# The Survey of Prevalent HIV Infections Diagnosed (SOPHID)

## 1 What is SOPHID?

The Survey of Prevalent HIV Infections Diagnosed (SOPHID) began in 1995 and is a cross-sectional survey of all persons with diagnosed HIV infection who attend for HIV care at an NHS site in England, Wales and Northern Ireland (E, W & NI). Scottish data is collected by Health Protection Scotland. Paediatric data (children<16 years of age) were from the Institute of Child Health (ICH). All those data are incorporated to produce United Kingdom (UK) totals.

## 2 What is being measured?

Key output from SOPHID include:

1. Number of individuals living with a diagnosed HIV infection in the UK by age, sex, probably route of HIV transmission, ethnicity, ART status, CD4 cell count, region of residence and region of providing care
2. Diagnosed HIV prevalence among population aged 15 – 59 years old in Local Authorities in England
3. Together with New HIV and AIDS Diagnoses and Deaths Database and CD4 Surveillance Scheme, proportion of individuals diagnosed late (CD4 cell count less than 350 cells/mm$^3$ in Local Authorities are generated.
4. Quality of care indicators are generated annually to measure
   - Early access to HIV testing,
   - Prompt integration into care
   - Access to ART
   - Achieving viral load suppression
   - Preventing symptomatic infection
   - Retention into care

## 3 Why is it being measured?

Timely data from SOPHID provide essential information on the changing profile of people living with diagnosed HIV infection for public health monitoring, prevention monitoring and the commissioning of HIV-related services.

Three national decisions based on information from SOPHID include:

1. HIV testing guidelines (2008) indicate that primary care trusts with a diagnosed HIV prevalence over 2/1000 population should expand HIV testing from traditional GUM settings to community settings. SOPHID provides this information to commissioners and policy makers.
2. In 2010, data collected through SOPHID was used to develop the HIV outpatient care pathway as part of the Department of Health's payment by results initiative. Work is underway to ensure data fields to support the resultant tariff will be collected through SOPHID.
3. SOPHID is continually used by HIV and sexual health commissioners to fund and allocate HIV services.

## 4 How is the indicator defined?

Data were de-duplicated to ensure that a patient has a record of attendance in each calendar year. Patients can be tracked between clinics and over time.

## 5 Who does it measure?

All HIV-diagnosed individuals receiving NHS-funded HIV care within the survey period.

## 6 When does it measure it?

The survey is run bi-annually in London (includes Brighton, Hastings and Eastbourne) covering attendances from January to June and from July to December, whereas the survey outside London covers attendances for the whole calendar year.

## 7 Will it measure absolute numbers or proportions?

Results can be presented in absolute numbers, proportions or rate per 1,000 population.

