NHS National Services Scotland

# Statistical Disclosure Control Protocol

## Version 3.0

| Version control | | | |
|---|---|---|---|
| **Version** | **Date issued** | **Description** | **Distribution** |
| 0.1 | 15 August 2008 | First (incomplete) draft. To be considered as framework, and to aid discussion of key points of ISD policy. | ISD Disclosure Group, for 18/08/08 meeting. |
| 0.2 | 15 September 2008 | Revised draft, includes flowchart on whether to apply SDC. | PJ/JJ/LH discussion. |
| 0.3 | 16 September 2008 | Further revised draft, includes decision flowchart. | PJ/JJ/LH discussion. |
| 0.4 | 16 September 2008 | Further revised draft, includes decision flowchart. | To ISD Disclosure Group, for 22/09/08 meeting. |
| 0.5 | 15 October 2008 | Revisions following 22/09/08 meeting of ISD Disclosure Group. | Selected ISD staff involved in testing of protocol against IRs/publication. |
| 0.6 | 25 November 2008 | Revisions following testing and project team review. | ISD Disclosure Group Short Life Working Group |
| 0.7 | 18 December 2008 | Revisions following discussion at ISD Disclosure Group Short Life Working Group. | Disclosure Project Team |
| 0.8 | 3 February 2009 | Final draft for approval by SMT. | SMT |
| 1.0 | 1 March 2009 | First release | All ISD |
| 1.1 | 5 November 2009 | First revision of first release. | Disclosure Control Team |
| 1.2 | 27 November 2009 | Revisions for review by Short Life Working Group | Selected ISD staff involved in reviewing updates of protocol |
| 1.3 | 8 December 2009 | Draft for approval by Head of Statistics | Head of Statistics Team |
| 1.4 | 18 February 2010 | Revisions following Head of Statistics, Caldicott Guardian and Data Protection and FOI Lead feedback and review by Short Life Working Group, | Head of Statistics Team and select ISD staff |
| 1.5 | 5 February 2010 | Draft for approval by Statistical Advisory Group | Head of Statistics Team |
| 1.6 | 24 March 2010 | Revisions following discussion and subsequent comments from Statistical Advisory Group. | Head of Statistics Team |
| 1.7 | 19 April 2010 | Revisions following discussion with ISD Head of Programme and Project Manager | Head of Statistics Team |
| 2.0 | 1 June 2010 | Second Release | All ISD |
| 2.1 | 1 February 2011 | Update instructions regarding Disclosure Form | All ISD |
| 2.2 | 19 January 2012 | Updated links to Disclosure Control Protocol | All ISD |
| 2.3 | 31 May 2012 | Updated following ISD re-organisation | All ISD |
| 2.4 | 09 July 2015 | Draft update for NSS | Disclosure subgroup |
| 2.5 | 15 September 2015 | Draft update for NSS – awaiting sign off at Statistical Advisory Group meeting | Disclosure subgroup, SAG |
| 2.6 | 20 October 2015 | Updated following feedback from SAG meeting | Disclosure subgroup, SAG |
| 3.0 | 26 November 2015 | Third release | All NHS NSS |

# Contents

# 1 Executive Summary

This NSS Statistical Disclosure Control (SDC) Protocol sets out guidance and practices for NSS 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 NSS working group, chaired by the PHI 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. NSS 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 NSS can exert on the use of the data once released. Data shared within NSS 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 topic in question (i.e. how 'sensitive' the topic is)
- the cell values and table design
- 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 NSS 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 NSS Disclosure form (Annex E).

**For all NSS publications a Disclosure form should be completed.**

**For NSS 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 11.1**]**;

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 the NSS Statistical Governance Team nss.nssstatsgov@nhs.net.

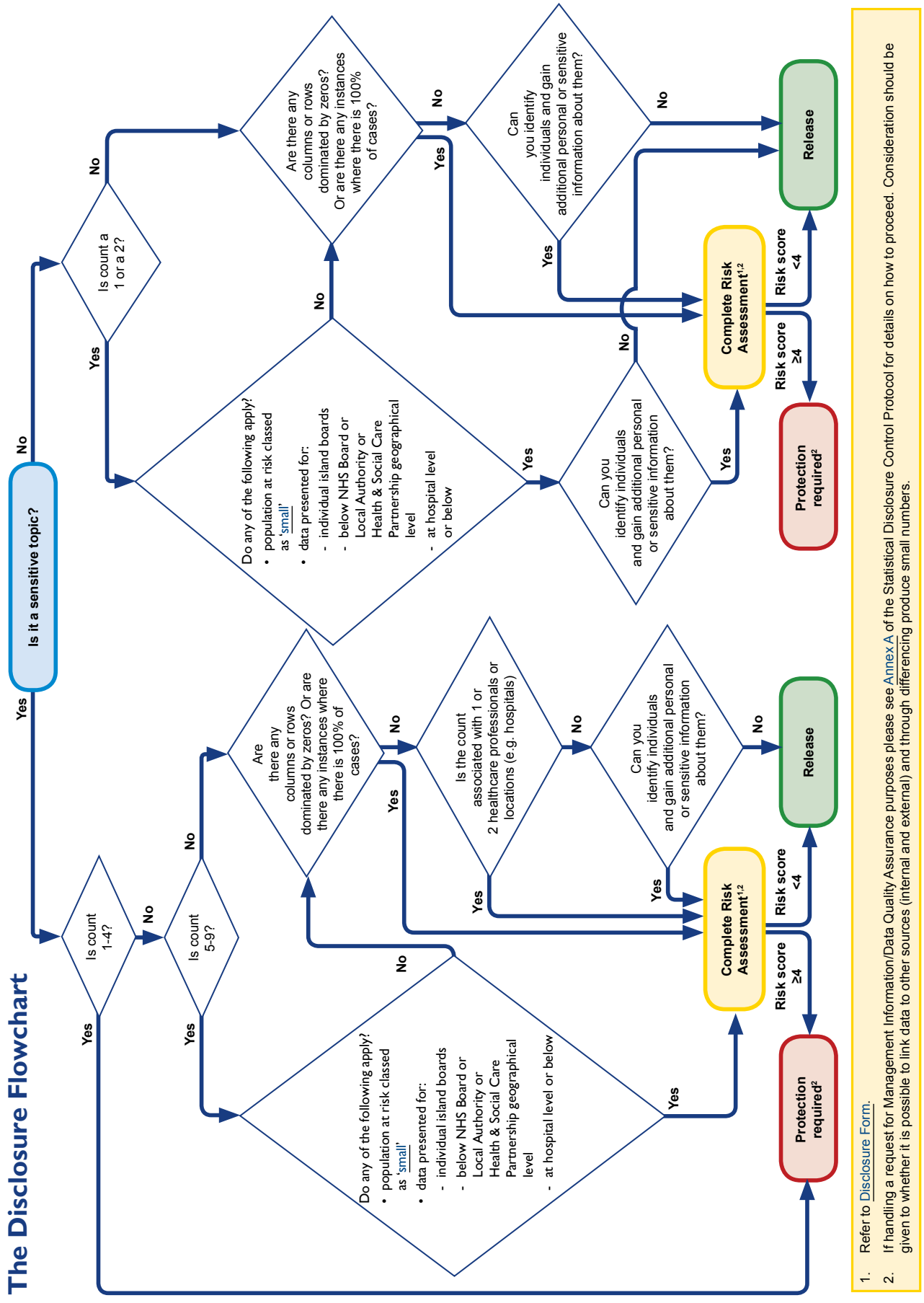
Note that different methods of SDC can result in varying levels of usefulness of the final data to the customer. For example table redesign may provide greater utility than a heavily suppressed table. Therefore the decision on which SDC method to use should take into account the use of the statistics 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 the NSS Statistical Governance Team nss.nssstatsgov@nhs.net.

This protocol mainly covers tabular data (and charts based on tabular data), and there is also a small section that covers micro data.

