### ISD Statistical Disclosure Control Protocol

© NHS National Services Scotland/Crown Copyright 2009

Brief extracts from this publication may be reproduced provided the source is fully acknowledged. Proposals for reproduction of large extracts should be addressed to:

ISD Scotland Publications
Information Services
NHS National Services Scotland
Gyle Square
1 South Gyle Crescent
Edinburgh EH12 9EB

Tel: +44 (0)131-275-6233

Designed and typeset by:
ISD Scotland Publications

<table>
<thead>
<tr>
<th>Document Control</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Version</strong></td>
</tr>
<tr>
<td><strong>Date Issued</strong></td>
</tr>
<tr>
<td><strong>Comments</strong></td>
</tr>
<tr>
<td><strong>Author + Owner</strong></td>
</tr>
<tr>
<td><strong>QA</strong></td>
</tr>
<tr>
<td><strong>Status</strong></td>
</tr>
<tr>
<td><strong>Configuration Control</strong></td>
</tr>
<tr>
<td><strong>Comments to</strong></td>
</tr>
<tr>
<td><strong>Other relevant documents</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Document History</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Version</strong></td>
</tr>
<tr>
<td>V1.0</td>
</tr>
<tr>
<td>1.1</td>
</tr>
<tr>
<td>1.2</td>
</tr>
<tr>
<td>1.3</td>
</tr>
<tr>
<td>1.4</td>
</tr>
<tr>
<td>1.5</td>
</tr>
<tr>
<td>1.6</td>
</tr>
<tr>
<td>1.7</td>
</tr>
<tr>
<td>2.0</td>
</tr>
</tbody>
</table>
Contents

1 Executive Summary .................................................................1
2 Background ........................................................................4
3 Types of Disclosure ............................................................6
4 Recognising Potentially Disclosive Data .............................8
5 The Disclosure Flowchart ....................................................13
6 Risk Assessment ................................................................18
7 Disclosure Control Methods .................................................20
8 Information released for Management or Data Quality Assurance Purposes
   (including Peer Review) ..........................................................24
9 Information released for Research Purposes ........................25
10 Documenting Disclosure Decisions .......................................26
11 Micro-data (including data extracts) ......................................28
12 Databases (e.g. HEAT, Navigator, PRISMS, ACADME) ..........29
References ..............................................................................30
Index ......................................................................................31
1 Executive Summary

This ISD Statistical Disclosure Control (SDC) Protocol sets out guidance and practices for ISD staff to follow on ‘statistical disclosure control’. The protocol describes considerations of risk that should be applied when data is being released, including into the public domain e.g. publications, Parliamentary Questions (PQs) and information requests under the Freedom of Information (Scotland) Act.

The protocol follows Office for National Statistics (ONS) guidance on dissemination of health statistics and has been overseen by an ISD working group, chaired by the ISD Head of Statistics.

For the purposes of this protocol the term ‘disclosure’ is used to describe the communication of personally-identifiable information about a data subject, where information is made public through a statistical output such as a graph or table. The most important consideration is adherence to data protection legislation. ISD staff are required to strictly adhere to relevant NSS data confidentiality guidelines; this protocol aims to be consistent with these guidelines and should be considered in conjunction with the confidentiality rules at all times.

This version builds on the first (March 2009) version and includes a list of the main changes in Annex H.

It should be noted that the consideration of disclosure risk may differ between publications and information requests depending on, for example, the degree of control ISD can exert on the use of the data once released. Data shared within ISD does not require SDC to be applied however NSS' confidentiality rules should be followed in these circumstances. The person providing the data should also highlight any potentially disclosive data and, if external release of the data is planned, advise on SDC.

This protocol sets out ‘guidelines’ for risk assessment of disclosure arising from a statistical release. It is important to note that this protocol does not set out a particular formula that provides a measure of risk for every scenario. Rather, the emphasis is on the need for judgement to be made, on a case-by-case basis, of the risk and this protocol provides guidance on how best to assess the risk. The risk will be based on the following:

- the cell values and table design
- the topic in question (i.e. how ‘sensitive’ the topic is)
- populations, geographies and institutions involved
- the likelihood of an attempt to identify an individual
- the level of impact of any disclosure

A flowchart has been designed to help ISD staff assess the risk of disclosure and decide on whether disclosure control is necessary. When assessing the risks of disclosure in data for management information purposes the same considerations will apply as for published data. Although it is not possible to summarise, within a few lines, all the scenarios set out in the flowchart, key points are as follows:

for a ‘sensitive’ topic:

- cell values 1 - 4 should not be shown
- for values of 5+, whether a cell value should be released or protected will depend on a number of its characteristics, however a key factor is whether the value is <10

for a ‘non sensitive’ topic:

- whether a cell value should be released or protected will depend on a number of its characteristics, however a key factor is whether the value is <3
- potentially disclosive data in relation to the ‘working lives’ of NHS staff may be made available for purposes associated with the management and delivery of health services
It is important to document the reasoning used for decisions on whether to apply SDC. This should be carried out via an ISD Disclosure form (Annex E).

For all ISD publications a Disclosure form should be completed.

For ISD information requests:

- it is **not necessary** to complete a Disclosure form where the outcome, following the Disclosure Flowchart, is to 'release' (that is without having had to 'complete a risk assessment')

- for **all other information requests** a Disclosure form should be completed to document those where potential disclosure issues were identified, any risk assessment and what, if any, disclosure control methods were applied prior to release. Note that where data relates solely to the customer’s own organisation or limited other situations e.g. groups with national responsibilities, in most cases it will not be necessary to complete a Disclosure form. In these circumstances any release of potentially disclosive information should be accompanied by the standard 'disclosure' text [see Section 10.1].

Where SDC is required, then the following approach should be followed:

- consider firstly **table redesign** (e.g. grouping or aggregating cells) and/or cell suppression for unsafe cells

- if table redesign or cell suppression is not considered appropriate then consider **rounding** (this method should not be applied without prior discussion with ISD Head of Statistics)

Note that different methods of SDC can result in varying levels of usefulness of the final data to the customer. For example table redesign may provide greater utility than a heavily suppressed table. Therefore the decision on which SDC method to use should take into account the use of the statistics in each individual case. Where possible, discussion should take place with the customer to help decide on the most appropriate SDC method.

Rounding, data perturbation (e.g. Barnardisation) and database modification (record-swapping) should not be applied without prior discussion with ISD Head of Statistics.

This protocol covers tabular data (and charts based on tabular data), but not micro-data (i.e. individual records). Guidance on SDC regarding micro-data will be issued at a later stage.

This protocol is likely to further evolve, following ISD ‘case law’ development and as a result of UK-wide work on SDC, and legislative and health policy changes/developments.

For further guidance on disclosure control the first point of contact should be immediate line managers.
The Disclosure Flowchart

1 Refer to Disclosure Form.

2 If handling a request for Management Information/Data Quality Assurance purposes please see Section 8 of the Statistical Disclosure Control Protocol for details on how to proceed.

Consideration should be given to whether it is possible to link data to other sources (internal and external) and through differencing produce small numbers.

Executive Summary
2 Background

Reliable health statistics are a pre-requisite for well-informed decision-making and to support health improvement. It is widely accepted however that where statistics provide information on small numbers of individuals Information Services Division (ISD) have a duty, under the Data Protection Act, to avoid directly or indirectly revealing any personal details. For more information on this please refer to the Legal and Policy Background Section in Annex C.

2.1 The Office of National Statistics (ONS)

The Office of National Statistics (ONS) issued new guidance on confidentiality practices in relation to the dissemination of health statistics in 2006. ONS are in the process of assessing the application of this guidance throughout the Government Statistical Service (and related organisations). Application of this guidance is an important element of the Code of Practice for Official Statistics, which ISD aims to adhere to. It should be noted that this Protocol may be revised in the future following any developments in case law.