## 8 Where does the data come from?

SOPHID returns collected and collated by the HIV & STI Department, PHE, Colindale.
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>8</strong></td>
<td><strong>How accurate and complete are the data?</strong></td>
<td>All patients receiving care in NHS-funded clinics are reported. Upon receipt of data it is checked using a validation process which outputs a list of records with missing or inconsistent data. The queries are ranked in 3 groups by order of importance: 1. missing or incorrect key fields which we cannot accept unless corrected 2. inconsistent data or data that has changed from a previous survey 3. data which we want to confirm is missing or unavailable These are then returned to the provider for correction and return. This process continues until the data is of sufficiently high quality to be accepted. Essential variables (patient identifiers and risk for HIV) are 100% complete and major variables (e.g. patient residence, ethnicity) are over 98% complete.</td>
</tr>
<tr>
<td><strong>9</strong></td>
<td><strong>Are there any known caveats/problems/weaknesses within the data?</strong></td>
<td>The open access nature of HIV clinics means HIV patients can and do access multiple sites, and the confidential nature of HIV means NHS numbers are not collected. Thus, a patient using different identifiers in different clinics cannot be identified.</td>
</tr>
<tr>
<td><strong>10</strong></td>
<td><strong>Are particular tests needed?</strong></td>
<td>SOPHID data are freely available to the public on our website. Due to the confidential nature of the data, local outputs are restricted to the level of detail appropriate to the data user and disseminated in accordance to data confidentiality guidelines: Website - SOPHID surveillance tables are presented at UK and regional level and also nationally by risk groups. In addition, HIV prevalence data calculated from the latest ONS Population Estimates are presented for every Local Authority in England. SOPHID aggregated tables at LA and PHE Centre level are available on the secure web portal to Local Authorities to provide aggregated breakdown of diagnosed HIV-infected individuals in their area. These tables are restricted only to the relevant LA. LAs are informed of the availability of these tables by their local PHE Centre. Disaggregate and anonymised SOPHID data are produced for each PHE Centre and published on the secure PHE web portal made only available to regional scientists and to appropriate contacts in Wales and Northern Ireland. Data requests – any analyses that are not already available via other means will be undertaken by the SOPHID team providing there are no disclosure issues and the time required to produce the data can be justified by it’s usefulness. Further information can be found at <a href="http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/HIV/">http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/HIV/</a> or please contact <a href="mailto:sophid.sophid@phe.gov.uk">sophid.sophid@phe.gov.uk</a>.</td>
</tr>
<tr>
<td><strong>11</strong></td>
<td><strong>Data availability</strong></td>
<td>Consultant: Valerie Delpech Scientific lead: Sarika Desai</td>
</tr>
<tr>
<td><strong>12</strong></td>
<td><strong>Frequency of data updates</strong></td>
<td>Annually updated</td>
</tr>
<tr>
<td><strong>13</strong></td>
<td><strong>Indicator lay summary</strong></td>
<td>In 2011, there were 73,659 people living with a diagnosed HIV infection (49,083 men and 24,576 women) who received HIV care in the UK, representing a 6.2% increase on the number seen in 2009 (69,348). The number and proportion of persons receiving ART has increased over the past decade. In 2011, 84% (61,275/73,376) of persons seen for HIV care were prescribed ART compared with 71% (20,577/29,162) in 2002. Since 2006, the number of persons receiving HIV care resident outside London (within England) has been greater than the number resident in London. This continues to be the case in 2011 (46%, 31,147 living in London and 54%, 36,548 living elsewhere in England).</td>
</tr>
</tbody>
</table>
Frequently Asked Questions

How does SOPHID complement other national HIV surveillance systems?

**Difference between newly diagnosed and prevalent diagnosed infections.**
The HIV and AIDS new diagnoses database collects information on individuals with newly diagnosed HIV and AIDS. Neither annual, nor cumulative reports of new HIV diagnoses can give a measure of individuals who are currently living with diagnosed HIV in the UK since these cannot take into account movement out of the UK or unreported deaths. SOPHID collects reports of individuals living with HIV who have been in contact with the NHS and is therefore a good measure of the annual prevalent HIV diagnosed population and the demand on services.

**Estimating the number of HIV-infected but undiagnosed individuals**
Some individuals infected with HIV and living in the UK have not yet been diagnosed. These individuals cannot be detected by either surveillance of new diagnoses or SOPHID. The total estimated national number of HIV-infected individuals (diagnosed and undiagnosed) can be estimated using multiple sources of information including SOPHID, unlinked anonymous surveys, the National Survey of Sexual Attitudes and Lifestyles and census data. Annual estimates of the total number of individuals living with (diagnosed and undiagnosed) HIV in the UK are available in the annual HIV report.

Why does the number of records I reported to SOPHID differ from the number of records in the tables I received back?
The following procedures may often cause final SOPHID figures to differ from those originally reported:

- The SOPHID survey can be used to allocate an individual to a LA if the residence information provided by the site of treatment maps to one LA only. Records where LA cannot be ascertained are marked as ‘not known’ and not allocated to an area of residence. This may result in a discrepancy between the number of individuals reported and final number included into LA/regional tables.
- A single site of treatment sometimes reports duplicate records of the same individual - these are identified and removed.
- If insufficient information was provided for an individual to enable de-duplication (no soundex, date of birth or sex) then the record is removed following consultation with the reporting site.
- If an individual has been reported but later found to have actually been seen outside of the survey period then the individual is removed following consultation with the reporting site.
- More than one record of the same individual are often reported from more than one site of treatment - these are identified and removed.