This protocol is likely to further evolve, following NSS '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
Is it a sensitive topic?
Yes
No
Is count 1-4?
Yes
No
Is count 5-9?
Yes
No
Is count a 1 or a 2?
Yes
No
Do any of the following apply?
• population at risk classed as 'small'
• data presented for:
- individual island boards
- below NHS Board or Local Authority or Health & Social Care Partnership geographical level
- at hospital level or below
Yes
No
Are there any columns or rows dominated by zeros? Or are there any instances where there is 100% of cases?
Yes
No
Is the count associated with 1 or 2 healthcare professionals or locations (e.g. hospitals)
Yes
No
Can you identify individuals and gain additional personal or sensitive information about them?
Yes
No
Can you identify individuals and gain additional personal or sensitive information about them?
Yes
No
Complete Risk Assessment[1,2]
Risk score ≥4
Risk score <4
Protection required[2]
Release
1. Refer to Disclosure Form.
2. If handling a request for Management Information/Data Quality Assurance purposes please see Annex A 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.

# 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 National Services Scotland (NSS) 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 NSS 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 NSS.

## 2.2 NSS's SDC Guidance

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

The third version of the protocol has seen the ISD protocol become an NSS wide version, following work overseen by a disclosure subgroup. The following content has been updated in this release.

This protocol aims to:

- Changed from ISD to NSS version
- Updated disclosure flowchart
- Minor updates to disclosure form
- Updated content for mortality data (section 4.6)
- Updated content for information released for research purposes (section 9)
- Added content for survey / sampled data (section 10)
- Updated content for micro-data (section 12)
- Added content for intermediate outputs (Annex A, section 1.2)
- Changes to geography and populations content and updated information on geographical differencing
- Updated contact information:

- For general disclosure queries, contact the NSS Statistical Governance Team on nss.nsstatsgov@nhs.net

- For disclosure queries relating to geography and populations, contact the PHI Geography, Population and Deprivation Team on NSS.isdGPD@nhs.net

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 practice 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 NSS’ 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 NSS 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**

| Treatment type | Age Group | | | | Total |
|---|---|---|---|---|---|
| | <12 | 12-15 | 16-19 | >19 | |
| Type 1 | 1 | 0 | 7 | 1 | 9 |
| Type 2 | 0 | 0 | 18 | 19 | 37 |
| Type 3 | 0 | 12 | 5 | 0 | 17 |
| **Total** | **1** | **12** | **30** | **20** | **63** |

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, NSS 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.

Another example is The Drug Related Hospital Statistics (DRHS) dashboard contains data on hospitals stays, patients and new patients with a drug related diagnosis in any of the relevant diagnosis positions. The background data is available to download. There were several aggregations, e.g. NHS Board, stays per drug type, which in some cases meant that data may be disclosive (i.e. between 1 and 4 stays, patients or new patients). These were removed in SPSS, but in some cases there was only one instance in a table where a figure was suppressed. In these cases, the analysts chose another figure to suppress (most often the smallest remaining figure). For instance, if a figure for NHS Orkney is the only suppressed figure in a table, analysts would suppress data from another NHS Board. This would mean that the sum of NHS Boards (minus NHS Orkney) would not equal [Scotland - NHS Orkney]. In suppressing another figure, the team have made it impossible to calculate the actual number of either NHS Board.

Consideration should be given to differencing between different geographies and also changes in geographical configurations, such as:

- NHS Board boundary changes 2014
- Data Zone redraw

Further problems arose with the combined dataset, in particular with stays. The sum of SMR01 and SMR04 stays should always equal the combined number of stays. The same is not true for patients and new patients, as one patient may have been admitted to an acute hospital and a psychiatric hospital in the same financial year. To preserve patient confidentiality the team cross-referenced across tables from SMR01, SMR04 and Combined datasets and suppressed any figures that may have were potentially disclosive. For instance, if South Lanarkshire had a suppressed figure in SMR04 but not in either of the other datasets, the analysts would suppress a corresponding figure from SMR01 or Combined for the reasons outlined above.

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 NSS'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 NSS 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 topic in question (i.e. how 'sensitive' is the topic)
- the cell values and table design
- 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 NSS 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 NSS 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 NSS, 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 5,000 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 5,000 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 NSS (within the same publication or Information Request, or indeed a previous IR) or from another source (e.g. National Records of Scotland) 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 2

**Table 2A**

| Age | < 16 | 16-19 | 20-24 | ≥ 25 | Total |
|---|---|---|---|---|---|
| Frequency | 5 | 26 | 13 | 16 | 60 |

**Table 2B**

| Age | < 17 | 17-19 | 20-24 | ≥ 25 | Total |
|---|---|---|---|---|---|
| Frequency | 6 | 25 | 13 | 16 | 60 |

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 NRS 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.

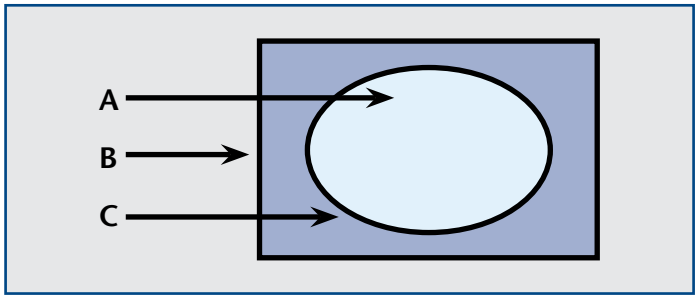
It may, however, be safe to publish directly and indirectly standardised rates. Calculating original counts from such rates would generally require a degree of specialised technical expertise in the associated field which could lower the likelihood of anyone attempting or successfully attempting to calculate 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 NSS, as many of the geographies reported on are not coterminous (for example West Lothian council area had an overlap with Lanarkshire NHS Board using the 2006 NHS Board boundaries).

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 zone 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, including details of the most recent geographical changes and the impact this has regarding disclosure.

To discuss any specific queries further, please contact the PHI Geography, Population and Deprivation (GPD) Team on NSS.isdGPD@nhs.net.

## 4.5 Workforce Statistics

### 4.5.1 NHS Staff

The majority of data published by NSS 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 NSS or NRS sourced data. Analysts are advised to follow the NSS Disclosure Control Protocol when releasing such information in each of the following scenarios:

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

If releasing NRS-sourced data then analysts should be aware of the possibility of differencing with already published NRS data or data that NRS 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.

Guidance on releasing mortality data can be found here: http://genss.nss.scot.nhs.uk/portal/page?_pageid=515,3762564&_dad=portal&_schema=PORTAL

## 4.7 Disclosure Guidance for Parliamentary Questions (PQs)

NSS'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 NSSs own guidance and policies and so SDC should always be considered.

# 5 The Disclosure Flowchart

A flowchart has been designed to help NSS 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 NSS 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 NSS Statistical Governance Team nss.nsstatsgov@nhs.net.

On occasion, NSS will incorporate data previously published by other organisations (for example, Scottish Government, NRS) into our own publications/IRs. Specific guidance on mortality data is contained in Section 4.6. For any other situations NSS 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 detail and an example are provided in Section 3.1.2.

### 5.1.4 Is the Count Associated with 1 or 2 healthcare professionals or locations (e.g. 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 NSS 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**

| | NHS Board X | NHS Board Y | NHS Board Z |
|---|---|---|---|
| **All Abortions** | **625** | **1550** | **3407** |
| *Rate per 1000 live births* | *183.6* | *253.8* | *247.7* |
| *Rate per 1000 women aged 15-44* | *10.6* | *14.7* | *13* |
| **Age of Woman** | | | |
| Under 20 | 174 | 382 | 818 |
| 20 - 24 | 174 | 418 | 1098 |
| 25 - 29 | 121 | 326 | 699 |
| 30 - 34 | 73 | 215 | 410 |
| 35 - 39 | 50 | 155 | 278 |
| 40+ | 33 | 54 | 104 |
| **Grounds for abortion** | | | |
| A | **0** | **150** | **100** |
| B | 0 | 100 | 100 |
| C | 625 | 1000 | 3000 |
| D | 0 | 100 | 100 |
| E | 0 | 100 | 50 |
| Emergency | 0 | 100 | 57 |

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**

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

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 11.1 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 NSS and, due to sensitivity or profile, would necessarily involve senior NSS staff. The NSS Statistical Governance Team nss.nssstatsgov@nhs.net 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 NSS information, applied to the flowchart, are shown in Annex D.

