 to 579)	
41-50	4493	455(232 to 661)	
51-60	1737	465(244 to 671)	
61-70	480	459(283 to 635)	
71-80	99	480(293 to 647)	
81-90	4	571(389 to 614)	

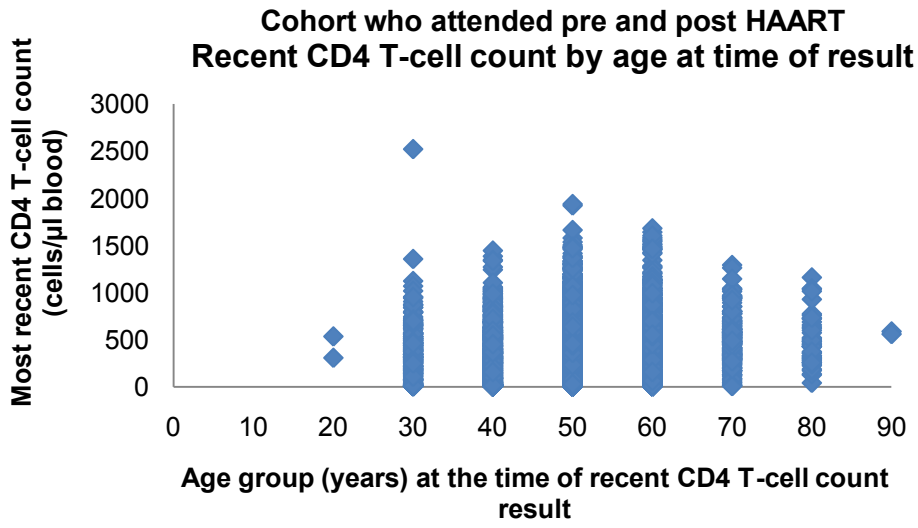
Patients who attended for HIV care
1988 – 1995 only
N=2,750

Excluded subjects for whom no CD4 T-cell count available



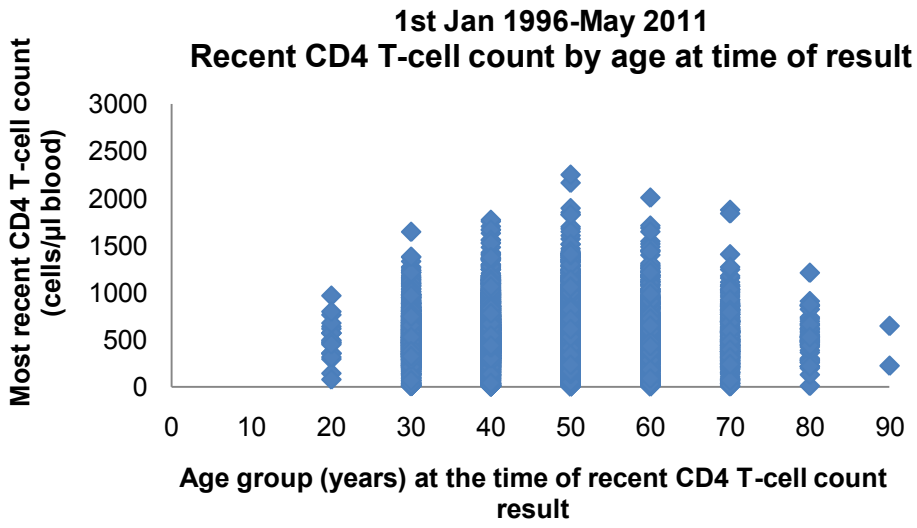
	Total=2,750 n	Median (IQR)	p-value
Age			
18-20	4	128(13 to 395)	<0.001
21-30	636	186(17 to 433)	
31-40	1197	30(8 to 198)	
41-50	681	24(8 to 120)	
51-60	194	24(9 to 75)	
61-70	37	32(8 to 72)	
71-80	1	4	
81-90	-	-	

Patients who attended for HIV care
pre HAART and since HAART
N=3,445
Excluded subjects for whom no CD4 T- cell count available



	Total=3,445 n	Median (IQR)	p-value
Age			
18-20	2	419(305 to 532)	<0.001
21-30	224	338(153 to 512)	
31-40	912	288(69 to 508)	
41-50	1299	468(258 to 696)	
51-60	746	494(297 to 716)	
61-70	213	481(319 to 616)	
71-80	47	429(282 to 654)	
81-90	2	557,585	

Patients who attended for HIV care
since HAART only
N=7,618
Excluded subjects for whom no CD4 T- cell count available

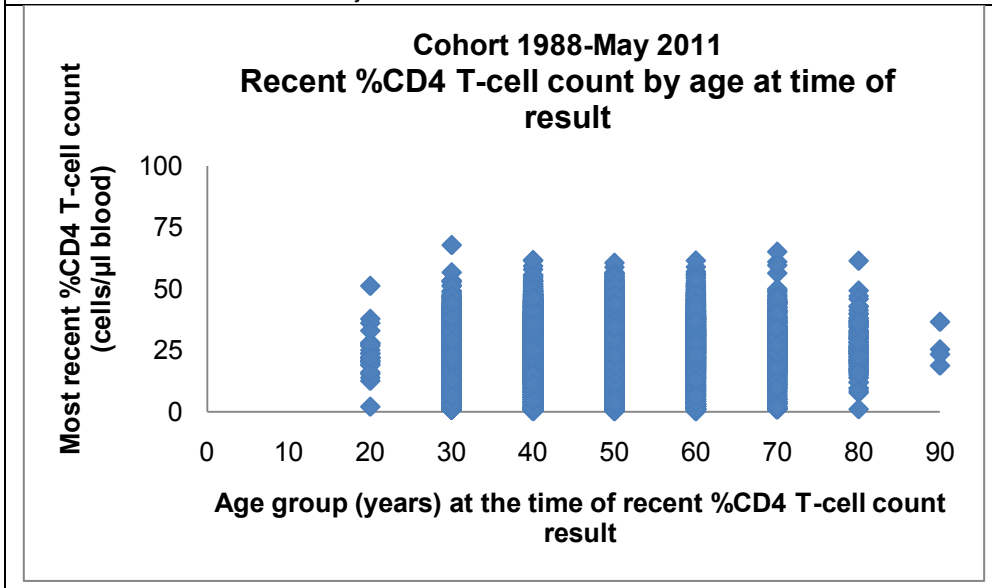


	Total=7,618 n	Median (IQR)	p-value
Age			
18-20	21	483(352 to 613)	<0.001
21-30	1137	480(349 to 634)	
31-40	2867	488(349 to 656)	
41-50	2513	516(365 to 696)	
51-60	797	506(346 to 683)	
61-70	230	479(315 to 693)	
71-80	51	505(362 to 647)	
81-90	2	221,643	

Figure A4.1.5: Distribution %CD4 T-Cell Count by Age at Time of Result by Cohort Strata

