Disclosure Control Procedure

Published January 2017

Information and technology
for better health and care

Document Management

Details

Document filename:	**Disclosure Control Procedure**		
Directorate / Programme	**Information and Analytics**	Project	**<insert>**
Document Reference	**<insert>**		
Project Manager	**Chris Roebuck**	Status	**Approved**
Owner	**Chris Roebuck**	Version	0.1
Author	**Adam Little**	Version issue date	**18/01/2017**

Revision History

Version	Date	Summary of Changes
1.0	1 November 2013	Approved
1.1	17 January 2014	Revised to reflect change to directorate
1.2	18 January 2017	Revised to reflect changes to directorate, organisation and brand. Updated Terms of Reference for the Disclosure Control Panel. Included greater detail about the NHS Anonymisation Standard.

Reviewers

This document must be reviewed by the following people:

Reviewer name	Title / Responsibility	Date	Version
Andy Sutherland	Head of Profession for Statistics	1 November 2013	1.0
Andy Sutherland	Head of Profession for Statistics	17 January 2014	1.1
Chris Roebuck	Head of Profession for Statistics	18 January 2017	1.2
Julie Stroud	Chair, Disclosure Control Panel	18 Janaury 2017	1.2

Approved by

This document must be approved by the following people: author to indicate approvers

Name	Signature	Title	Date	Version
Chris Roebuck				
Julie Stroud				

Glossary of Terms

Term / Abbreviation	What it stands for

Document Control:

The controlled copy of this document is maintained in the NHS Digital corporate network. Any copies of this document held outside of that area, in whatever format (e.g. paper, email attachment), are considered to have passed out of control and should be checked for currency and validity.

Contents

1 Introduction 5

1.1 Purpose of Document 5

2 Introduction 5

2.1 Protecting and processing personal information 5
2.2 Anonymisation 5
2.3 The assessment of risk in statistical outputs 6
2.4 The requirements of the Code of Practice 6
2.5 Guidance from the National Statistician 7
2.6 NHS Digital assessment of risk 7
2.7 Risk assessment process 8

3 The Disclosure Control Procedure 9

3.1 Meeting users' needs while protecting confidentiality 9
3.2 Ensuring access to non-disclosive statistics 9
3.3 The procedure 10

4 Small Numbers Threshold and Disclosure Control Panel 13

4.1 Small Numbers Threshold 13
4.2 The Disclosure Control Panel 13

5 Risk Assessment 14

5.1 Documentation requirements 14
5.2 Template 14
5.3 Risk assessment process 15

6 Disclosure Control Panel Proposal Template 16

Purpose of this paper 16

Background 16
Statistical report format 16
Issues and proposals 16

7 Risk Assessment 17

1 Introduction

1.1 Purpose of Document

To describe the process to be used in NHS Digital to manage the risk of disclosure of personal information from outputs with published data that are available to all. This includes outputs published under the Statistics and Registration Service Act[1] as Official Statistics[2] (whether planned or adhoc), other scheduled and adhoc publications, Freedom of Information requests[3], and data sharing activities not covered by specific confidentiality agreements.

2 Introduction

2.1 Protecting and processing personal information

According to the Data Protection Act 1998[4], personal data can be defined as data which relates to a living individual who can be identified from those data, or from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller. Protecting personal information is a legal requirement under this Act.

There are a range of complex legal and professional obligations that limit, prohibit or set conditions in respect of the management, use and disclosure of information, and similarly a range of statues that permit or require information to be used or disclosed; these provisions, which relate to NHS and social care information, are outlined in the document 'NHS Information Governance – Guidance on Legal and Professional Obligations'[5].

The required standards of practice concerning confidentiality and patients' consent to use their data are set out in the 'Confidentiality: NHS Code of Practice'[6] document.

2.2 Anonymisation

"Anonymisation" is the term given to techniques which convert personal data into anonymised data - data that, provided the risks are assessed properly, does not itself identify any individual and that is unlikely to allow any individual to be identified through its combination with other data. There is a distinction between anonymisation techniques used to produce aggregate information, and those used to produce individual-level anonymised data – this technique is referred to as "pseudonymisation".