# 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 NSS. 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 (Section 11.1 describes particular situations where this **may not** be the case).

## 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
- the level of **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 NSS 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. NSS 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 NSS 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 the NSS Statistical Governance Team nss.nssstatsgov@nhs.net.

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 NSS 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 the NSS Statistical Governance Team nss.nssstatsgov@nhs.net.

Similarly database modification (record-swapping) should be discussed with the NSS Statistical Governance Team prior to implementing.

Further information on these techniques can be found at:

http://www.ons.gov.uk/ons/guide-method/best-practice/disclosure-control-of-health-statistics/working-paper-3--risk-management.pdf

For NSS compendium publications, advice from the originating NSS data providers should be sought to ensure that the appropriate levels of SDC are applied, and that data is consistent between NSS 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 NSS 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 NSS Data Formatting and Presentation Guidance (http://genss.nss.scot.nhs.uk/pls/portal/url/ITEM/1D3CC471FD3D6203E0440003BA08EF28) lists the symbols which should be used across NSS 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.
- If only 1s and 2s have been suppressed and there are marginal totals then secondary suppression may be necessary to shield their value.
- 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 NSS 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 cannot be identified. Care must therefore be taken to avoid this whilst ensuring trends that do not actually exist cannot 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 NSS's statistics who are familiar with historical NSS 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 NSS and should not be applied without prior discussion with the NSS Statistical Governance Team. 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)

NSS 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. NSS 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 and tailored analyses

Currently requests for the release of patient identifiable information may require a confidentiality statement to be signed or an application made to the Public Benefit and Privacy Panel for Health and Social Care.

Further details can be found here: http://www.informationgovernance.scot.nhs.uk/

For other requests and pending specific guidance on information released for research purposes NSS's SDC protocol should be followed; this may also require discussion with senior managers, the NSS Caldicott Guardian and the NSS Information Governance team.

For guidance on the release of intermediate outputs for researchers, please see section 1.2 of annex A.

# 10 Survey / Sampled Data

When considering disclosure control for reporting results produced from survey data, there is no set protocol that covers all outputs. It should be acknowledged that disclosure risk must be reviewed on a case by case basis. Examples of points to consider include:

- The topic of the survey and any sensitivities associated with the responses
- Any commitments / assurances that were made to survey respondents in respect of how the results would be reported
- The sample size of the overall survey and any results published at sub sample level (e.g. gender, age group, designations, geography)
- When reporting free text responses, is it possible to identify individuals?
- Survey results may require restricted outputs for quality purposes e.g. avoid reporting estimates with potentially high sampling errors.

**Where surveys are run by PHI on behalf of an external organisation, consultation in relation to disclosure protocol should be discussed with the customer.**

For any further enquiries, please contact the NSS Statistical Governance Team nss.nsstatsgov@nhs.net.

# 11 Documenting Disclosure Decisions

The issue of disclosure concerns all of NSS'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 NSS publications (including contributions to 'non NSS' publications)**
- **certain Information Requests, including FoI and PQs (**see Section 11.1)

Note that information released by NSS 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 or**
- **to undertake a risk assessment (with the resultant outcome being to either protect or release).**

(Section 11.1 below describes particular situations where this **may not** be the case.)

## 11.1 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.

## 11.2 Guidance for Completion of Disclosure Form

### 11.2.1 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, the Service Manager/Information Consultants and the Statistical Governance team.

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*

    a. 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*.

- Exactly what data is requested? *Data sources, variables, time periods etc*
- Who is requesting the data? *Named contact, position/role, organization*
- What is the intended use of the data? *e.g. inform committee meeting, FOI, PQ, research paper, publication, SG policyWho 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 organizations*

  a. 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.

### 11.2.2 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 values were 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 an Associate Director, Head of Service, Service Manager, Information Consultant or Consultant in Public Health Medicine for sign off. This should be done electronically and the form then stored by the analyst for future reference.

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

Completed disclosure forms provide evidence for future assessment of NSSs

practice and evolution of guidance. It is also a good reference for future releases.

# 12 Micro-data (including data extracts)

Micro-data is data collected at an individual level, such as patient level data or data collected for a survey or for the Census.

The disclosure protocol does not apply for micro-data as different processes apply and a confidentiality agreement is required.

For more information on the release of micro-data, please contact the NSS Statistical Governance Team nss.nssstatsgov@nhs.net.

# 13 Databases (e.g. HEAT, Navigator, PRISMS, ACADME)

This guidance refers to 'databases' (or 'warehouses', or 'datamarts', etc) that are made accessible to NSS and non NSS staff for analysis e.g. the extraction of reports. **There will be a range of types of databases made available to non-NSS 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 NSS - 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 NSS 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 NSS's Statistical Disclosure Control Protocol http://www.isdscotland.org/Products-and-Services/Data-Protection-and-Confidentiality/#smallNumbers. 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.*

# Annex A[1] Information released for Management or Data Quality Assurance Purposes (including Peer Review)

If handling a request for management information or information for data quality assurance purposes users will be directed from the Disclosure Flowchart to this section of the Protocol for further guidance on how to proceed. Following this guidance will in some cases result in a different outcome to that of the flowchart. A risk assessment form is not always necessary and the guidelines below should be read to determine whether or not a form should be completed.

## 1.1 Information released for Management Purposes

Management information is information provided for management (or 'operational' or 'clinical management') purposes. One of NSS's key functions is to assist partner organisations (mainly NHS Boards, but also others) plan, monitor and evaluate health and care services; this often requires supporting data. Sometimes this data would not be deemed suitable for general publication so NSS needs to minimise the risks associated with this data while still providing partner organisations with the data they need to operate, plan and monitor services.

As for published data, the risks of statistical disclosure still need to be considered, and the principles involved in risk assessment are the same. However, in general, given the recipients and their intended use of the data, the risks of disclosure will be much lower than for general publications, in particular the risk of an attempt to disclose. Risks can often be reduced by taking certain steps such as marking the information released as **management information only** and discussing the risks of disclosure with the recipient. This is what allows data that would be deemed potentially disclosive in the public domain to be released to the health service and key partners.

Often the greatest risk that needs to be considered is not the risks of disclosure by the recipients themselves but the risks posed by onward distribution or inadvertent release into the public domain. Focus therefore will often be on the actions necessary to reduce this risk.

While an exact definition of 'management information' is not possible, the following list should be taken as a guide:

- data to monitor and improve a service;
- reports and analysis provided to programme boards or steering groups;
- data to monitor targets;
- service planning
- quality control/improvement
- performance management/targets
- routine reports to NHS Boards etc.;

Often these purposes will require data that may contain small numbers or be disclosive in other ways, even within just a small proportion of the outputs. Sometimes the user requires the full unadjusted output, due, for example, to the inclusion of data for small NHS Boards or for tracking financial flows or target performance. Often management information will be in the form of routine standard reports which could, on occasion, contain some cells with a higher risk of disclosure due to small numbers.

1 For internal reference only.

A key principle however, as for published data, is to release data only at the level of detail needed to fit the purposes of the user and avoid releasing unnecessary disaggregations that may pose a higher risk of disclosure. **It is important to discuss needs with the user when the request involves potentially disclosive data.**

Regardless of who is requesting the information or the intended audience of the data, if the decision has been made to release potentially disclosive data for management information purposes it is NSS's responsibility to highlight this to the recipient and the following rules should be applied:

- Each table within the release should be labelled as 'Management Information only, not for onward distribution'. This statement should appear prominently at the top of each table.
- The **standard text** below should be attached as a footnote to each table:
  *This information has been released for management information purposes only. The data have not been adjusted to protect against potential disclosure risks and 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 NSS's Statistical Disclosure Control Protocol*
  http://www.isdscotland.org/Products-and-Services/Data-Protection-and-Confidentiality/#smallNumbers
  Please contact xxx@nhs.net if you have any queries regarding this.
- If information is presented in meetings, NSS staff should reiterate verbally that it is for management information purposes only. In such cases, NSS staff should be aware of the attendees and the organisations/groups they are representing, prior to sharing the information.