Patients who attended for HIV care
1988 – May 2011
N=13,541

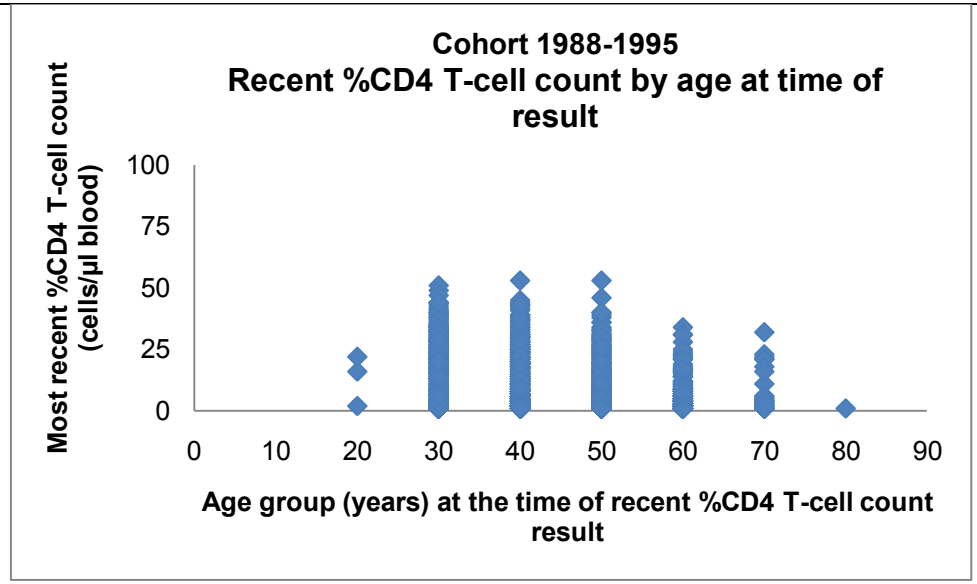
Excluded subjects for whom no %CD4 cell count available



	Total=13,541 n	Median (IQR)	p-value
Age			
18-20	25	22(19 to 27)	<0.001
21-30	1922	23(15 to 31)	
31-40	4855	23(10 to 31)	
41-50	4429	23(10 to 31)	
51-60	1728	26(16 to 33)	
61-70	479	26(18 to 32)	
71-80	99	25(20 to 32)	
81-90	4	24(21 to 31)	

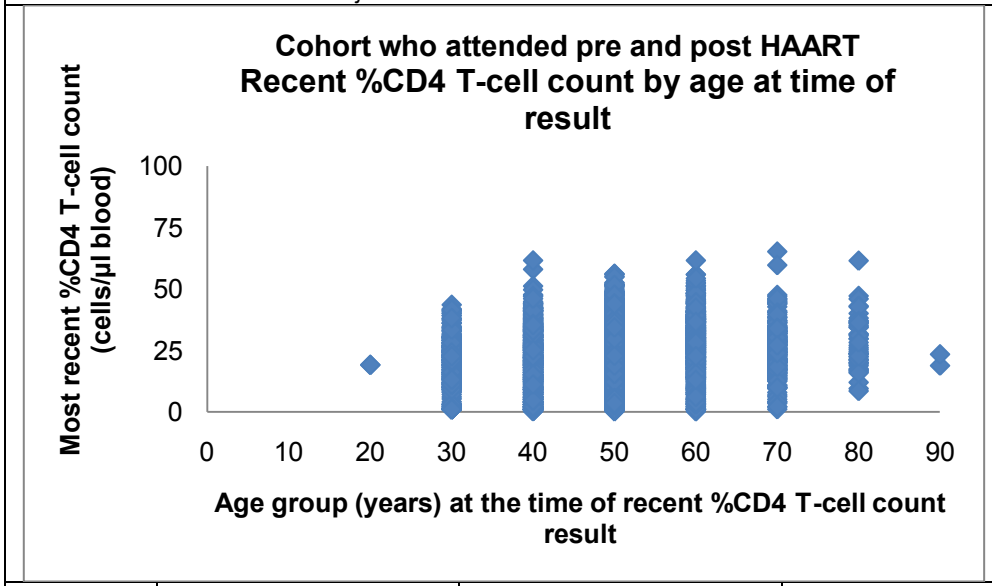
Patients who attended for HIV care
1988 – 1995 only
N=2,559

Excluded subjects for whom no %CD4 T-cell count available



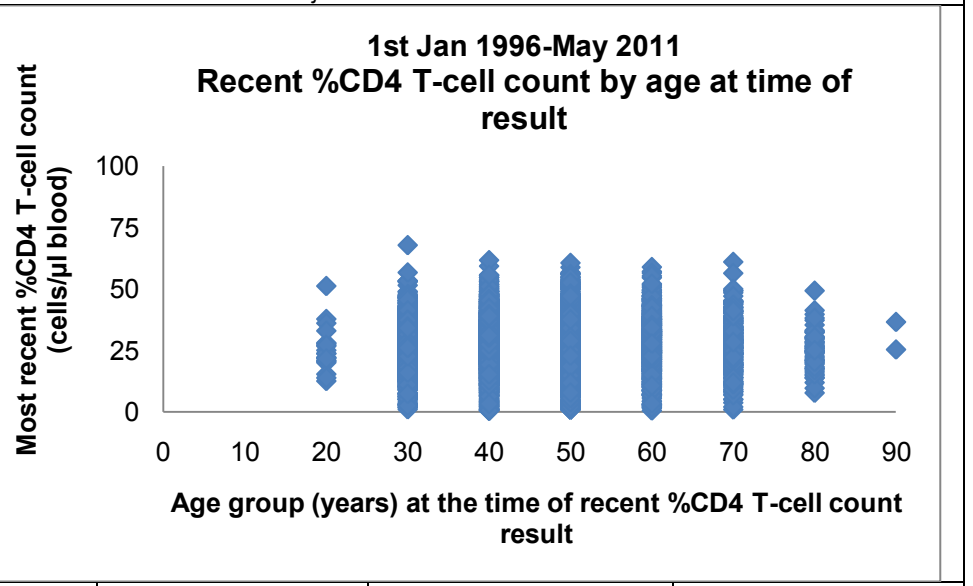
	Total=2,559 n	Median (IQR)	p-value
Age			
18-20	3	16(2 to 22)	<0.001
21-30	576	13(3 to 23)	
31-40	1118	4(2 to 13)	
41-50	635	4(1 to 10)	
51-60	190	4(2 to 8)	
61-70	36	4(2 to 6)	
71-80	1	1	
81-90	-	-	

Patients who attended for HIV care
pre HAART and since HAART
N=3,393
Excluded subjects for whom no %CD4 cell count available



	Total=3,393 n	Median (IQR)	p-value
Age			
18-20	2	19,19	<0.001
21-30	214	21(13 to 27)	
31-40	882	18(8 to 27)	
41-50	1290	26(17 to 34)	
51-60	743	27(18 to 34)	
61-70	213	25(19 to 32)	
71-80	47	25(20 to 36)	
81-90	2	19,23	

Patients who attended for HIV care
since HAART only
N=7,589
Excluded subjects for whom no %CD4 cell count available

