



GUMCAD and HIV Data Sharing Policies

GUMCAD Data Sharing Policy	Pages 1 - 7
HIV and AIDS New Diagnoses and Deaths Database Data Sharing Policy	Pages 8 - 14
SOPHID Data Sharing Policy	Pages 15 - 21
CD4 Surveillance Data Sharing Policy	Pages 22 - 28

Further Information

GUMCAD

Email: gumcad@hpa.org.uk

Web: http://www.hpa.org.uk/web/HPAweb&HPAwebStandard/HPAweb_C/1201265888302

HIV and AIDS New Diagnoses and Deaths Database

Email: HARSQueries@hpa.org.uk

Web: <http://www.hpa.org.uk/web/HPAweb&HPAwebStandard/Page/1201094588891>

SOPHID

Email: SOPHID@hpa.org.uk

Web: <http://www.hpa.org.uk/web/HPAweb&HPAwebStandard/Page/1201094588844>

CD4 Surveillance

Email: Alan.Hunter@hpa.org.uk

Web: <http://www.hpa.org.uk/web/HPAweb&Page&HPAwebAutoListName/Page/1201094588994>



GUMCAD DATA SHARING POLICY:

Document Version Control

Version:	6.1
Document Creation Date:	19 th September 2008
Document Approved Date:	10 th December 2008
Authors:	Gwenda Hughes, Andrew Chronias, Barry Evans, Fortune Ncube
Approved By:	GUMCAD Steering Group
Next date for review:	31 st June 2010 or as required.

Revision date	Version	Summary of Changes
31 st October 2008	2	Greater detail on use and sharing of data extracts and more background information on how data are used.
4 th November 2008	3	Minor editorial changes.
9 th December 2008	4	Minor editorial changes.
10 th December	5	Addition of further information on levels of breach and reporting protocols. Also consideration given to publication of anonymised PCT level data on a case by case basis. Minor editorial changes.
12 th March 2009	6	Change in policy on small cell sizes at SHA and national level in accordance with small cell sizes consultation document.
21 st July 2009	6.1	Minor editorial changes.

Contents

1. BACKGROUND TO POLICY	3
2. LEVELS OF ACCESS TO GUMCAD DATA	3
3. DATA SHARING GUIDELINES	4
3.1. <i>Principles for storing and accessing 'current' patient-level data</i>	4
3.2. <i>Download and analysis of patient-level data extracts</i>	5
3.3. <i>Presentation of local-level aggregated data.....</i>	5
3.4. <i>Presentation and publication of high-level aggregated data.....</i>	6
3.5. <i>Acknowledging data sources</i>	6
4. RESPONSIBILITIES FOR ADHERING TO POLICY	7

1. Background to policy

- 1.1. The Genitourinary Medicine Clinic Activity Dataset (GUMCAD) is used to monitor trends in new diagnoses of sexually transmitted infections (STI) and other sexual health problems and to determine which specific groups are at particular risk. This information is used to inform the public health response by:
- Improving the planning and management of services.
 - Developing, adapting and refining interventions.
 - Monitoring the effectiveness of sexual health policies.
- 1.2. GUMCAD extracts are submitted to the Health Protection Agency (HPA) Centre for Infections for processing and analysis. The GUMCAD return includes patient demographic details collected at patient registration at their first attendance at a GUM clinic, and clinical and risk factor data collected during the patient consultation. The data are pseudonymised i.e. they contain the patient's clinic (hospital) number but they do not contain patient-identifiable information such as name, date of birth, postcode etc. However they are sensitive patient-level data and their storage and access to them requires strict control. Currently only HPA has approval from the Patient Information Advisory Group (PIAG) to handle patient-level GUMCAD data.
- 1.3. In addition, following the recent ONS guidance on data disclosure, the Small Cell Sizes subgroup of the HPA Caldicott Group have produced general guidance for the publication of HPA data for consultation¹. As a result, HPA rules on the publication or release of aggregated STI data including GUMCAD (on the website or otherwise) were changed.
- 1.4. The GUMCAD Steering Group and the Small Cell Sizes subgroup of the HPA Caldicott Group have agreed the following policy with respect to storing, access, sharing and use of patient level and aggregated GUMCAD data which is in line with Caldicott principles, ONS guidance and the HPA Caldicott Group consultation paper. This policy will be reviewed annually and updated as required.

2. Levels of access to GUMCAD data

- 2.1. GUMCAD data can be accessed at different levels according to purpose. The different levels of access are outlined below.
- 2.1.1. **High-level aggregated data.** These data are presented in tables, graphs or maps which summarise trends and diagnosis rates by various patient characteristics at the national, regional or Strategic Health Authority (SHA) level. These data are usually made available publicly on the HPA website.

¹ Health Protection Agency. Internal consultation on policy on sharing and dissemination of datasets with small cell sizes. January 2009.