The core of the ONS guidance is the assessment and management of risk of ‘disclosure’ occurring when data are being released. The risk is a function of the design of a table and the impact of disclosure. Importantly, judgement is required from those who have a detailed understanding of the statistics – it is not considered possible to produce a simple formula which will provide a precise and consistent assessment of risk for every scenario. This judgement should also take into account the public’s trust in statistics produced by ISD.

2.2 ISD’s SDC Guidance

This protocol aims to set out specific guidance and practice for ISD staff to follow on ‘statistical disclosure control’. It is based on ONS guidance but adds the necessary detail to provide support for ISD staff to make decisions on statistical disclosure. The first version of the protocol was produced by the Head of Statistics team and overseen by an ISD Disclosure Working Group (with Scottish Government statistician input). This second version has been overseen by a smaller working group and has been ratified by the ISD Statistical Advisory Group. It is important to note that ISD guidance is expected to continue to evolve as a result of developing ISD experience of disclosure risk and use of disclosure control, and following UK-wide collaboration on policy and best practice.

For the purposes of this protocol the term ‘disclosure’ is used to describe the communication of personally-identifiable information about a data subject, where information is made public through a statistical output such as a graph or table. ‘Disclosure control’ is the practise of reducing the risk of disclosure. The protocol aims to ensure that the right balance is struck between maximising data utility (including meeting customer requirements) and the management of data confidentiality risks.

The initial disclosure control protocol covered ISD publications (including contributions to ‘non ISD’ publications) and Information Requests (including Freedom of Information requests and Parliamentary Questions). This revised protocol refines guidance following experience developed since the release of the first version.

Although data shared within ISD does not require SDC to be applied, NSS’ confidentiality rules should be followed. However the person providing the data should highlight any potentially disclosive data and, if external release of the data is planned, advise on SDC.

This protocol aims to:

- provide background information on statistical disclosure control
- provide guidance on assessing the risk of disclosure of personal information
- set direction on the application of disclosure control, including advice on selecting methods to protect released tables of statistics
- ensure that the most important consideration is maintaining confidentiality while recognising that decisions must also accommodate the need for clear, consistent and practical solutions that can be implemented within a reasonable time and using available resources
- promote the consistent use of methods that will balance the potential loss of information in outputs against the likelihood of individuals’ information being disclosed
- promote openness and transparency in the processes used, and documentation of decisions and the risk assessment process so that these can be reviewed.
3 Types of Disclosure

3.1 Attribute Disclosure

General attribute disclosure arises when someone who has some information about a statistical unit or individual could, with the help of data from the table, discover details that were previously unknown to them.

3.1.1 Individual Attribute Disclosure

Individual attribute disclosure arises when a data subject/individual can be identified and previously unknown information gained about them from a table.

Disclosure may arise if there is a count of 1 in a marginal row or column total of a table. For example, on examining Table 1 below, if we knew that an individual under the age of 12 in the NHS Board had received a particular treatment, we would now know that this was treatment type 1. Note that attribute disclosure can also occur from a marginal total of 2 or more, where one or more of the individuals could potentially identify information about the other and hence disclose additional information.

3.1.2 Group Attribute Disclosure

Group Attribute Disclosure arises when additional information about a certain group of people can be identified.

Disclosure can also arise from tables with larger values, where they appear in rows or columns dominated by zeros. A zero indicates that no-one in that population has that particular attribute. This can be seen in Table 1 below where all 12 to 15 year olds had treatment type 3. The risk from many zeros in a table may not be significant but in certain situations may need to be protected. Specific care should be taken if analysis shows that no one in a selected population has a particular attribute. This in itself can be disclosive about the selected population e.g. a value of zero was obtained for cancer group A in a particular health board.

Table 1 Treatment Type by Age group for NHS Board X

<table>
<thead>
<tr>
<th>Treatment type</th>
<th>Age Group</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;12</td>
<td>12-15</td>
</tr>
<tr>
<td>Type 1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Type 2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Type 3</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1</strong></td>
<td><strong>12</strong></td>
</tr>
</tbody>
</table>

Source: Office of National Statistics (ONS)

It may also be the case whereby 100% rates are considered disclosive and is another example of group attribute disclosure. For instance, if a table were to show every girl from a particular school year in a specific board had the Human Papilloma virus (HPV) immunisation then this would provide information about each female pupil that may have previously been unknown. Although circumstances such as these may not always present information which is considered sensitive or personal, ISD staff should be aware of the risks this presents and consider applying SDC.
3.2 Identification and Self Identification

Where a table contains small cell values, particularly if there are counts of 1, more consideration is needed as identification or self-identification can lead to the discovery of rare or even unique characteristics in a population.

For certain types of information, rarity or uniqueness may encourage others to seek out the individual. The threat or reality of this could cause distress to the individual, or may lead them to claim that the statistics are inadequate to protect them, and therefore others.

For example, a table showing attendance at a drug misuse clinic by age and sex has a count of 1 for a particular CHP. The individual may in fact be the only person who knows who this ‘1’ is, but they may feel exposed by the statistic. If this fear is communicated to their peers the result may be a lack of trust in the confidentiality of the clinic.

In order to protect against unique identification or self-identification, cells with values of 1 or 2 are usually considered potentially unsafe. Although direct identification / self-identification is not necessarily a significant risk in itself, protection is often required since this could lead to attribute disclosure if other tables have been produced from the same data source and these contain additional information about an individual.

3.3 Residual Disclosure (or ‘Differencing’)

Residual disclosure (or differencing) occurs where outputs from the same or different sources can be combined to reveal information about an individual or a group. This can occur in a publication with many tables, for example, where the same data is cut in different ways, or from combining data from similar information requests.

For example, a recent enquiry from a journalist asked about the number of plastic surgery procedures carried out on teenagers under 18 years of age. A follow-up enquiry from a different journalist at the same organisation asked for information on one procedure for the age group aged under 17 years. Combining the two sets of figures provides the small number of 17 year olds who had this particular procedure.

Further guidance on differencing is provided in Section 4.4.

3.4 ‘The Motivated Intruder’

When releasing data, it should be borne in mind that our data could be combined with that from other local sources to identify individual(s) and disclose further personal details about them.

This situation may arise when small cell values are presented for small geographies. In larger populations, the effort and expertise required to discover more details about an individual may be considered disproportionate, but when the base population is decreased (for example consider data for small geographies such as council area or CHP), it will, in many cases, become easier to find additional information about individuals.

Although locally sourced data may reveal the identity of an individual, it may be ISD’s publication that prompts the motivated intruder to start an investigation. A motivated intruder is someone who deliberately tries to gain information about some person or business e.g. potentially the media or ‘nosy neighbour’. It may not always be necessary, or feasible for ISD to consider all local sources of data, but it is necessary to consider the information likely to be available to third parties, and assess the likelihood and risk of an intruder being motivated enough to track down individual(s).
4 Recognising Potentially Disclosive Data

An assumption sometimes made is that ‘disclosure’ largely relates to the risk of revealing details about an individual as a consequence of a cell in a table having a value close to 1. This assumption oversimplifies the risk as it avoids a scenario, for example, where all the individuals in a table appear within one cell, thus potentially revealing some personal detail about them all (Group Attribute Disclosure). In some cases it is also possible to reveal a detail (or an absence of a detail) about one or more individuals through the appearance of zeros in cells within a table. If different sources are compared, and someone is sufficiently determined to piece together data from these sources, then it is possible that that person could obtain information that is not evident from a single data source alone.

