

# Disclosure control guidance for birth and death statistics

## Briefing note (revised Jan 2014) on the publication of tabular data

### **IMPORTANT NOTICE (added March 2016)**

ONS decided in February 2016 to begin a further revision of this guidance. This document may continue to be used, subject to the following advice, until a new version is published.

The availability of an increasing range of detailed 'open' data products across government and from other sources, combined with constantly improving technologies for internet searching, data linkage and exploitation, and the ubiquity of online social networking, has brought threats as well as new opportunities for official statistics.

In particular, there is a need to reconsider the level of disclosure control which might be needed to prevent potentially unlawful access to personal information caused by the increased technical ability of an intruder to associate unique instances in a large statistical table with identifiable individuals, with the aid of other readily available data sources.

As a result, ONS recommends that a cautious approach should be taken when a very sparse table affords possible opportunities for attribute disclosure by, for example, linkage to news reports about deaths or other publicly available information. To avoid this, as a general rule, it is recommended that cell counts of less than 3 should be suppressed in tabulations at geographical levels below region (former GOR), where attribute disclosure is possible.

It should be noted that in the light of the above, the combinations of variables and geographies shown as 'green' in Annexes 1A and 1B of this document should not be relied on without further consideration of the disclosure risk in each case.



# Disclosure control guidance for birth and death statistics

**Briefing note (revised Jan 2014) on the publication of  
tabular data**

Version No.: **3.0**  
Editor: **Emma Gordon**  
**Office for National Statistics**

## A National Statistics publication

National Statistics are produced to high professional standards set out in the Code of Practice for Official Statistics. They are produced free from political influence.

## About us

### The Office for National Statistics

The Office for National Statistics (ONS) is the executive office of the UK Statistics Authority, a non-ministerial department which reports directly to Parliament. ONS is the UK government's single largest statistical producer. It compiles information about the UK's society and economy, and provides the evidence-base for policy and decision-making, the allocation of resources, and public accountability. The Director-General of ONS reports directly to the National Statistician who is the Authority's Chief Executive and the Head of the Government Statistical Service.

### The Government Statistical Service

The Government Statistical Service (GSS) is a network of professional statisticians and their staff operating both within the Office for National Statistics and across more than 30 other government departments and agencies.

## Contacts

### This publication

For information about the content of this publication, contact Emma Gordon

Tel: 01633 45 5866

Email: [Emma.Gordon@ons.gsi.gov.uk](mailto:Emma.Gordon@ons.gsi.gov.uk)

### Other customer enquiries

ONS Customer Contact Centre

Tel: 0845 601 3034

International: +44 (0)845 601 3034

Minicom: 01633 815044

Email: [info@statistics.gsi.gov.uk](mailto:info@statistics.gsi.gov.uk)

Fax: 01633 652747

Post: Room 1.101, Government Buildings,  
Cardiff Road, Newport, South Wales NP10 8XG

[www.ons.gov.uk](http://www.ons.gov.uk)

### Media enquiries

Tel: 0845 604 1858

Email: [press.office@ons.gsi.gov.uk](mailto:press.office@ons.gsi.gov.uk)

## Copyright and reproduction

© Crown copyright 2011

You may re-use this information (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence.

To view this licence, go to:

[www.nationalarchives.gov.uk/doc/open-government-licence/](http://www.nationalarchives.gov.uk/doc/open-government-licence/)

or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU

email: [psi@nationalarchives.gsi.gov.uk](mailto:psi@nationalarchives.gsi.gov.uk)

Any enquiries regarding this publication should be sent to: [info@statistics.gsi.gov.uk](mailto:info@statistics.gsi.gov.uk)

This publication is available for download at:

[www.ons.gov.uk](http://www.ons.gov.uk)

# Contents

1. Executive summary .....	2
2. Introduction and background .....	4
3. Legal framework .....	8
4. Types of data.....	11
5. Geographic and population size considerations.....	14
6. Publishing tables for individual or grouped communal establishments.....	18
7. Designing a table for a statistical publication .....	19
8. Designing a bespoke, ad hoc table .....	21
9. Answering a Freedom of Information request .....	21
10. Answering a Parliamentary Question .....	22
11. Designing an open dataset .....	24
12. Sensitive topics.....	25
13. Types of disclosure.....	25
14. Selecting methods of statistical disclosure control .....	27
15. Definitions .....	31
Annex 1A: Summary guidance on disclosure control – birth registrations .....	35
Annex 1B: Summary guidance on disclosure control – death registrations .....	36
Annex 2: Risk assessment .....	37

# 1. Executive summary

## 1.1 Introduction

This document sets out guidance for the release and publication of tables based on ONS birth and death registration data for England and Wales, and replaces the previous briefing note on the disclosure control policy for birth and death statistics. It is intended for use by ONS staff and those in other organisations involved in publishing birth and death registration data, either in the form of a statistical publication or research article, a response to a customer request, Freedom of Information request or Parliamentary Question, or if designing an open dataset.

This document does not cover access to individual record data (microdata), or the use of data derived from linked datasets.<sup>1</sup> For further information about the release of microdata see the [Approved Researcher accreditation section](#) on the ONS website.

The previous briefing note has been revised in response to a requirement in the UK Statistics Authority assessment report on ONS mortality statistics ([Assessment Report number 200](#)) to “Review the disclosure policies for the mortality statistics, and how the policies are applied in the production of the statistical releases, to ensure that the policies are not unduly limiting the practical utility of the statistics.”

Death registration information is collected under the same legislation as birth registration, so it is logical to extend the remit of the revised guidance to cover births data as well.

This revised guidance has been written specifically to support greater use of existing data (within ONS and by our customers), thereby supporting collaboration with others (public bodies, academic institutions and other customers). This guidance will be useful for our customers (and ONS staff) to:

- Understand the legal context behind why these data are collected and how they can be disseminated,
- Understand the disclosure control rules in relation to births and deaths registration data,
- Be able to apply the appropriate disclosure control rules to microdata derived from birth and death registration data.

In producing the revised guidance, ONS has carefully considered the balance between the government’s Open Data agenda and the need to protect confidentiality.

Also, in line with Statistics Suppliers and Users Group (Stat SUG) recommendations, this revised guidance is based on the concept of balancing the risk of disclosure against the utility of the statistics. So, we now require those designing outputs to consider if there is sufficient uncertainty to protect confidentiality rather than insisting on there being absolutely zero risk in outputs. This guidance is intended to help inform that decision making process.

---

<sup>1</sup> With the exception of linked birth/death registration datasets, where no information derived from NHS datasets is included.

## 1.2 What tables can be published, without the need for disclosure control

At the request of users of the previous guidance (internal and external to ONS), this version provides more explicit advice about the range of tables that can be produced using a range of different geographical boundaries, without the need for disclosure control of small numbers. Annex 1A and 1B present summaries of the disclosure guidance to illustrate when disclosure control of birth and death registrations data respectively will not/may be/will be required.