---

[1] Statistics and Registration Service Act: http://www.legislation.gov.uk/ukpga/2007/18/contents

[2] Types of Official Statistics: https://www.statisticsauthority.gov.uk/national-statistician/types-of-official-statistics/

[3] Freedom of Information Act requests can be refused by a statistical producer if the information to be released would breach the Data Protection Act. For further information see: http://www.legislation.gov.uk/ukpga/2000/36/contents

[4] Data Protection Act 1998: http://www.legislation.gov.uk/ukpga/1998/29/contents

[5] NHS Information Governance – Guidance on Legal and Professional Obligations: http://systems.NHS Digital.gov.uk/infogov/codes/lglobligat.pdf

[6] NHS Code of practice on confidential information: http://systems.NHS Digital.gov.uk/infogov/codes

The “Anonymisation: Managing Data Protection Risk Code of Practice[7]” states that the effective anonymisation of personal data is possible, desirable and can help society to make rich data resources available whilst protecting individuals’ privacy. The Code helps to identify issues which producers of statistics need to consider in order to ensure that anonymisation of personal data is effective.

The Code acknowledges that determining what other information is ‘out there’, who it is available to, and whether it is likely to be used in a re-identification process can clearly be extremely problematic and that the risk will increase as data linkage techniques and computing power develop (as well as the assessing the certainty of availability of data in future). Whilst persons carrying out unlawful re-identification methods are accountable to the Information Commissioner, the onus is on producers of statistical information to carry out as thorough risk assessment as possible, to release only the anonymised data necessary for a particular purpose, and to guard against the risk of re-identification by applying appropriate disclosure controls.

2.3 The assessment of risk in statistical outputs

NHS Digital must ensure that all information published – i.e. made generally available – by the organisation avoids the risk of disclosing personal information[i]. In particular, NHS Digital must ensure that outputs which may be used to deduce personal information have been properly risk assessed, and the risk of such disclosure controlled; this applies for all outputs including those already anonymised or aggregated. Great care must be taken to ensure that information released which is not inherently disclosive cannot be used in combination with other information to work out personal information.

2.4 The requirements of the Code of Practice

NHS Digital follows the Code of Practice for Official Statistics[8], which requires that:

- 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 (Principle 5 of the Code)

and requires statistical producers to ensure that:

- Official Statistics do not reveal the identity of an individual or organisation, or any private information relating to them, taking into account other relevant sources of information (Principle 5, Practice 1)
- 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
- details of such arrangements are published (Principle 5, Practice 4).

[7] Anonymisation: Managing Data Protection Risk Code of Practice: https://ico.org.uk/media/1061/anonymisation-code.pdf

[8] https://www.statisticsauthority.gov.uk/monitoring-and-assessment/code-of-practice/

2.5 Guidance from the National Statistician

The National Statistician has also issued guidance[9] on how the requirements of The Code of Practice can be met. Relevant extracts are:

What, then, is 'private information'? This principle [Principle 5] applies to information that:

- Relates to an identifiable legal or natural person
- Is not in the public domain, or common knowledge
- If disclosed would cause them damage, harm or distress.

In particular, producers of official statistics should be aware of the expectation individuals may have when their information is used to produce statistics. Information relating to an individual should be considered by a producer of statistics to be 'private' if it was:

- Provided with the expectation that the information would be kept out of the public domain.

When does information identify an individual?

The opinion of the Working Party [European Article 29 Working Party Opinion on the concept of personal data[10]] was that account should be taken of the means likely reasonably to be used to identify an individual. Thus the hypothetical but remote possibility of identification is not something that automatically makes a statistic disclosive. The design and selection of intruder scenarios should be informed by the means likely reasonably to be used to identify an individual in the statistic, and therefore these will vary according to the topic of the statistic, its uses, and other factors.

- Principle 5, Practice 1 requires producers of official statistics to take account of other sources of information when considering disclosure risk. These sources may be public or private but the relevance of them is determined by whether they are likely reasonably to be used to identify an individual and reveal information about them. Guidance on determining the relevance of another source of information is included in the advice issued to members of the Government Statistical Service (GSS).
- Principle 5, Practice 4 suggests that the design of a statistic should achieve the obligation to protect against disclosure but then should be optimised to include as much detail in the statistic as is reasonably possible. This is a change from the former Code (National Statistics Code of Practice), which required that the disclosure control settings should be as extensive as possible whilst still meeting specific user needs.

2.6 NHS Digital assessment of risk

NHS Digital assesses the risk that its statistics might be used to identify an individual about whom information might be learned and which could harm that individual as a result.

If NHS Digital statistics are presented in such a way that the requirements of 2.3 above are met and the guidance in 2.4 is taken into consideration, no further action need be considered as this meets the requirements of the Code.