This protocol sets out ‘guidelines’ for risk assessment of disclosure arising from a statistical release. It is important to note that this protocol does not set out a particular formula that provides a measure of risk for every scenario. Rather, the emphasis is on the need for judgement to be made, on a case-by-case basis, of the risk and provides guidance on how best to assess the risk. The risk will be based on the following:

- the cell values and table design
- the topic in question (i.e. how ‘sensitive’ is the topic)
- populations, geographies and institutions involved
- the likelihood of an attempt to identify an individual
- the level of impact of any disclosure

Note that the likelihood of disclosure can be different depending on the intended use of the data, who has requested it and the degree of control ISD has over its further use once released. Data released on a website for example is freely available to anyone to use as they wish. Data provided to a core customer as an information request for a specific purpose such as a statistical analysis may have a lower risk of attempted disclosure.

4.1 Sensitive Topic

In this context ‘sensitive’ refers to topics where disclosure of personal information is considered likely to cause a relatively high impact for example distress or embarrassment to an individual. It is not considered possible to produce a comprehensive list of ‘sensitive topics’. For example, there are thousands of different diagnoses that can be analysed (e.g. for hospital admissions, GP presentations) and it would not be feasible to categorise all topics into those considered ‘sensitive’ and ‘not sensitive’. It is also possible that any one topic could change ‘sensitivity’ over time depending on public perceptions and the degree of impact. However, the following broad information areas have been identified by ISD as being ‘sensitive topics’ and should be used as a guide (it is important to note that within these topics there may be some specific subjects which might be considered ‘non sensitive’).

- sexually transmitted infections
- abortions
- suicides, self harm
- pregnancies under 16 years of age
- alcohol or drugs misuse
- mental health diagnoses and treatments
- prescriptions for contraceptives, mental health or any ‘sensitive’ condition
- crime related statistics – e.g. gunshot injuries, assault, stabbings
- other sensitive diagnoses or treatments
The classification of these as ‘sensitive topics’ is important in terms of how disclosure risk is assessed and handled. On occasion the sensitivity of the topic may be a difficult decision to make and may require discussion with colleagues and senior managers. It may also be useful to consider whether the individuals the data represents would consider it to be sensitive.

When assessing sensitivity of a topic the potential political impact should not be a factor.

**Note that not all ‘sensitive’ topics will be sensitive at all times, and vice versa.** It is possible to have a topic which is considered sensitive (for example, Drugs Misuse) where certain statistics are considered non sensitive (for example high level waiting times for treatment for drug misuse).

## 4.2 Populations, Geographies and Institutions

Small populations and geographies increase the likelihood of disclosure under the scenarios described in Section 3.

For purposes of SDC within ISD, the following guidance is given for ‘small’ population and geography:

- **small population** – it is emphasized that there is no definitive threshold below which a population can be considered to be small. The population threshold should be dependent on the situation and should be based on the identifiable population at risk. In general, if the population at risk is more than 5000 then the likelihood of disclosure is considered to be low. However you should still assess the population at risk for each output, taking into account factors that affect the likelihood of disclosure such as sensitivity, geography (rural versus urban), etc.

Where the population at risk is smaller than 5000 a more detailed assessment of the likelihood of disclosure should be undertaken to determine a minimum population threshold adequate to provide protection. This should take account of the scenarios described in Section 3 by which individuals could be identified. For example, this might involve considering the minimum number of households or schools to deter an intruder from trying to identify an individual household or school or the effect on individuals of statistics being released in local media.

The ‘population at risk’ is the denominator for a cell in question; this may not be the entire population for a particular table or a particular cell – an example would be where numbers of births are shown for ‘<20’, the population at risk would be ‘females aged 15-19’ not ‘females aged <20’). In some cases the base population itself may require discussion and judgement but commonly will be based on age and gender.

- **small geography** – the considerations for geography are similar to those for population and very often the two will be inter-linked. However the sparsity of the population within an area could be a factor affecting the risk of disclosure in some cases. In general a small geography would mean data presented for individual island NHS Boards, Local Authorities (or below, e.g. CHPs) or data presented below NHS Board level.

- **Institution level** - data presented at hospital level or below (e.g. clinics, GP practices) can increase the risk of disclosure of patients.

Annex B contains further information on geographies and populations.

## 4.3 Table/Chart Design

The design of a table (or chart) clearly impacts on the risks of disclosure. Where tables show the following then there is generally a higher risk of statistical disclosure:

- sensitive topics (see Section 4.1 above) with cells of values 1 – 4 and 5-9 for smaller populations, geographies and institutions

- any cells with a value of 1 or 2

- rows or columns dominated by 0s
4.4 Residual Disclosure or Differencing

4.4.1 Differencing (to Produce Small Numbers)

Any sets of tables that are being released should be checked to see if they can be combined so that by inference or differencing, between rows and columns of two or more tables, disclosive cells cannot be derived. This applies to other tables produced by ISD (within the same publication or IR, or indeed a previous IR) or from another source (e.g. GROS) which could be linked to the analysis that has been produced and by differencing produce numbers that may be disclosive.

When linked tables are produced from the same dataset it is not sufficient to consider the protection for each table separately. If a cell requires protection in one table then it will require protection in all tables, otherwise the protection in the first table could be undone.

Tables 2A and 2B below are generated from the same dataset and provide counts for a particular characteristic by age group. The counts for 16 year olds can easily be calculated by differencing the frequencies for age bands 16-19 and 17-19. The counts for these age bands may be considered safe but the difference reveals a small and therefore potentially disclosive count for 16 years olds (one person).

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Table A</strong></td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Frequency</td>
</tr>
<tr>
<td><strong>Table B</strong></td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Frequency</td>
</tr>
</tbody>
</table>

It may also be possible to derive disclosive cells from information in one table. For example, in a recent table on abortions it was possible to derive that a group of eight women had had a surgical termination at less than ten weeks in a particular NHS board, from the overall total and the percentage breakdowns of the estimated gestation and method of termination.

Where tables provide data in terms of rates or percentages, the figures themselves may not be disclosive. However, if the rate or percentage is based on an unsafe cell and it is possible by linking with other tables to recover the original count, then the cell with the rate or percentage is itself unsafe.

Some protection can be provided by the use of rounding rates, or percentages but care still needs to be taken to avoid disclosure. Protection will be provided if the base population from which the rate or percentage is calculated is sufficiently large, since the implied count could be a range of values (however this range must be large enough to satisfy disclosure rules and thresholds).

It is also important to consider if denominators are easily or publicly available (from another table in the publication or perhaps populations which can be obtained from the GROS website). If they are, then cells must be considered unsafe and SDC should be applied. If denominators are not known to be available from another source then figures can be considered safe. Note that crude rates may need to be handled differently, considering the ease of recalculating original counts.

It may, however, be safe to publish directly and indirectly standardised rates. Calculating original counts from such rates would generally require a degree of specialised technical expertise in the associated field which could lower the likelihood of anyone attempting or successfully attempting to calculate the exact numbers. In such cases other factors that should be considered are the availability of population counts, the time period...
involved (are data aggregated over a number of years) and any other known/unknown variables that may be required to calculate original figures. Each case should be examined individually before deciding if some form of SDC is necessary.

4.4.2 Geographical Differencing

Suppose that figures are produced for the two geographical areas A and B as shown in Figure 1, where A is a subset of B. Data for geographical area C could easily be produced by subtraction (differencing). If two tables are produced for different geographies from the same dataset then disclosure by differencing can occur even if the two tables have been protected independently.

Figure 1

Source: Office of National Statistics (ONS)

Consideration should be given when releasing data at different geographical levels in ISD, as many of the geographies reported on are not coterminous (for example West Lothian council area has an overlap with Lanarkshire NHS Board).

Data can be provided at a geographical organisational level (i.e. NHS Board, Local Authority, CHP) and also at a physical geography level (i.e. postcode sector, datazone or intermediate geography levels). Very few of these areas have a one to one mapping and often the use of “best fit” geography is required when ‘building’ a higher geographical area from small areas. This does not always provide exactly the same physical geographical region as the true region and therefore differences will occur in the data reported.