Tables based on birth registrations data presenting 'fact of live birth' or 'fact of stillbirth' by year, standard age groups (see section 4.1) and either local authority (LA), clinical commissioning group (CCG) or local health board (LHB) do not need to be disclosure controlled, although small numbers in tables presenting figures by other standard geographic boundaries (for example, parliamentary constituency, PC or middle super output area, MSOA) do need to be disclosure controlled, because of the risk of disclosure by differencing.

Tables based on death registrations data presenting figures by year, sex, standard age groups (see section 4.2), International Classification of Diseases (ICD) code and any sub-national geography do not need to be disclosure controlled, as long as the underlying resident population size is estimated to be  $\geq 5,000$ . So small numbers and zeros presented in tables by LA, CCG, LHB, PC or MSOA can be released without suppression.<sup>2</sup>

The guidance is deliberately more cautious for birth registration data, since details about stillbirths and mother's age are not on the public register.<sup>3</sup> See section 5.1 for more information.

Where small cell counts do need to be disclosure controlled, only cell counts of  $< 3$  need to be protected (0, 1 and 2), for birth (stillbirth and live birth) and death registration data. This differs from the previous guidance, which required that all counts of  $< 5$  should be protected for any data relating to stillbirths.

There is a vast range of ways to design a table using birth or death registration data. This guidance does not attempt to cover every possibility. Instead, advice is presented on how to make a judgement about whether there is sufficient uncertainty to protect confidentiality, as well as scenarios when some form of disclosure control (such as suppression of small numbers) definitely needs to be carried out.

Links to example cases are presented in sections 7 – 10, to illustrate the way in which risk of disclosure can be balanced against the utility of the information being presented. Also, a list of terms and definitions are presented in section 15, to help the user to understand and interpret the guidance.

This version of the guidance does not present a list of 'sensitive causes of death'. Therefore, deaths which may have previously been considered 'sensitive' for example suicides, HIV-related deaths or maternal deaths should be treated in exactly the same way as all other causes of death. In these cases, ONS considers the risks of disclosure or of causing harm or distress by publishing small numbers are now outweighed by the utility of these statistics for monitoring purposes.

---

<sup>2</sup> With the exception of the Isles of Scilly Unitary Authority (UA), which should be combined with Cornwall UA, or small numbers suppressed, or a full risk assessment carried out demonstrating that the risk of disclosure is small.

<sup>3</sup> Some personal information collected at birth/death registration is available (on application and for a charge) from the General Register Office (GRO), via the supply of individual certificate copies. This information is considered to be on the public register. For more information, see the [GRO website](#).

In summary, this guidance is designed to help those producing tabular data derived from birth and death registration data, to determine what can be disseminated lawfully. If there is any doubt around what tabular data can be shared or published, please contact ONS Statistical Disclosure Control branch (see section 2.2 below for contact details).

## 2. Introduction and background

### 2.1 Summary

This document sets out guidance for the release and publication of tables based on ONS birth and death registration data, for England and Wales. It is intended for use by ONS staff and those in other organisations involved in publishing birth and death statistics, either in the form of a statistical publication or research article, a response to a customer request, Freedom of Information request or Parliamentary Question, or if designing an open dataset. This document does not cover the release or publication of individual record data (microdata) or data derived from linked datasets.<sup>1</sup>

The guidance is also intended for use by researchers who have access to individual records of ONS birth and death registration data, or microdata (individual level record data), whether this is supplied directly by ONS or via another organisation (for example, via the Health & Social Care Information Centre [data linkage and extract service](#)).

The guidance will also be useful to users of published birth and death statistics for England and Wales and to those who request bespoke tables, as it will help the user understand the factors that were considered when the table was designed. Information is presented in section 4 about the types of statistical information collected as part of the birth and death registration processes, which is then available via publications and to customers on request.

ONS also releases microdata files for statistical purposes, under certain conditions. For further information about the release of microdata see the [Approved Researcher accreditation section](#) on the ONS website.

Information about the broad legal framework under which birth and death registration data are collected and held by ONS is presented, as is information about the types of disclosure control that can be applied to tables and how information may be published and/or released (see sections 3, 14 and 15 respectively).

The guidance is presented so that if users want to see examples of how to apply the guidance in relation to a specific type of output, they can go directly to the appropriate section, for example:

- Published statistical outputs,
- Customer requests,
- Responses to Freedom Of Information (FOI) requests,
- Parliamentary Questions,
- Open data.

In each case, links to example cases are presented to illustrate the way in which risk of disclosure can be balanced against the utility of the information being presented.

A list of terms and definitions are also presented, to help the user's understanding of the guidance, as it is recognised that some terms used in this guidance do not have the same meaning across all organisations.

Annex 1a presents a summary of when information collected at birth registration requires disclosure control when presented in a table and when it does not, to guide users through the decision making process.

Annex 1b presents a comparable summary for the tabular release of death registration data. These are intended to help inform decision making for those involved in designing tables.

It is recognised that although the guidance will help in the majority of cases, occasionally there will still be a requirement for a full risk assessment to be carried out in collaboration with ONS Statistical Disclosure Control experts (see section 2.2 for contact details), before a table can be released. Examples of this type of risk assessment are presented in Annex 2.

This guidance complies with the UK Statistics Authority Code of Practice for Official Statistics ([Code of Practice](#)).

## 2.2 Background

This guidance replaces the [ONS Disclosure Control Policy for Birth and Death: Briefing Note](#), which was last updated in 2010. In line with Statistics Suppliers and Users Group (Stat SUG) recommendations, this revised guidance is based on the concept of balancing the risk of disclosure against the utility of the statistics. So, we now require those designing outputs to consider if there is sufficient uncertainty to protect confidentiality rather than insisting on there being absolutely zero risk in outputs. This guidance is intended to help inform that decision making process.

There are also legal and policy obligations that must be respected (see section 3). The [Statistics and Registration Service Act 2007](#) places a legal obligation on the need to protect the confidentiality of personal information, as does the [Data Protection Act 1998](#). The National Statistics [Code of Practice](#) sets out principles for protecting confidentiality, including the principle that:

'The National Statistician will set standards for protecting confidentiality, including a guarantee that no statistics will be produced that are likely to identify an individual unless specifically agreed with them.'

Publishing detailed tables of data carries a risk that individuals or other entities (for example families or businesses) could be identified and confidential information about them could be released. Confidentiality needs to be protected for a number of reasons. We make pledges that certain information collected at birth and death registration will only be used for statistical purposes. If we do not honour our pledge there is a potential risk that people will no longer be willing to provide it – which would impact on data quality. Such trust cannot be maintained unless the privacy of individuals' information is protected.

If ONS supply individual records of birth and/or death registration data (microdata) to individuals in external organisations for statistical purposes (for example, academic organisations or other public bodies), the same confidentiality principles must be maintained by these organisations when handling the data. This is a requirement of the Statistics and Registration Service Act 2007.