---

[9] https://www.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-confidentiality-of-official-statistics_tcm97-27560.pdf - Link to 'National Statistician's Guidance on the Code of Practice

[10] http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf

However, if this is not the case, NHS Digital will review the legal requirements. For example:

- Is there permission / consent to publish identifiable information?
- Would a public interest test show that the risk of distress or harm caused to a patient by publishing identifiable information was negligible?

2.7 Risk assessment process

In 2006 the Office for National Statistics (ONS) conducted a review into the dissemination of health statistics to ensure the principles of the then National Statistics Code of Practice were being upheld. The guidance[11] produced from this review is intended for anyone in the health community involved in the publication of health statistics. This procedure sets out how NHS Digital:

- Implements the ONS guidance and the Anonymisation Standard for Publishing Health and Social Care Data
- Ensures each output is risk assessed to risk the likelihood of disclosing information that could identify an individual
- Make provisions for a Disclosure Control Panel which is a group of experts available to discuss potentially disclosive situations and to approve small number thresholds (see below) for each NHS Digital data source.

[11] http://www.ons.gov.uk/ons/guide-method/best-practice/disclosure-control-of-health-statistics/index.html

3 The Disclosure Control Procedure

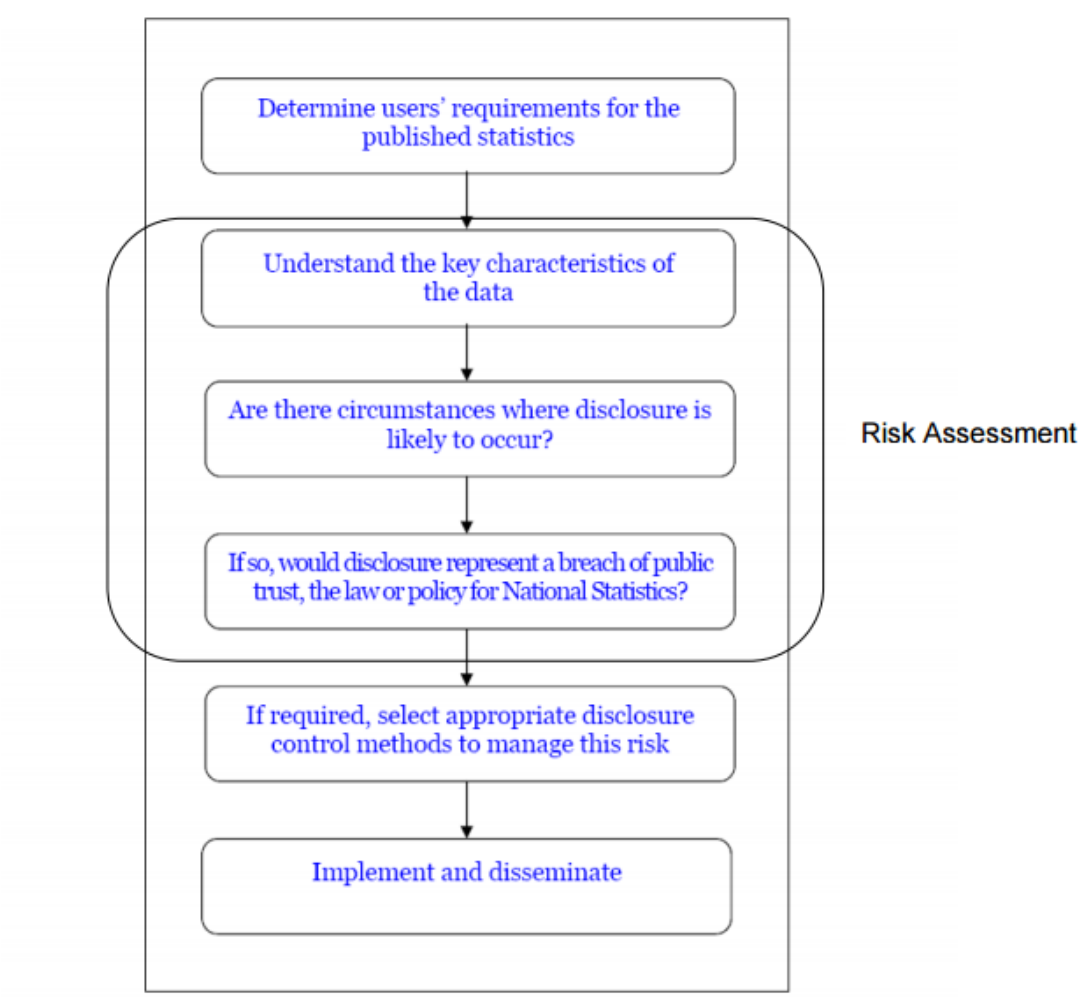
3.1 Meeting users' needs while protecting confidentiality