Annex B contains further information on geographies and populations.

4.5 Workforce Statistics

4.5.1 NHS Staff

The majority of data published by ISD relates directly to patients or other recipients of care, organisations or care providers. However disclosure control should also be considered for data relating to the NHS Workforce.

Personal data are processed and published on NHS staff for workforce planning purposes and their use is governed by the Data Protection Act 1998. It is recognised, however, that personal data on public authority employees - which specifically relate to individuals’ duties as public authority employees - may be made available for purposes associated with the management and delivery of health services. Examples may include information relating to staff grade or certain types of care provided. There may be the exception, however, where information could be learned that could infringe on individuals’ private, home and family lives, for example, where clinicians could potentially be identified as providing certain types of care which may be considered sensitive. In these instances SDC should be considered.

There are other aspects of Workforce data that also require consideration of SDC including personal information that is not related to an employee’s role within the NHS, such as ethnicity and sexual orientation.
4.5.2 NHS Contractors

Contractors providing general medical services, general dental services, general ophthalmic services or pharmaceutical services under the NHS (Scotland) Act 1978 are covered by Scottish freedom of information legislation in respect of information relating to the provision of those services, and such information will, generally speaking, be able to be released without the need for SDC (insofar as the information is not potentially identifiable in relation to patients). However, the freedom of information legislation contains some exemptions that may permit the withholding of data in certain limited circumstances. These exemptions include the confidentiality of personal data, Data Protection obligations and the likelihood of substantial prejudice to contractors’ commercial interests.

In relation to confidentiality of personal data, the guidelines that apply to staff who work directly within the NHS (details on NHS Staff available in Section 4.5.1) should also be followed for employees working for contractors who provide ‘NHS services’. For example, within a GP practice contracted to the NHS, personal information of all individual GPs working in that practice would be protected by the same Data Protection obligations that apply to all NHS staff. Similar also to NHS staff, information relating specifically to individuals’ duties in providing ‘NHS services’ may also be made available for purposes associated with the management and delivery of health services.

Specific to NHS Contractors, where the release of certain information is likely to cause real and actual harm to a relevant commercial, financial or economic interest, SDC should be considered. This may involve the release of data whereby private information regarding an aspect of a business could potentially be gained by a contractor’s competitors. SDC should always be a consideration in these instances.

Therefore potentially disclosive data relating to ‘NHS services’ provided by contractors may be released without the need for SDC. However the risk of disclosure should always be assessed as it may be that the information is not released due to FOI exemptions such as those described above. Further guidance on handling this category of statistics may be included in any future updates of this Protocol.

4.6 Mortality Data

While information relating to deceased individuals is not covered by the Data Protection Act (1998) the NHS regard information relating to deceased patients still to be protected by a duty of confidentiality. Health records of the deceased are protected from disclosure under the Freedom of Information (Scotland) Act for a period of 100 years from the date of last entry and the Access to Health Records Act 1990 provides a limited right of access to health records.

Information in relation to deaths will generally come from either ISD or GROS sourced data. Analysts are advised to follow the ISD Disclosure Control Protocol when releasing such information in each of the following scenarios:

- ISD-sourced death data (e.g. deaths recorded on SMR hospital records)
- Linked ISD and GROS data
- GROS-sourced death data only (including published and unpublished data)

If releasing GROS-sourced data then analysts should be aware of the possibility of differencing with already published GROS data or data that GROS would release. This risk should be assessed on a case by case basis and if necessary then the analyst should seek advice from their line manager.

4.7 Disclosure Guidance for Parliamentary Questions (PQs)

ISD’s role in the PQ process is to provide information and advice to assist the Scottish Government in answering PQs. However, we should treat the information we provide as being potentially publicly released and the guidance contained in this Protocol should be followed. For example, where the disclosure flowchart advises to ‘protect’ or ‘risk assess’ information prior to release, this should be done for PQs prior to forwarding draft answers to the Scottish Government. Any information released in response to a PQ should always be subject to ISDs own guidance and policies and so SDC should always be considered.
5 The Disclosure Flowchart

A flowchart (see page 17) has been designed to help ISD staff assess the risk of disclosure and decide on whether disclosure control is necessary. The key points in the flowchart are as follows:

for a ‘sensitive’ topic:
- values of 1 to 4 should not be shown
- for values of 5+, whether a cell should be released or protected will depend on a number of its characteristics, though a key factor is whether the value is less than 10

for a ‘non sensitive’ topic:
- whether a cell should be released or protected will depend on a number of its characteristics, however a key factor is whether the value is under 3

Note that it may be necessary to follow the flow chart more than once for an individual table to ensure that the different value ranges have been captured. This means, for example, if “Protection Required” is reached that this may not be the final step as other outcomes may be possible for different cells within the table. Similarly, if “Release” is reached when following the flowchart for one range of numbers, “Protection Required” or “Complete Risk Assessment” may be the outcome for another set of values. It is also an important part of the decision process to keep in mind all the tables in any particular publication/IR when referring to the flowchart, as data could be linked between tables. This will include consideration of linking to other tables previously published or released by ISD or indeed from other sources.

As part of the decision process, analysts should initially seek guidance from within their team and line manager/programme lead. Further guidance is available from the Assistant Heads of Groups/Programme Principals. The Head of Statistics team can also be contacted for advice.

On occasion, ISD will incorporate data previously published by other organisations (for example, Scottish Government, GROS) into our own publications/IRs. Specific guidance on mortality data is contained in Section 4.6. For any other situations ISD should discuss what should be published with the other relevant organisation, bearing in mind the other organisations guidelines and any risk of differencing.

5.1 Steps in Flowchart

5.1.1 Is it a Sensitive Topic?

This is the first question in the flow chart and is an important part of the decision process. It is not considered possible to provide a comprehensive list of ‘sensitive’ topics and the topics listed in Section 4.1 should be taken as a guide. However, the classification of a topic as ‘sensitive’ or ‘not sensitive’ is important in that it can result in a different decision being made on whether to release or to protect the data in question. It should be noted that the answer to this question may differ from table to table, potentially from cell to cell. If the topic is initially deemed ‘non-sensitive’ but involves small numbers then the sensitivity should be carefully considered. Others may consider the topic to be of a sensitive nature and so further discussion may be required with colleagues and managers.

When assessing sensitivity of a topic the potential political impact should not be a factor.

5.1.2 Population, Geography, Institution

Small populations and geographies, and institution level data, can increase the risk of disclosure and depending on the cell values and other factors can dictate whether or not the data requires protection. (See Section 4.2 for further information).
5.1.3 Rows or Columns Dominated by Zeros

Tables which contain rows or columns dominated by zeros may lead to attribute disclosure (see Section 3.1) and therefore need to be considered when thinking about whether data is disclosive. For example, if a row in a table contains zeros in every cell apart from one, then it is possible to see that everyone in the row has a particular attribute. This risk may not be significant but in some instances disclosure would occur and therefore protection would be required. Further details and an example is provided in Section 3.1.2.

5.1.4 Is the Count Associated with 1 or 2 Practitioners or Hospitals?

This question has been added following its inclusion in ONS guidance (and flowcharts). The question aims to provide coverage for situations where, for example, the data shown might not identify individual(s) receiving a certain type of care for a certain condition, but might help identify individual(s) providing certain types of care. A scenario might be where a particular surgical operation is carried out by only one clinician (or in only one hospital) in a certain NHS Board – release of such data might disclose information particular to the individual clinician (or institution), particularly if it is of a sensitive nature. It should be noted that this does not conflict with Section 4.5.1 as the information that could potentially be disclosed may be of a personal nature to an individual and could infringe on their private, home or family life.