The information supplied to a registrar by an informant at birth or death registration falls into three categories:

- Information required under registration legislation, for the public birth and death registers<sup>3</sup>,
- Information required under population statistics legislation, which is not recorded in the public registers,

- Information provided to the registrar on a voluntary (consent) basis, for statistical purposes only.

In certain circumstances deaths may be investigated by a coroner. In these cases, the coroner sends additional information directly to the registrar. Registration and Population Statistics legislation enables the Registrar General to compile information from across all categories and to supply this to ONS for statistical purposes.

Specimen copies of birth registration forms and details of the procedures relating to birth registration are available from the ONS website in the [births metadata](#). Comparable information relating to death registration is available in the [mortality metadata](#). It is clear from these specimen copies which data items are collected for which purpose.

A copy of the information held on the public birth and death register can be obtained from the Registrar General by searching the appropriate publically available index to the birth or death register and paying a fee to obtain a certified copy of the entry in register. For more information, see the [GRO website](#).

The primary purpose of any statistical data collection is to facilitate the use of the data collected. The government is also committed to making better use of data collected for [administrative purposes](#). ONS recognises the importance of making the maximum possible use of data without compromising confidentiality, while also maintaining a reasonable balance between disclosure risk, data utility and data access. This balance is a difficult one. The previous guidance was criticised by some users who believed ONS was exercising excessive caution on the side of risk avoidance. ONS recognises the need to reassess this balance, while still ensuring that we (and others) remain within the law and the Code of Practice. Ultimately, the identifiability of the data within a table still needs to be protected, but the resulting tables also need to be useful to the user.

This updated guidance puts more emphasis on the significant value of birth and death registration data for statistical purposes, presenting clearer guidance around the level of detail at which it is acceptable to release data, so analysts can confidently operate within the law while also producing useful statistical outputs. ONS also proposes that the approaches that have been used to revise this guidance could be extended to other data collections used across the Government Statistical Service (GSS) and the Government Social Research Service (GSR), and potentially the National Health Service (NHS).

Where questions or ambiguities remain, the ONS Statistical Disclosure Control (SDC) branch is happy to give advice to users and data providers on an ad hoc basis. Both users and data providers may also request information on training in statistical disclosure control from ONS SDC. Contact details for SDC advice are as follows:

Statistical Disclosure Control branch

Office for National Statistics

Segensworth Road

Titchfield

Fareham

Hants PO15 5RR

E-mail: [sdc.queries@ons.gsi.gov.uk](mailto:sdc.queries@ons.gsi.gov.uk)

Special extracts and tabulations of birth and death registration data for England and Wales are available to order (subject to legal frameworks, disclosure control, resources and [agreements of costs](#), where appropriate). Contact details for bespoke extracts of births and deaths data are as follows:

Vital Statistics Outputs Branch

Office for National Statistics

Segensworth Road

Titchfield

Fareham

Hants PO15 5RR

E-mail: [vsob@ons.gsi.gov.uk](mailto:vsob@ons.gsi.gov.uk)

Telephone: +44 (0)1329 444110

Information on the release of births and deaths data for Scotland and Northern Ireland respectively are available from the links below:

[ISD Scotland Statistical Disclosure Control Protocol](#)

Information about what National Records of Scotland may provide (including some considerations that would apply and conditions that would have to be met) is given in the "[Enquiries for Vital Events statistics](#)" section of the GRO Scotland website.

[Statistical Policy Statement on Confidentiality and Access \(Northern Ireland\)](#)

## 2.3 Scope

All information collected as part of the birth and death registration process is within the scope of this guidance. This includes data collected for both administrative and statistical purposes. It also includes all information submitted to ONS by coroners.

For some outputs published by ONS (and other organisations), statistical information from birth and/or death registration is linked to other data sources. One example of this is birth registration records linked to 'NHS Numbers for Babies' (NN4B) data, to produce gestation-specific infant mortality statistics. Another example is cancer registration data (derived from NHS sources) linked to death registration data, to produce cancer survival statistics. Disclosure guidance related specifically to the release of tables produced from these linked data sources is out of scope of this guidance. However, ONS will collaborate with other organisations to advise, if there are defined user requirements that are not being met.

## 2.4 The registration process

Birth and death registration is a statutory duty under the Registration of Births and Deaths Act, 1953. Information about the disease(s) or condition(s) that led to a death is provided either by a medical practitioner or coroner. Medical practitioners are required to complete the Medical Certificate of Cause of Death (MCCD) to the best of their knowledge and belief. They are asked to include on the death certificate every disease or condition which, in their clinical judgement, contributed to the death either directly or indirectly.

Registrars of deaths are under a legal duty to report certain categories of deaths about which they have been informed to the coroner before they can be registered. The coroner will then decide whether to order a post-mortem and/or to hold an inquest; if they do, it is the coroner who certifies the cause of death. If the coroner decides not to investigate, then an MCCD is completed by a medical practitioner instead.

Other information collected at birth and death registration is provided by the person registering the birth or death (the informant), for example place of birth/death and occupation.

Further information about the death certification process can be found on the [General Register Office website](#). Links to specimen copies of birth and death registration forms and details of the procedures relating to birth and death registration are available in section 2.2 above.

### 3. Legal framework

**The data items collected on birth and death registration are collected under different legislation. This impacts on how decisions can be taken about risk of disclosure in table production, since some data items are explicitly collected as confidential data items for the preparation and supply of statistical information and others are not. Further details about how key pieces of legislation impact on the collection and release of birth and death registration are presented in this section.**

#### 3.1 The Data Protection Act 1998

The [Data Protection Act 1998](#) gives individuals rights over how their personal information is collected, stored, managed and used. It also places obligations on Data Controllers to process personal information according to the Data Protection Principles. Personal data that are processed exclusively for statistics and statistical research purposes are exempt from some elements of the Data Protection Principles.

Data protection legislation also provides a regulatory framework for the processing of personal data, which has been taken into account in the production of this disclosure guidance. Of specific relevance is whether information in a statistical table or analytical report identifies a particular individual (directly or indirectly). Where identification of an individual is likely reasonably to occur, the rights and obligations of the Data Protection Act are engaged.

#### 3.2 The Population (Statistics) Acts 1938 and 1960

[The Population \(Statistics\) Act 1938](#) gave the Registrar General power to collate any information obtained by registrars in the process of birth and death registration, which is needed for statistical purposes. Some amendments were made by the [Population \(Statistics\) Act 1960](#). The information includes confidential items regarding a birth or death which do not appear in the public register<sup>3</sup>, and may be used 'only for the preparation and supply of statistical information'.

An information paper entitled 'How changes to the Population Statistics Act will affect birth statistics' explains recent amendments to the Population (Statistics) Act 1938 (following those amendments made in 1960), to improve the statistical information collected at birth registration in England and Wales. This is available from the [ONS website](#).

### 3.3 The Statistics and Registration Service Act 2007

The [Statistics and Registration Service Act 2007](#) (SRSA) came into force on 1st April 2008. Section 39 of the SRSA governs the confidentiality of personal information held by the United Kingdom Statistics Authority and its executive office, The Office for National Statistics.