To help NSS staff determine whether a request for potentially disclosive data can be released, there is a list of considerations contained within the risk assessment – these apply to all types of information to be released, including that intended for management purposes. It is important to note that this protocol does not set 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 which also applies to all types of information to be released, including that intended for management purposes.** There are several key points that should specifically be considered when handling a request for management information:

- is the data considered sensitive or non-sensitive
- is an organisation requesting their own data or that of another organisation (if that of another organisation then the requesting organisation should be able to demonstrate that they have the other's approval)
- what is the intended audience or distribution e.g. restricted internal, likely to be widely circulated or presented.

## Non-sensitive data

If the management information request is for data that is considered **non-sensitive** and the intended audience is restricted to the recipient's own organisation then this is unlikely to pose a high risk of disclosure. The data should be labelled as for management information only (described above) and the standard text attached.

Where it is possible that the audience is not restricted to the recipient's own organisation then in addition to labelling the data as for management information only and attaching the standard text further steps are needed to reduce the risks of disclosure as much as possible. These might include discussion with the recipient to ensure they are aware of the potential risk of disclosure and the issues surrounding that risk and drawing attention to the standard text to make the recipient fully aware of their responsibility.

In both instances it is not necessary to complete a Disclosure Form.

## Sensitive data

If the request is for data that is **sensitive** but the intended audience is restricted to a recipient's own

organisation then this is unlikely to pose a high risk of disclosure. The data should be labelled as management information only and the standard text attached. Steps should be taken to reduce the risks of disclosure as much as possible. These could include discussion with the recipient to ensure they are aware of the potential risk of disclosure and drawing attention to the standard text to make the recipient fully aware of their responsibility. In this instance a disclosure form is not required.

Where it is not known if the audience is restricted to the recipient's own organisation a risk assessment should be completed and a disclosure form submitted. Depending on the outcome of this assessment disclosure control may need to be applied. In these instances discussion with the user is required to determine whether requirements can be changed or onward distribution controlled while still meeting the user's needs. The data should be labelled as for management information only and the standard text attached.

Sensitive data to be released to a wide or undefined audience across a number of organisations may pose a high risk of disclosure even if the intended purpose is for management. In some cases where this information is needed unadjusted, it may be necessary to treat the request in the same way as requests for confidential data (e.g. in some cases a confidentiality statement may be required).

### Summary:

Judgement is required and as a guide:

#### Non-sensitive Data

- Where the audience is restricted to recipients own organisation - attach management information label and standard text.
- Where it is possible that the audience is not restricted to the recipients own organisation – attach management information label, standard text and take steps to reduce the risks of disclosure.

In both instances a disclosure form is not required.

#### Sensitive Data

- Where the audience is restricted to recipients own organisation – attach management information label, standard text and take steps to reduce the risks of disclosure. A disclosure form is not required.
- Where it is not known if the audience is restricted to the recipients own organisation - complete risk assessment and disclosure form, disclosure control may be required, attach management information label and standard text.
- Where it is a wide or undefined audience across a number of organisations this may pose a high risk of disclosure. Where information is needed unadjusted it may be necessary to treat the same as requests for confidential data.

## 1.2 Information released for Data Quality Assurance (QA) purposes (including Peer Review)

Quality assurance, in this instance, refers to the sharing of data with others other than the original data provider, though it may also include them. This refers to outputs, e.g. publications and reports, and not the data collection and maintenance process. A common example is the QA of figures as part of the preparation for public release. When data is shared for quality assurance purposes it may involve organisations that are not the original data provider. If the information is for quality assurance purposes and the data is potentially disclosive then the guidelines provided above in 1.1 which apply to information released for management purposes should also be followed when releasing information for QA purposes.

The label and standard text attached to the data should also be amended for quality assurance purposes as follows:

- Each table within the release should be **labelled** as 'Data Quality Assurance only, not for onward

distribution'. This statement should appear prominently at the top of each table.

- The **standard text** below should be attached as a footnote to each table:
  This information has been released for data quality assurance purposes only. The data have not been adjusted to protect against potential disclosure risks and 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 NSS's Statistical Disclosure Control Protocol http://www.isdscotland.org/Products-and-Services/Data-Protection-and-Confidentiality/#smallNumbers
  Please contact xxx@nhs.net if you have any queries regarding this.
- If information is presented in meetings, NSS staff should reiterate verbally that it is for data quality assurance purposes only. In such cases, NSS staff should be aware of the attendees and the organisations/groups they are representing, prior to sharing the information.

**Information released to a researcher as an intermediate output for the purpose of discussion with colleagues and not for publication**

Researchers carrying out data analysis in a safe haven using potentially disclosive data may require the release of intermediate outputs during the course of the project. Access to a safe haven is strictly limited and some research colleagues will have no access. Intermediate outputs are used in discussion with supervisors and colleagues within a research group to make decisions about the direction of a research project. This is particularly useful when the researcher is a PhD student. If an output (which could, for example, be a table, a graph or regression output from a statistical package) is cleared for release for this purpose then standard text equivalent to that used above for QA purposes will be attached as a footnote (see below). In particular, it must be made clear to the researcher that the output is not for publication. If an intermediate output is required at the end of the project for a publication then it must go through the disclosure process again because more stringent criteria may be applied. This would be the case if other outputs requested at the end of the project might, in combination with this one, result in disclosure.

- The **standard text** below should be attached as a footnote to all intermediate outputs released to a researcher:
  *This information has been released for internal use by the research team only. The data have not been adjusted to protect against potential disclosure risks and 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 to the named individuals in the research team (as listed in the data access application form) and that patient confidentiality is not compromised. This information must not be disseminated beyond the research team and is not for publication.*

# Annex B[2] Geographies and Populations

Smaller sized geographies or populations will commonly increase the likelihood of disclosure of information about an individual or group of individuals, and this should be taken into account when assessing the risk of disclosure for any analysis. In addition to the size of the reporting geography or population, NSS analysts also need to be aware of residual disclosure or differencing.

Disclosure by differencing occurs whereby the comparison of two or more tables reveals information that is not available in any single table. Another form of disclosure by differencing is geographical differencing, which may be an issue when the boundaries of two geographic areas overlap.

For example, there used to be overlap between health board (1st April 2006 configuration) and local authority boundaries. Following the health board boundary changes on 1st April 2014 this is no longer an issue; however the following example highlights the issue if the 2006 health board configuration is used for analyses.

If one table contains data relating to a Local Authority and another contains data relating to a Health Board (2006), it is possible that table values may be subtracted to reveal confidential information about the overlapping geography. This is an important issue for NSS to bear in mind as information is frequently published from the same source at differing geographical levels.

There are many different geographical classifications available for analysis in NSS - from postcode level up to Scotland level. It should be noted that not all these geographical classifications are co-terminous (i.e. share the same boundaries) and hence gaps or overlaps can occur to create “slivers” or differencing issues. These slivers will potentially have small populations and so care should be taken when analysing and publishing data at varying geographical levels to reduce the risk of any disclosure.

The document ISD Geographical Reporting and Disclosure Control Issues [PDF] was compiled in 2009 and contains information on geographies and populations **relevant at that time**. The document:

- Identifies the main geographical classifications used by analysts in NSS and defines them thoroughly;
- Identifies ‘slivers’ or areas of concern in Scotland, where the main geographical boundaries are not coterminous;
- Assesses the extent of the populations concerned, where ‘slivers’ occur.

**Please note:** as this document was produced several years ago, some of the information is out of date and no longer applies.

**For up-to-date advice regarding geography and populations, please contact the PHI Geography, Population and Deprivation (GPD) team on NSS.isdGPD@nhs.net.**

Examples of recent changes regarding geographies:

**(i) NHS Board Boundary changes on 1st April 2014**

NHS Board boundaries changed on 1st April 2014 and now align with those of local authorities. As the boundaries are now aligned, this eliminates the ability to identify small numbers by differencing that previously existed between local authority and 2006 NHS Board boundaries.