This question has not hitherto been part of general ISD disclosure risk assessment.

5.1.5 Can you Identify Individuals (or a Group of Individuals) and Gain Additional Personal or Sensitive Information About Them?

It is necessary to ask this question to avoid inadvertent disclosure that might arise following answers to previous questions in the flowchart. This question may be particularly useful in helping avoid disclosing information for groups of individuals.

Identifying individuals and gaining additional personal or sensitive information: Table 3 shows that there are 625 women from NHS Board X that had an abortion, and from this table we learn that all of these were on ‘Ground C’. The meaning of Ground C will be further explained in the footnote to the table so therefore we are gaining additional sensitive or personal information about these individuals.

### Table 3 Abortions by NHS Board, age and grounds

<table>
<thead>
<tr>
<th>Grounds for abortion</th>
<th>NHS Board</th>
<th>NHS Board</th>
<th>NHS Board</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X</td>
<td>Y</td>
<td>Z</td>
</tr>
<tr>
<td>All Abortions</td>
<td>625</td>
<td>1550</td>
<td>3407</td>
</tr>
<tr>
<td>Rate per 1000 live births</td>
<td>183.6</td>
<td>253.8</td>
<td>247.7</td>
</tr>
<tr>
<td>Rate per 1000 women aged 15-44</td>
<td>10.6</td>
<td>14.7</td>
<td>13</td>
</tr>
<tr>
<td>Age of Woman</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Under 20</td>
<td>174</td>
<td>382</td>
<td>818</td>
</tr>
<tr>
<td>20 - 24</td>
<td>174</td>
<td>418</td>
<td>1098</td>
</tr>
<tr>
<td>25 - 29</td>
<td>121</td>
<td>326</td>
<td>699</td>
</tr>
<tr>
<td>30 - 34</td>
<td>73</td>
<td>215</td>
<td>410</td>
</tr>
<tr>
<td>35 - 39</td>
<td>50</td>
<td>155</td>
<td>278</td>
</tr>
<tr>
<td>40+</td>
<td>33</td>
<td>54</td>
<td>104</td>
</tr>
</tbody>
</table>

**Note**: The above table shows artificial data for the purposes of this guidance only.
Identifying individuals and not gaining any additional personal or sensitive information:

From the table below on type of vouchers of General Ophthalmic Services we can see that there was one bifocal complex payment on the Western Isles but we do not learn anything additional about this person.

Table 4  NHS vouchers GOS(S)3 by type and NHS board
Year ending 31 March 2008

<table>
<thead>
<tr>
<th>Number</th>
<th>Scotland</th>
<th>Orkney</th>
<th>Shetland</th>
<th>Tayside</th>
<th>Western Isles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single vision :</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>305,366</td>
<td>602</td>
<td>1,088</td>
<td>19,855</td>
<td>1,521</td>
</tr>
<tr>
<td>B</td>
<td>56,444</td>
<td>70</td>
<td>170</td>
<td>3,273</td>
<td>204</td>
</tr>
<tr>
<td>C</td>
<td>3,410</td>
<td>3</td>
<td>5</td>
<td>193</td>
<td>9</td>
</tr>
<tr>
<td>D</td>
<td>1,706</td>
<td>2</td>
<td>5</td>
<td>104</td>
<td>5</td>
</tr>
<tr>
<td>Bifocal :</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>67,439</td>
<td>71</td>
<td>140</td>
<td>5,160</td>
<td>270</td>
</tr>
<tr>
<td>F</td>
<td>13,719</td>
<td>4</td>
<td>24</td>
<td>1,031</td>
<td>36</td>
</tr>
<tr>
<td>G</td>
<td>373</td>
<td>0</td>
<td>1</td>
<td>30</td>
<td>2</td>
</tr>
<tr>
<td>H</td>
<td>279</td>
<td>0</td>
<td>0</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>Complex payment :</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>single</td>
<td>1,403</td>
<td>2</td>
<td>0</td>
<td>142</td>
<td>5</td>
</tr>
<tr>
<td>bifocal</td>
<td>577</td>
<td>1</td>
<td>0</td>
<td>65</td>
<td>1</td>
</tr>
<tr>
<td>All voucher types</td>
<td>450,716</td>
<td>755</td>
<td>1,433</td>
<td>29,877</td>
<td>2,053</td>
</tr>
</tbody>
</table>

Source: OPTIX, ISD Scotland

The decision of whether information is sensitive enough to apply SDC lies with the analyst team involved. The team may consider that the disclosure of additional information is not sufficiently personal or sensitive enough an issue to apply SDC, however, there could be a condition or data item that is considered sensitive enough for SDC to be applied to the data.

5.1.6 Complete Risk Assessment:

Where the flowchart advises ‘complete risk assessment’ then the following guidance should be followed:

- undertake a risk assessment referring to Section 1 of the Disclosure form (please refer to Section 10.1 for particular situations where a risk assessment may not be required). See Section 6 for guidance on risk assessment.
- in general, if the risk score on the Disclosure form is 4+ then apply SDC.
- however, if the topic is considered particularly sensitive then SDC may be considered appropriate where the risk assessment score is 3.

5.2 Flowchart Guidance

The guidance that is set out in the flowchart can be described as follows:

For a ‘sensitive’ topic:

- for values 1- 4, protect
- for values of 5-9, if the population/geography/institution is ‘small’, protect
for values of 5-9 and the population/geography/institution is not 'small', consider if there are any rows/columns dominated by zeros, any counts associated with 1 or 2 practitioners/hospitals or if you are able to identify individuals and gain additional personal or sensitive information about them. If the answer is yes to any of these, complete a risk assessment. If the answer is no to all, release.

for values 10+, consider if there are any rows/columns dominated by zeros, any counts associated with 1 or 2 practitioners/hospitals or if you are able to identify individuals and gain additional personal or sensitive information about them. If the answer is yes to any of these, complete a risk assessment. If the answer is no to all, release.

for a 'non sensitive' topic:

- for values of 1 or 2, if the population/geography/institution is 'small' and additional personal or sensitive information about identifiable individuals can be obtained, protect.

- for values of 1 or 2, if the population/geography/institution is 'small' and the data does not reveal anything additional personal or sensitive about individuals, release.

- for values of 1 or 2, if the population/geography/institution is not 'small', consider if there are any rows/columns dominated by zeros or if you are able to identify individuals and gain additional personal or sensitive information about them. If the answer is yes to any of these, complete a risk assessment. If the answer is no to all, release.

- for values 3+, consider if there are any rows/columns dominated by zeros or it is possible to identify individuals and gain additional personal or sensitive information about them. If the answer is yes to any of these, complete a risk assessment. If the answer is no to all, release.

It is important to note that there may be situations where it is not appropriate to follow precisely the guidance that is contained in this protocol. An example might be where the 'public interest' may be judged to be best satisfied by release of statistics, rather than adherence to this protocol which would result in disclosure control being applied.

Such situations are expected to be rare within ISD and, due to sensitivity or profile, would necessarily involve senior ISD staff. ISD Head of Statistics should be consulted on such decisions.

It is important to document any decisions made with regards to SDC so that the outcomes of any discussions can be reviewed for future releases.

Examples of ISD information, applied to the flowchart, are shown in Annex D.
Figure 2: Disclosure Flowchart

The Disclosure Flowchart

1 Refer to Disclosure Form.

2 If handling a request for Management Information/Data Quality Assurance purposes please see Section 8 of the Statistical Disclosure Control Protocol for details on how to proceed.