All information held by ONS and which relates directly or indirectly to a particular person (whether living or dead) is protected by section 39 of the 2007 Act. Disclosure of identifying information is an offence, unless an exemption to that offence applies. For example, the disclosure of personal information to Approved Researchers is an exemption to the non-disclosure rule. For information about how to apply to become an Approved Researcher, see the [Approved Researcher accreditation section](#) on the ONS website.

Section 42 of the SRSA created a new legal gateway between the Registrar General and ONS, enabling the Registrar General to provide ONS with any information entered in any births and deaths register, as well as any other information received by the Registrar General in relation to any birth or death. This includes all categories of information collected as part of the birth and death registration process (as specified in section 2.2).

Section 42(4) of the SRSA (as amended by the Health and Social Care Act 2012) includes provision for the ONS to supply information on individual births and deaths for the purpose of assisting the Secretary of State or the Welsh Ministers, or any one of a list of health-related organisations,<sup>4</sup> to enable them to produce statistics or carry out statistical analysis. This means that disclosive personal information of the specified type can be passed by ONS to the NHS or other health bodies, including local authorities when acting in their health role only, provided the information is used only for the purpose of producing and analysing statistics. Any of the listed bodies may use the information to assist any of the other listed bodies in their functions, as well as in their own functions relating to the health service. However, onward disclosures by those bodies of this information to non-listed bodies are not authorised by the SRSA.

Disclosures of personal information derived from birth and death registration data, other than those covered by the exemptions described above is a criminal offence which carries a maximum penalty of up to two years' imprisonment. The only statutory defence available to a person who makes an unlawful disclosure is that they had "reason to believe" the disclosure did not contain any personal information. The proper use of this guidance would contribute to such a defence.

One particular exemption to the non-disclosure rule – disclosures where the personal information has already lawfully been made available to the public – is covered in more detail in section 3.4.

### 3.4 The Code of Practice for Official Statistics

The United Kingdom Statistics Authority Code of Practice for Official Statistics (or the Code of Practice) applies to all UK bodies that are responsible for the production of official statistics. Compliance with the

---

<sup>4</sup> The full list consists of: the Secretary of State (the Department of Health and its Agencies: Public Health England), the Welsh Ministers (relevant departments and agencies of the Welsh Government), the NHS Commissioning Board (now called NHS England), a commissioning consortium, a local authority, a Local Health Board, an NHS Trust established under section 18 of the National Health Service (Wales) Act 2006, the National Institute for Health and Care Excellence, the Health and Social Care Information Centre, a Special Health Authority, the Care Quality Commission, and such other persons as the appropriate authority (i.e. the Secretary of State or Welsh Ministers) may specify in a direction given for the purposes of this section.

Code is a statutory requirement on bodies that produce statistics that have already been designated as National Statistics. The current series of statistics produced by ONS on births and deaths have all been designated as National Statistics, following assessment by the United Kingdom Statistics Authority.

Principle 5 in the United Kingdom Statistics Authority Code of Practice covers confidentiality, and states 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.’

Practice 1 of Principle 5 states that producers of Official Statistics must:

‘Ensure that official statistics do not reveal the identity of an individual or organisation, or any private information relating to them.’

The Code of Practice reference to the obligation to protect ‘private information’ implies that information that is not private does not automatically require protection under the Code. As discussed above, there is a conditional right in law for any member of the public to obtain a copy of any certificate of a birth or death<sup>3</sup>. This raises the matter of whether disclosures of personal information that have already lawfully been available to the public are exempt under S39(4) of the SRSA. However, legal opinion obtained by ONS has advised that there is a distinction in the relevant legislation between ‘Information relating to births and deaths’ and ‘information consisting of statistics’. The categories are mutually exclusive, because they are treated separately and distinctly in the relevant legislation.

In summary, the legal framework is such that a birth or death statistic produced by ONS (or any other organisation supplied with birth/death microdata by ONS), cannot lawfully contain information identifying a particular individual’s birth or death, even if some details of that person’s birth or death is discoverable through the certificate copy service of the GRO, because there is no power for ONS to disclose such information within the legal framework.

### 3.5 Summary of the ONS position on release of data

Release of personal birth and death information by ONS, other than to Approved Researchers or to a health-related body (as specified in the Act<sup>4</sup>), is potentially a criminal act. The Code of Practice states that all relevant data sources must be taken into account when protecting the identification of an individual. This implies that the existence of private data sources must be borne in mind when releasing and publishing statistics. For this reason, the ONS policy is that microdata (or individual level data) that are potentially identifiable may not be published (see Annex 2 for details of the risk assessment process). The uses of private data sources to identify an individual are prevented by the use of data access agreements (DAAs) between ONS and the external customer (see information about section 39(4) of the Statistics & Registration Service Act, above).

For those producing tabular data derived from birth and death registration data, this guidance is designed to help the decision making process to determine what can lawfully be disseminated. If there is any doubt around what tabular data can be shared or published, please contact ONS Statistics Disclosure Control branch (see ‘Background’, section 2.2 above for contact details).

## 4. Types of data

With respect to defining disclosure control guidance, it is helpful to categorise the types of information collected at birth and death registration and stored in the analytical datasets into four kinds: **identifiable**, **confidential**, **other routinely used variables** and **other variables available for statistical analysis**.

The '**identifiable**' category includes variables that may identify an individual. These might be direct identifiers such as name and address, or indirect identifiers, which are variables in a dataset that assist with identification of an individual without directly referring to them. For example combinations of age, sex and full postcode in a table could allow an individual to be identified with a great degree of confidence.

The '**confidential**' category includes variables collected under the Population Statistics Acts, for statistical purposes only. The information in this category is not available on the public record.

The '**other routinely used variables**' are those that are used in standard table production for ONS outputs, and those produced by other organisations (such as the Health & Social Care Information Centre). There is no legal definition of the types of variables that fall into this category; rather, this definition is a convenience to allow guidance to be issued for standard table production, where decisions can be informed by previous risk assessments (see Annex 2) and by precedence.

Similarly (but conversely), the '**other variables available for statistical analysis**' category defines those variables that are available for statistical analysis, but are used less commonly, so there are fewer examples on which to base risk assessment decisions. Information presented in sections 5.1 – 5.4 below use these categories to explain how to interpret the disclosure guidance.

### 4.1 Births

#### Identifiable variables

- Names,
- Addresses,
- Full postcodes (if presented in combination with other variables),
- Exact date of birth (if presented in combination with other variables),
- Specific occupation (if presented in combination with other variables).

#### Confidential variables

- Whether the child was stillborn,
- Gestational age (stillbirths only)<sup>5</sup>
- Birthweight of baby (stillbirths only).<sup>5</sup>

---

<sup>5</sup> For stillbirths, details of the gestational age and weight of the foetus are supplied as part of the registration process via a certificate completed by a doctor or midwife, which is taken to the registrar by the informant. If information is missing at registration, but the registration is linked to the NHS birth notification, then the information from the notification is taken. Gestational age and birthweight data for live births are only available via linkage between registration data and birth notification (NHS) data.