- 2.1.2. **Local-level aggregated data.** These data are presented in tables, graphs or maps which summarise trends and diagnosis rates by various patient characteristics at the Primary Care Trust (PCT), Local Authority (LA), Lower Super Output Area Level (LSOA) and GUM clinic level. These data are distributed, in confidence, to a range of stakeholders within the NHS, Department of Health and HPA for the purposes of planning and managing services, developing interventions and monitoring the effectiveness of health policies. These data may sometimes be made available publicly on the HPA website, although restrictions apply (see section 3.3.2-3.3.3).
 - 2.1.3. **Patient-level data extracts** containing data on individual patient episodes. Data extracts are not updated and are generally used by epidemiologists, data analysts and statisticians for epidemiological or statistical analyses.
 - 2.1.4. **'Current' patient-level data** held in the GUMCAD database or on the mirror server. These data are regularly updated and may be accessed by the HPA database administrator for data management, information officers and epidemiologists for running queries and producing data extracts for analysis, and by software engineers developing standard reports which run from the database.
- 2.2. The agreed policy for accessing GUMCAD data at these different levels is detailed in the following guidelines.

3. Data sharing guidelines

3.1. Principles for storing and accessing 'current' patient-level data

- 3.1.1. GUMCAD data will be directly uploaded to the HPA by GUM clinic staff via the HPA MESH (Microbiology and Epidemiology of STIs and HIV) web portal and stored in an SQL database held on a secure server at the Colindale site.
- 3.1.2. Access to 'current' patient-level GUMCAD data will be restricted to a limited list of named HPA staff in Cfi and LRS (nominated users). Such 'full access' users will each have a personal user name and password*.
- 3.1.3. Users will be nominated only by the appropriate data custodian: the Head (or deputy) of STI Surveillance at Cfi, and by the Regional Epidemiologist (or deputy) with responsibility for STIs and HIV in LRS. The list of nominated users will be jointly reviewed by the data custodians each year*.
- 3.1.4. Nominated users will access data relevant to their respective areas of responsibility. For example, Regional Epidemiologists will be able to access patient-level data on patients attending GUM clinics in their region as well as on patients residing in their region who travelled elsewhere to access a clinic.
- 3.1.5. Regional nominated users will directly access patient-level data held on the mirror server at the Colindale site.

- 3.1.6. Non-HPA public health staff (e.g. Public Health Observatories, Primary Care Trusts etc.) with legitimate reasons for analysing the data may access the data in collaboration with and under the supervision of the appropriate local, regional or national HPA body. Access will be restricted to a nominated user who is accountable to HPA. Any work undertaken by a non-HPA body will be governed by a memorandum of understanding between both organisations.
- 3.1.7. 'Current' patient-level GUMCAD data must never be accessed by non-nominated users.
- 3.1.8. Joint posts between the HPA and other public health organisations seeking to undertake analyses should actively be sought to facilitate collaborative working.

*To obtain a username and password please contact your regional HPA office.

3.2. Download and analysis of patient-level data extracts

- 3.2.1. Data extracts of patient-level information for epidemiological analysis must be downloaded by nominated users to a password-protected network drive on a secure server. Downloaded files must be password protected.
- 3.2.2. Data extracts should never be held on computer hard disks, laptops or any transportable storage media such as CDs or memory sticks.
- 3.2.3. Data extracts should be deleted immediately after the purpose for which they were downloaded has been completed. Where extensive analyses are required, nominated users must request permission to keep data extracts beyond 1 year from the appropriate data custodian. Data extracts must not be held for longer than 2 years.
- 3.2.4. Data extracts must not be shared with non-nominated users. If complex statistical analysis of the data by non-nominated users is required, patient GUM clinic numbers must be removed from the data file and replaced with dummy numbers prior to analysis by the elected non-nominated user. Such stripped down files must be managed according to principles 3.2.1-3.2.3 above.
- 3.2.5. Analyses of GUMCAD data will be strictly for the purpose of improving public health.

3.3. Presentation of local-level aggregated data

- 3.3.1. STI data tables at the level of GUM clinic, LSOA, LA and PCT will be distributed in confidence to relevant organisations within the NHS, Department of Health, local government and HPA. Data will also be accessed using GUMCAD automated reports via the HPA MESH web portal *.

*To obtain a username and password please contact your regional HPA office.

- 3.3.2. Requests to publish LA and PCT level data, in hard copy or on the website, should be reviewed on a case by case basis by the appropriate data custodian in consultation with the CfI or relevant regional Caldicott Guardian and the Head of STI Surveillance at CfI.
- 3.3.3. STI data tables at the level of LA and PCT may only be published provided small cell sizes are suitably anonymised. Cells with values between 1 and 4 inclusive must be anonymised with an asterix. In addition, where the anonymised cell could be deduced from the total, the next smallest cell size in the same row and/or column must also be anonymised.
- 3.3.4. HPA will review the operation of the ONS policy within the HPA as required.
- 3.3.5. When PCT or clinic level data are requested by other organisations (such as for Freedom of Information requests or Parliamentary Questions), small cell sizes must be anonymised according to the protocol described in 3.3.3, above.
- 3.3.6. Maps of STI rates by PCT of residence, where rates are grouped into categories, may be published in reports and on the HPA website.

3.4. Presentation and publication of high-level aggregated data

- 3.4.1. STI data tables at the level of Strategic Health Authority (SHA) and above may be published by the HPA, in hard copy or on the website. Anonymisation would not usually be required but may be considered appropriate in some cases.