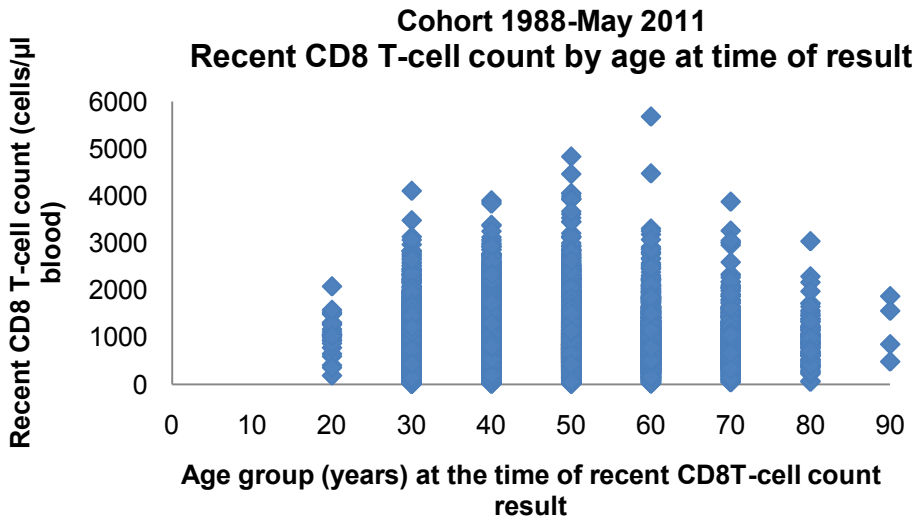


	Total=7,589 n	Median (IQR)	p-value
Age			
18-20	20	24(21 to 28)	<0.001
21-30	1132	27(21 to 33)	
31-40	2855	28(21 to 34)	
41-50	2504	29(22 to 35)	
51-60	795	28(21 to 35)	
61-70	230	28(20 to 34)	
71-80	51	25(19 to 30)	
81-90	2	25,37	

Figure A4.1.6: Distribution CD8 T-Cell Count by Age at Time of Result by Cohort Strata

Patients who attended for HIV care
1988 – May 2011
N=13,278

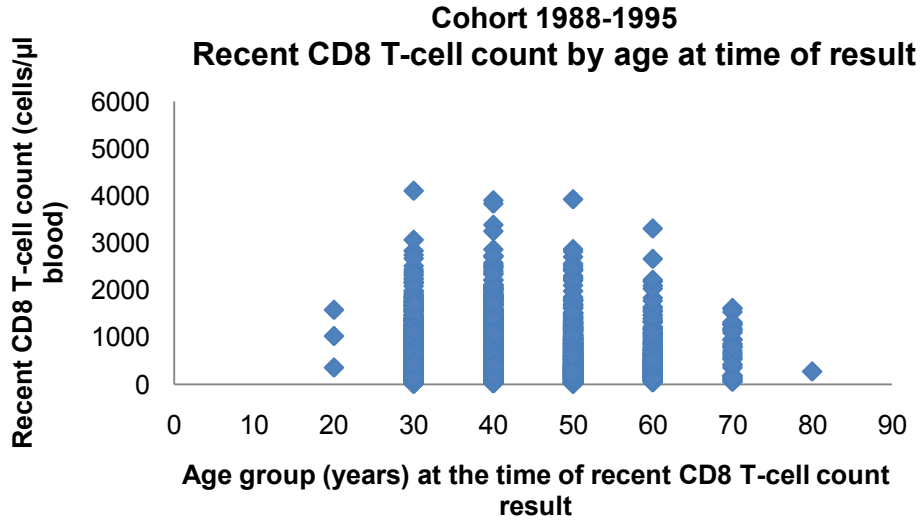
Excluded subjects for whom no CD8 T-cell count available



	Total=13,278 n	Median (IQR)	p-value
Age			
18-20	24	1043(820 to 1274)	<0.001
21-30	1826	847(569 to 1149)	
31-40	4771	772(502 to 1088)	
41-50	4371	787(536 to 1104)	
51-60	1710	801(544 to 1130)	
61-70	476	804(557 to 1164)	
71-80	97	859(522 to 1158)	
81-90	4	1201(662 to 1710)	

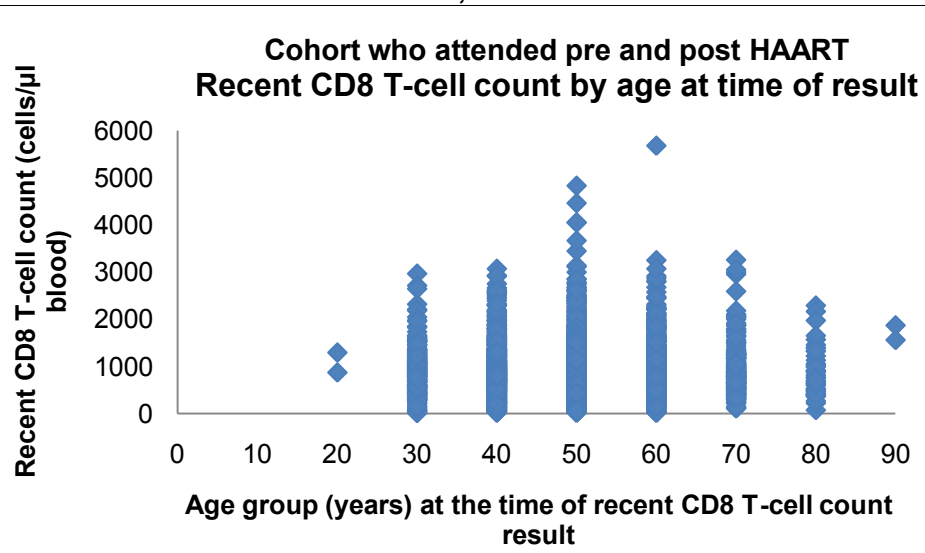
Patients who attended for HIV care
1988 – 1995 only
N=2369

Excluded subjects for whom no CD8 T-cell count available



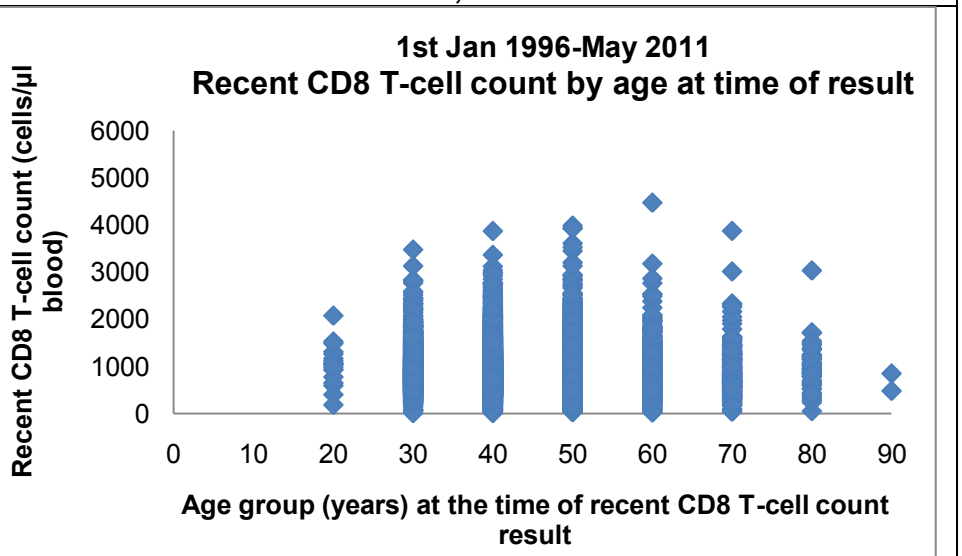
	Total=2,370 n	Median (IQR)	p-value
Age			
18-20	3	1020(352 to 1575)	<0.001
21-30	502	713(360 to 1099)	
31-40	1057	511(224 to 922)	
41-50	595	457(234 to 853)	
51-60	178	460(256 to 782)	
61-70	34	745(348 to 1156)	
71-80	1	268	
81-90	-	-	