However, due to this boundary change consideration must now be given to differencing between NHS Board level statistics previously produced based on the 2006 NHS Board boundaries compared with the same statistics now being produced based on the 2014 NHS Board boundaries.

If differencing in this way causes a disclosure risk, then please contact the GPD team for further advice.

2 For internal reference only.

### (ii) Data Zone redraw on 6th November 2014

Data Zones were first created in 2004 and were based on the 2001 Census. On 6th November 2014, Data Zone boundaries were redrawn by the Scottish Government to take changes in population since the first edition into account and ensure a more consistent population size. The new redrawn Data Zones have been based on the 2011 Census, and are known as '2011 Data Zones', with the original Data Zones being known as '2001 Data Zones'.

As a result of this, consideration must be given to Data Zone level statistics. Differencing can occur when comparing statistics for a 2001 Data Zone and its closest equivalent 2011 Data Zone. It is possible to work out the difference between the two and assign this to the small non-overlapping area between the two sets of boundaries.

Information relating to Data Zone matching and the degree of fit between the 2 different sets of boundaries can be found on the Scottish Government's website: http://www.gov.scot/Topics/Statistics/sns/SNSRef/DZMatchingQGuide.

If differencing in this way causes a disclosure risk, then please contact the GPD team for further advice.

# Annex C[3] Legal and Policy Background

## 1.1 Background

In its 2006 'Working Paper 1: Confidentiality Protection – Legal and Policy Considerations' the Office for National Statistics states:

> 'For health statistics, it is essential that any published statistic respects the privacy of the information shared by individuals with health and statistics professionals. Failure to respect this privacy might result in harm or distress to a specific individual. Such a breach may have a wider effect. A breach of privacy in Official Statistics could damage the relationship of trust between private individuals and health and statistics professionals. Thus the public interest is served when statistical records are kept strictly confidential.'

However, ONS is clear that not only is there a 'public interest in confidentiality of individual information', but also that measures to protect confidentiality must not be excessive so as to harm the usefulness of Official Statistics.

The ONS guidance has helped form part of the Code of Practice for Official Statistics (UK Statistics Authority, 2009)

NSS aims to meet the standards set by ONS and the UK Statistics Authority, and, equally, be careful to ensure that it meets all legal obligations associated with the handling of information, including those of the Freedom of Information (Scotland) Act 2002.

## 2. Relevant Law

The confidentiality of identifiable personal health information is legally protected in a number of key ways, listed below:

### 2.1 The Human Rights Act 2000

This requires public authorities to comply with obligations set out in the European Convention on Human Rights (ECHR) to safeguard human rights in a number of ways. Article 8 requires that individuals' right to private and family life is respected, and that this right is only 'interfered' with where it is lawful, necessary and proportionate for certain purposes.

### 2.2 The Data Protection Act

This Act is the UK enactment of the European Union Data Protection directive and it sets standards for, and safeguards individuals' rights in relation to, how identifiable personal information is used and disclosed. The definition of 'personal data' is broadly drawn:

'personal data' means data which relate to a living individual who can be identified-

(a) from those data, or

(b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller

and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual'

3 For internal reference only.

With some limited exemptions, personal data must be used and disclosed in a way that meets the standards known as the 'Data Protection Principles' set by the Act. These include the requirements of fairness and lawfulness, appropriate data quality and security and the right of access.

### 2.3 Common law duty of confidence

For information to be confidential in law, it must have been imparted in circumstances importing an obligation of confidence and it must have the "necessary quality of confidence", this means:

- The information is in fact confidential i.e. not in the public domain, not common knowledge or not easily available by other means. That is not to say it must be a secret, but is not widely available.
- The information must be worthy of protection i.e. is neither useless nor trivial.
- The public interest that confidences should be preserved outweighs some other public interest, which favours disclosure.

The unauthorised use of confidential information may give cause for an action for damages through a breach of legal duty.

### 2.4 Protection of these within the Freedom of Information (Scotland) Act 2002

The important legal requirement imposed on public authorities by the Freedom of Information (Scotland) Act 2002 (FOISA) to disclose information when asked does not override the legal requirements outlined above. Any disclosure of data deemed to be personal data in terms of the Data Protection Act may only proceed if it complies with the requirements set down by the Data Protection Act. Also, 'confidential' information is protected from disclosure under FOISA so long as the requirements set out in section 36 of the Act are met.

### 2.5 Confidentiality in NSS

NSS staff, in common with those of the rest of NSS and the NHS more widely, are contractually obliged to protect confidentiality. NSS staff are required to read, understand, and adhere to the NSS Confidentiality Guidelines which are published on geNSS.

Information governance matters generally, including those of confidentiality, are led in NSS through its Information Governance group, which is chaired by NSS's Caldicott Guardian.

### 2.6 Code of Practice for Official Statistics

NSS are required to follow the Code of Practice for Official Statistics, which were issued by the UK Statistics Authority in 2009. NSS's publications are assessed by the UK Statistics Authority, against the Code of Practice. The Code of Practice includes the following:

*Principle 5: Confidentiality*

*Private information about individual persons (including bodies corporate) compiled in the production of official statistics is confidential, and should be used for statistical purposes only.*

*Practice 4. Ensure that arrangements for confidentiality protection are sufficient to protect the privacy of individual information, but not so restrictive as to limit unduly the practical utility of official statistics. Publish details of such arrangements.*

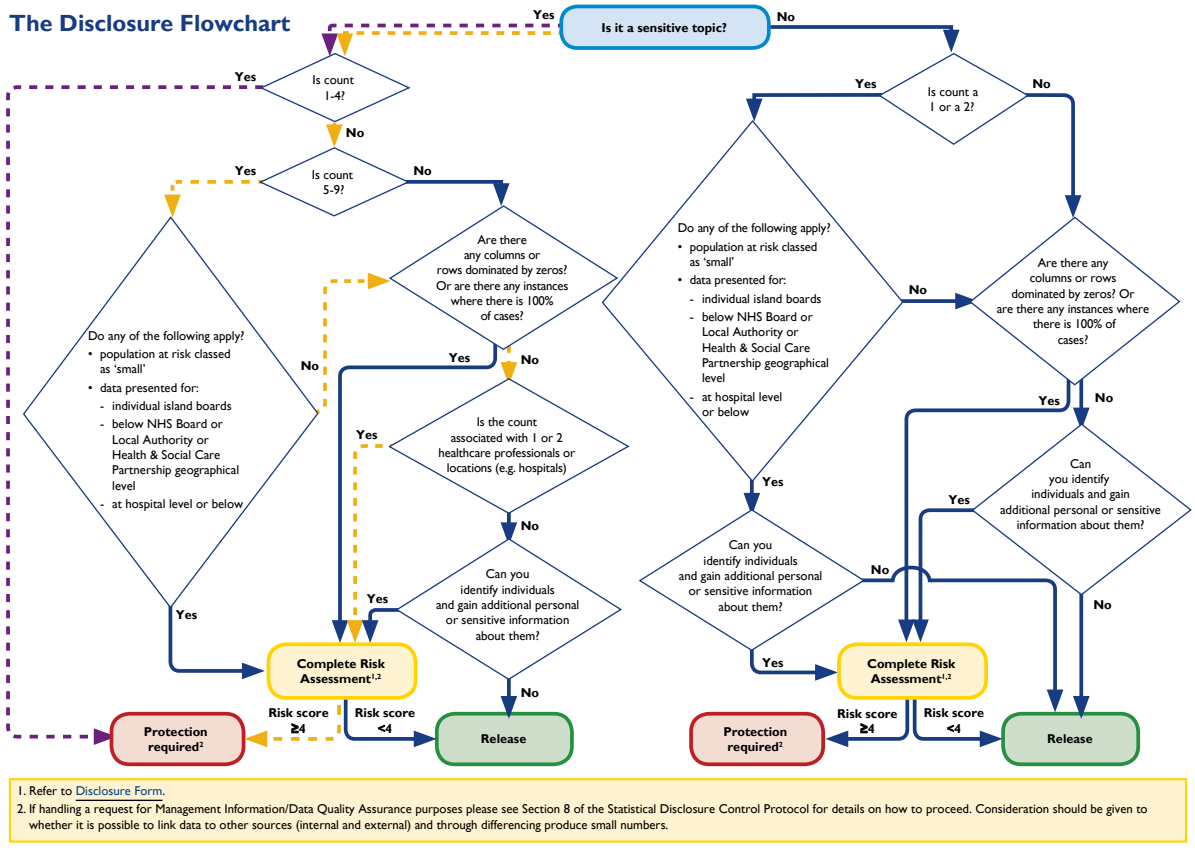
# Annex D[4] Examples Using the Disclosure Flowchart

1. The table below is an extract from a recent publication on abortions. The **purple** route illustrates that a count of four would lead to 'Protection Required'. The **yellow** route on the flowchart below demonstrates when other factors including population size and number of hospitals have to be considered. In this example a risk assessment is undertaken and the decision reached is that protection **is** required.

