

The creation of a large UK-based cohort of HIV infected individuals: the Seven Centre HIV Database Project

Caroline A Sabin and Teresa Hill on behalf of the Seven Centre HIV Database Project Steering Committee

Royal Free and University College Medical School, Rowland Hill Street,
London, England

Tel: +44 207 830 2239 ext. 4752. Fax: +44 207 794 1224. E-mail: c.sabin@pcps.ucl.ac.uk

Background

- **Information on the HIV epidemic in the UK is currently available through a number of sources (Table 1)**
- **Most of these sources of data are restricted in that they provide information only on selected groups of patients (eg. seroconverters, haemophilic men), or the data are limited to that needed for surveillance purposes or are collected for other reasons (eg. For assessing the costs of HIV care)**
- **No data currently exist which provide information on key issues such as the impact of new therapies on the natural history of disease in the population of HIV-infected individuals in the UK**

Table 1 - Current sources of HIV data in the UK

- PHLS AIDS and STD Centre, CDSC**
- Scottish Centre for Infection and Environmental Health (SCIEH)**
- Survey of Prevalent HIV Infections that are Diagnosed (SOPHID)**
- UK Register of HIV Seroconverters**
- European CASCADE Study**
- UK Haemophilia Register**
- Cohorts from single clinical sites (eg. Royal Free, Chelsea and Westminster, St. Mary's)**
- National Prospective Monitoring Scheme (NPMS)**

Aim and objectives

To create a large database of HIV-infected individuals seen at one of seven large clinical centres in the UK since 1st January 1996, with the following broad objectives:

- **To describe the characteristics of individuals currently under care in the UK**
- **To describe changes in the spectrum of HIV disease and survival over time**
- **To describe virological, immunological and clinical response to treatment**

Centres currently involved

Chelsea and Westminster	- B Gazzard
Royal Free	- M Johnson, A Mocroft
St. Mary's	- J Weber, G Scullard
Mortimer Market	- R Gilson
Brighton NHS Trust	- M Fisher
Kings	- P Easterbrook
St. Thomas'	- B Peters

MRC Clinical Trials Unit (K Porter, D Dunn, A Babiker, R Matthias),

PHLS (J Mortimer)

Royal Free and University College Medical School (C Sabin, A Phillips, T Hill)

Data collected

Centres have been asked to provide an electronic copy of all clinical and laboratory information for any patients seen since 1/1/1996. The data collected includes the following:

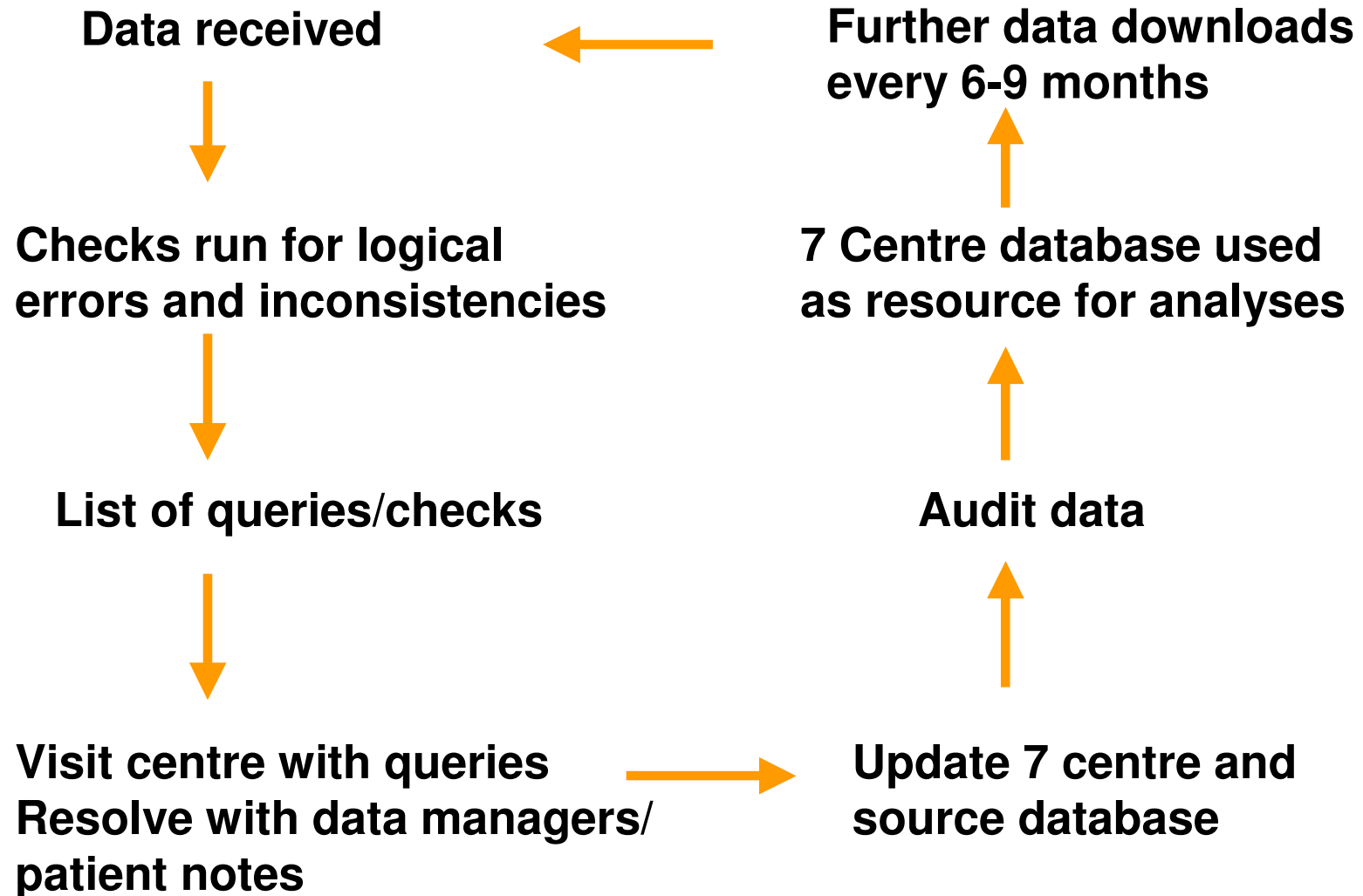
Demographics - clinic ID, date of birth, soundex code, initial, sex, dates of HIV positive and negative tests, dates when first and last seen at clinic, exposure group, ethnicity, country of birth

Clinical events - dates and type of all AIDS events, date of death

Treatment - dates of starting/stopping all antiretroviral drugs

Laboratory data - dates and results of all CD4/CD8 counts, CD4 percentages and viral loads

Database development - processes involved



Identifying patients who attend more than one centre

In order to preserve confidentiality, no names have been transferred. The only identifiers on the database are soundex code, first initial and date of birth.

Three categories of possible 'matches' have been defined:

- 1) Patients with the same soundex code and date of birth**
- 2) Patients with the same initials and date of birth**
- 3) Patients with the same date of birth (less likely to be genuine matches but necessary because of a high frequency of false names)**

Other information is then used to assess whether or not the patients are thought to be genuine matches. The information used in this process was :

- 1) Date of first HIV positive test**
- 2) Country of birth (where non-UK)**
- 3) Duration of follow-up at either centre**
- 4) Centres transferred to or from**
- 5) Last seen and first seen dates**
- 6) Known dates of death**

Where this information was not available, other data (eg. CD4 counts, treatment data) was used to see if any information conflicted. Where there was no conflicting information, two patients were assumed to be a genuine match.