Patients who attended for HIV care
pre HAART and since HAART
N=3,404



	Total=3,404 n	Median (IQR)	p-value
Age			
18-20	2	1077(866 to 1288)	0.002
21-30	211	877(544 to 1178)	
31-40	895	780(474 to 1102)	
41-50	1293	815(565 to 1186)	
51-60	741	852(587 to 1193)	
61-70	214	825(573 to 1167)	
71-80	46	723(477 to 1126)	
81-90	2	1555,1865	

Patients who attended for HIV care
since HAART only
N=7,505

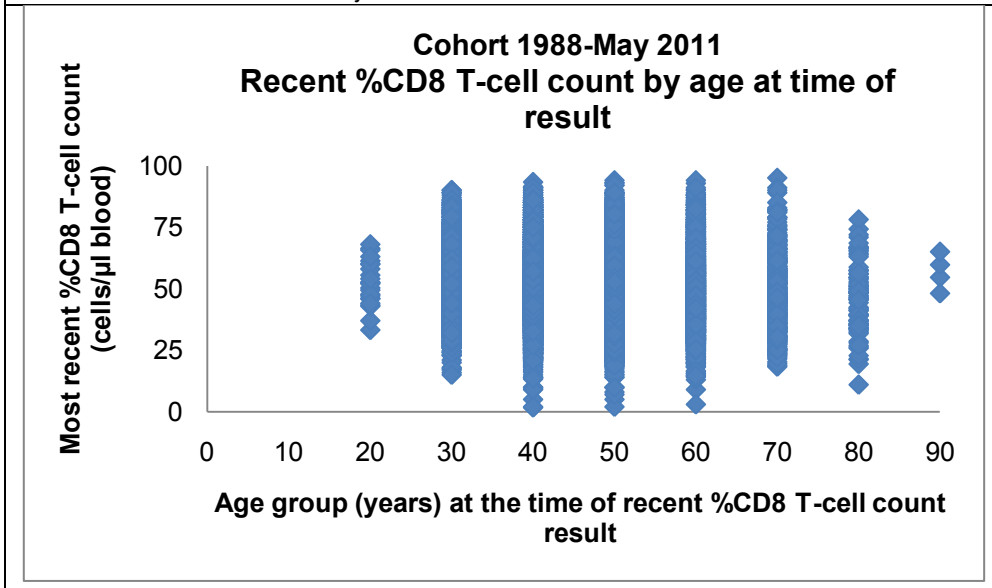


	Total=7,505 n	Median (IQR)	p-value
Age			
18-20	19	1046(773 to 1260)	0.001
21-30	1113	878(640 to 1163)	
31-40	2819	837(604 to 1124)	
41-50	2483	824(587 to 1110)	
51-60	791	822(580 to 1109)	
61-70	228	802(557 to 1162)	
71-80	50	923(645 to 1192)	
81-90	2	478, 846	

Figure A4.1.7: Distribution %CD8 T-Cell Count by Age at Time of Result by Cohort Strata

Patients who attended for HIV care
1988 – May 2011
N=13,316

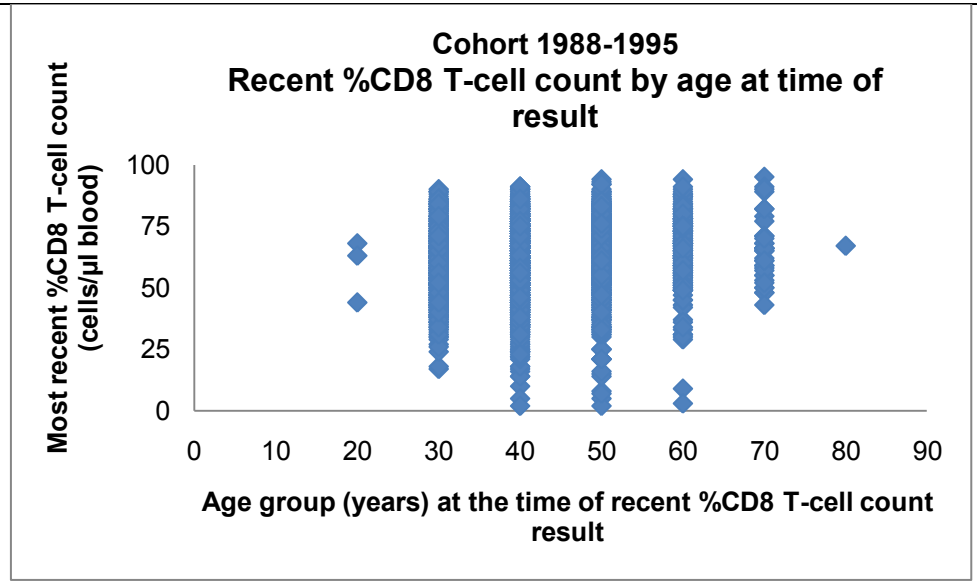
Excluded subjects for whom no %CD8 cell count available



	Total=13,316 n	Median (IQR)	p-value
Age			
18-20	24	52(46 to 60)	<0.001
21-30	1846	53(45 to 63)	
31-40	4778	52(42 to 63)	
41-50	4379	48(39 to 59)	
51-60	1710	48(39 to 59)	
61-70	478	47(39 to 58)	
71-80	97	47(36 to 55)	
81-90	4	57(51 to 62)	

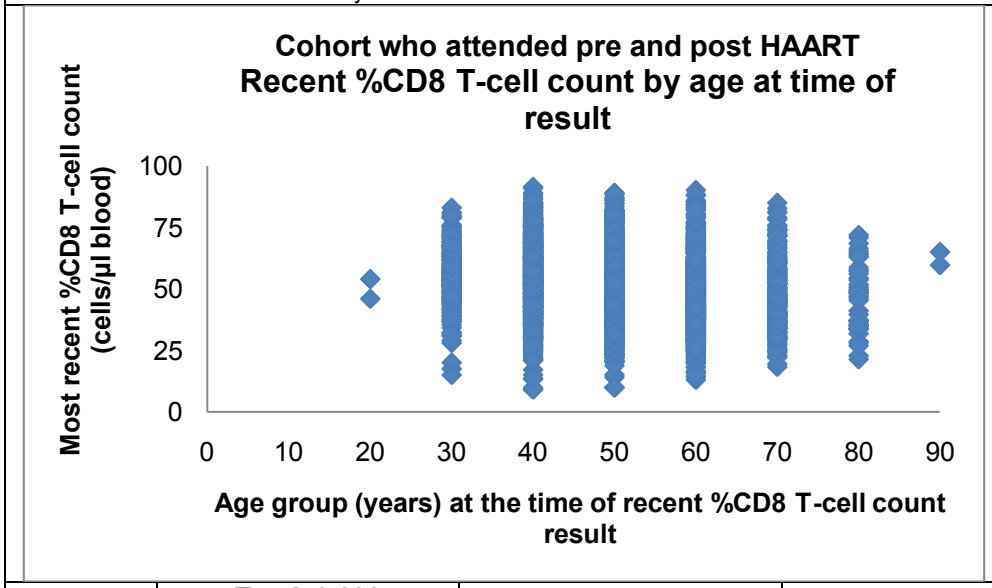
Patients who attended for HIV care
1988 – 1995 only
N=2,422