## Abortions 1 performed in Scotland - NHS board analysis

| | All Areas | Area of residence | |
|---|---|---|---|
| | | Ayrshire & Arran | Borders |
| **All Abortions** | **13 703** | **804** | **183** |
| ***Rate per 1000 live births*** | ***237.2*** | ***208.2*** | ***172.8*** |
| ***Rate per 1000 women aged 15-44*** | ***13.0*** | ***11.4*** | ***10.4*** |
| **Age of Woman *** | **Number** | | |
| Under 20 | 3 548 | 234 | 51 |
| 20 - 24 | 4 116 | 234 | 54 |
| 25 - 29 | 2 732 | 149 | 36 |
| 30 - 34 | 1 640 | 96 | 31 |
| 35 - 39 | 1 219 | 69 | 4 |
| 40+ | 448 | 22 | 7 |

Note The above table shows artificial data for the purposes of this guidance only.



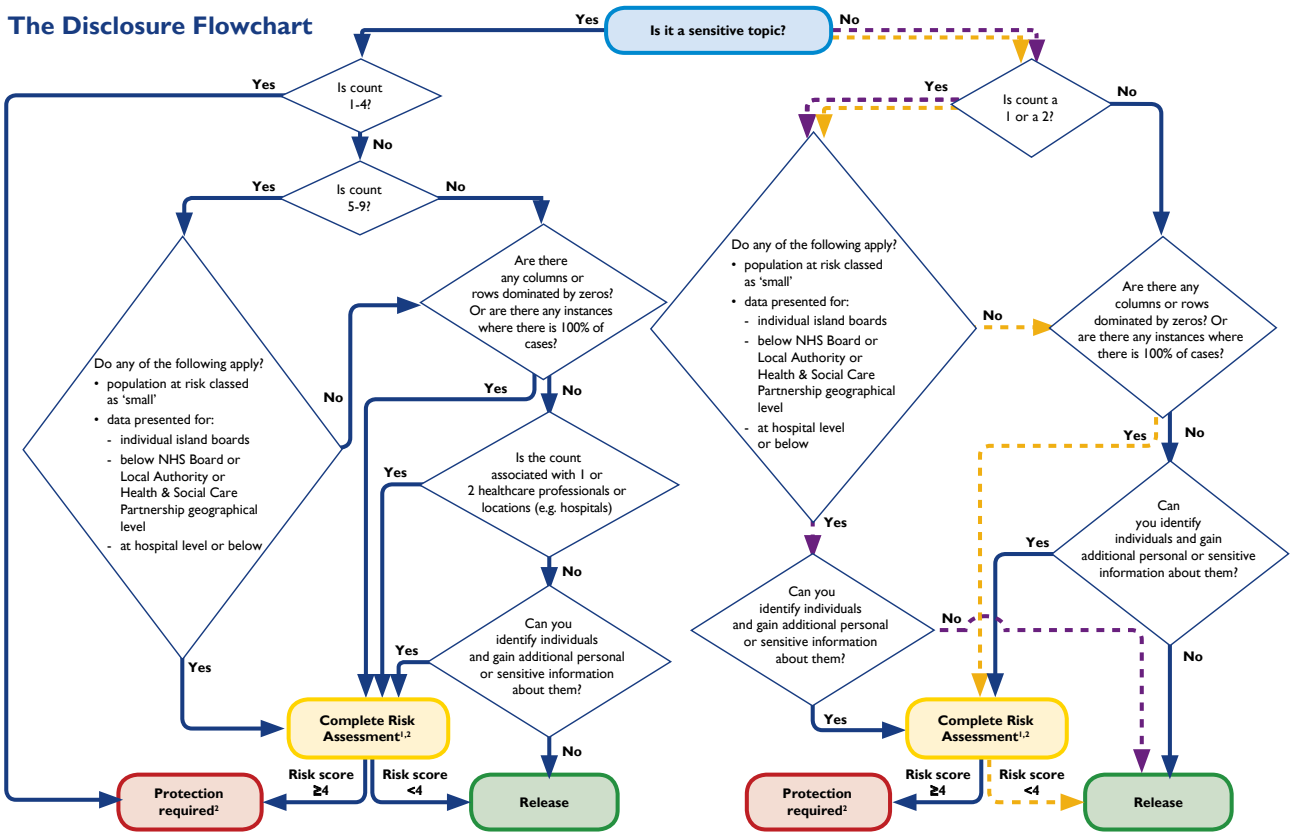

4 For internal reference only.

2. The table below is an extract from a recent publication on delayed discharges where it is possible to determine that the two people delayed in NHS Shetland were both 'waiting to go home'. Following the **purple** route on the flowchart below the decision is made to release the data, having assessed the information as neither personal nor sensitive enough to warrant protection. If this information had been in a larger NHS board, e.g. NHS Lothian, the **yellow** path illustrates when a risk assessment would need to be undertaken because of the domination of zeros which, in this example, results in the decision to release the data.

## Table 3 NHS Delayed Discharges

**Number of NHS delayed discharges: Principal reason for delay: NHS board area of treatment: as at October 2008 census**

| NHS Board area of treatment | Total | Number outwith the six week discharge planning period | Principal reason group | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Community Care Assessment | Patients waiting to go home | Awaiting funding for a care home placement | Awaiting place availability in a care home | Healthcare Assessment Arrangements | Patient exercising statutory right of choice | Legal and Financial | Other |
| **Scotland** | **678** | **92** | **298** | **64** | **51** | **224** | **13** | **11** | **5** | **12** |
| Orkney | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Shetland | 2 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 |




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.