- Age of mother/father/second female parent at birth,
- Date of parents' marriage or civil partnership if applicable,
- Whether the mother has been previously married or in a civil partnership (only if she is currently married or in a civil partnership),
- Whether the mother has ever been married or in a civil partnership (only if she is not currently married or in a civil partnership),
- Number of previous live born children,
- Number of previous stillborn children.

#### **Other routinely used variables<sup>6</sup>**

- Year of birth,
- Month of birth,
- Sex of baby,
- Type of registration (inside marriage/civil partnership, joint registration by parents living at same address, joint registration by parents living at different addresses, sole registration),
- Area of usual residence of the mother (if the underlying population size is large enough for resulting table to have low risk of disclosure),
- Whether baby was born as part of a multiple birth,
- Place of birth (categorised - e.g. NHS hospitals, non-NHS hospitals, at home, elsewhere),
- Age of the mother, when presented in standard age groups (<20, 20-24, 25-29 etc., 40 and over),<sup>7</sup>
- National Statistics Socio-Economic Classification<sup>8</sup> (NS-SEC),
- Country of birth of the mother,
- Country of birth of the father.

#### **Other variables available for statistical analysis<sup>9</sup>**

- Individual communal establishment where the birth occurred (if relevant).

## **4.2 Deaths**

### **Identifiable variables**

- Names,

<sup>6</sup> Defined as variables used routinely in many tables published by ONS.

<sup>7</sup> These age groups are used in the majority of ONS publications presenting birth data by age of mother. More detailed age groups are used in some publications, although any resulting small numbers are suppressed and secondary suppression is also carried out to avoid disclosure by differencing.

<sup>8</sup> NS-SEC is derived from information about 'occupation' and information submitted voluntarily by informants about industry type and employment status. Standard presentations include 5- and 8-class versions.

<sup>9</sup> Defined as variables used less commonly in tables published by ONS, but available in the standard analytical dataset.

- Addresses,
- Full postcodes (if presented in combination with other variables),
- Exact date of death (if presented in combination with other variables),
- Exact date of birth (if presented in combination with other variables),
- Exact age at death (if presented in combination with other variables),
- Specific occupation (if presented in combination with other variables),
- Raw text information collected as part of a coroner's death registration (which may identify individuals other than the deceased).

### **Confidential variables**

- Marital/civil partnership status of the deceased,
- Date of birth of spouse/civil partner (if married or in a civil partnership),
- Final (underlying) cause of death.

### **Other routinely used variables<sup>10</sup>**

- Year of death (occurrence, or registration),
- Sex of deceased,
- Area of usual residence of the deceased (if the underlying population size is large enough for resulting table to have a low risk of disclosure),
- Place of death (categorised - e.g. NHS hospitals, non-NHS hospitals, at home, elsewhere),
- Original (underlying) cause of death,
- Age at death, when presented in standard age groups (<1, 1-4, 5-9 etc., 95+)<sup>11</sup>,
- National Statistics Socio-Economic Classification (NS-SEC).<sup>8</sup>

### **Other variables available for statistical analysis<sup>12</sup>**

- Month of death,
- Coded 'mentions' of other conditions or diseases that contributed to the death,
- Country of birth,
- Whether the death was related to employment,
- Whether the death was referred to a coroner,
- Individual communal establishment where the death occurred (if relevant),

---

<sup>10</sup> Defined as variables used routinely in many tables published by ONS.

<sup>11</sup> Previously, standard age groups for death registration data were: <1, 1-4, 5-9 etc. up to 85+. With the adoption of the new European Standard Population in 2014, the upper age group range will be extended to 85-89, 90-94 and 95+. For further information see the [ONS website](#).

<sup>12</sup> Defined as variables used less commonly in tables published by ONS, but available in the standard analytical dataset (with the exception of coded information derived from text collected as part of a coroner's death registration).

- Other text coded for statistical purposes, derived from information collected as part of the death registration process (for example, mentions of specific drugs associated with a death).

## 5. Geographic and population size considerations

There are a number of different scenarios under which a table including births or deaths registration data might be designed. The standard ones are presented below, in sections 7 – 11. Although the disclosure control guidance that should be applied is the same in each scenario, there are other considerations that need to be taken into account in each case. As the issues of geographic detail and underlying population size are relevant to each scenario, guidance on these is presented first.

### 5.1 Small populations

**In a large population (for example, a country or region), the effort and expertise required to discover more details about an individual person (or a private business) is generally considered to be disproportionate. As the size of the base population is decreased by moving to greater geographic detail or sub-populations, it becomes easier to identify individuals and discover information. However, there is still a balance to be struck between the risk of individuals actually being identified, and the utility of the statistics.**

In drafting this revised guidance, ONS has specifically considered the requirement to be able to publish birth and death registration statistics by both clinical commissioning groups (CCG, the official geography of NHS England) and local authorities (LA), where responsibility for public health in England now resides.

Before CCG boundaries were finalised on 1 April 2013, ONS worked closely with the Department of Health to ensure that differences between final CCG boundaries and existing LA boundaries would not cause disclosure issues related to small geographic ‘sliver’ areas (where different geographical boundaries overlap, potentially identifying small populations). This was a known issue previously, as some slivers between LA and the previous primary care organisation (PCO) boundaries were very small. This hindered the publication of some tables by PCO, where confidential variables (or NHS data) were used in the production of these statistics.

As the smallest sliver between final CCG boundaries and existing LA boundaries is above an estimated population size of 5,000, we confirm that routine statistics such as the number of stillbirths and live births by standard age group of the mother, and deaths by cause, age group and sex, can be published using both boundary sets. This judgement was made as a result of assessing the risk of disclosure with the utility of the resulting statistics.<sup>13</sup>

The concept of being able to publish death statistics by sex, age group and cause down to underlying population sizes of  $\geq 5,000$  can be extended further. For example, local health board (LHB), parliamentary constituency (PC), middle super output area (MSOA) and ward-level statistics can also be published for all areas with a population size greater than 5,000. For those areas with populations below this level, these can be aggregated with neighbouring areas. For the production of routine statistics, standard aggregations of areas should be defined in collaboration with ONS SDC (see section 2.2 for contact details). Users can

---

<sup>13</sup> Figures for the Isles of Scilly UA should be aggregated with the Cornwall UA, or small numbers suppressed, or a full risk assessment carried out demonstrating that the risk of disclosure is small.

also define their own geographic areas, for example by using aggregations of lower super output areas (LSOAs), where each new unit has an aggregated population size of  $\geq 5,000$ .