Health statistics provide an important public benefit and their value is often enhanced by greater granularity e.g. by identifying high risk population groups and localities. However, when statistics are released at a detailed level the risk of disclosing information about individuals is likely to be increased, and this is particularly true of tables which include small numbers (counts). It is therefore critical that we demonstrate to data subjects that our statistical information is necessary, relevant and trustworthy with a clear benefit to themselves and others (personally and as citizens) but that they do not risk being identified.

3.2 Ensuring access to non-disclosive statistics

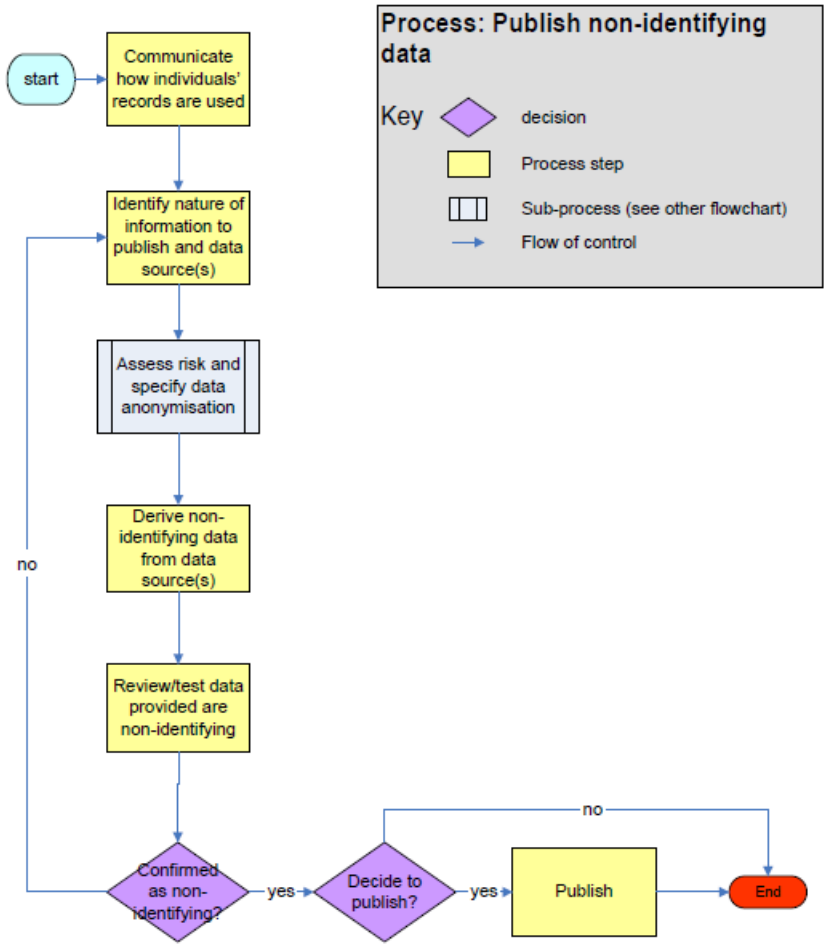
3.2.1 ONS Process

The figure below, taken from the ONS review, shows the main steps to be taken in considering disclosure control in relation to tables of health data.



3.2.2 NHS Anonymisation Standard

The diagram below shows the process for publishing non-identifying data from the NHS Anonymisation Standard[12], which goes into more detail and describes the process of assessing risk of disclosure, and of establishing a small numbers threshold (the Standard uses the term “k-anonymity”). These processes form the basis for the Disclosure Control Procedure.



3.3 The procedure

This section describes the procedure, using headings from the ONS figure, supplemented by detail taken from the NHS Anonymisation Procedure (process steps in italicised text below).

3.3.1 Determine users’ requirements for the published statistics

The first step involves establishing the user requirement for the particular health statistics and the level of detail they require. This is to ensure that the design of the output is relevant and the amount of disclosure protection used has the least possible adverse impact on the usefulness of the statistics.