3.5. Acknowledging data sources

- 3.5.1. Any analysis undertaken using GUMCAD data which is published in reports, peer-reviewed journals or on the website must acknowledge the data source. “Data from the Genitourinary Medicine Clinic Activity Dataset (GUMCAD), Health Protection Agency”.
- 3.5.2. Academic research presenting data from individual clinics must not be published in peer-reviewed journals without either the prior consent or the collaboration of the lead consultants at those clinics (preferably, the latter).

4. Responsibilities for adhering to policy

- 4.1.1. The GUMCAD data custodian at CfI is responsible for ensuring that patient-level GUMCAD data are held, managed and accessed at the Colindale site in line with Caldicott principles and according to these guidelines.
- 4.1.2. Regional GUMCAD data custodians are responsible for ensuring that patient-level GUMCAD data for their region are managed and accessed in line with Caldicott principles and according to these guidelines.
- 4.1.3. The GUMCAD data custodian at CfI is responsible for ensuring that high-level aggregated data are appropriately anonymised prior to publication on the HPA website or in national reports.
- 4.1.4. Regional GUMCAD data custodians are responsible for ensuring that high-level aggregated data are appropriately anonymised prior to publication on regional websites or in regional reports.
- 4.1.5. Any breaches of this policy should be reported immediately, as follows:
 - 4.1.5.1. Serious breaches, such as loss of patient-level GUMCAD data due to storage on non-permitted media (e.g. CDs, memory sticks etc.), should be reported to the relevant HPA (i.e. CfI or regional) and hospital, community or primary care trust Caldicott Guardians. Reports of serious breaches should be copied to all CfI and regional Caldicott Guardians and to the STI Surveillance lead at CfI.
 - 4.1.5.2. Breaches involving the publication of tabular aggregated data at PCT level or below which has not been anonymised should be reported to the appropriate CfI or regional Caldicott Guardian and copied to the STI Surveillance lead at CfI.
 - 4.1.5.3. Minor breaches such as the publication of anonymised tabular aggregated data at PCT level or below without permission from the relevant data custodian should be reported to the appropriate CfI or regional Caldicott Guardian and copied to the STI Surveillance lead at CfI.



HIV AND AIDS NEW DIAGNOSES AND DEATHS DATABASE

DATA SHARING POLICY:

Document Version Control

Version:	2
Document Creation Date:	16 th October 2009
Document Approved Date:	5 th November 2009
Authors:	Brian Rice, Sonia Ribeiro
Approved By:	Valerie Delpech
Next date for review:	5 th November 2010 or as required.

Revision date	Version	Summary of Changes
19/10/2009	2	Minor editorial changes
12/11/2009	2	Minor editorial changes

Contents

1. BACKGROUND TO POLICY	10
2. LEVELS OF ACCESS TO NEW HIV AND AIDS DIAGNOSES DATA ...	11
3. DATA SHARING GUIDELINES	12
3.1. <i>Principles for storing and accessing patient-level data.....</i>	<i>12</i>
3.2. <i>Analysis of patient-level data extracts</i>	<i>12</i>
3.3. <i>Presentation of local-level aggregated data.....</i>	<i>13</i>
3.4. <i>Presentation and publication of high-level aggregated data.....</i>	<i>13</i>
3.5. <i>Acknowledging data sources</i>	<i>13</i>
4. RESPONSIBILITIES FOR ADHERING TO POLICY	14

1. Background to policy

1.1. The HIV and AIDS new diagnoses and deaths database collects information from voluntary laboratory and clinician reports on new diagnoses of HIV, first AIDS diagnosis, and deaths in HIV-infected individuals aged 15 years and over at diagnosis made in England, Wales and Northern Ireland. Information on children is collected separately by the Institute of Child Health (ICH), and data for Scotland is collected by Health Protection Scotland (HPS) and collated at the Health Protection Agency (HPA) Centre for Infections biannually. The HIV and AIDS new diagnoses database is used to monitor trends in new HIV and AIDS diagnoses and deaths and to determine which specific groups are at particular risk. This information is used to inform the public health response by:

- Improving the planning and management of HIV testing and services.
- Developing, adapting and refining interventions.
- Monitoring the effectiveness of sexual health policies.

1.2. New HIV and AIDS diagnoses data are submitted to the HPA Centre for Infections for processing and analysis. Epidemiological information is collected to describe the characteristics of those newly diagnosed with HIV. The data are part-pseudonymised i.e. they contain the patient's Soundex code (4-character code of surname), date of birth and clinic number but do not contain their name or residence postcode. Access to and storage of these highly sensitive patient-level data are under strict control.

Information collected includes:

- Demographics - age, sex, ethnicity, and country of birth;
- Patient's probable route and country of infection;
- Contact's probable route and country of infection;
- Reasons for seeking an HIV test;
- Clinical details – AIDS events, cause of death, CD4/VL count at diagnosis.

1.3. In addition, following the recent ONS guidance on data disclosure, the Small Cell Size subgroup of the HPA Caldicott Group have produced general guidance for the publication of HPA data for consultation². As a result, HPA rules on the publication or release of aggregated HIV data were changed.

1.4. The Small Cell Sizes subgroup have agreed the following policy with respect to storing, access, sharing and use of patient level and aggregated HIV data which is in line with Caldicott principles, ONS guidance and the HPA Caldicott Group consultation paper. This policy will be reviewed annually and updated as required.

² Health Protection Agency. Internal consultation on policy on sharing and dissemination of datasets with small cell sizes. January 2009.