The concept of being able to publish birth statistics down to other underlying population sizes of  $\geq 5,000$  (other than LA, LHB and CCG) cannot be extended further to include PCs, MSOA's or wards, because the most commonly used variables (for example, fact of stillbirth in combination with age of mother) may be used to attempt to identify an individual. Therefore, more consideration needs to be taken of geographic slivers between boundary sets and the potential to inadvertently disclose personal information that is not freely available elsewhere. Birth statistics using these other boundary sets can still be produced, but small numbers would need to be disclosure controlled, or a full risk assessment carried out that demonstrated the risk of disclosure of personal information was small.

When assessing the risk of disclosure, ideally the population at risk would be considered rather than the entire resident population of an area. However, while the estimated resident population of a standard geographic area is a straightforward figure to access (for example, by accessing sub-national mid-year population statistics from the [ONS website](#)), the population at risk can be much more difficult to access or estimate. If this need to be accurately defined each time there was a new request for a bespoke table, then very quickly the guidance would become unusable.

Ultimately, as stated in the introduction, this guidance is designed to help users reduce the risk of disclosure when publishing tables – not to remove the risk altogether. Therefore, this guidance has specifically been designed around considering what tabular statistics can be released with a low risk of disclosure (considering what information is on the public record and what is confidential) with an underlying population size of 5,000 or more, and what can be released at more granular levels.

Annex 1a and 1b present summaries of when disclosure control is not, may not be or is required at different geographic levels, for tables based on birth and death registrations respectively. For combinations of variables where disclosure control is required, if a full risk assessment demonstrated that there would actually be a low risk of disclosure, then the table can still be published without the need for disclosure control. These summaries are a guide for users of microdata and customers requesting bespoke tables to illustrate when this should definitely be done, and when it is not required.

## 5.2 National and regional level tables

**At national or regional level, counts of live births, stillbirths and deaths (including by cause of death) can be published by year, standard age group of the mother (births) or of the deceased person (deaths) and sex, by a range of other variables, with no disclosure control required.<sup>14</sup>**

See Annex 1A and 1B for a summary of what can be published. For example, a table producing counts of live births and/or stillbirths by year of registration and NS-SEC of a parent would be acceptable, with no requirement for any disclosure control of small numbers. Likewise, a table of deaths by year of registration, sex, age group, cause of death and place of death (categorised) would not need to be disclosure controlled.

---

<sup>14</sup> Standard age groups for birth registration data are age of mother: <20 years, 20-24, 25-29, 30-34, 35-39, 40 years and over. Previously, standard age groups for death registration data were: <1, 1-4, 5-9 etc. up to 85+. With the adoption of the new European Standard Population in 2014, the upper age group range will be extended to 85-89, 90-94 and 95+. For further information see the [ONS website](#).

In this context, 'national' means data for England and Wales combined in a table, or presented separately. As the population of Wales is comparable with that of some English regions, it is reasonable to have the same guidance for regions as for England and Wales separately. In this context, the term 'region' refers to regions of England, previously known as Government Office Regions.

Birth and death registration data can also be published by more detailed breakdowns at national or regional level. However, care needs to be taken if using variables in the 'identifiable' or 'confidential variables' categories, in combination with those in the 'other standard variables' or 'other non-standard variables' category. For example, a table containing counts of live births presented by year and month of birth, country of birth of mother (categorised), age of mother, specific occupation and place of birth (categorised) may become disclosive, even at regional level, and should therefore have cells with counts of < 3 suppressed, even if the population threshold is  $\geq 5,000$ .

In general, tables using confidential variables should be designed in such a way as to avoid cell counts of less than three, to avoid the risk of disclosure. Where this is not possible, counts of less than three (including 0's) should be suppressed. Secondary suppression should also be applied to avoid disclosure through differencing, if required.

The summary information presented in Annexes 1a and 1b does not cover every possible combination of variables collected, instead focussing on the more routinely used (or requested) combinations. If users of this guidance have questions remaining about a specific data request, ONS staff (including SDC staff) would be happy to work with individuals to investigate possible table designs to meet specific needs, while keeping the requirement for statistical disclosure control to a minimum.

By their very nature, identifiable variables do not lend themselves to being incorporated into tables. If such a table were required, a full risk assessment would need to be carried out first, in collaboration with ONS SDC (see section 2.2 for contact details).

### 5.3 Sub-national areas (category 1)

**Tables based on death registrations data presenting figures by year, sex, quinary age group, International Classification of Diseases (ICD) code and any sub-national geography do not need to be disclosure controlled if the underlying resident population size is estimated as being  $\geq 5,000$ . So small numbers presented in tables by local authority (LA),<sup>9</sup> local health board (LHB), clinical commissioning group (CCG) or middle super output area (MSOA) do not need to be suppressed.**

**Tables based on birth registrations data presenting 'fact of live birth' or 'fact of stillbirth' figures by year, standard age groups (see section 4.1) and either LA, LHB or CCG also do not need to be disclosure controlled.**

Further breakdowns are also possible, with no requirement for disclosure control, for example by month of birth or death. See Annex 1A and 1B for a summary of what can be published.

Where small cell counts do need to be disclosure controlled, only cell counts of < 3 need to be suppressed (0, 1 and 2). The previous guidance required that all counts of < 5 should be suppressed for any data relating to stillbirths.

The 'sub-national areas (category 1)' grouping includes all geographical areas where the estimated underlying population size is  $\geq 5,000$  people for each unit, for tables based on **death registrations data**. This therefore includes:

- All local authorities in England and Wales (with the exception of the Isles of Scilly<sup>15</sup>) (LA),
- NHS England regions, area teams and clinical commissioning groups (CCG),
- Local health boards in Wales (LHB),
- Middle Super Output Areas (MSOA),
- Any non-standard geographical area, if the estimated resident population size is greater than or equal to 5,000 people (for example, postcode sectors) and the variables being presented are on the public register.

For **birth registrations data**, the 'sub-national areas (category 1)' grouping only includes the following:

- All local authorities in England and Wales (with the exception of the Isles of Scilly<sup>12</sup>) (LA),
- NHS England regions, area teams and clinical commissioning groups (CCG),
- Local health boards in Wales (LHB).

The guidance is deliberately more cautious for birth registration data, for the reasons described in section 5. The ONS Statistical Disclosure Control group have carried out a specific risk assessment on 'fact of live birth' or 'fact of stillbirth' figures by year, standard age group of mother and LA / CCG and have confirmed these tables have a low risk of disclosure. The release of tables based on other geographic boundaries (for example, MSOA, or non-standard boundaries) would therefore need to be considered on a case by case basis.

Disclosure control may be required for other combinations of variables; this will depend on how detailed a breakdown is required. If specific combinations of variables are required that are not covered explicitly within the guidance presented in Annexes 1A and 1B, please contact ONS for advice about what table could be released, with or without disclosure control (see above for contact details).