Linking to death registry

- **Whilst date of death is available for many patients, this information may be incomplete if patients are lost to follow-up.**
- **In order to identify any deaths that have occurred in patients who are thought to be alive, the database has been linked to the national death register (provided by the Office of National Statistics) which contains information on date and cause of death**
- **There are a number of limitations with this: matching has to be performed on the basis of soundex code and date of birth and ONS data does not include information on deaths occurring outside the UK, which may represent a large proportion of deaths in this cohort**

Preliminary results

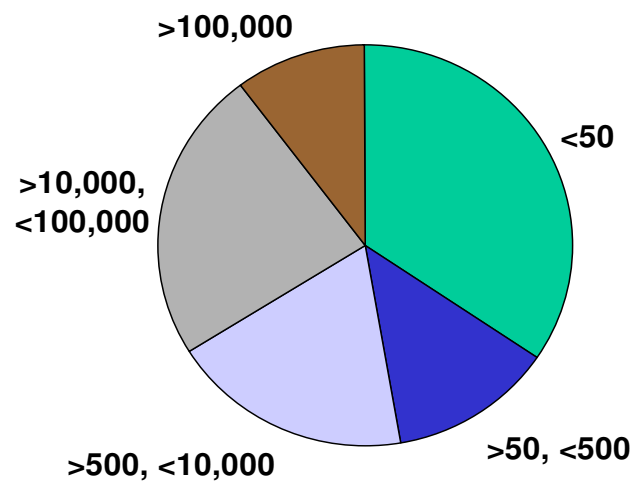
- **To date, data is available from 6 of the 7 cohorts; information from the final centre is expected in the next three months.**
- **The initial dataset contained 15224 patient records; after linking information on patients believed to be the same individual, 13825 unique individuals have been identified who have attended one of the six centres at least once during the study period.**
- **The process of cleaning and auditing the data continues, but some preliminary results on the demographics of the patients under follow-up are available (Table 2).**

Table 2 - demographics of patients under care

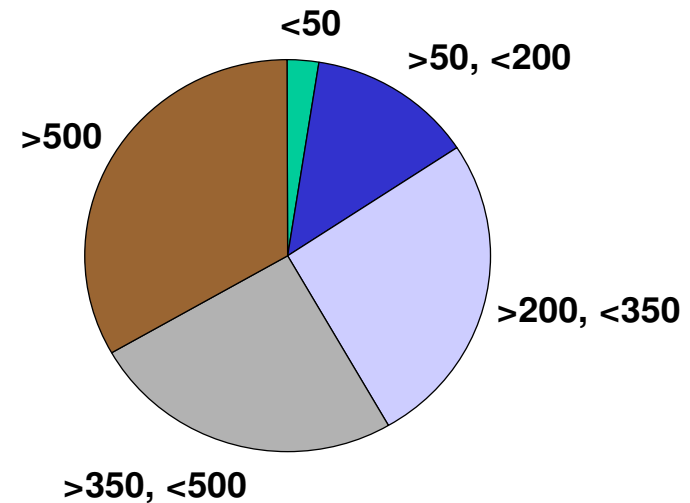
		n	(%)
Number of patients:		13825	(100.0)
Sex:	Male	11345	(82.1)
	Female	2480	(17.9)
Risk group:	Homo/bisexual	8500	(61.5)
	IDU	566	(4.1)
	Heterosexual	3074	(22.2)
	Other	146	(1.6)
	Not known	1539	(11.1)
Ethnicity:	Caucasian	7579	(54.8)
	Black African	2141	(15.5)
	Other Black	562	(4.1)
	Asian/oriental	817	(5.9)
	Not known	2480	(17.9)
Date of birth:	Median (range)	May 63	Sept 22 – Dec 83
Date of positive HIV test:	Median (range)	June 95	Jan 80 – Dec 01

- **CD4 and HIV RNA data are currently available from five of the six centres. Of the 7668 patients attending these centres who were under follow-up in 2001, 7459 (97.3%) had at least one CD4 count and 5686 (74.2%) at least one RNA measurement**

Distribution of most recent laboratory markers among those under follow-up in 2001