Excluded subjects for whom no %CD8 T-cell count available



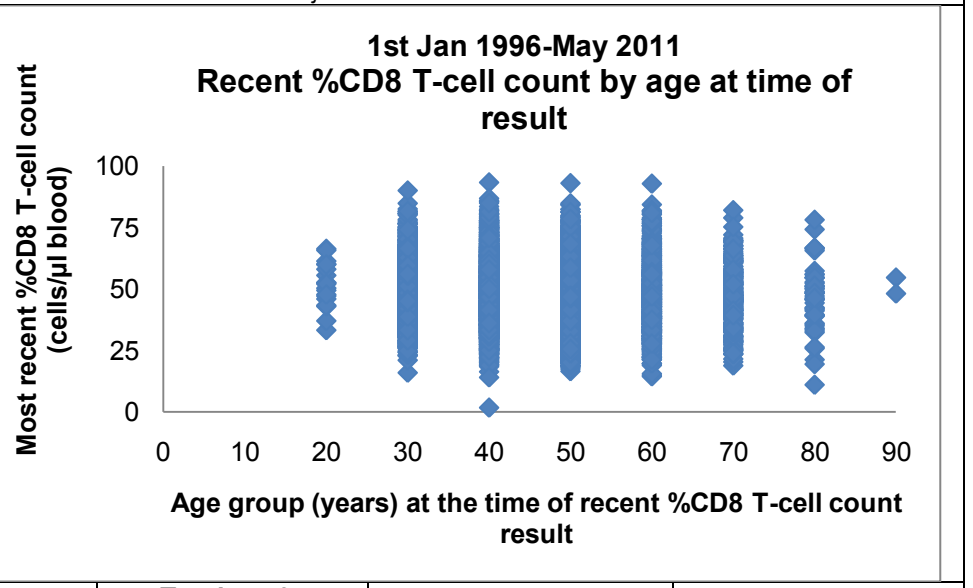
	Total=2,422 n	Median (IQR)	p-value
Age			
18-20	3	63(44 to 68)	<0.001
21-30	522	61(51 to 69)	
31-40	1073	64(53 to 73)	
41-50	606	65(54 to 75)	
51-60	181	68(57 to 76)	
61-70	36	66(56 to 71)	
71-80	1	67	
81-90	-	-	

Patients who attended for HIV care
pre HAART and since HAART
N=3,399
Excluded subjects for whom no %CD8 cell count available



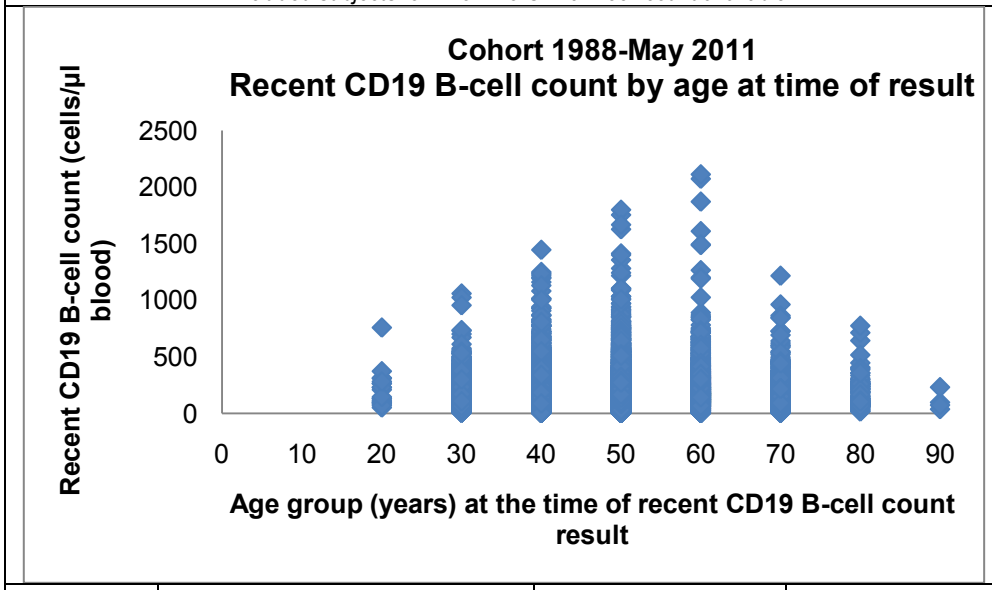
	Total=3,399 n	Median (IQR)	p-value
Age			
18-20	2	46,54	<0.001
21-30	212	56(49 to 64)	
31-40	891	55(45 to 66)	
41-50	1292	48(40 to 59)	
51-60	740	47(38 to 58)	
61-70	214	46(36 to 56)	
71-80	46	47(35 to 56)	
81-90	2	60,65	

Patients who attended for HIV care
since HAART only
N=7,495
Excluded subjects for whom no %CD8 cell count available



	Total=7,495 n	Median (IQR)	p-value
Age			
18-20	19	52(46 to 60)	<0.001
21-30	1112	50(42 to 58)	
31-40	2814	48(40 to 57)	
41-50	2481	46(38 to 54)	
51-60	789	46(38 to 55)	
61-70	228	47(39 to 55)	
71-80	50	47(39 to 53)	
81-90	2	48,55	

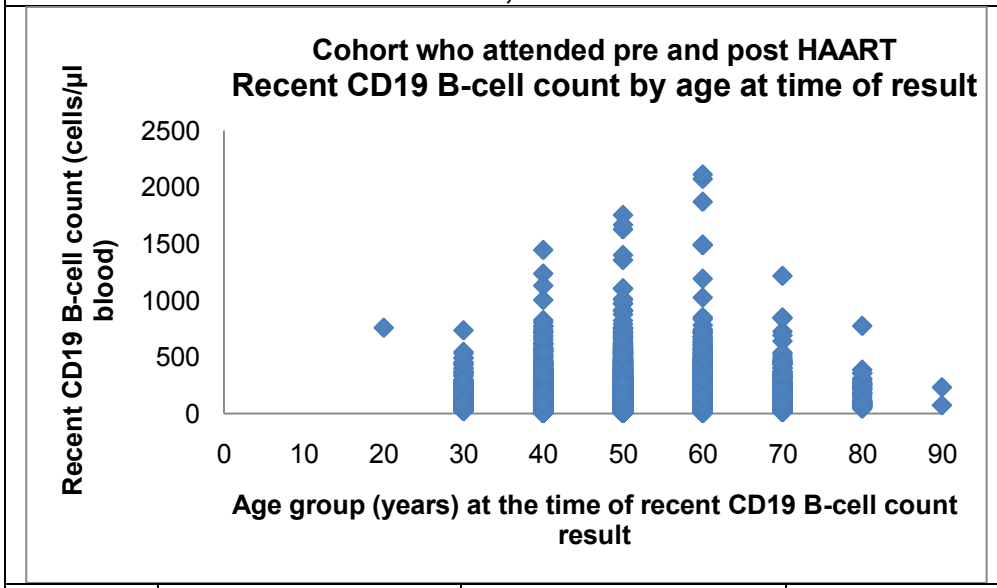
Figure A4.1.8: Distribution CD19 B-Cell Count by Age at Time of Result by Cohort Strata

Patients who attended for HIV care 1988 – May 2011 **N=10,557** *Excluded subjects for whom no CD19 T-cell count available*	**Patients who attended for HIV care** 1988 – 1995 only **N=0** *Excluded subjects for whom no CD19 T-cell count available*
	**CD19 not available pre HAART**

