NHS
Business Services Authority

Statistics at NHSBSA
Statistical Disclosure Control Protocol

Document Release Note

Document Name: NHSBSA Statistical Disclosure Control Protocol

Document Details Name	Version Number	Description
NHSBSA Statistical Disclosure Control Protocol	V1.1	Supporting Strategy Document

Revision Details Revision Number	Revision Date	Revision Description	Page Number	Previous Page Number	Action Taken	Addenda/ New Page
1.1	28 January 2020	Corrected example A Spelling	all 10-11 6	all 10-11 6	Updated styles to match other stat policies Corrected example A Replace “deanonymising” with “could be used to re-identify anonymous data”	
1.2	10 August 2020	Revised thresholds from 10 to 5 in line with confidentiality standard / PQ/FOI	10 12 1, 17		Added ref to Open Data Portal 9 changed to 4 Examples adjusted to less than 5 Removed rounding from some examples Added flexibility around cell suppression Refers to “SDC” broadly now, not calling out Rounding for consistnecy	

About this document

This document is outlines our initial approach in this area, we are continuing to consult with Office for National Statistics’ (ONS) methodologists on our Statistical

Disclosure Control approach, as well as our colleagues in NHS Digital, and the Information Services Division (ISD) of National Services Scotland, a part of NHS Scotland. The approach outlined in this document is therefore subject to change. Further amendments we make will be preannounced with this document being revised and updated accordingly. Aside from ad hoc updates this document will be reviewed on an annual basis.

Information in this document has been organised as follows:

Chapter:

Table of contents

1. Purpose

This document sets out the NHS Business Services Authority (NHSBSA) policy on the application of methods of disclosure control to statistics. The NHSBSA Anonymisation and Pseudonymisation standard, referred to in our Data Protection and Confidentiality Policy, as well as our Data Governance Policy, will be updated to reflect this protocol shortly. This protocol describes the areas of risk that should be considered when data is released; whether into the public domain through statistical publications, Parliamentary Questions (PQs), requests under the Freedom of Information Act (FOIs), or requests for data not available through Freedom of Information. Disclosures relating to Parliamentary Questions and Freedom of Information Act (FOI) will follow this policy with the exception that thresholds for suppression and rounding will be determined on a case by case risk basis.

We will apply statistical disclosure control where there is a reasonable risk of sensitive personal information being identified by a motivated intruder who could use such information to cause damage, harm, embarrassment, anxiety or distress to an individual(s) or organisation(s).

How we approach the risk of disclosure differs depending on the degree of control the NHSBSA can exert on the use of the data once released. Data shared within the NHSBSA itself, which is made available via ePACT2, eDEN, and provided to our partners within the NHS and across Government working in the planning and delivery of patient care e.g. the Department of Health and Social Care, does not require statistical disclosure control to be applied, however confidentiality rules will always be followed. The person providing any such data must always highlight any aspect that risks disclosure of personal information and, if external release is later planned, advise on the application of this statistical disclosure control protocol.

1.1. Definitions

For the purpose of this document, the following definitions are used:

Barnardisation: A method of disclosure control for tables of counts that involves randomly adding or subtracting 1 from some cells in the table.

Disclosure is used to describe the communication of personally identifiable information (PII) about an individual, where information is made public through a release such as a statistical output.

Disclosure control is the process of reducing the risk of disclosure. It aims to ensure an appropriate balance of data usability for our customers and the management of data confidentiality risks.

Cell suppression is the process where cells that contain disclosive information are suppressed and replaced by a special character, for example an asterisk (“*”), to indicate a suppressed value.

Cell swapping is the process of creating pairs of records with similar attributes and swapping records that are partially matched on a set of key variables but differ in other respects, for example, they may describe in different geographical locations.

Inferential disclosure is where disclosure occurs if an intruder is able to determine the value of some characteristic of an individual or organization more accurately with the released data than otherwise would have been possible.

Microdata is data on the characteristics of units of a population, such as individuals, households, or establishments, collected by a transactional activity, survey, or experiment.

Motivated Intruder(s) are person(s) (an individual or an organisation) who wishes to identify the individual from data released.

Noise addition is the process on adding or multiplying a randomised number to the

original values to protect data.

Residual disclosure (or differencing) can occur where outputs from the same or different sources can be combined to reveal information about an individual or a group. This can happen inadvertently 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.

Rounding is the process of replacing a number with the nearest number that is a whole multiple of a specified value, for example the nearest ten, hundred, thousand, etc.

Personally identifiable information (PII) is any data that could potentially identify a specific individual. Any information that can be used to distinguish one person from another, and could be used to re-identify anonymous data should be considered PII

2. Background

Trustworthy statistics and the data behind them are an important part of well informed decision making, and are vital to support improvement across the wider health and social care system. It is accepted, however that where statistics provide information on small numbers of individuals, the NHS Business Services Authority have a duty, under data protection law, to avoid directly or indirectly revealing any personal details. In addition, NHSBSA staff members are required to adhere to relevant NHS 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 policy has been developed in accordance with the UK Statistics Authority's (UKSA) Code of Practice for Statistics 2.0, specifically principle T6, which describes effective data governance. This principle includes control over data access and protection of confidential information; it states:

T6 Data governance

Organisations should look after people's information securely and manage data in ways that are consistent with relevant legislation and serve the public good.