2. Levels of access to new HIV and AIDS diagnoses data

2.1. New HIV and AIDS diagnoses data can be accessed at different levels according to purpose. The different levels of access are outlined below.

2.1.1. **High-level aggregated data.** These data are presented in tables or graphs which summarise trends and number of diagnoses by various patient characteristics at the country, regional or Strategic Health Authority (SHA) level. These data are available publicly on the HPA website.

2.1.2. **Local-level aggregated data.** These data are not available to the public and requests for this type of data should be made to the local Health Protection Unit. This is subject to Terms and Conditions of Use (see 2.1.4).

2.1.3. **Patient-level data extracts.** Disaggregate data may only be made available under the following circumstances:

- To those from whom data was received and only for the purpose of improving the completeness and accuracy of data from that source;
- To those who have sought and received approval from the reporting sites from which the data is being requested (documentation of this agreement to be forwarded with the collaborative research proposal form);
- To regional Health Protection Units for the purpose of pre-agreed local analyses. In this instance the data should be kept and used with the same level of security as at the HPA Centre for Infections and should be destroyed as soon as analyses are complete.

2.1.4. **Ad-hoc (aggregate) data request.** If routine data outputs do not satisfy needs aggregate data can be requested upon completion of a data request form. When requesting data please consider the following:

- Whether or not confidentiality is compromised;
- Whether the data request is appropriate to the expressed purpose;
- Whether the variable(s) requested have a low completion rate;
- Whether the length of time extracting the data is justified by its usefulness (considered on a case by case basis);
- Whether a new variable needs to be created to satisfy the data request (such a data request can be included as a Collaborative Research Proposal)
- Whether the data request is clearly defined to avoid further requests in a short period of time;

2.2. The agreed policy for accessing new HIV and AIDS diagnoses data at these different levels is detailed in the following guidelines.

3. Data sharing guidelines

3.1. Principles for storing and accessing patient-level data

- 3.1.1. New HIV and AIDS diagnoses data will be directly uploaded to the HPA by clinic staff via the HPA MESH (Microbiology and Epidemiology of STIs and HIV) web portal and stored in an SQL database held on a secure server at the Colindale site.
- 3.1.2. Access to new HIV and AIDS diagnoses data will be restricted to a limited list of named HPA staff at the Colindale site. Such 'full access' users will each have a personal user name and password.
- 3.1.3. Patient level new HIV and AIDS diagnoses data must never be accessed by non-nominated users.
- 3.1.4. Joint posts between the HPA and other public health organisations seeking to undertake analyses should actively be sought to facilitate collaborative working.

3.2. Analysis of patient-level data extracts

- 3.2.1. Data extracts of patient-level information for epidemiological analysis must be extracted by nominated users to a password-protected network drive on a secure server. Files must be password protected.
- 3.2.2. Data extracts should never be held on computer hard disks, laptops or any transportable storage media such as CDs or memory sticks.
- 3.2.3. Data extracts should be deleted immediately after the purpose for which they were extracted has been completed. Where extensive analyses are required, nominated users must request permission to keep data extracts beyond one year from the appropriate data custodian. Data extracts must not be held for longer than two years.
- 3.2.4. Data extracts must not be shared with non-nominated users. If complex statistical analysis of the data by non-nominated users is required, patient identifiers (soundex, date of birth and clinic numbers) must be removed from the data file prior to analysis by the elected non-nominated user. Such stripped down files must be managed according to principles 3.2.1-3.2.3 above.
- 3.2.5. Analyses of new HIV and AIDS diagnoses data will be strictly for the purpose of improving public health.

3.3. Presentation of local-level aggregated data

- 3.3.1. Primary Care Trust (PCT) level new HIV and AIDS diagnoses data should be requested from the local Health Protection Unit. This is subject to Terms and Conditions of Use (see 2.1.4).
- 3.3.2. New HIV and AIDS diagnoses data tables at the level of PCT may only be released provided small cell sizes are suitably anonymised. Cells with values between one and four inclusive must be anonymised. In addition, where the anonymised cell could be deduced from the total, the next smallest cell size in the same row and/or column must also be anonymised.
- 3.3.3. The operation of the ONS policy within the HPA will be reviewed as required.
- 3.3.4. When PCT or clinic level data are requested by other organisations (such as for Freedom of Information requests or Parliamentary Questions), small cell sizes must be anonymised according to the protocol described in 3.3.2.

3.4. Presentation and publication of high-level aggregated data

- 3.4.1. New HIV and AIDS diagnoses data tables at the level of country and SHAs may be published by the HPA, in hard copy or on the website. Anonymised data would not usually be required but may be considered appropriate in some cases.

3.5. Acknowledging data sources

- 3.5.1. Any analysis undertaken using new HIV and AIDS diagnoses data which is published in reports, peer-reviewed journals or on the website must acknowledge the data source. “Data from the New HIV and AIDS Diagnoses and Deaths database, Health Protection Agency Centre for Infections”.
- 3.5.2. Academic research presenting data from individual clinics must not be published in peer-reviewed journals without either the prior consent or the collaboration of the lead consultants at those clinics (preferably, the latter).