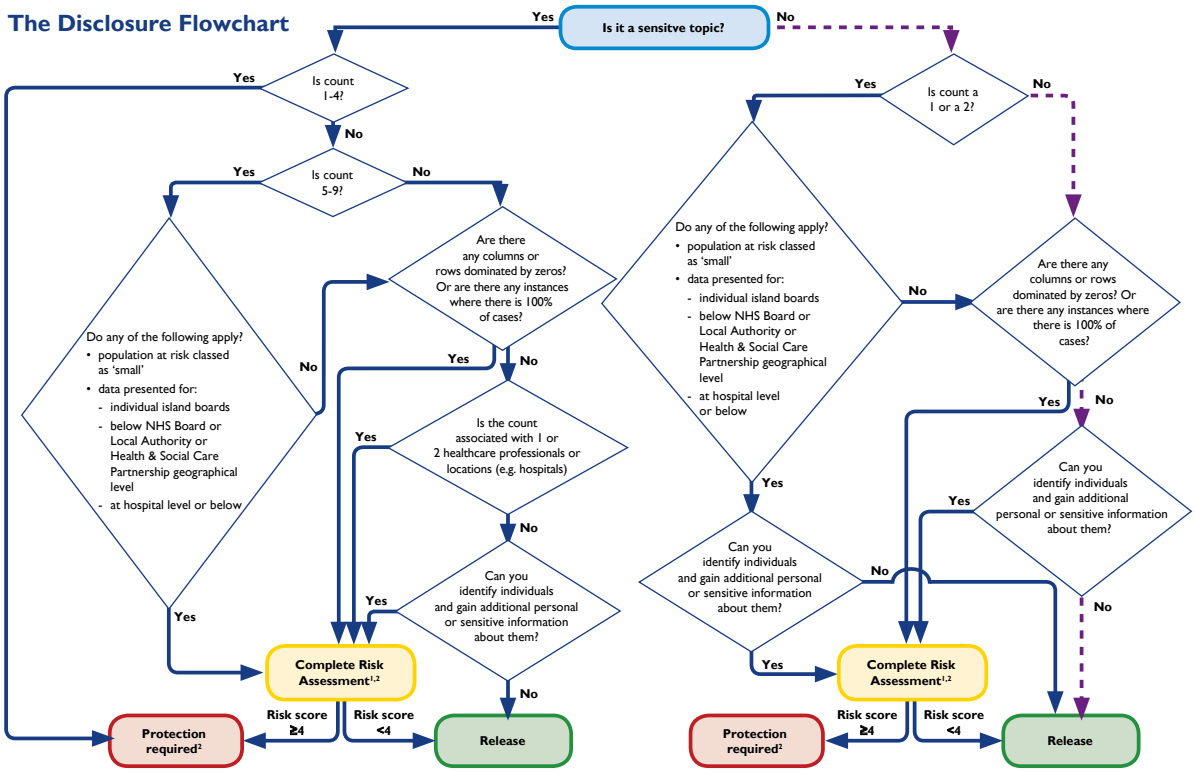
3. The table below on primary immunisation uptake rates by 12 months old children provides an example of when the risk of disclosure is low and, as the **purple** route on the flowchart below shows, the data is released. Note that there are no counts of 1 or 2, nor are any of the percentages based on a count of 1 or 2.

## Primary Immunisation Uptake Rates by 12 months old

**Evaluation Period: 1 April 2007 - 31 March 2008**
**Born 1 April 2006 to 31 March 2007**

| NHS Board | Number in Cohort | % completed primary course by 12 months | | | | | |
|---|---|---|---|---|---|---|---|
| | | D | T | P | Pol | Hib | MenC |
| Ayrshire & Arran | 3881 | 97.3 | 97.3 | 97.3 | 97.2 | 97.1 | 97.6 |
| Borders | 1125 | 97.2 | 97.2 | 97.2 | 96.9 | 97.0 | 96.6 |
| Dumfries & Galloway | 1533 | 98.1 | 98.1 | 98.1 | 98.0 | 97.7 | 97.9 |
| Fife | 4035 | 96.7 | 96.7 | 96.7 | 96.6 | 96.7 | 96.5 |
| Forth Valley | 3285 | 96.8 | 96.8 | 96.8 | 96.8 | 96.7 | 96.9 |
| Grampian | 5962 | 96.9 | 96.9 | 96.9 | 96.9 | 96.9 | 95.9 |
| Greater Glasgow & Clyde | 13092 | 96.5 | 96.5 | 96.5 | 96.4 | 96.1 | 96.2 |
| GG&C (part 1) | 9611 | 96.3 | 96.3 | 96.3 | 96.2 | 95.8 | 95.9 |
| GG&C (part 2) | 3481 | 97.2 | 97.2 | 97.2 | 97.1 | 97.0 | 97.0 |
| Highland | 3038 | 95.1 | 95.1 | 95.1 | 95.1 | 94.7 | 95.1 |
| Highland (part 1) | 2251 | 94.3 | 94.3 | 94.3 | 94.2 | 93.7 | 94.4 |
| Highland (part 2) | 787 | 97.5 | 97.5 | 97.5 | 97.5 | 97.5 | 97.0 |
| Lanarkshire | 6723 | 96.9 | 96.9 | 96.9 | 96.8 | 96.8 | 97.1 |
| Lothian | 9073 | 97.3 | 97.3 | 97.3 | 97.1 | 96.8 | 96.1 |
| Orkney | 207 | 76.3 | 76.3 | 76.3 | 74.9 | 75.8 | 74.9 |
| Shetland | 273 | 97.4 | 97.4 | 97.4 | 97.4 | 97.4 | 96.7 |
| Tayside | 4115 | 97.1 | 97.1 | 97.1 | 97.1 | 97.0 | 96.5 |
| Western Isles | 293 | 92.8 | 92.8 | 92.8 | 92.2 | 92.2 | 92.2 |
| NHS Board unknown (Former A&C) | 37 | .. | .. | .. | .. | .. | .. |
| **Scotland** | **56672** | **96.7** | **96.7** | **96.7** | **96.6** | **96.5** | **96.3** |

## 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.

# Annex E[5] Disclosure Form

## Disclosure Form

This form should be completed for all NSS publications and IRs where the Disclosure Flowchart has instructed to undertake a **Risk Assessment and/or Protect**. The completed form, once signed off, should be stored locally for further reference. Please use the Disclosure Protocol for guidance on completing the form.

**Pub. Title or IR Number:** _____

**Name:** _____

**Date:** _____ **Team:** _____

**PART ONE: RISK ASSESSMENT -** *only complete as instructed on Flowchart*

Indicate below what you think the likelihood of disclosure is and the level of impact that this disclosure would cause:

| **Likelihood of an attempt to disclose** | | **Impact of disclosure** |
|---|---|---|
| ☐ High (3) | | ☐ High (3) |
| ☐ Medium (2) | **X** | ☐ Medium (2) |
| ☐ Low (1) | | ☐ Low (1) |

Questions to consider when undertaking a risk assessment include:

- Exactly what data is requested*? Data sources, variables, time periods etc*
    - *a.* 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).*

If by multiplying the *likelihood* by *impact* the score is 4 or above, disclosure control methods should always be applied 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 Two of this form. In some circumstances, you may also wish to consider applying disclosure control for lower scores especially if the topic is sensitive.

**Now complete part two of this form.**

VERSION: Nov 2015

**PART TWO: DETAILS OF DISCLOSURE CONTROL** - *complete for **all NSS publications** and **IRs where a Risk Assessment has been undertaken OR protection of data is instructed as indicated on the Flowchart.***

**1. Did you apply any Disclosure Control techniques prior to releasing this data?**

☐ Yes, go to question 2

☐ No, go to question 5 to explain why no disclosure control was applied

**2. Indicate what the potential disclosure risks were with this data (tick all that apply):**

☐ Cells with values of 1 - 4

☐ Cells with values of 5 - 9

☐ Sensitive topic

☐ Identify individuals & gain additional personal or sensitive information

☐ Rows/Columns dominated by zeros

☐ 1 or 2 hospitals/practitioners

☐ Small population/geography or institution level

☐ Other, please specify:

_____

**3. Indicate which disclosure control techniques you applied:**

☐ Table Redesign ☐ Cell suppression

☐ Other, please specify (Head of Service, Statistics Support must approve other methods):

_____

**4. Explain below the impact of applying disclosure control on the various tables of data. If this is different for each table please document details for each. You should also explain why the method(s) were chosen.**

| |
|---|
| |

**5. Additional comments (please use the space below to provide any further relevant information. Use this space to document why you did not decide to apply any disclosure control prior to release.)**

| |
|---|
| |

**6. Sign off: (should be undertaken by a Associate Director, Head of Service, Information Consultant, Service Manager or CHPM)**

| **Name:** | | **Date:** | |
|---|---|---|---|

5 For internal reference only.

# Annex F[6] τ-ARGUS Software

τ-Argus (Tau-argus) is a software tool, developed by Statistics Netherlands, which is designed to assist the user in producing 'safe' tables, and protect against the risk of disclosure.

Data (either tabular or micro-data) is read in to the τ-Argus package, and from disclosure rules (that the user sets themselves) tables are created and scrutinised for potentially disclosive cells. Unsafe cells are highlighted and can be protected by modifying the table(s) so that it contains less detailed information.

τ-Argus allows for several modifications of a table: a table can be redesigned (by combining rows or columns); sensitive cells can be suppressed (with primary and secondary suppression); and controlled rounding is possible.