It is important that the range of statistical users is taken into account, and consideration is given to future needs. In the next steps we should endeavour to ensure that publishing a statistical tabulation now will not undermine the utility of another produced in future due to

[12] NHS Anonymisation Standard:

the small numbers threshold being too high and enabling the tables to be cross-referenced to reveal identifying information.

3.3.2 Understand the key characteristics of the data

This second step involves gaining an understanding of the data that will underpin the statistics (*Communicate how individuals' records are used; Identify nature of information to publish and data source(s)).* The characteristics of the data will affect any disclosure risks. In particular, risk increases as statistics become more detailed (in terms of geography, number of fields (categories), and whether they are based on aggregate or individual level records) and also as the dimensions of the statistical table grow. Risks are higher if the distribution of the counts is skewed or the data are considered sensitive.

To check understanding of the data source, the statistical production team needs to ask questions such as "are there cells in this table that could identify an individual?"

3.3.3 Are there circumstances where disclosure is likely to occur?

This step involves assessing the threat level associated with the data and its release, and then the risk of extra information being used to try to reveal identity.

The threat level associated with the release of health and social care data is usually considered to be normal, but is high where the data relate to a sensitive subject e.g. abortion, FGM.

The risk of extra information being used to try to reveal identity is normal, unless:

- The threat level above has been identified as high.
- there is extra motivation for users to try to identify individuals, e.g. where individuals in the public eye are amongst the data subjects, or where the data relate to individuals in protected accommodation;
- other data could be used in combination with your data source to identify an individual. This includes public records e.g. the electoral register and social media sites; and other published especially relevant information;
- there is a skewed distribution of prevalence in a population e.g. sickle cell anaemia by ethnic group.

3.3.4 If so, would disclosure represent a breach of public trust, the law, or policy for National Statistics?

Where a risk is identified above, producers then need to quantify the impact of such a disclosure and determine whether it would constitute a breach of:

- public trust
- a legal obligation
- a national or international policy standard for Official Statistics.

Small Numbers Threshold

Once the characteristics of each data source and tabulation have been understood and the threat level, risk of extra information being used to try to reveal identity, and impact of disclosure have been considered, a small numbers threshold should be set for that source.

This will identify which cross tabulations from that source should be treated with caution and the level at which numbers in cells could become disclosive.

3.3.5 If required, select appropriate disclosure control methods to manage the risk

After this consideration an appropriate disclosure control method must be selected to manage the risk effectively (*Assess risk and specify data anonymization*). Where this risk is high, relatively little intrinsically identifiable information may be released and a combination of methods such as aggregation, suppression and statistical rounding may be required. Where the risk of extra information being used to try to reveal identity is normal, it may be possible to release more information.

The Anonymisation Standard specifies broad disclosure control plans according to the assessed risk of extra information being used to try to reveal identity, and the population size each cell relates to. After selecting a plan, an appropriate disclosure control technique must be chosen. The various techniques of disclosure control have different advantages and disadvantages and must be chosen bearing in mind the users, uses and characteristics of the data.

Particular care should be taken to ensure that:

- Small number values cannot be deduced by subtracting from totals and/or subtotals
- Information about an individual cannot be inferred from new information about a group. Note that there is a difference between information between information that individuals have because that person is known well to them, and information that can be deduced by a motivated ‘intruder’ who does not know the subject. However consideration must be used in both cases to prevent someone being able to learn something new about the subject (e.g. an individual may not wish their family to learn certain medical information).

3.3.6 Implement and disseminate

The final stage in the process is implementation of the disclosure control method(s) selected above and dissemination (publication) of the statistics (*Derive non-identifying data from data source(s); Review/test data provided are non-identifying; Confirm as non-identifying; Publish*). User must be taken into account to ensure that the tabulation(s) is still useful, and will not prevent the release of important information in the future

4 Small Numbers Threshold and Disclosure Control Panel

4.1 Small Numbers Threshold

As mentioned above, a small numbers threshold should be established for each statistical release series. This will be discussed and agreed with the Disclosure Control Panel. This threshold will identify the levels at which cells in tabular output may become unsafe and which, if any, variables within the data source are highly sensitive. Once established, the threshold will be used to determine what can be published. If all cells in a table are above the threshold then the table can be published, subject to checking for the requirement for secondary suppression (e.g. where there is a general risk that small numbers can be deduced by subtraction of unsuppressed numbers from totals). When cells are below the threshold they must be suppressed.

There is a threshold where, without further work and having checked for the secondary suppression requirement, the risk of determining further information is insignificant. The threshold will generally be larger than ideal, meaning that some useful and non-disclosive information could be suppressed.