4. Responsibilities for adhering to policy

- 4.1.1. The new HIV and AIDS diagnoses data custodian at the Centre for Infections is responsible for ensuring that patient-level new HIV and AIDS diagnoses data are held, managed and accessed at the Colindale site in line with Caldicott principles and according to these guidelines. The data custodian is also responsible for ensuring that high-level aggregated data are appropriately anonymised prior to publication on the national website and/or in national reports.
- 4.1.2. Regional new HIV and AIDS diagnoses data custodians are responsible for ensuring that patient-level new HIV and AIDS diagnoses data for their region are managed and accessed in line with Caldicott principles and according to these guidelines. The regional custodian is also responsible for ensuring that high-level aggregated data are appropriately anonymised prior to publication on regional websites or in regional reports.
- 4.1.3. Any breaches of this policy should be reported immediately, as follows:
 - 4.1.3.1. Serious breaches, such as loss of patient-level new HIV and AIDS diagnoses data due to storage on non-permitted media (e.g. CDs or memory sticks), should be reported to the relevant HPA (i.e. Cfi or regional) and hospital, community or PCT Caldicott Guardians. Reports of serious breaches should be copied to all Cfi and regional Caldicott Guardians and to the HIV Surveillance lead at Cfi.
 - 4.1.3.2. Breaches involving the publication of tabular aggregated data at PCT level or below which has not been anonymised should be reported to the appropriate Cfi or regional Caldicott Guardian and copied to the HIV Surveillance lead at Cfi.
 - 4.1.3.3. Minor breaches such as the publication of anonymised tabular aggregated data at PCT level or below without permission from the relevant data custodian should be reported to the appropriate Cfi or regional Caldicott Guardian and copied to the STI Surveillance lead at Cfi.



SOPHID DATA SHARING POLICY:

Document Version Control

SOPHID DATA SHARING POLICY:	
Document Version Control	
Version:	1
Document Creation Date:	9 th November 2009
Document Approved Date:	11 th November 2009
Authors:	Alison Brown
Approved By:	Valerie Delpech
Next date for review:	11 th November 2010 or as required.

Revision date	Version	Summary of Changes
12/11/2009	1	Minor editorial changes

Contents

1. BACKGROUND TO POLICY	17
2. LEVELS OF ACCESS TO SOPHID DATA	18
3. DATA SHARING GUIDELINES	19
3.1. <i>Principles for storing and accessing patient-level data.....</i>	<i>19</i>
3.2. <i>Analysis of patient-level data extracts</i>	<i>19</i>
3.3. <i>Presentation of local-level aggregated data.....</i>	<i>20</i>
3.4. <i>Presentation and publication of high-level aggregated data.....</i>	<i>20</i>
3.5. <i>Acknowledging data sources</i>	<i>20</i>
4. RESPONSIBILITIES FOR ADHERING TO POLICY	21

1. Background to policy

1.1. The Survey of prevalent HIV infections diagnosed (SOPHID) collects data on patients with diagnosed HIV infection attending treatment and care services in England, Wales and Northern Ireland. Information on children with diagnosed HIV infection is collected separately by the Institute of Child Health (ICH), and data for Scotland is collected by Health Protection Scotland (HPS) and collated at the Health Protection Agency Centre for Infections annually. The data are used to census the number of people living with diagnosed HIV infection, and to describe them in terms of their: age group, risk group, ethnicity and residence. Data are used to:

- Estimate and characterise the number of people living with HIV infection in the UK
- Improve the commissioning, planning and management of HIV testing and services.
- Develop, adapting and refining interventions.
- Develop clinical outcomes to audit the quality of services
- Monitor the effectiveness of sexual health policies.

1.2. SOPHID data are submitted to the Health Protection Agency (HPA) Centre for Infections for processing and analysis. Epidemiological information is collected to describe the characteristics of diagnosed HIV-infected individuals accessing treatment and care. The data are part-pseudonymised i.e. they contain the patient's Soundex code (4-character code of surname), date of birth, postcode and clinic number but does not contain their name. They are highly sensitive patient-level data and their storage and access to them requires strict control.

Information collected includes:

- Demographics - age, sex, and ethnicity, country of birth, residence;
- Patient's exposure
- ARV treatment
- Markers of clinical progression – CD4 count, viral load and most advanced clinical stage reached

- 1.3. Following the recent ONS guidance on data disclosure, the Small Cell Sizes subgroup of the HPA Caldicott Group have produced general guidance for the publication of HPA data for consultation³. As a result, HPA rules on the publication or release of aggregated HIV data were changed.
- 1.4. The Small Cell Sizes subgroup of the HPA Caldicott Group have agreed the following policy with respect to storing, access, sharing and use of patient level and aggregated HIV data which is in line with Caldicott principles, ONS guidance and the HPA Caldicott Group consultation paper. This policy will be reviewed annually and updated as required.