What data are collected for the SOPHID survey?
Fields we require on each individual are as follows:

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Details</th>
<th>Description / comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDEX</td>
<td>Soundex</td>
<td>Code of surname used to link reports from the same patient</td>
</tr>
<tr>
<td>DOB</td>
<td>Date of birth</td>
<td>Used to link reports from the same patient and to calculate age (dd/mm/yyyy)</td>
</tr>
<tr>
<td>SEX</td>
<td>Sex</td>
<td>Male or female</td>
</tr>
<tr>
<td>INIT</td>
<td>Initial</td>
<td>Used to link reports from same patient</td>
</tr>
<tr>
<td>CLINID</td>
<td>Clinic identification code</td>
<td>Treatment centre’s GUM clinic id code - used to link reports from the same patient</td>
</tr>
<tr>
<td>PCTres</td>
<td>PCT of residence</td>
<td>Primary Care Trust code - if postcode not available</td>
</tr>
<tr>
<td>LA/UAres</td>
<td>LA/UA of residence</td>
<td>Local/Unitary Authority Code - if postcode not available</td>
</tr>
<tr>
<td>POSTCODE</td>
<td>Postcode (last letter may be removed for added confidentiality)</td>
<td>Used for derivation of LA, Upper tier LA, PHE Centre and Region of residence.</td>
</tr>
<tr>
<td>SITE</td>
<td>Site of care</td>
<td>Place where patient received HIV-related care</td>
</tr>
<tr>
<td>PEXP</td>
<td>Infection route</td>
<td>How infection was probably acquired</td>
</tr>
<tr>
<td>DATEAIDS</td>
<td>Date of most recent AIDS</td>
<td>Date of diagnosis of most recent AIDS defining illness in survey period (not defined by CD4 count)</td>
</tr>
</tbody>
</table>
**ETHN**
Ethnic group classification (NHS classification accepted)

**ARV**
Anti-retroviral therapy

**ARVSTART**
Date of start of ARV

**CD4**
CD4 cell count per micro litre at date last seen.

**VL**
Most recent viral load

**VLDATE**
Date of most recent viral load

**DATEPOS**
Date first positive on site or date of first attendance on site

**PREVCARE**
Previous care at another site

**DLSEEN**
Date patient last seen at this site or date of death in the survey period

---

**Is reporting to SOPHID voluntary?**
Participation in SOPHID is not compulsory. However Department of Health allocations for each clinical commissioning group (CCG) include an adjustment for HIV-infected residents, which are calculated using SOPHID data: therefore it is in each CCG’s interest to respond. In addition, many London trusts have Service Level Agreements that contractually require timely SOPHID reporting.

**What about individuals who are seen at more than one centre? What is de-duplication?**
At the end of the year when all the data are received multiple reports of an individual are removed through the process of “de-duplication”. These are identified based on matched clinic-attributed patient ID number, soundex, date of birth and sex. Individuals who are seen at more than one centre are assigned to the centre at which they were seen most recently. All routine survey summaries are then prepared using this de-duplicated dataset.

**Are children included in the final data?**
Children reported by ICH include infants reported to have been born to HIV-infected mothers during the survey year regardless of their infection status. These infants will have received HIV-related care while their infection status was being established (many of whom will subsequently be confirmed as uninfected). Children who are uninfected or whose infection status has not yet been confirmed will not be included in the final numbers of individuals living with diagnosed HIV infection but are included in separate national tables.

**What confidentiality procedures are in place?**
We ask that all data be sent to us electronically via the secure HIV & STI web portal. All staff are briefed on Caldicott guidelines and are aware of the sensitive nature of the data. When ad hoc queries come in we do not provide any breakdowns below LA level.

To maintain patient confidentiality, soundex codes are used instead of names. All data are stored on restricted and secure databases, with strict adherence to the Data Protection Act and Caldicott Guidelines. SOPHID has approval under the section 60 regulations of the Health and Social Care Act 2001. For confidentiality reasons, data is not broken down within a region smaller than a LA and all outputs are presented in aggregate form. No identifiers are provided.

For further information on safeguarding the confidentiality of patient information whilst protecting public health please see: [http://www.hpa.org.uk/confidentiality/](http://www.hpa.org.uk/confidentiality/)
What is the HIV & STI Web Portal and how do I use it?
The web portal is a secure Internet site recommended by PHE, which allows transfer of data discreetly and confidentially to a secure server located within PHE. The SOPHID data can be placed into the SOPHID folder which is accessible only by SOPHID staff and themselves. For instructions on using the web portal please contact the SOPHID team (sophid.sophid@phe.gov.uk).

Can you provide data showing an individual's soundex?
No, no information with identifiers will be disseminated.