Consideration should be given to whether it is possible to link data to other sources (internal and external) and through differencing produce small numbers.
6 Risk Assessment

The introduction, via this protocol, of a described Disclosure Flowchart effectively removes the need for a ‘risk assessment’ (as defined below, i.e. based on ‘likelihood’ and ‘impact’) for a considerable proportion of statistical releases. However, for those releases that do require such a risk assessment the following guidance is considered important in helping ensure the most appropriate outcome and to provide consistency across ISD. It should also be noted that, having followed the Disclosure Flowchart, if the outcome is “Complete Risk Assessment” then both parts one and two of the Disclosure Form must be completed.

6.1 Assessing the Likelihood of an Attempt to Disclose

Assessing the disclosure risk, and therefore the need for disclosure protection, involves undertaking a risk assessment. The risk assessment considers the:

- **likelihood** of an attempt to associate the information being released with an individual person, and
- **impact** that would arise from this disclosure

This terminology has been adopted from relevant ONS guidance but staff should be aware that the likelihood of an attempt to disclose information should include the possibility of inadvertent disclosure as well as deliberate.

6.1.1 Calculation of the Risk Score

To help with this risk assessment, both ‘likelihood’ and ‘impact’ are measured in terms of ‘high’ (risk score = 3), ‘medium’ (score = 2) or ‘low’ (score = 1). The scores for ‘likelihood’ and ‘impact’ are then multiplied.

Where the resulting score is 4 or greater then disclosure control methods should be applied. In addition, if the topic is considered particularly sensitive then SDC may be considered appropriate where the risk assessment score is 3.

The risk score will be particularly important for some scenarios where the score will impact on the decision on whether to release or to protect data.

When considering the **likelihood** of an attempt to disclose the following should be considered:

**Low Risk:** Low risk would mean that it would be difficult for someone to identify disclosive information from the release. This would be the case where there is little chance of differencing between tables as there are not many tables released from the dataset or if the topic was not sensitive, the cell value high and presented at Scotland level. Releases could also be considered low risk if the data is from a large population, however this would depend on the variables used in the released tables.

**Medium Risk:** This level of risk might occur if there was data which was not in the ‘sensitive’ health-related data list but had cells of value 1 and 2, columns or rows dominated by zeros, or data presented for a small population or geography.

**High Risk:** In many instances, topics identified as ‘sensitive’ will be categorised as high risk, however not all scenarios classed as sensitive will necessarily be scored as high risk. Examples may be where there are specific local issues or where the topic concerned is currently the subject of particular media attention (see Section 4.1).

Note that the likelihood of disclosure can be different depending on the intended use of the data, who has requested it and the degree of control ISD has over its further use once released. Data published on the website for example is freely available to anyone to use as they wish. Data provided to a core customer as an information request for a specific purpose such as a statistical analysis may have a lower risk of an attempt to disclose.
The likelihood of an attempt to disclose can also be affected by the timeline associated with the data. For instance, if the information is aggregated over a number of years rather than single years then this could lower the risk involved as identifying individuals may be more difficult. This may also be the case for information which is not particularly recent and so the lapse in time from the period in which the figures are based to when the data is released should also be a consideration when determining the risk.

6.2 Assessing the Impact of Disclosure

The assessment of the impact of disclosure can be considered to be somewhat subjective. The assessment should be undertaken by a member of staff with experience of the information being released. The assessment of impact should take into account those who may have an interest in the data being released, the views of patients and carers and potentially disclosive situations which could occur through disclosure. As a simple rule, all ‘sensitive’ health-related data should generally be considered as having a ‘high impact’ of disclosure.

Consideration should also be given to the impact on the data of any disclosure control applied. It is important to balance risk and utility and reach a position where risk is minimised and utility maximised. If no disclosure control is applied then risk may be very high. Alternatively the application of any disclosure control technique which results in, for example, a table primarily consisting of suppressed cells would have little value. Wherever possible, discussion should take place with the customer to decide on the most appropriate SDC method.

6.3 The Disclosure Form – Section 1: Risk Assessment

The Disclosure Form has been amended from that contained in Version 1.0 of this protocol to reflect updates to the SDC guidance however the process remains the same. Section 1 of the Disclosure Form aims to aid analysts to assess risk and document their decision making. If a risk assessment has been carried out then Part 2: Details of Disclosure Control must also be completed.
7 Disclosure Control Methods

Where the need for SDC is identified then there are a range of methods that can be considered for use. The choice of method should balance uses to be made of the information and simplicity of approach. These methods can be divided into three categories, i.e.:

- those that determine the design of the table
- those that modify the values in the table
- those that adjust the data before tables are designed

Alternative methods for presenting data can be considered as an approach for providing users access to information without disclosing the underlying data. In some cases this will provide a more robust analysis than reliance on the accuracy of small cell values – for example, these could include presenting data graphically with limited detail in scale.

In applying certain types of SDC – for example, table redesign and cell suppression – Tau Argus software can assist. ISD has access to Tau Argus software and has a certain degree of experience in its use. Further information is given in Annex F.

7.1 ISD policy

Where SDC is required, then firstly consider:

- table redesign (e.g. grouping or aggregating cells) and/or
- cell suppression (primary and secondary) for unsafe cells

If table redesign or cell suppression is not thought most appropriate then rounding can be considered (please note this method should not be applied without prior discussion with ISD Head of Statistics).

The decision on which SDC method to be used should take into account the use of the statistics in each individual case. Where possible, discussion should take place with the customer to help decide on the most appropriate SDC method. Annex G contains an extract from the Scottish Government’s Practical Guide to Statistical Disclosure Control and provides illustrative examples of how table redesign and suppression can be performed. It also shows an ISD example of where rounding could possibly be used as a disclosure control technique while also highlighting the issues surrounding this method.

Data perturbation methods (e.g. Barnardisation) should not be applied without prior discussion with ISD Head of Statistics. Similarly database modification (record-swapping) should be discussed with ISD Head of Statistics prior to implementing.

Further information on these techniques can be found at:


For ISD compendium publications, advice from the originating ISD data providers should be sought to ensure that the appropriate levels of SDC are applied, and that data is consistent between ISD publications.
7.1.1 Table Redesign

Table redesign is recommended as being a relatively simple method that will minimise the number of unsafe data and preserve original counts. However, the use of this method should be balanced against consistency in table design and publication plans.

**Description:** Remove unsafe cells by, for example
- grouping categories within a table
- aggregating to a higher level geography or for a larger population sub-group
- aggregating tables across a number of years/months/quarters

**Advantages:**
- original counts in the data are not damaged
- easy to implement
- easily understood by user

**Disadvantages:**
- detail in the table will be reduced
- may be policy or practical reasons for requiring a particular table design

*Annex G* provides further information and example.

Depending on the nature of the request and/or certain circumstances ISD staff may wish to explain that the design of the table has been influenced by disclosure. For example, table presentation may differ from that contained in previous publications. In this instance, the following could be used, amending as required.

The design of a number of the tables presented in this publication has been revised from previous editions. These changes attempt to minimise the risk of disclosure and to help maintain patient confidentiality.

7.1.2 Cell Suppression

**Description:** Unsafe cells are not released. They are suppressed and replaced by a '*' (an asterisk) to indicate a suppressed value. Such suppressions are called **primary** suppressions. To make sure that the primary suppressions cannot be derived by subtraction, it may be necessary to select additional cells for **secondary** suppression.

**Advantages:**
- original counts in the data that are not suppressed are not adjusted
- can provide protection for zeros
- allows original/requested structure to be maintained
- depending on number of cells ‘at risk’, can be preferable to table re-design

**Disadvantages:**
- most of the information about suppressed cells will be lost
- secondary suppressions will hide information in safe cells (this could include totals)
- information loss may be high if more than a few suppressions are required
- any potentially disclosive zeros would need to be suppressed
- does not always protect against disclosure by differencing
Past experience has shown that it is good practice to present tables with totals. If totals are not included then a customer could return to ask for totals. This must then be considered in conjunction with any cell suppression applied to the original table and may result in some totals being suppressed to ensure previously suppressed figures cannot be calculated through differencing.