	Total=10,557 **n**	**Median (IQR)**	**p-value**
Age			
18-20	20	175(98 to 271)	<0.001
21-30	1256	160(103 to 228)	
31-40	3526	168(109 to 252)	
41-50	3709	194(126 to 292)	
51-60	1506	190(119 to 301)	
61-70	440	173(104 to 274)	
71-80	96	137(81 to 254)	
81-90	4	83(53 to 162)	

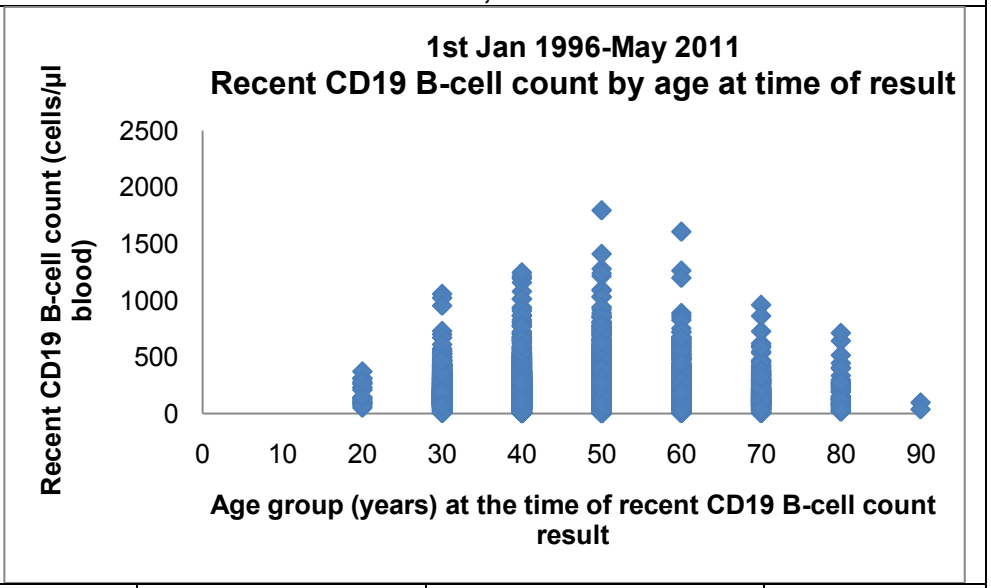
	Total=0 **n**	**Median (IQR)**	**p-value**
Age			
18-20	Not available pre HAART		
21-30			
31-40			
41-50			
51-60			
61-70			
71-80			
81-90			

Patients who attended for HIV care
pre HAART and since HAART
N=3,063



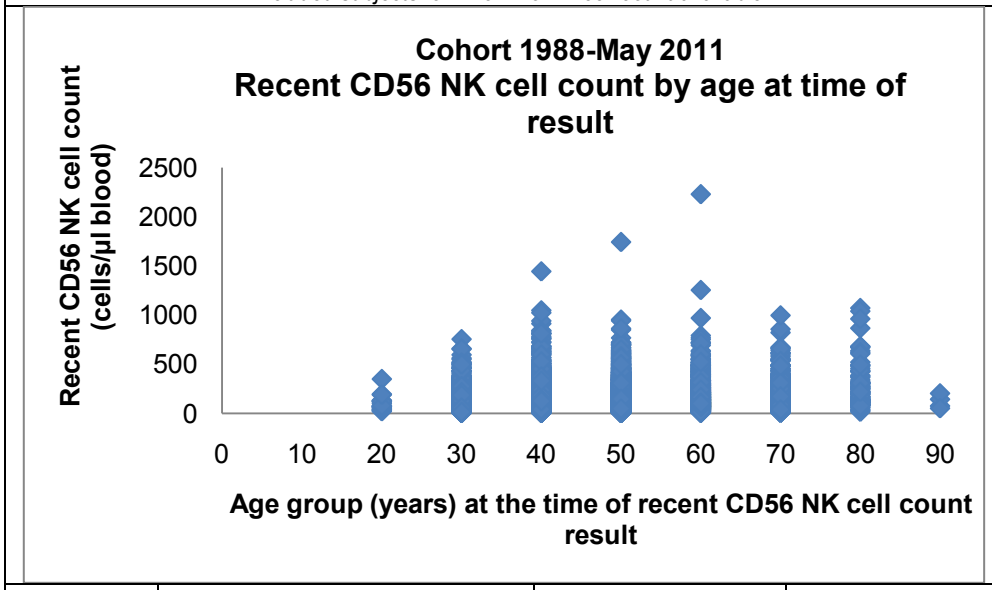
	Total=3,063 n	Median (IQR)	p-value
Age			
18-20	1	756	<0.001
21-30	144	144(86 to 224)	
31-40	713	151(84 to 250)	
41-50	1228	195(120 to 304)	
51-60	717	207(117 to 312)	
61-70	212	186(114 to 277)	
71-80	46	138(86 to 245)	
81-90	2	71,229	

Patients who attended for HIV care
since HAART only
N=7,494



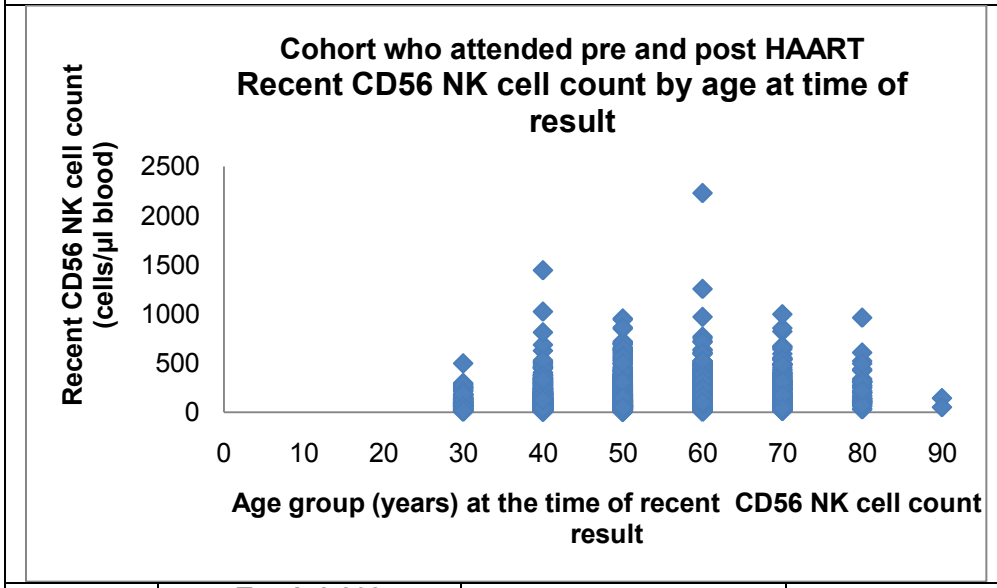
	Total=7,494 n	Median (IQR)	p-value
Age			
18-20	19	143(96 to 262)	<0.001
21-30	1112	161(106 to 229)	
31-40	2813	170(114 to 253)	
41-50	2481	194(128 to 288)	
51-60	789	182(121 to 285)	
61-70	228	154(97 to 267)	
71-80	50	137(79 to 258)	
81-90	2	35,95	

Figure A4.1.9: Distribution CD56 NK Cell Count by Age at Time of Result by Cohort Strata