NSS has a certain degree of experience in τ-Argus (Tau-Argus) software, which can help with the above methods of SDC and NSS staff should consider utilising τ-Argus.

Contact the NSS Statistical Governance Team on nss.nsstatsgov@nhs.net if you would like further information or wish to use this package.

# Annex G[7] Examples of Statistical Disclosure Control Methods

**Extract based on the Scottish Government's Practical Guide to Statistical Disclosure Control (the categorisation in Example B has been amended from salaries to ethnicity to reflect current NSS practice)**

## (a) Table Redesign

Example A illustrates the process of table redesign. The first table shows information about the number of people in a local authority who are suffering from illnesses A, B and C by age group. We shall assume that, because the data is sensitive, the data owner considers cell values of less than 5 to be disclosive. There are five such cells in the table, shown in boxes.

In order to protect the table without actually altering the data, the age groups could be combined to form 10-year intervals instead of 5-year intervals. It can be seen that changing the spanning variables in this way has protected the sensitive data and produced a table which can safely be released into the public domain. It is important to be consistent in groupings within variables, between tables produced, to avoid disclosure by differencing.

### Example A

| | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | 50-54 | 55-59 | Total |
|---|---|---|---|---|---|---|---|---|---|
| **A** | 5 | 9 | 14 | 11 | 20 | 3 | 14 | 10 | 86 |
| **B** | 1 | 6 | 4 | 2 | 8 | 9 | 3 | 11 | 44 |
| **C** | 24 | 28 | 19 | 22 | 35 | 39 | 28 | 27 | 222 |

| | 20-29 | 30-39 | 40-49 | 50-59 | Total |
|---|---|---|---|---|---|
| **A** | 14 | 25 | 23 | 24 | 86 |
| **B** | 7 | 6 | 17 | 14 | 44 |
| **C** | 52 | 41 | 74 | 55 | 222 |

## (b) Suppression

The table in Example B contains information about peoples' ethnicity in Local Authority X, according to their occupation. Assume that the data owner recommends that any cell value of 5 or less should be suppressed. The table shows that one nurse in X is ethnicity B and four teachers are ethnicity D - these cells are disclosive and must therefore be suppressed. Additional suppressions are also required because both row and column totals are shown in the table and can easily be used to work out the missing values.

7 For internal reference only.

### Example B

| | 50-<60 | 60-<70 | 70-<80 | 80-<90 | 90-<100 | Total |
|---|---|---|---|---|---|---|
| **Lawyer** | 16 | 7 | 11 | 21 | 6 | **61** |
| **Nurse** | 10 | 1 | 6 | 17 | 16 | **50** |
| **Police** | 11 | 13 | 19 | 9 | 15 | **67** |
| **Teacher** | 9 | 8 | 12 | 4 | 14 | **47** |
| **Total** | **46** | **29** | **48** | **51** | **51** | **225** |

If we first consider the single nurse earning between £60k and £70k, it can be seen that two secondary suppressions are required to protect this cell - one from the occupation row and another from the income column. The same is also true of the disclosive 'teacher' cell. However, due to the positioning of these two disclosive cells, it is in fact possible to protect them simultaneously using only two secondary suppressions instead of four, thus preserving more of the original data.

| | 50-<60 | 60-<70 | 70-<80 | 80-<90 | 90-<100 | Total |
|---|---|---|---|---|---|---|
| **Lawyer** | 16 | 7 | 11 | 21 | 6 | **61** |
| **Nurse** | 10 | * | 6 | * | 16 | **50** |
| **Police** | 11 | 13 | 19 | 9 | 15 | **67** |
| **Teacher** | 9 | * | 12 | * | 14 | **47** |
| **Total** | **46** | **29** | **48** | **51** | **51** | **225** |

The second table shows how the secondary suppressions have successfully protected the data by removing two cell values, thus preventing disclosure by subtraction.

In practice, there are usually more than two disclosive cells, which can cause the number of secondary suppressions required to rise substantially.

## NSS Example of Using Controlled Rounding as a Statistical Disclosure Control Method

### (c) Controlled Rounding

### Example C

Consider the following table given in response to an information request submitted by a member of the public which shows the number of NHS hospital discharges for alcohol poisoning aged under 16 at four hospitals:

| **Hospital** | **2007- 08** |
|---|---:|
| Hospital A | 1 |
| Hospital B | 7 |
| Hospital C | 9 |
| Hospital D | 976 |
| **Total** | **993** |

Note This table shows artificial data for the purposes of this guidance only.

As this topic was considered 'sensitive', and there are values of less than 10 for individual hospitals, then it might be considered that disclosure control would need to be applied before the data could be released. The following table shows how the data would be presented if suppression were used:

| **Hospital** | **2007- 08** |
|---|---:|
| Hospital A | * |
| Hospital B | * |
| Hospital C | * |
| Hospital D | 976 |
| **Total** | **993** |

The data that can be made available is fairly limited as all numbers less than ten need to be suppressed. Table redesign is not a likely option as the customer only needs data for one year and it is the breakdown of each hospital that particularly interests them. In this type of circumstance controlled rounding could be considered as a possibility where, if applied, the following table could be produced:

| **Hospital** | **2007- 08** |
|---|---:|
| Hospital A | 0 |
| Hospital B | 10 |
| Hospital C | 10 |
| Hospital D | 975 |
| **Total** | **995** |

*Rounding using base 5*

The benefit of supplying this information to the customer is that a figure close to that of the real number can be shown which may prove to be of more use to the data requestor rather than if the numbers were suppressed. A disadvantage, however, is exact numbers may be required by the customer and so clarification of this would be very important before releasing the data in this format. This highlights the need for communication with the customer to ensure that data provided by NSS meets the user's needs as closely as possible.

# Annex H Revisions to Protocol in Version 3.0

A number of revisions and additions have been made to the Statistical Disclosure Control Protocol with the release of version 3.0. The main changes are listed below:

a. Changed from ISD to NSS version

b. Updated disclosure flowchart

c. Minor updates to disclosure form

d. Updated content for mortality data (section 4.6)

e. Updated content for information released for research purposes (section 9)

f. Added content for survey / sampled data (section 10)

g. Updated content for micro-data (section 12)

h. Added content for intermediate outputs (Annex A, section 1.2)

i. Changes to geography and populations content

j. Updated contact information:

   i. For general disclosure queries, contact the NSS Statistical Governance Team on NSS.nssstatsgov@nhs.net

   ii. For disclosure queries relating to geography and populations, contact the PHI Geography, Population and Deprivation Team on NSS.isdGPD@nhs.net

# 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/portal/page?_pageid=514,1056303,514_1056314&_dad=portal&_schema=PORTAL

Government Statistical Service, GSS / GSR Disclosure Control Policy for Tables Produced from Administrative Data Sources
http://www.ons.gov.uk/about-statistics/best-practice/disclosure-control-policy-for-tables/disclosure-control-policy-for-tables-produced-in-admin-data-sources.pdf

ONS (2008), revised from 2005, Disclosure Review for Health Statistics Guidance for Abortion Statistics (draft).

ONS (2006) Review of the Dissemination of Health Statistics : Confidentiality Guidance;
http://www.statistics.gov.uk/about/data/disclosure/downloads/confidentiality-guidance.pdf

ONS (2006) Review of the Dissemination of Health Statistics : Confidentiality Guidance; Working Paper 2 : Risk Assessment;
http://www.ons.gov.uk/ons/guide-method/best-practice/disclosure-control-of-health-statistics/index.html

ONS (2006) Review of the Dissemination of Health Statistics : Confidentiality Guidance; Working Paper 3 : Risk Management;
http://www.ons.gov.uk/ons/guide-method/best-practice/disclosure-control-of-health-statistics/index.html

ONS (2005) Disclosure Review for Health Statistics, 1st Report – Guidance for Abortion Statistics
http://www.statistics.gov.uk/downloads/theme_health/abortion_stag_final.pdf

Scottish Government (2009) Scottish Government Practical Guide to Statistical Disclosure Control.

# Index