## 5.4 Sub-national areas (category 2)

**This includes all geographical areas where the estimated underlying population size is  $< 5,000$  people for each unit. At these small geographic levels, the risk of disclosure for births and deaths data is much higher because most table designs would be presentations of very small cell counts, which could be linked to other datasets with a reasonable level of accuracy. This means that a full risk assessment should always be undertaken to consider the risk of disclosure, in collaboration with ONS SDC (see section 2.2 for contact details).**

The 'sub-national areas (category 2)' includes the following geographic boundaries:

- Some electoral wards in England and Wales,
- Lower Super Output Areas (LSOA's),

---

<sup>15</sup> Figures for the Isles of Scilly UA should be aggregated with the Cornwall UA, or small numbers suppressed, or a full risk assessment carried out demonstrating that the risk of disclosure is small.

- Any non-standard geographical area, if the estimated resident population size is less than 5,000 people (for example, groups of postcodes).

This does not mean that birth or death registration statistics may never be released at these small geographic levels; only that small numbers should be suppressed or a full risk assessment must be carried out before making the decision to release the table. For example, presentation of fact of death at ward level, by sex, may be considered to be 'low risk', although the same breakdown either at a more detailed geographic level such as postcode or a more detailed 'time' level such as month or quarter of birth or death registration may not.

If customers request birth or death registration data at detailed geographic levels where the resident population size is < 5,000 and the risk assessment confirms that the tables can be released, ONS will publish these on the ad hoc customer request section of the ONS website (see Section 8 for further details). In this way, a range of statistics for small areas will be put into the public domain, and a body of 'good practice' examples will build up over time.

## 6. Publishing tables for individual or grouped communal establishments

**The guidance around publishing birth and death registration statistics for individual or grouped communal establishments varies depending on whether the establishment type is publicly or privately owned, since figures for individual private establishments cannot be disclosed. Also, the concept of 'resident population' is less helpful here, since this is not as easy to define as the resident population in a defined geographic area.**

In the context of birth and death registration data, the term 'Communal establishments' covers a broad range of places; for example, managed residential accommodation such as care homes, as well as hospitals, hospices and prisons.

The standard ONS definitions used to present 'place of death' statistics (which incorporate communal establishment groupings) are as follows:

- Home,
- Care Home: Local Authority, Non-Local Authority,
- Hospices: NHS, Non-NHS,
- Hospitals (acute or community not psychiatric): NHS, Non-NHS,
- Other communal establishments (includes schools, convents and monasteries, nurses' homes, university and college halls of residence, young offender institutions, secure training centres, detention centres; prisons and remand homes),
- Elsewhere (includes all places not covered above and people who are pronounced dead on arrival at hospital).

Care must be taken when presenting figures by individual communal establishments, as private organisations cannot be identified, because personal information under s39 of the SRSA also applies to the identification of 'a body corporate'. This means that care must also be taken if presenting figures using all of the above standard groupings below the level of country, or region, as some local authorities may only have one (privately owned) hospice, for example.

Tables presenting birth and death registration data by individual publicly owned communal establishments can be published. However, if presenting statistics by variables not on the public register, then small cell counts (<3) should be suppressed.

## 7. Designing a table for a statistical publication

**Statistics based on birth and death registration data support a wide range of users and uses, as they provide essential information for government, business, academia and the community. When designing tables for publication, producers of statistics need to balance the utility of the statistics against the need to protect confidentiality, while also considering the quality of the underlying data.**

When designing a statistical output, the main considerations are:

### 7.1 Does the table present the information in a way that is understandable to users?

Very detailed tables may provide the user with large amounts of data, but this may not be the best way of getting across a key message. For example, the ONS definition of avoidable deaths is an aggregation of statistics on underlying cause from across a number of disease chapters in the International Classification of Diseases, tenth revision (ICD-10). In this case, the value is in the aggregated statistics, not in the counts of deaths by individual, 4-digit ICD-10 code. One solution to this problem is to make both the summary and the detailed statistics available. This is why ONS releases both summary tables and charts and a more detailed downloadable (non-disclosive) dataset as part of a number of releases.

To see (and download) examples of both types of data release, follow the link to see the [avoidable mortality statistical bulletin](#).

Table and charts presented within the body of the publication are designed to help interpretation. The downloadable dataset presents the number of deaths by each cause included in the avoidable mortality definition, sex, 5-year age group, year (2001 to most recent available) and country (England and Wales combined and separately). This type of dataset will be useful to those wanting to analyse the impact of different conditions and diseases, by age, on the avoidable mortality definition.

### 7.2 Are the underlying data of sufficient quality to warrant publication in an official publication?

The quality of the underlying data needs to be considered when designing a table for either an official statistical output (for example, as Experimental, Official or National Statistics) or a research article. For example, ONS publishes National Statistics on drug-related deaths, by region, cause, age group and sex, because an assessment has been made that the data are of sufficient quality to warrant this level of detail. A more detailed geographic breakdown is not published within the statistical bulletin, because of known geographical variation in the quality of these data. For instance, the majority of drug-related deaths are known to be registered by coroners rather than medical practitioners. It is the text supplied to ONS by the coroner that is coded for each mention of a drug, where the underlying cause was drug-related. From this detailed data, we know that some coroners covering specific geographic areas routinely mention every substance tested for and found in the deceased's system, whereas another coroner may routinely not submit this level of detail. If the data were published as National Statistics at a more detailed geographical level than 'region', a user of the statistics may assume that a low count of deaths from certain drugs in an

area meant that this area did not have a problem with people dying in relation to this type of drug. However, it may just be that the coroner local to this area does not submit this level of detail on a death certificate. Instead, these statistics may be released separately from the official publication, for example as part of a customer request for a bespoke table.

As an example of both types of data release, follow the link to see the [drug-related deaths statistical bulletin](#). The table presenting [drug misuse deaths for local authorities in Wales](#) is published separately on the ad-hoc data analysis section of the ONS website. Comparable figures for [drug-related deaths in England](#) are also available.

Another example of when it may not be in the best interest of the user to publish very detailed tabular data is if the statistics are not stable over time. For example, if life expectancy estimates for local authority areas were published for individual years, then there would be a lot of instability in the estimates over time. This would reduce their value for planning and policy development purposes. Therefore, all figures presented in the ONS '[Life expectancy at birth and at age 65 for local areas](#)' bulletin are three-year averages, produced by aggregating the number of deaths and mid-year population estimates across each three-year period to provide large enough numbers to ensure that the figures presented are sufficiently robust.

In summary, whether the data in a table are disclosive or not should not be the only consideration when deciding whether to publish. The quality and stability of the underlying data should also be considered, particularly if the output may be used for planning purposes or as evidence in the development of policy.

### 7.3 Does the table present the information in a way that is useful to users?

Some key variables are routinely presented in certain formats, to allow comparisons to be made across datasets and to help promote their further use. For example, as population estimates are routinely published for 5-year age groups and sex, it aids the reuse of statistics if registration statistics are published in the same format (or aggregations of 5-year age groups, if small numbers are an issue).