T6.1: All statutory obligations governing the collection of data, confidentiality, data sharing, data linking and release should be followed. Relevant nationally and internationally endorsed guidelines should be considered as appropriate. Transparent data management arrangements should be established and relevant data ethics standards met.

T6.2: The rights of data subjects must be considered and managed at all times, in ways that are consistent with data protection legislation. When collecting data for statistical purposes, those providing their information should be informed in a clear and open way about how that information will be used and protected.

T6.3: Organisations, and those acting on their behalf, should apply best practice in the management of data and data services, including collection, storage, transmission,

access, and analysis. Personal information should be kept safe and secure, applying relevant security standards and keeping pace with changing circumstances such as advances in technology.

T6.4: Organisations should be transparent and accountable about the procedures used to protect personal data when preparing the statistics and data including the choices made in balancing competing interests. Appropriate disclosure control methods should be applied before releasing statistics and data. Appropriate protocols should be applied to approved researchers accessing statistical microdata.

T6.5: Regular reviews should be conducted across the organisation, to ensure that data management and sharing arrangements are appropriately robust.

3. Scope and Implementation

This protocol will apply to all new NHSBSA statistical publications, management information releases, Parliamentary Questions (PQs), Freedom of Information requests (FOIs) and release of data not available through Freedom of Information from October 2019. Existing regular releases will be brought into compliance incrementally when resources allow, and in tandem with the on-going review of all our products as part of our Publication Strategy. Data released via the NHSBSA's Open Data Portal (ODP), Information Services Portal (ISP), as well the NHSBSA's public insight portal Catalyst is currently considered to be out of scope for this protocol until we fully explore the most appropriate technical solution.

While the examples in this document primarily relate to patient counts, it is acknowledged that patient counts may not be the only form of disclosive information that could be released. Some values which initially may appear not to relate to patients, such as a count of items prescribed, the total quantity prescribed, number of packs, and values relating to other measures of activity such as the delivery of a service to a person could be equivalent to releasing personal information. For example, if data is released that states a single pharmacy contractor has performed between 1 and 4 Medicine Use Reviews for sexual health, it can be assumed that these reviews were carried out on individual patients and therefore would be disclosive. Any data therefore that could reveal information relating to a number of patients by inference, inferential disclosure, that is below the disclosure threshold will be suppressed.

This protocol will also apply to calculations we produce, the most common being percentages or rates. Unless the denominator for a given percentage is high, it may be possible to work out the numerator and denominator because only one possible pair of values would give a percentage, disclosure control will therefore be applied to percentages and rates we release. This extends to simple calculations such as the mean, median or mode where a number calculated represents a number of individuals.

4. Statistical Disclosure Control Methods Used

Where the need for statistical disclosure control (SDC) is identified then the following range of methods listed will be used to balance risk of disclosure with data utility. As part of our commitment to be transparent and release information in a useable and accessible format we will always state clearly where statistical disclosure control has been applied, and make reference to this document for further information. This will be in addition to providing an explanation of how to interpret any symbols and or numbers used. Specific notes where applicable, such as where data may not appear as normally expected, for example ensuring notes accompany a table where totals may not sum due to rounding will be included.

What follows is our standard approach; this may evolve in time as we consult our users and the relevant experts in this field. We may also on occasion apply a statistical disclosure control method beyond what is outlined here always explaining our rationale where we have done so.

4.1. Table Redesign

Table redesign will always be considered as a first step to provide useful information while protecting confidentiality. This is a simple method that aims to minimise the number of potentially disclosive values released through grouping and aggregation. Ways in which table redesign can be used range from, grouping or collapsing categories within a table of data released, aggregating to a higher level geography or for a larger population sub-group, and or aggregating tables across a number of years, quarters and or months

Example A, below, illustrates the process of table redesign. The first table (i) shows information about the number of people in a Clinical Commissioning Group (CCG) who have been prescribed drugs A, B and C by age group, prior to any statistical disclosure control. Cell values of between 1 and 4 are considered to be disclosive. There are five such cells in the table, shown in shaded boxes. Table iii shows the result of statistical disclosure control being applied to this table, in order to protect the table the age groups could be combined to form 10-year intervals instead of 5-

year intervals. This can be used as an alternative to providing a table which contains suppressed "*" values, table (ii). It can be seen that changing the range of the age groups in this way has protected the disclosive data and produced a table which can safely be released into the public domain.