Patients who attended for HIV care 1988 – May 2011 **N=9,524** **CD56 NK cell count results available since 1997** *Excluded subjects for whom no NK cell count available*	**Patients who attended for HIV care** 1988 – 1995 only **N=0** *Excluded subjects for whom no CD19 T-cell count available*
	**CD56 NK cell count not available pre HAART**

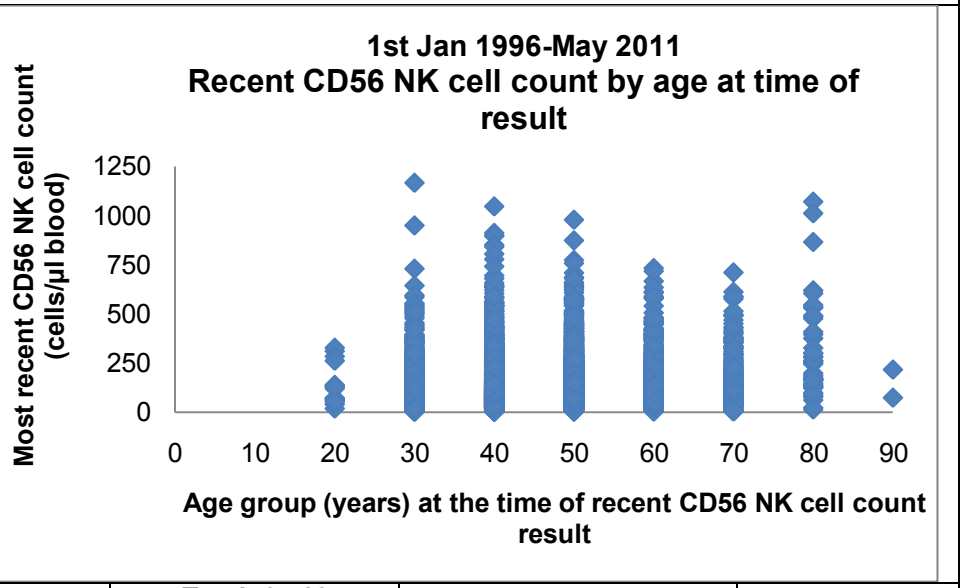
	Total=9,524 **n**	**Median (IQR)**	**p-value**		**Total=0** **n**	**Median (IQR)**	**p-value**
Age							
18-20	22	73(44 to 122)	<0.001		-		
21-30	1048	97(55 to 165)					
31-40	3177	97(54 to 170)					
41-50	3456	116(65 to 193)					
51-60	1335	138(77 to 222)					
61-70	393	157(91 to 266)					
71-80	89	197(113 to 319)					
81-90	4	107(61 to 171)					

Patients who attended for HIV care
pre HAART and since HAART
n=2,808
CD56 NK cell count results available since 1997



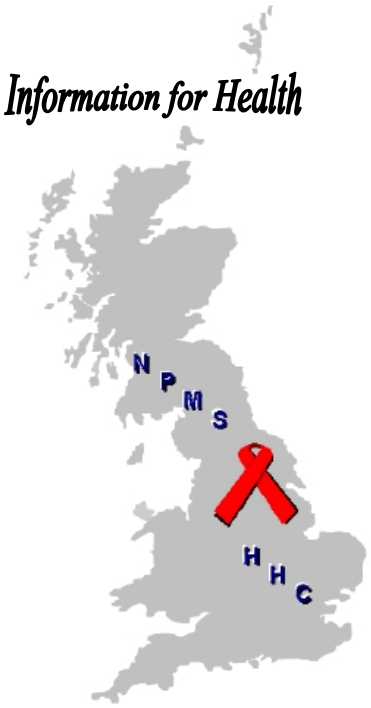
	Total=2,808 n	Median (IQR)	p-value
Age			
18-20	-	-	<0.001
21-30	99	92(56 to 144)	
31-40	621	91(47 to 160)	
41-50	1201	114(60 to 188)	
51-60	647	144(80 to 230)	
61-70	195	175(104 to 299)	
71-80	43	193(106 to 305)	
81-90	2	49,141	

Patients who attended for HIV care
since HAART only
N=6,716
CD56 NK cell count results available since 1997



	Total=6,716 n	Median (IQR)	p-value
Age			
18-20	22	73(44 to 122)	<0.001
21-30	949	98(55 to 166)	
31-40	2556	98(56 to 171)	
41-50	2255	117(67 to 195)	
51-60	688	133(74 to 213)	
61-70	198	146(81 to 233)	
71-80	46	216(123 to 426)	
81-90	2	73,201	

A 5.1 NPMS-HHC: LIST OF REQUIRED FIELDS



NPMS-HHC

List of Variables
to be collected for NPMS-HHC
on the use, cost and outcome
of HIV service provision in U.K hospitals

CONTENTS

PAGE

NOTES

If any of our variables have a different format or use different codes from the ones that you use, we will write a program to convert them to the format we require. We ask that you include with your data a list of formats or codes that you have used.

Please include in **each file**:

- SOUNDEX
- DATE OF BIRTH
- GENDER
- A UNIQUE IDENTIFIER YOU USE OR ONE THAT IS CONSISTENT THROUGH ALL FILES. **This should not be patient identifiable that is used in the clinic/hospital. For example do not provide hospital ID number.**

Demographic File	
REQUIRED FIELD	**NOTES**
SOUNDEX	SOUNDEX coding program, J Nash 11/07/87 version 2 based on Surname
DATE OF BIRTH	DD/MM/YY e.g. 12/03/97 – same for all dates
GENDER	Male/Female
ETHNIC GROUP	We will map your codes to the following: 1 = White 2 = Black – Caribbean 3 = Black – African 4 = Black – Other 5 = Black – Unspecified 6 = Indian Sub-Continent 7 = Chinese 8 = Asian – Other 9 = Other 10 = Unknown
NATIONALITY AT FIRST ATTENDANCE	We will map your codes to the following: 1 = British 2 = EU Other than British 3 = North American 4 = South American 5 = Caribbean 6 = Sub-Saharan African 7 = North African – Middle East 8 = African – Unspecified 9 = Asian 10 = Australasian 11 = Other 12 = Unknown
SOURCE OF REFERRAL AT FIRST ATTENDANCE	We will map your codes to the following: 1 = On your own accord (self-referral) 2 = Referral by GP 3 = Referral by Family Planning Clinic 4 = Referral by consultant within hospital 5 = Referral by consultant elsewhere 6 = Other 7 = Unknown
DHA OF PATIENT AT FIRST ATTENDANCE	We will map your codes to the CDSC codes
LOCAL AUTHORITY OF PATIENT AT FIRST ATTENDANCE	We will map your codes to the CDSC codes
IS PATIENT REGISTERED WITH GP AT FIRST ATTENDANCE	Y/N
(IF YES) GP ID NUMBER	GP name if ID unknown
DATE FIRST HIV POSITIVE TEST	-

