

Methodology

All laboratories carrying out CD4 cell counts in England, Wales and Northern Ireland in 2007 that are registered with the National External Quality Assessment Scheme for Leukocyte Immunophenotyping (UK NEQAS) were invited to participate. The list of sites reporting to the CD4 surveillance Scheme is regularly checked with the NEQAS list and new sites contacted. Clinicians as well as the laboratory approve for the release of the data to be used.

Information requested consists of local **patient identifiers** (soundex codes of the patients' surname, date of birth, sex, and clinic/hospital/laboratory identification numbers) and details of **test results** (absolute CD4 count, CD4 test date, CD4 percentage and total lymphocyte count). Data are reported bi-annually, quarterly, or monthly to suit individual laboratories.

Counts from the same patient in each laboratory are linked using laboratory identification numbers. A process of **deduplication** is then carried out using soundex and date of birth to identify and link reports of patients that have had tests in more than one laboratory.

Patient records on the CD4 database are then matched to patients on the databases for new HIV diagnoses and the Survey of Prevalent HIV Infections Diagnosed (SOPHID) using hierarchical combinations of patient identifiers. This results in a large dataset that includes epidemiological information from the new diagnoses database (such as infection route, ethnicity etc), and SOPHID database (region of residence and treatment information) as well as a history of CD4 cell counts from the CD4 database.

The flow chart in figure 1 outlines this process.

Figure 1: Flow chart outlining the data collection and processing methodology.