The comparison of data from numerous tables must also be considered (including previously released data) to ensure against differencing and so suppressing data can be time consuming and complicated (see Section 4.4).

The following rules should be applied for suppression (primary and secondary):

- replace both primary and secondary suppressed cells with "*" (an asterisk). This symbol should not be used for any other value. Do not use different symbols for primary and secondary suppressions. The ISD Data Formatting and Presentation Guidance (http://genss.nss.scot.nhs.uk/portal/page?_pageid=514,1073415,514_1090450&_dad=portal&_schema=PORTAL) lists the symbols which should be used across ISD when presenting data.

- Values of zero should not automatically be selected for primary suppression. On some occasions suppression of zeros may be required for secondary suppression or where rows and/or columns are dominated by zeros. Care must be taken with any secondary suppression of data. Normally the next smallest number would be selected for secondary suppression. However this is not always the best option. Selecting another larger number may lead to less cell suppression within the table, thereby maximising utility.

- A footnote advising that cells have been suppressed should be added to all relevant tables (and not only specified in an attached email, for example) and be consistent for all tables within a publication. The footnote should not detail the values suppressed e.g. <5. The example below can be used.

* Indicates values that have been suppressed due to the potential risk of disclosure and to help maintain patient confidentiality

See Annex G for further information and example.

7.1.3 Rounding

There are a range of methods of applying rounding as a method of SDC, including: controlled rounding, deterministic rounding and random rounding. If rounding is to be used the method currently recommended for ISD is controlled rounding.

**Description:** Involves adjusting the values in all cells in a table to a specified base, so as to create uncertainty about the real value for any cell, while adding a small but acceptable amount of distortion to the data. The base for rounding can be chosen with common choices being 3 or 5. All rounded values (other than zeros) are then integer multiples of 3 or 5 respectively.

**Advantages:**

- if the number of unsafe cells is large then the table can be protected while still providing counts for all cells and without altering the design of the table
- will protect zeroes without removing them since, within a table rounded to base 5 for example, a zero could represent any count between 0 and 4
- cells rounded to a common base in such a way as to preserve additivity to totals within table (unlike random rounding where all figures including totals are rounded randomly and so may not be additive)
- fully protects against disclosure by differencing
Disadvantages:

- difficulties in disguising cells in which the count can be associated with either 1 or 2 practitioners/hospitals whom it may be necessary to protect
- if user requires exact counts rounded values would not be appropriate
- if population size is small then rounding may not offer enough protection against identification
- can at times distort data to such a degree that original trends can not be identified. Care must therefore be taken to avoid this whilst ensuring trends that do not actually exist can not be wrongly interpreted
- may be prone to effects of data revisions, for example updates to figures that are contained in future editions of a publication series may require a different pattern of rounding than that used in previous presentation of the figures
- may not be helpful to those users of ISD’s statistics who are familiar with historical ISD methods of presentation (including SDC)
- totals may be adjusted, thereby altering ‘headline’ figures

Due to these various issues, rounding is currently not a preferred method of SDC within ISD and should not be applied without prior discussion with ISD Head of Statistics. Any use of rounding should be carefully considered, for example there may be an impact on the use and interpretation of any rounded figures. It is therefore essential to ensure that any information provided is not misleading to the user.

See Annex G for more detail and an illustrative example.
8 Information released for Management or Data Quality Assurance Purposes (including Peer Review)

ISD has a duty, laid out in statutory orders, to provide the NHS and the Scottish Government with data and analysis to allow the proper management of the health service. ISD need to actively seek ways to facilitate this in a way that does not compromise its duties under data protection rules.

Annex A contains guidance on handling requests for information for management purposes and also information for data quality assurance purposes.
9 Information released for Research Purposes

Further guidance is currently being produced in relation to information released for research purposes and is planned for future release.

Currently requests for the release of patient identifiable information may require a confidentiality statement to be signed or an application made to the Privacy Advisory Committee (PAC). PAC acts as an advisory committee to the Board of NHS National Services Scotland (NSS) and the Registrar General on the correct balance between protecting personal data and making data available for research, audit and other important uses and ensures that any information releases are carefully controlled.

For other requests and pending specific guidance on information released for research purposes ISD’s SDC protocol should be followed; this may also require discussion with senior managers and the NSS Caldicott Guardian.
10 Documenting Disclosure Decisions

The issue of disclosure concerns all of ISD’s outputs (including publications and Information Requests). To document decisions taken on disclosure issues a Disclosure form should be completed. A Disclosure form should be completed for:

- all ISD publications (including contributions to ‘non ISD’ publications)
- certain Information Requests, including FoI and PQs (see 10.1)

Note that information released by ISD in response to a PQ should be treated as being publicly released and should always be subject to our own guidance and policies. SDC should therefore always be considered before releasing information in response to a PQ.

A Disclosure form should be completed when the disclosure flowchart has instructed:

- to protect
- to undertake a risk assessment (with the resultant outcome being to either protect or release).

10.1 When to Complete a Disclosure Form for Information Requests (IRs)

A Disclosure form should be completed for all IRs where the end point by using the Disclosure Flowchart is “Protection Required”. You should also complete a Disclosure form if the Disclosure Flowchart instructs you to undertake a “Risk Assessment”. Note that where data relates solely to the customer’s own organisation or limited other situations e.g. groups with national responsibilities, in most cases it will not be necessary to complete a Disclosure form - provided the standard text in Annex A is attached to data.

You should not complete a Disclosure form where you are instructed using the Disclosure Flowchart to “release” data.

Prior to the release of information, where possible, the appropriateness of the customer requesting and receiving the information should always be checked to ensure that it is appropriate for them to receive the data, including without any SDC applied. It should also be noted that data contained in Information Requests should be released in line with NSS’ Confidentiality Rules. Data should not be emailed unless this is via NHS.net or is encrypted using approved software. See geNSS for further information:

http://genss.nss.scot.nhs.uk/portal/page?_pageid=514,1056303,514_1056418&_dad=portal&_schema=PORTAL

10.2 Guidance for Completion of Disclosure Form

10.2.1 Part One: Risk Assessment

A Risk Assessment should only be undertaken when instructed on the Disclosure Flowchart (see Section 5).

The risk assessment involves assigning a score of Low (1), Medium (2) or High (3) to the ‘likelihood of an attempt of disclosure’ and the ‘impact of any disclosure’. See Section 6 of this protocol for further guidance on risk assessment. The risk assessment may involve discussion with colleagues and managers within your team and/or group, the Programme Principle and the Assistant Head or Head of Statistics.

Questions have been included around suitability, access and use to aid assessment of the risk. These include:

- Exactly what data is requested? Data sources, variables, time periods etc
  - How sensitive is the topic area considered? For data initially deemed non-sensitive, if small numbers are present then the topic sensitivity should be given careful thought. Could the information be considered sensitive by others?
b. What are the sizes of the geographies / populations / institutions involved?

c. Consider the size of cell values / is the table design most appropriate? See protocol for guidance / also take guidance from previous publications or requests.

- Who is requesting the data? Named contact, position/role, organisation
- What is the intended use of the data? e.g. inform committee meeting, FOI, PQ, research paper, publication, SG policy
- Who will have access to the data? Named contact(s)/groups/organisation, position/role, NHS board, NHS Steering Group, Scottish Government policy makers, non-NHS partner organisations
  
  a. Will the data be in the public domain? e.g. Information request to media, PQ, FOI
  
  b. Will there be controlled access?

- What measures are there in place to protect the information? e.g. none (info will be in public domain); info will be distributed at meeting only; info will be distributed within SG only.

These questions should be considered before information is released.