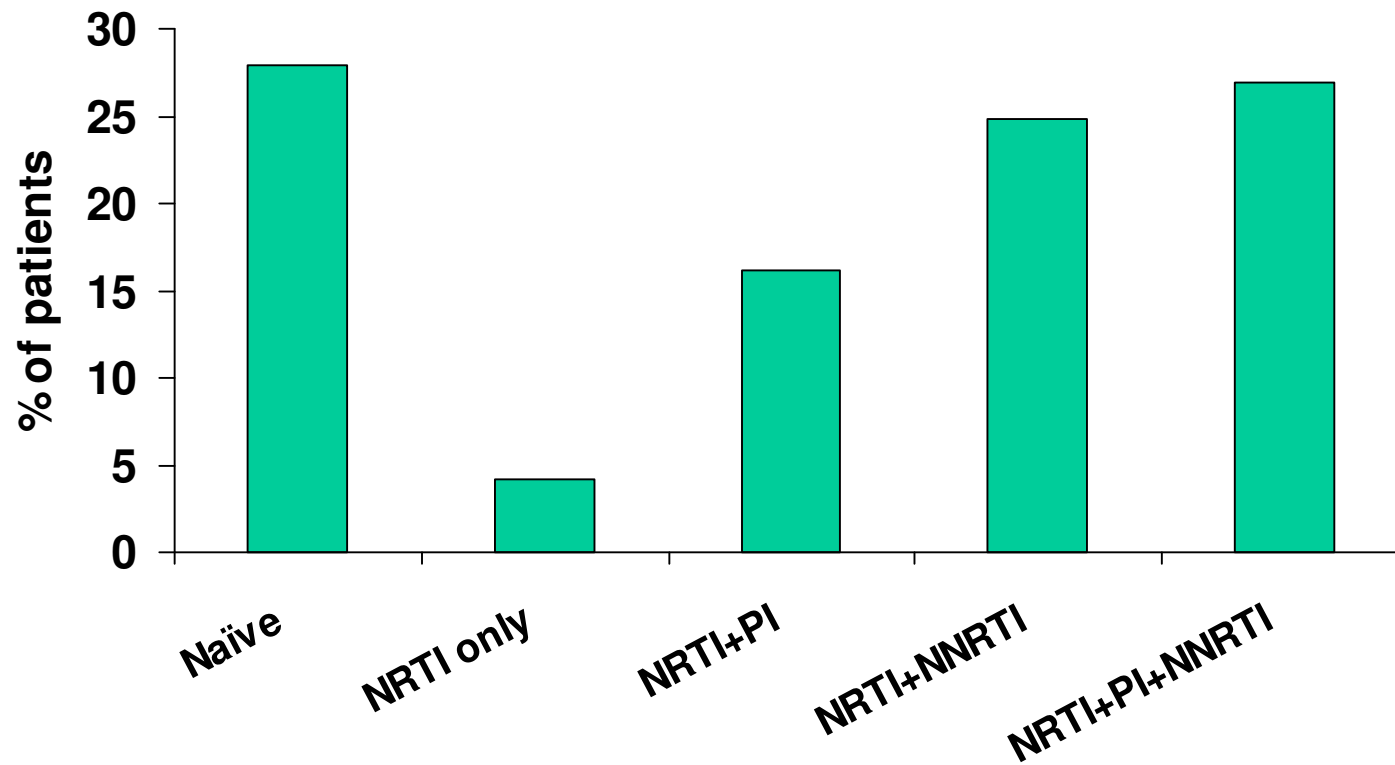
HIV RNA (copies/ml)



CD4 count (cells/mm³)

- Among patients under follow-up in 2001, 6620 (72.1%) had been exposed to at least one ARV drug.

Classes of drugs received



Comments

- **The process of cleaning and auditing this database continues and is expected to take a further three months.**
- **However, preliminary results from these cohorts have already provided interesting information on the characteristics of patients infected with HIV in the UK and their exposure to antiretroviral therapy**
- **After the initial database has been merged, it is anticipated that the dataset will be updated every 6-9 months to ensure that the information is as up-to-date as possible, and it is hoped that the database will ultimately expand to include other cohorts in the UK.**