2. Levels of access to SOPHID data

- 2.1. SOPHID data can be accessed at different levels according to purpose. The different levels of access are outlined below.
 - 2.1.1. **High-level aggregated data.** These data are presented in tables or graphs which summarise trends and number of diagnoses by various patient characteristics at the country, regional or Strategic Health Authority (SHA) level. These data are available publicly on the HPA website.
 - 2.1.2. **Local-level aggregated data.** These data are not available to the public and requests for this type of data should be made to the local Health Protection Unit. This is subject to Terms and Conditions of Use (see 2.1.4).
 - 2.1.3. **Patient-level data extracts.** Disaggregate data may only be made available under the following circumstances:
 - To those from whom data was received and only for the purpose of improving the completeness and accuracy of data from that source;
 - To those who have sought and received agreement (such agreement to be forwarded with the collaborative research proposal form from the reporting sites for which data is being requested);
 - To regional Health Protection Units for the purpose of pre-agreed local analyses. In this instance the data should be kept and used with the same level of security as at the Health Protection Agency Centre for Infections and should be destroyed as soon as analyses are complete.
 - 2.1.4. **Ad-hoc (aggregate) data request.** If routine data outputs do not satisfy needs aggregate data can be requested upon completion of a data request form. When requesting data please consider the following:
 - Whether or not confidentiality is compromised;
 - Whether the data request is appropriate to the expressed purpose;
 - Whether the variable(s) requested have a low completion rate;

³ Health Protection Agency. Internal consultation on policy on sharing and dissemination of datasets with small cell sizes. January 2009.

- Whether the length of time extracting the data is justified by its usefulness (considered on a case by case basis);
 - Whether a new variable needs to be created to satisfy the data request (such a data request can be included as a Collaborative Research Proposal)
 - Whether the data request is clearly defined to avoid further requests in a short period of time;
- 2.2. The agreed policy for accessing SOPHID data at these different levels is detailed in the following guidelines.

3. Data sharing guidelines

3.1. Principles for storing and accessing patient-level data

- 3.1.1. SOPHID data will be directly uploaded to the HPA by clinic staff via the HPA MESH (Microbiology and Epidemiology of STIs and HIV) web portal and stored in an SQL database held on a secure server at the Colindale site.
- 3.1.2. Access to SOPHID data will be restricted to a limited list of named HPA staff in CfI. Such 'full access' users will each have a personal user name and password.
- 3.1.3. Patient level SOPHID data must never be accessed by non-nominated users.
- 3.1.4. Joint posts between the HPA and other public health organisations seeking to undertake analyses should actively be sought to facilitate collaborative working.

3.2. Analysis of patient-level data extracts

- 3.2.1. Data extracts of patient-level information for epidemiological analysis must be extracted by nominated users to a password-protected network drive on a secure server. Files must be password protected.
- 3.2.2. Data extracts should never be held on computer hard disks, laptops or any transportable storage media such as CDs or memory sticks.
- 3.2.3. Data extracts should be deleted immediately after the purpose for which they were extracted has been completed. Where extensive analyses are required, nominated users must request permission to keep data extracts beyond 1 year from the appropriate data custodian. Data extracts must not be held for longer than 2 years.
- 3.2.4. Data extracts must not be shared with non-nominated users. If complex statistical analysis of the data by non-nominated users is required, patient

identifiers (soundex, date of birth and clinic numbers) must be removed from the data file prior to analysis by the elected non-nominated user. Such stripped down files must be managed according to principles 3.2.1-3.2.3 above.

- 3.2.5. Analyses of SOPHID data will be strictly for the purpose of improving public health.

3.3. Presentation of local-level aggregated data

- 3.3.1. PCT level SOPHID data should be requested from the local Health Protection Unit. This is subject to Terms and Conditions of Use (see 2.1.4).
- 3.3.2. SOPHID data tables at the level of PCT may only be released provided small cell sizes are suitably anonymised. Cells with values between 1 and 4 inclusive must be anonymised with an asterisk. In addition, where the anonymised cell could be deduced from the total, the next smallest cell size in the same row and/or column must also be anonymised.
- 3.3.3. HPA will review the operation of the ONS policy within the HPA as required.
- 3.3.4. When PCT or clinic level data are requested by other organisations (such as for Freedom of Information requests or Parliamentary Questions), small cell sizes must be anonymised according to the protocol described in 3.3.2.

3.4. Presentation and publication of high-level aggregated data

- 3.4.1. SOPHID data tables at the level of country and Strategic Health Authority (SHA) may be published by the HPA, in hard copy or on the website. Anonymisation would not usually be required but may be considered appropriate in some cases.

3.5. Acknowledging data sources

- 3.5.1. Any analysis undertaken using SOPHID data which is published in reports, peer-reviewed journals or on the website must acknowledge the data source. “Data from the SOPHID: Health Protection Agency”.
- 3.5.2. Academic research presenting data from individual clinics must not be published in peer-reviewed journals without either the prior consent or the collaboration of the lead consultants at those clinics (preferably, the latter).