If by multiplying the ‘likelihood’ by ‘impact’ the score is 4 or above then some form of disclosure control should be applied to the data prior to release. If the risk score is less than 4, then the data can be released however you should complete the relevant questions in Part 2 of this form. In some circumstances, it may be judged that disclosure control should be applied when the risk score is 3 (for example if the topic is considered ‘sensitive’).

If a risk assessment has been carried out then Part Two of the disclosure form should also be completed.

10.2.2 Part Two: Details of Disclosure Control

Part Two of the disclosure form should be completed when the disclosure flowchart has instructed to protect or undertake a risk assessment (following completion of Part One).

**Question 1:** Indicate whether or not any disclosure control was applied to the data prior to publication.

**Question 2:** Indicate all of the potential disclosure risks e.g. sensitive topic, counts of 1-4 etc.

**Question 3:** Identify the methods used for disclosure control.

**Question 4:** Describe the effect on the output of applying the disclosure control techniques. An example might be:

‘All cells values of 1 and 2 in Tables 1 to 10 of this publication were suppressed because these were values based on a ‘small’ population. Secondary suppression was also required within these tables to ensure primary suppressed values could not be calculated. Within Tables 11 to 15, age groups were combined to aggregate small numbers.’

Provide as much detail as possible here. This information will be used as a reference for future publications or similar IRs.

**Question 5:** This space should be used to document any other information you feel appropriate and also to document why you did not decide to apply any disclosure control prior to release.

The completed Disclosure form should be passed to a Head of Group (HoG), Assistant HoG (AHoG), Head of Programme (HoP) or Programme Principal (PP) to sign off. This should be done electronically with the HoG/AHoG/HoP/PP emailing the form to the ISD Disclosure Team mailbox. For publications this should be submitted before pre-release access is provided (normally the 2nd last Tuesday of the month of publication) and for Information Requests, immediately after release.

**Only once the Disclosure form is signed off can the data be released.**

Submitted Disclosure forms will provide evidence for future assessment of ISDs practice and evolution of guidance. It is also a good reference for future releases.
11 Micro-data (including data extracts)

*Section to be added in future update*
12 Databases
(e.g. HEAT, Navigator, PRISMS, ACADME)

This guidance refers to ‘databases’ (or ‘warehouses’, or ‘datamarts’, etc) that are made accessible to ISD and non ISD staff for analysis e.g. the extraction of reports. There will be a range of types of databases made available to non-ISD staff and for some (e.g. ACADME, PRISMS) there may be wider confidentiality issues. It is important that confidentiality arrangements that are specific to individual databases are followed.

A formal procedure for authorising access to each database should exist. The decision on whether an individual requires access is the responsibility of the organisation (e.g. NHS Board) accessing the database. These organisations – in conjunction with ISD - are responsible for ensuring that only appropriate staff have access to the database. They must also make employees fully aware of their responsibilities in relation to disclosure control and ensure they adhere to confidentiality principles and comply with Data Protection Act obligations.

The ISD database manager, or equivalent, is responsible for

- ensuring that appropriate access controls are in place regarding access to the database
- maintaining an up-to-date list of who has access (and where appropriate the level of access).

A prominent note should exist for each database which contains the statement below or similar:

The data presented have not been adjusted to protect against potential disclosure risks and are released within this system for management information purposes. The data presented may contain information which enables (perhaps with the aid of further knowledge of the topic) an individual patient or member of staff to be identified. Please ensure circulation is restricted and that patient confidentiality is not compromised. For further guidance see ISD’s Statistical Disclosure Control Protocol [http://www.isdscotland.org/isd/4489.htm](http://www.isdscotland.org/isd/4489.htm). Please contact xxx@nhs.net if you have any queries regarding this.

Information made available in these types of formats might commonly be considered ‘management information’; the specific guidance on ‘management information’ in this protocol should therefore be considered
References

Freedom of Information Guidance (on geNSS – the NSS intranet)
http://genss.nss.scot.nhs.uk/portal/page?_pageid=514,1056303,514_1056318&_dad=portal&_schema=PORTAL

ISD Guidance (Nov 2007) Applying ONS Confidentiality Guidance and Disclosure Control for ISD Statistical Outputs (Publications & IRs)

NSS Confidentiality Rules (on geNSS – the NSS intranet)
http://genss.nss.scot.nhs.uk/portal/page?_pageid=514,1056303,514_1056418&_dad=portal&_schema=PORTAL

NSS Data Protection Policy (on geNSS – the NSS intranet)
http://genss.nss.scot.nhs.uk/pls/portal/docs/PAGE/GENSS_DOCUMENT_LIBRARY/GENSSFUNLIBINF/GV/TAB1050124/REVISED%20NSS%20DATA%20PROTECTION%20POLICYV220071.2DOC.DOC

Government Statistical Service, GSS / GSR Disclosure Control Policy for Tables Produced from Administrative Data Sources


Index

A
Aims of Protocol, 4
Assess Risk, 8
Attribute Disclosure, see Disclosure Types

B
Background, 4
Barnardisation, 20

C
Cell Suppression, 21
-footnote, 22
-primary suppression, 21, 22
-rules, 22
-secondary suppression, 21, 22
-totals, 22

D
Databases, 29
Database Modification, 20
Denominators, availability of, 10
Differencing, see Disclosure Types
Disclosure, term, 4
Disclosure Control, 4
Disclosure Control Methods, 20
Disclosure Flowchart, 13, 17
-consider all tables, 13
-consider all values, 13
-exceptions, 13
-guidance, 13, 15
Disclosure Form, 26, 27
-exceptions, 26
-guidance, 26
-information requests, 26
-signing off, 27
Disclosure Team Contact, 28
Disclosure Types, 6
-attribute disclosure, 6
-differencing, 7, 10, 11
-geographical differencing, 11
-group attribute disclosure, 6
-identification, 7
-individual attribute disclosure, 6
-motivated intruder, 7
-residual disclosure, 7, 10
-self identification, 7

F
Freedom of Information
-exemptions 12

G
Gaining Additional Information, 14
Geographies, small, 9, 13
GROS Data, 10, 12, 13
Group Attribute Disclosure, see Disclosure Types

H
Hospitals, 14

I
Identification, see Disclosure Types
Impact of disclosure, 19
Individual Attribute Disclosure, see Disclosure Types
Information released for
-management information purposes, 24
-quality assurance purposes, 24
-research purposes, 25
Information Requests, 8, 10, 13, 26, 27
Institutions, small, 9, 13

L
Likelihood of an attempt to disclose, 18
Linking data, 10, 13, 22, also see differencing

M
Management Information, 24
Micro-data, 28
Mortality Data, 12
Motivated Intruder, see Disclosure Types

N
NHS Contractors, 12
NHS Staff, 11
Non-sensitive Topic, 13, 16

O
ONS Guidance, 4
Other Organisations Data, 13, 26
Other Sources of Information, 8, 10, 13
**P**
Parliamentary Questions, 12, 26
Population, small, 9, 13
Practitioners, 14
Protecting Data, see Cell Suppression and Table
Redesign
Publications, 4, 13, 20, 27

**Q**
Quality Assurance, 24

**R**
Rates
  - 100%, 6
Research
  - data request, 25
Residual Disclosure, see Disclosure Types
Risk Assessment, 15, 18, 26
  - impact, 18, 19, 25
  - likelihood, 18, 26
  - public interest, 16
  - score, 18
Rounding, 22

**S**
Self-identification, see Disclosure Types
Sensitive Topic, 8, 9, 13
  - example, 14
Small Numbers, 4, 7, 10, 13
Standardised Rates, 10
Statistical Disclosure Control Methods, 20
Suppression, see cell suppression

**T**
Table/Chart
  - design, 9
  - redesign, 20, 21
Tau Argus, 20

**W**
Workforce Statistics, 11

**Z**
Zeroes, 6, 14