LOCATION WHERE FIRST POSITIVE TEST WAS MADE	We will map your codes to the following: 1 = At this clinic 2 = Other local service 3 = Other UK clinic – London 4 = Other UK clinic – Non-London 5 = Other UK clinic – Unspecified 6 = Clinic outside UK 7 = Unknown
demographic file contd..	
DATE FIRST AIDS DIAGNOSIS	-
LOCATION WHERE FIRST AIDS DIAGNOSIS WAS MADE	We will map your codes to the following: 1 = At this clinic 2 = Other local service 3 = Other UK clinic – London 4 = Other UK clinic – Non-London 5 = Clinic outside 6 = Unknown
DATE OF DEATH	-
STAGE IF HIV ON 1ST JAN 1996 ***(IF PROVIDING DATA FROM 1996 ONWARDS)***	We will map your codes to the following: 1 = Asymptomatic 2 = Symptomatic Non AIDS 3 = AIDS

Self Reported Risk Factors File	
REQUIRED FIELD	**NOTES**
RISK (for each patient any number of the following can be selected – but only one of the first three)	We will map your codes to the following: 1 = Homosexual 2 = Heterosexual 3 = Bisexual 4 = IDU 5 = Haemophiliac 6 = Blood transfusion recipient before October 1985 (UK) or abroad 7 = Prostitute 8 = Child of Seropositive mother 9 = Sexual intercourse with partner from high endemic area 10 = Sexual intercourse with partner from Africa 11 = Sexual intercourse with partner from SE Asia 12 = Sexual intercourse with partner from W Indies 13 = Sexual intercourse with partner from USA 14 = Sexual intercourse with partner from S America 15 = Sexual intercourse with partner from Europe 16 = Sexual intercourse with partner from other high endemic area 17 = Sexual intercourse in a high endemic area 18 = Sexual intercourse in Africa 19 = Sexual intercourse in SE Asia 20 = Sexual intercourse in W Indies 21 = Sexual intercourse in USA 22 = Sexual intercourse in S America 23 = Sexual intercourse in Europe 24 = Sexual intercourse in another high endemic area 25 = Other risk factor apart from the above 26 = Unknown

Visit File	
REQUIRED FIELD	**NOTES**
For each patient we would like the following information for **every** visit i.e. inpatient, outpatient or day case.	
VISIT TYPE	We will map your codes to the following: O = Outpatient I = Inpatient D = Dayward
DATE OF VISIT **(or admission date if inpatient)**	-
LATEST DHA OF PATIENT	We will map your codes to the CDSC codes
LATEST LOCAL AUTHORITY OF PATIENT	We will map your codes to the CDSC codes

Inpatient File
(a patient who has been admitted to hospital and has spent at least one night in hospital)

REQUIRED FIELDS	**NOTES**
DATE OF VISIT (or admission date)	-
NATURE OF VISIT	We will map your codes to the following: 1 = Planned: planned test/procedure duration of stay which includes at least one night in hospital 2 = Emergency: where the patient is admitted to hospital when admission is unpredictable and at short notice because of clinical need 3 = Unknown
DATE OF DISCHARGE	Date of admission already recorded in Visit File
DISCHARGE VENUE	We will map your codes to the following: 1 = Home alone 2 = Home with support 3 = Staying at friends or relatives 4 = Hospice 5 = Other hospital 6 = Unknown
OUTCOME OF EPISODE	We will map your codes to the following: 1 = Improved 2 = Stable 3 = Deteriorate 4 = Death 5 = Unknown
ITU ADMISSION DATE	-
ITU DISCHARGE DATE	-

Outpatient File **(a patient attending the outpatient clinic for consultation)**	
REQUIRED FIELDS	**NOTES**
DATE OF VISIT	-
NATURE OF VISIT	We will map your codes to the following: 1 = Planned: planned or booked in advance 2 = Emergency: to deal with acute problems, not booked 3 = Unknown
HAS PATIENT RECEIVED CARE FOR HIV FROM GP?	Y/N
HAS PATIENT RECEIVED CARE FOR HIV FROM INPATIENT AT ANOTHER CLINIC?	Y/N
HAS PATIENT RECEIVED CARE FOR HIV FROM OUTPATIENT AT ANOTHER CLINIC?	Y/N

Professionals File **(for both Outpatient and Day Ward files)**	
REQUIRED FIELDS	**NOTES**
DATE OF VISIT	-
WHICH PROFESSIONAL DID PATIENT SEE DURING OUTPATIENT/DAY WARD VISIT?	We will map your codes to the following: 1 = HIV Consultant team 2 = Clinic Nurse 3 = Surgeon 4 = Neurologist 5 = Respiratory Physician 6 = Gastroenterologist 7 = Haematologist 8 = Psychiatrist 9 = Oncologist 10 = Psychologist 11 = Social Worker 12 = Dietician 13 = Home Support Team Nurse 14 = Physiotherapist 15 = Occupational Therapist 16 = Health Advisor 17 = Dermatologist 18 = ENT 19 = Cardiologist 20 = Opthamologist 21 = Dentist 22 = Other 23 = Unknown

Day Ward	
REQUIRED FIELD	NOTES
DATE OF VISIT	-
NATURE OF VISIT	We will map your codes to the following: 1 = Day case: a patient who is pre-booked for a procedure and not expected to stay overnight in hospital and who returns home the same day as expected 2 = Day attendance: a patient admitted regularly and electively for treatment over a planned sequence of days, being discharged each evening 3 = Unknown
LOCATION OF DAY TREATMENT/PROCEDURE	We will map your codes to the following: 1 = Outpatients 2 = Inpatients

Tests **(for inpatient, outpatient and day ward files)**	
REQUIRED FIELDS	NOTES
	At least collected when first seen, at first AIDS diagnosis and subsequent AIDS diagnosis
DATE OF TEST	-
TYPE OF TEST	1 = Haemoglobin (mg/dl) 2 = Lymphocytes (x 10^9 /1) 3 = CD4 count (x 10^6/1) 4 = Viral Load (RNA copies/ml)
RESULT (in units stated)	-

Drugs **(for inpatient, outpatient and day ward files)**	
REQUIRED FIELDS	**NOTES**
DATE PRESCRIBED	-
DRUG NAME	Use free text – codes to be finalised
DOSE	-
ROUTE OF ADMISSION	Oral Intravenous (IV) Intramuscular (IM) Other; including rectal, topical, inhaled, nebulised, subcutaneous

If possible we would prefer to receive the ARV drug history in the following format instead:

ARV combination **(for inpatient, outpatient and day ward files)**

Soundex Gender DOB	**Date started therapy**	**Date stopped therapy**	**ARV 1**	**ARV 2**	**ARV 3**	**ARV 4**	**ARV 5**	ARV 6	**Arv 7**	**ARV 8**	**ARV 9**
e.g. Patient 1	12/3/97	13/01/98	AZT	DDI	NFV						
e.g.. Patient 1	14/01/98	Ongoing at current data download	AZT	3TC	EFV						

Refer to page 16 for ARV drug codes if necessary

Diagnoses File **(For inpatient, outpatient and day ward files)**	
REQUIRED FIELDS	NOTES
DATE OF VISIT (or diagnosis)	-
HIV DIAGNOSES	Provide details of all opportunistic infections
NON-HIV DIAGNOSES	READ/ICD10

Please note when you do extract the Diagnosis file from your clinics database, please could you download this particular file from when the patient was first seen even if this means that the data is prior to 1996. The reason for this is we will be calculating the patients stage of HIV from their first attendance at your clinic and then see when they change clinical stage status.

Procedures File **(for Day Ward file)**	
REQUIRED FIELD	NOTES
DATE OF PROCEDURE	-
PROCEDURES	OPCS Coding – Tabular list of the Classifications of Surgical Operations and Procedures – 4th Revision