4.2 The Disclosure Control Panel

The Disclosure Control Panel, which will agree the threshold for each data source, will consist of subject matter and area representatives. Full details are included in the Terms of Reference for the Disclosure Control Panel[13].

Requests for access to information below the threshold (including the creation of additional publication products) will need to be approved by the Disclosure Control Panel on a case by case basis. This means that too large a threshold will require an excess of ad-hoc work.

A template is provided in Section 6 on which to submit casework to the Disclosure Control Panel. A separate version of this with worked examples is available on the NHS Digital intranet.

Casework must be sponsored by the Responsible Statistician for the publication, although anyone from the team may attend the Panel session to present the case

[13] Disclosure Control Panel casework template (available to NHS Digital only): https://hscic365.sharepoint.com/InformationandAnalytics/Pages/Statistical%20Services/Disclosure-Control.aspx

5 Risk Assessment

5.1 Documentation requirements

A written risk assessment covering each release series must be completed by the Responsible Statistician, and this must be signed off by the NHS Digital Head of Profession for Statistics. It must be reviewed and updated at least on an annual basis, and whenever significant changes are made.

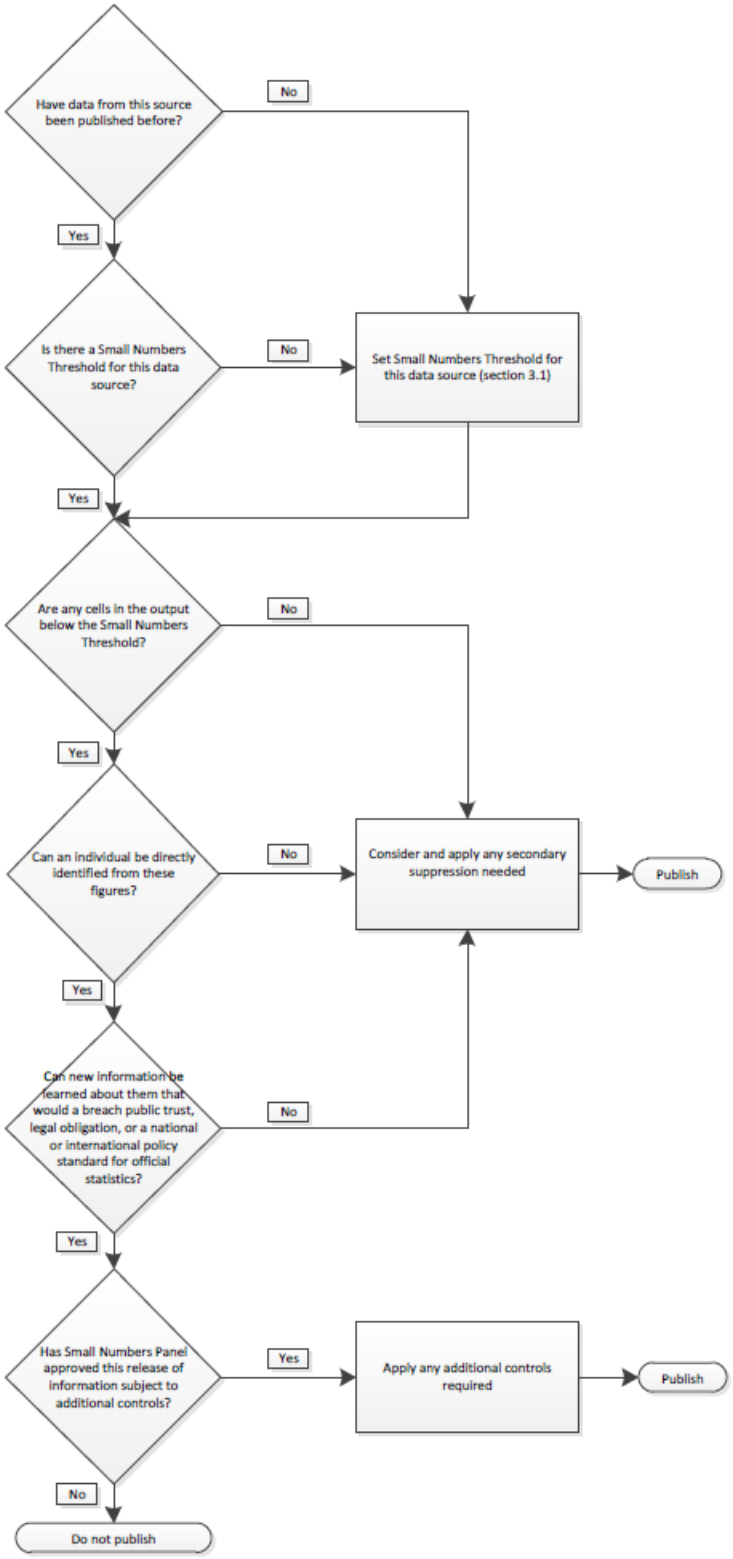
The risk assessment provides evidence that a thorough risk assessment of the material to be published has been carried out in accordance with the published legislation, standard, and guidance; that suitable disclosure control methods have been implemented and that any decisions made by the Disclosure Control Panel have been acted upon.