Example A (i) (original table produced, no suppression, unrounded)

Age Group	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	Total
Drug A	20	31	16	11	10	3	40	17	148
Drug B	4	14	4	35	26	15	21	12	131
Drug C	3	18	25	2	30	4	15	18	115

Example A (ii) (values 1-4 with suppressed "*")

Age Group	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	Total
Drug A	20	31	16	11	10	*	40	17	148
Drug B	*	14	*	35	26	15	21	12	131
Drug C	*	20	30	*	30	*	15	18	115

Example A (iii) (redesigned table, combined age groups)

Age Group	20-29	30-39	40-49	50-59	Total
Drug A	51	27	13	57	148
Drug B	18	39	41	33	131
Drug C	21	27	34	33	115

We will always be consistent in groupings within variables, and between tables produced, to avoid disclosure by differencing (residual disclosure).

4.2. Cell Suppression

A method of protecting disclosive values we will utilise where a customer requires information in a format where redesign is not possible is cell suppression. This means that unsafe values are not published, but are suppressed and replaced with a placeholder or symbol, commonly 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 also be necessary to select additional cells for secondary suppression where further granularity is requested. The use of any placeholders will always be clearly stated where applied.

We will apply suppression at all geographical levels for patient counts between 1 and 9. This includes where prescribed items can be attributed but we have been unable to identify any patients. Cell suppression will apply to all patient counts, regardless of any age or gender breakdowns, and to any other fields supplied alongside patient count, which could be used to derive a more granular patient count. We will not provide a lower granularity breakdown than a national total where such a breakdown would in itself allow the identification of supressed values.

Example B, below, illustrates the process of cell suppression. The first table (i) shows information about the number of people prescribed Drug A, B and C at a national level prior to any statistical disclosure control being applied.

Cell values of between 1 and 4 are considered to be disclosive. There are two such cells in the table, shown in boxes. In order to protect the table these values are supressed, replaced with “*”.

Example B (i) (original table produced, no suppression, unrounded)

	National
Drug A	12
Drug B	4
Drug C	3
Total	19

Example B (ii) (Cell suppression applied)

	National
Drug A	12
Drug B	*
Drug C	*
Total	19

Where further breakdowns are requested at a lower level geography than national, or for a specific category, or sub group we will only do so where the higher level it is a part of is unsuppressed, and where residual differencing is not possible to prevent disclosure of personally identifiable information, for example if a national total is 10 and this was made up of 10 sub-national elements we would not provide this as it would be possible to determine the actual unsuppressed values.

Example C, below, illustrates the process of cell suppression at a lower geographical granularity than example A. The first table (i) shows the number of people prescribed Drug A in three areas without any statistical disclosure control.

Cell values of between 1 and 4 are considered to be disclosive. There is one such cell in table C (i) below. Even after these values are supressed, that is replaced with "*" this table remains disclosive as the value for Area 2 can be calculated by subtraction. Further disclosure control measures would be required to protect this table before it could be released publicly as discussed in the following sections.Example C(iii) in the rounding section shows how using rounded values could control the risk for this example.

Example C (i) (original table produced, no suppression)

	Drug A
Area 1	11
Area 2	1
Area 3	0
Total	12

Example C (ii) (Cell suppression applied, disclosure control ineffective)

	Drug A
Area 1	11
Area 2	*
Area 3	0
Total	12

4.3. Other Methods

We may also on occasion apply a statistical disclosure control method beyond what is outlined here always explaining our rationale where we have done so. Additional methods of protecting disclosive values we may utilise include but are not limited to, barnardisation, cell swapping, and noise addition, definitions for these methods can be found in section 1.1.

5. Rounding

Rounding can be a useful technique to improve data clarity when tabulated and presented for example. In many cases, the detail provided by releasing exact values is not necessary; in fact doing so may distort the main features of the data, making it more difficult for users to draw conclusions.

Throughout all our statistical products and data we release, values, rates and percentages will be subject to disclosure control methods unless there is a specific reason that exact values are required. While data we provide to our partners across Government working in the planning and delivery of patient care e.g. the Department of Health and Social Care, does not require statistical disclosure control it may be applied depending on the specific nature of the request. We will always clearly explain this for each request where it has been applied.

As a standard, we will mainly rely on table design, aggregation and cell suppression to control the disclosure risk. However in some specific cases such as Example C (ii) above we may round the values in a table to prevent disclosure by differencing values in a table.

The impact of this is shown in table C (iii) below which represents the data from example C (i).

Example C (iii) (Cell suppression applied, rounded to the nearest 5)

	Drug A	**Drug A Interpretation**
Area 1	10	8 to 12
Area 2	*	1 to 4
Area 3	0	0
Total	10	8 to 12

In such cases users must be advised to interpret the data as a range rather than the nominal values presented, as shown in the right hand column of table C (iii)

Contact us

Feedback is important to us, we welcome any questions and comments relating to the new disclosure control method we are implementing.

Please quote 'NHSBSA Disclosure Control Protocol' in the subject title of any correspondence via the contact methods listed below.

You can contact us by:

Email: nhsbsa.statistics@nhs.net

Telephone: 0191 203 5050

You can also write to us at:
NHSBSA – Statistics
NHS Business Services Authority
Stella House
Goldcrest Way
Newburn Riverside
Newcastle upon Tyne
NE15 8NY

END.