4. Responsibilities for adhering to policy

- 4.1.1. The SOPHID data custodian at CfI is responsible for ensuring that patient-level SOPHID data are held, managed and accessed at the Colindale site in line with Caldicott principles and according to these guidelines.
- 4.1.2. Regional SOPHID data custodians are responsible for ensuring that patient-level SOPHID data for their region are managed and accessed in line with Caldicott principles and according to these guidelines.
- 4.1.3. The SOPHID data custodian at CfI is responsible for ensuring that high-level aggregated data are appropriately anonymised prior to publication on the HPA website or in national reports.
- 4.1.4. Regional SOPHID data custodians are responsible for ensuring that high-level aggregated data are appropriately anonymised prior to publication on regional websites or in regional reports.
- 4.1.5. Any breaches of this policy should be reported immediately, as follows:
 - 4.1.5.1. Serious breaches, such as loss of patient-level SOPHID data due to storage on non-permitted media (e.g. CDs, memory sticks etc.), should be reported to the relevant HPA (i.e. CfI or regional) and hospital, community or primary care trust Caldicott Guardians. Reports of serious breaches should be copied to all CfI and regional Caldicott Guardians and to the HIV Surveillance lead at CfI.
 - 4.1.5.2. Breaches involving the publication of tabular aggregated data at PCT level or below which has not been anonymised should be reported to the appropriate CfI or regional Caldicott Guardian and copied to the HIV Surveillance lead at CfI.
 - 4.1.5.3. Minor breaches such as the publication of anonymised tabular aggregated data at PCT level or below without permission from the relevant data custodian should be reported to the appropriate CfI or regional Caldicott Guardian and copied to the STI Surveillance lead at CfI.



CD4 SURVEILLANCE DATA SHARING POLICY:

Document Version Control

Version:	1
Document Creation Date:	9 th November 2009
Document Approved Date:	11 th November 2009
Authors:	Alison Brown
Approved By:	Valerie Delpech
Next date for review:	11 th November 2010 or as required.

Revision date	Version	Summary of Changes
12/11/2009	1	Minor editorial changes

Contents

1. BACKGROUND TO POLICY	24
2. LEVELS OF ACCESS TO NEW HIV AND AIDS DIAGNOSES DATA ...	25
3. DATA SHARING GUIDELINES	26
3.1. <i>Principles for storing and accessing patient-level data.....</i>	26
3.2. <i>Analysis of patient-level data extracts</i>	26
3.3. <i>Presentation of local-level aggregated data.....</i>	27
3.4. <i>Presentation and publication of high-level aggregated data.....</i>	27
3.5. <i>Acknowledging data sources</i>	27
4. RESPONSIBILITIES FOR ADHERING TO POLICY	28

1. Background to policy

1.1. The CD4 surveillance scheme aims to measure the level of immunosuppression among patients with diagnosed HIV infection. Data on patient's CD4 counts are collected from over 60 laboratories in England, Wales and Northern Ireland. Data for Scotland is collected by Health Protection Scotland (HPS) and collated at the Health Protection Agency Centre for Infections annually. Data are used to:

- Monitor levels of immuno-suppression among patients living with diagnosed HIV infection in the UK, by risk group and over time
- In conjunction with data from the HIV and AIDS new diagnoses and deaths, monitor the proportion of newly HIV diagnosed patients who present at a late stage of infection (CD4 counts < 200 cells per mm³ within 90 days of diagnosis).

1.2. CD4 surveillance data are submitted to the Health Protection Agency (HPA) Centre for Infections for processing and analysis. The data are part-pseudonymised i.e. they contain the patient's Soundex code (4-character code of surname), date of birth, sex and laboratory number but does not contain their name or residence. They are highly sensitive patient-level data and their storage and access to them requires strict control.

Information collected includes:

- Demographics - age, and sex
- CD4 counts
- Data are routinely linked to SOPHID and HIV and AIDS new diagnoses and death data

1.3. Following the recent ONS guidance on data disclosure, the Small Cell Sizes subgroup of the HPA Caldicott Group has produced general guidance for the publication of HPA data for consultation⁴. As a result, HPA rules on the publication or release of aggregated HIV data were changed.

1.4. The Small Cell Sizes subgroup of the HPA Caldicott Group have agreed the following policy with respect to storing, access, sharing and use of patient level and aggregated HIV data which is in line with Caldicott principles, ONS guidance and the HPA Caldicott Group consultation paper. This policy will be reviewed annually and updated as required.

⁴ Health Protection Agency. Internal consultation on policy on sharing and dissemination of datasets with small cell sizes. January 2009.

2. Levels of access to New HIV and AIDS diagnoses data

- 2.1. CD4 data can be accessed at different levels according to purpose. The different levels of access are outlined below.
- 2.1.1. **High-level aggregated data.** These data are presented in tables or graphs which summarise trends and number of diagnoses by various patient characteristics at the country, regional or Strategic Health Authority (SHA) level. These data are available publicly on the HPA website.
- 2.1.2. **Local-level aggregated data.** These data are not available to the public and requests for this type of data should be made to the local Health Protection Unit. This is subject to Terms and Conditions of Use (see 2.1.4).
- 2.1.3. **Patient-level data extracts.** Disaggregate data may only be made available under the following circumstances:
- To those from whom data was received and only for the purpose of improving the completeness and accuracy of data from that source;
 - To those who have sought and received agreement (such agreement to be forwarded with the collaborative research proposal form from the reporting sites for which data is being requested);
 - To regional Health Protection Units for the purpose of pre-agreed local analyses. In this instance the data should be kept and used with the same level of security as at the Health Protection Agency Centre for Infections and should be destroyed as soon as analyses are complete.
- 2.1.4. **Ad-hoc (aggregate) data request.** If routine data outputs do not satisfy needs aggregate data can be requested upon completion of a data request form. When requesting data please consider the following:
- Whether or not confidentiality is compromised;
 - Whether the data request is appropriate to the expressed purpose;
 - Whether the variable(s) requested have a low completion rate;
 - Whether the length of time extracting the data is justified by its usefulness (considered on a case by case basis);
 - Whether a new variable needs to be created to satisfy the data request (such a data request can be included as a Collaborative Research Proposal)
 - Whether the data request is clearly defined to avoid further requests in a short period of time;
- 2.2. The agreed policy for accessing CD4 data at these different levels is detailed in the following guidelines.