5.2 Template

The template in section 7 below should be considered for each release, completed, and sent to the Statistical Governance Manager via the Statistical Governance Mailbox. This will be reviewed by the Manager and, if satisfactory, sent to the Head of Profession for assessment and signoff.

5.3 Risk assessment process

The flowchart below should be followed when completing an assessment.



6 Disclosure Control Panel Proposal Template

Title of publication:

Proposer name:

Proposal:

For DCP discussion on XX Month 201X

Purpose of this paper

Include summary of advice being requested, and why (is it a new publication; has there been a change in methodology and/or dissemination methods; has a new disclosure risk been identified etc.)

Background

Include background to data source and collection method, analytical methodology, and any other relevant information.

Statistical report format

Include a description of the reporting format and granularity.

Level of risk

With reference to the NHS Anonymisation Standard, identify areas where a disclosure risk is present and describe. Considerations include self-identification and identification by others, level of motivation by users to attempt to identify individuals (e.g. sensitive data, data where one or more subjects might be persons of public interest), other data sets which could be used in conjunction with yours to facilitate identification.

Issues and proposals

Use the section below to discuss each issue (add more if required) and the Responsible Statistician's proposal/options for risk mitigation for discussion and approval by the Panel. Include sample data and worked examples wherever possible, and any supporting information including advantages/disadvantages (/repercussions) of using each methodology. Use this section also to include any specific questions to be answered by the Panel.

Issue #1

Issue #2

Issue #3

Issue #4

7 Risk Assessment

Name of Output:	**Publication title**
Data source(s):	Name(s) of the source(s) of the data used
Output type:	e.g. csv, spreadsheet, pdf
Version:	1.0
Date:	Publication date
Author:	NHS Digital
Branch:	Team name
Division:	Division name
Responsible statistician:	RS name
E-mail:	enquiries@nhsdigital.nhs.uk
Telephone:	0300 303 5678

The flowchart below should be followed when completing a risk assessment.

Note: When proposing a release of data from an NHS Digital information asset, you should inform the relevant Information Asset Owner and Administrator, notify them when a proposal is to be brought before the Disclosure Control Panel and send them a copy of the data disclosure risk assessment when approved.

Guidelines are shown in the boxes below.

Delete these boxes when the template is completed.

1. Background to the data source

-
-
-

Short background paragraph

- Introduce the data source(s) used, e.g. administrative datasets or survey data.
- Mention any characteristics of the data that are relevant to a risk of disclosure e.g.
  - data sources are sensitive
  - data has been linked (or could be) to other sources potentially increasing risk
  - other publicly available datasets (e.g. with different design / aggregation) could increase risk

2. Legal issues (collection and dissemination)

Refer to any relevant statutory arrangements relating to the data (tick as applicable)

☐ This publication aims to comply with the Code of Practice for Official Statistics and follows the National Statistician's Guidance "Confidentiality for Official Statistics."

The following have also been referenced / considered:

☐ • Anonymisation Standard for Publishing Health and Social Care Data

• Data Protection Act, Principle One regarding Fair Processing and how personal data are being processed and used (http://www.hscic.gov.uk/privacy)

• Other regulations regarding disclosive or sensitive data to which this release will adhere, e.g. HES Analysis Guide (please specify)