3. Data sharing guidelines

3.1. Principles for storing and accessing patient-level data

- 3.1.1. CD4 surveillance data will be directly uploaded to the HPA by clinic staff via the HPA MESH (Microbiology and Epidemiology of STIs and HIV) web portal and stored in an SQL database held on a secure server at the Colindale site.
- 3.1.2. Access to CD4 surveillance data will be restricted to a limited list of named HPA staff in CfI. Such 'full access' users will each have a personal user name and password.
- 3.1.3. Patient level CD4 surveillance data must never be accessed by non-nominated users.
- 3.1.4. Joint posts between the HPA and other public health organisations seeking to undertake analyses should actively be sought to facilitate collaborative working.

3.2. Analysis of patient-level data extracts

- 3.2.1. Data extracts of patient-level information for epidemiological analysis must be extracted by nominated users to a password-protected network drive on a secure server. Files must be password protected.
- 3.2.2. Data extracts should never be held on computer hard disks, laptops or any transportable storage media such as CDs or memory sticks.
- 3.2.3. Data extracts should be deleted immediately after the purpose for which they were extracted has been completed. Where extensive analyses are required, nominated users must request permission to keep data extracts beyond 1 year from the appropriate data custodian. Data extracts must not be held for longer than 2 years.
- 3.2.4. Data extracts must not be shared with non-nominated users. If complex statistical analysis of the data by non-nominated users is required, patient identifiers (soundex, date of birth and clinic numbers) must be removed from the data file prior to analysis by the elected non-nominated user. Such stripped down files must be managed according to principles 3.2.1-3.2.3 above.
- 3.2.5. Analyses of CD4 surveillance data will be strictly for the purpose of improving public health.

3.3. Presentation of local-level aggregated data

- 3.3.1. PCT level CD4 surveillance data should be requested from the local Health Protection Unit. This is subject to Terms and Conditions of Use (see 2.1.4).
- 3.3.2. CD4 surveillance data tables at the level of PCT may only be released provided small cell sizes are suitably anonymised. Cells with values between 1 and 4 inclusive must be anonymised with an asterix. In addition, where the anonymised cell could be deduced from the total, the next smallest cell size in the same row and/or column must also be anonymised.
- 3.3.3. HPA will review the operation of the ONS policy within the HPA as required.
- 3.3.4. When PCT or clinic level data are requested by other organisations (such as for Freedom of Information requests or Parliamentary Questions), small cell sizes must be anonymised according to the protocol described in 3.3.2.

3.4. Presentation and publication of high-level aggregated data

- 3.4.1. CD4 surveillance data tables at the level of country and Strategic Health Authority (SHA) may be published by the HPA, in hard copy or on the website. Anonymisation would not usually be required but may be considered appropriate in some cases.

3.5. Acknowledging data sources

- 3.5.1. Any analysis undertaken using CD4 data which is published in reports, peer-reviewed journals or on the website must acknowledge the data source. “Data from CD4 surveillance: Health Protection Agency”.
- 3.5.2. Academic research presenting data from individual clinics must not be published in peer-reviewed journals without either the prior consent or the collaboration of the lead consultants at those clinics (preferably, the latter).

4. Responsibilities for adhering to policy

- 4.1.1. The CD4 surveillance data custodian at Cfi is responsible for ensuring that patient-level CD4 surveillance data are held, managed and accessed at the Colindale site in line with Caldicott principles and according to these guidelines.
- 4.1.2. Regional CD4 surveillance data custodians are responsible for ensuring that patient-level New HIV and AIDS diagnoses data for their region are managed and accessed in line with Caldicott principles and according to these guidelines.
- 4.1.3. The CD4 surveillance data custodian at Cfi is responsible for ensuring that high-level aggregated data are appropriately anonymised prior to publication on the HPA website or in national reports.
- 4.1.4. Regional CD4 surveillance data custodians are responsible for ensuring that high-level aggregated data are appropriately anonymised prior to publication on regional websites or in regional reports.
- 4.1.5. Any breaches of this policy should be reported immediately, as follows:
 - 4.1.5.1. Serious breaches, such as loss of patient-level CD4 surveillance data due to storage on non-permitted media (e.g. CDs, memory sticks etc.), should be reported to the relevant HPA (i.e. Cfi or regional) and hospital, community or primary care trust Caldicott Guardians. Reports of serious breaches should be copied to all Cfi and regional Caldicott Guardians and to the HIV Surveillance lead at Cfi.
 - 4.1.5.2. Breaches involving the publication of tabular aggregated data at PCT level or below which has not been anonymised should be reported to the appropriate Cfi or regional Caldicott Guardian and copied to the HIV Surveillance lead at Cfi.
 - 4.1.5.3. Minor breaches such as the publication of anonymised tabular aggregated data at PCT level or below without permission from the relevant data custodian should be reported to the appropriate Cfi or regional Caldicott Guardian and copied to the HIV Surveillance lead at Cfi.