_____

_____

_____

_____

3. Key characteristics of the output

-
-
-

- Identify user needs, which are key to the design and structure of the output and to understand the necessary balance between confidentiality and the utility of the release.
- Comment upon, for example:
  - sensitive variables
  - age of data
  - data quality
  - area coverage
  - population base
  - linked tables
  - structure of output e.g. level of geographical breakdown and sub-categories
  - how figures will be expressed e.g. aggregate data tables, figures and charts
  - identify units of measure such as individual, episode, organisation, procedure / activity, number of bed days, attendances etc. and ensure there is no identifying data
  - how results by age, gender, ethnicity will be handled …
    (for additional guidance here see "Review of the Dissemination of Health Statistics: Confidentiality Guidance" section 5).
- List output(s) e.g. executive summary, background quality statement, specification document, report, metadata, list of indicators

4. Evidence of risk in the output

-
-
-

Explain the issues considered and evidence of resulting risk

- Describe the nature of the information and the likelihood of its being disclosive. Assess the threat level – is it feasible within reasonable levels of expertise, time and resourcing that disclosure could occur?
  - State the nature of possible disclosure, e.g.
    - ♦ attribute – individual or group
    - ♦ identification / self-identification
    - ♦ residual disclosure by differencing, linking or deconstructing rates and percentages released elsewhere?
    - ♦ are other tables based on the source data already available?
  - could other datasets be linked to the output and allow disclosure?
  - motivated intruder – able to combine with extra information that is likely to be available etc.
  - what are the populations, geographies and organisations involved; are there small populations at risk?
  - is it a sensitive topic area?
  - are there potentially unsafe cells in the table? Consider, for example:
    - ♦ 0, 1, 2, <5, 100% and so on
    - ♦ statistical units that are in the table more than once?
- Assess the potential impact of disclosure: low, medium or high (e.g. in terms of distress or embarrassment)
- Identify the level of risk associated with this release: low, medium or high. (Risk is combination of likelihood and impact)
- Might additional protection be required e.g. due to :
  - skewed distribution of prevalence in population
  - special knowledge available about individuals in the dataset
  - other relevant information available
- Consider potential disclosure scenarios and their impact

(See sections 6 and 7 "Review of the Dissemination of Health Statistics: Confidentiality Guidance" for more detail).

Include:

- Any disclosure arrangements agreed with the data provider e.g. if data is provided by another Government Department
- Views of colleagues and peers

5. Disclosure Control Panel

Tick as applicable (one or more).

☐ The disclosure control proposal was approved by the Disclosure Control Panel on DD/MM/YYYY

☐ This is an update to an extant risk assessment. Review by the Disclosure Control Panel was not necessary.

☐ Review of the risk of disclosure by the Disclosure Control Panel was not required.

6. Proposals for mitigating risk in publishing output (see Section 8 of ONS guidance) including responses to ad-hoc requests

-
-
-

- List options (based upon risk level) e.g.
  - Table design e.g.
    - ♦ Group / aggregate
    - ♦ Reduce the level of detail
    - ♦ Exclude variables
    - ♦ Ensure each table is internally consistent - row and column totals do not allow disclosure by differencing
    - ♦ Totals are consistent across tables describing the same population or subset thereof
    - ♦ Aggregate totals at higher levels are the sum of cells at lower levels
  - Cell suppression – primary, secondary and marginal totals
    NB rows and columns dominated by 0s
  - Rounding
  - Data perturbation e.g. Barnardisation or targeted record swapping
  - K-anonymity for record level data
- **Confirm approach to disclosure control of data released in response to ad-hoc requests**
- Also consider:
  - ensure a consistent and practical approach that can be implemented within reasonable time limits and using appropriate level of resource.
  - Trade-off between risk and utility, i.e. balance the loss of information in the output against likelihood of information being disclosed
  - to promote openness and transparency the disclosure methodology should make sense to the user

7. Ad-hoc queries

Tick only one of the boxes below:

☐ I confirm that the approach to the disclosure control of ad-hoc query responses based on this publication follows that outlined in this document. Any responses not following this approach will have any alternative disclosure control approach approved by the Disclosure Control Panel.

or

☐ I confirm that all ad-hoc queries are responded to only after consultation with the Disclosure Control Panel and application of the agreed disclosure control method.

8. Review process

Tick below.

☐ I understand that this risk assessment must be reviewed at the agreed interval (e.g. monthly, quarterly), or at least on an annual basis. I understand that a new risk assessment is required whenever the data collection changes, following a methodological change or publication redesign.

Review frequency

☐ A risk assessment is produced for each release of this publication.

☐ This is an annual risk assessment unless more frequent review is required (see above).

[i] Personal data: The Data Protection Act defines personal data as "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.

See https://ico.org.uk/for-organisations/guide-to-data-protection/key-definitions/ for further detail.