Occasionally, detailed breakdowns of statistics actually hinder their usefulness, if by adding the detail it means small cell counts have to be disclosure controlled. One example of this is the number of live and stillborn babies born as part of a multiple birth, by the age group of the mother and the sex of the child. As details about a stillborn child are not in the public domain, but may be potentially disclosed by presenting them by characteristics of the mother, this level of detail cannot be published where there are small cell counts. A table presenting these statistics is included in the 'Characteristics of Birth 2, England and Wales – 2011' release available on the [ONS website](#) (see 'Table 6: All maternities: age of mother, multiplicity and type of outcome').

If the public health interest in these statistics is actually in how many live and still born children result from individual maternities, then simply removing the further breakdown by sex of the child would mean more statistics could be released into the public domain. This change was implemented in the November 2013 release of these statistics (see 'Characteristics of Birth 2, England and Wales - 2012' edition, available from the same link as above, for the previous edition).

## 8. Designing a bespoke, ad hoc table

Additional statistical tables and analyses not included in standard ONS publications are available on request (subject to legal frameworks, disclosure control, resources and agreement of costs, where appropriate). Tables requested by the media or those requested by others that are likely to be of wider public interest are then published on the [ONS website](#), so they are freely available to all.

When designing a bespoke table, for example for a customer, the main considerations are:

### 8.1 Does the table present the information in a way that is meeting the customer's needs?

In general, customers require as much detail as can be provided for each variable in question, without the need to suppress small cell counts (where there is a risk of disclosure). Examples are presented below of bespoke tables produced by ONS for customers, to illustrate how this balance might be struck in different scenarios.

**Example 1:** [Suicides in farmers](#). In the published table, figures for Cornwall, Devon, Dorset and Somerset were combined to reduce the risk of disclosure. This was the table design preferred by the customer, rather than the suppression of small cell counts.

**Example 2:** [Deaths by selected cause and single year of age](#), England and Wales, for deaths registered between 1959 and 1967. In the published table, figures are presented by single year of age as requested, even though this is not a standard table design for ONS outputs. As the figures were requested only at England and Wales level (with no further geographic breakdown), the figures were deemed to be not disclosive.

**Example 3:** The number of [drug poisoning deaths where helium was mentioned](#) on the death certificate in Manchester metropolitan district, for deaths registered between 2004 and 2011. In the published table, figures are combined across all years from 2004 to 2011. In this case it was more important for the customer to know the exact number of deaths from this very specific cause over the period, resulting in a table with a low risk of disclosure, rather than the numbers within each year, which would have resulted in a table with a higher risk of disclosure.

**Example 4:** The number of [deaths registered in each postcode sector in England and Wales](#), for each of the four quarters of 2011. In the published table, small numbers are presented without any suppression, even though the requested table design is not standard, because the risk of disclosure was considered to be low. Again, the figures presented are only fact of death, with no information about the characteristics of the person.

## 9. Answering a Freedom of Information request

The [Freedom of Information Act \(2000\)](#) gives individuals the right to ask any public sector organisation for all of the recorded information they have on any subject. The main principle behind it is that people have a right to know about the activities of public authorities, unless there is a good reason for them not to. There are a range of exemptions to the Act; for example, personal data cannot be released if releasing it would be contrary to the Data Protection Act.

As ONS has established mechanisms for individuals to request bespoke tables of birth and death registration data, we consider information relating to these data is reasonably accessible to the applicant by other means. However any bespoke service must be published on the ONS publication scheme prior to the receipt of any FOI request. Therefore, if individuals submit requests for these types of data via the Freedom of Information (FOI) route, then Section 21 of the Act is used as an exemption and FOI requests are answered with the following information:

“Information on the availability of [births or deaths data] can be obtained from the Life Events team [add e-mail address for the appropriate team]. ONS considers that S21(1) applies to this request and the information does not have to be supplied under the terms of the Freedom of Information Act. S21(1) is an absolute exemption and no consideration of the public interest test needs to be applied”.

Further information on Section 21 of the FOI Act is available from the [Information Commissioner's Office website](#).

Other public sector organisations holding birth and death registration data supplied by ONS may receive FOI requests relating to these data. In these circumstances, we would ask that the request is forwarded onto ONS, unless the request relates specifically to a table which that organisation has already produced.

Information about how to make an FOI request to ONS is provided on the [ONS website](#). Contact details for the FOI team in ONS are as follows:

Knowledge and Information Management  
Room 1100  
Office for National Statistics  
Segensworth Road  
Titchfield  
PO15 5RR  
Email: [foi.team@ons.gov.uk](mailto:foi.team@ons.gov.uk)

## 10. Answering a Parliamentary Question

**Parliamentary Questions are tools that can be used by Members of Parliament to seek information or to press for action. They oblige Ministers to explain and defend the work, policy decisions and actions of their Departments. All answers (and the original questions) are published in The Official Report ([Hansard](#)).**

When responding to a Parliamentary Question, the aim is to give all relevant information fully (taking the relevant disclosure guidance into account), and as concisely as possible. This is very similar to answering a customer request, except:

- It is not possible to discuss the proposed table layout of a table with the Member of Parliament requesting the data, to ensure that the resulting table is as useful as possible,

- Parliamentary Questions must be answered within the guidance on disproportionate cost,<sup>16</sup> whereas a customer will be asked to confirm they are willing to meet the costs of producing a table.

Occasionally a balance needs to be struck between producing a table that answers the MP's question, but does so using a method that allows the organisation to stay under the disproportionate cost threshold.

**Example 1:** [How many children and young people in England died from drowning in lakes, rivers or the sea](#), in each month since July 2008?

In the published answer, although the table is predominantly presenting small numbers (defined as counts of fewer than 3), this was deemed non-disclosive because information was requested for England as a whole.

**Example 2:** [\(a\) How many children under the age of three in the London Borough of Havering have died as a result of becoming entangled in window blind cords since 2008, and \(b\) how many children under the age of three have died \[in England and Wales\] as a result of becoming entangled in window blind cords since 2008?](#)

In the published answer, although the information presented gives small numbers for a specific age group and cause of death (and zeros, in the case of the specific geographical area), this was deemed to be a low disclosure risk, because no information was being released that could be used to identify an individual.

**Example 3:** [How many people died of alcoholic poisoning in each principal seaside town](#) in each of the last five years?

In the published answer, the table is presented with the small numbers unsuppressed, because although detailed cause and geographic information were requested, no information about sex or age group were requested. Without the addition of variables, the table was deemed non-disclosive.

**Example 4:** [Eight Parliamentary Questions asked by the same MP](#), combined into one Answer:

1. Which 10 (a) lower super output areas, (b) medium layer super output areas and (c) wards had the highest (i) number and (ii) proportion of sole registered births in each of the last five years?
2. Which 10 (a) lower super output areas, (b) medium layer super output areas and (c) wards had the highest (i) number and (ii) proportion of children born outside of marriage in each of the last five years?
3. Which 10 (a) lower super output areas, (b) medium layer super output areas and (c) wards had the highest (i) number and (ii) proportion of sole registered live births in each of the last five years?
4. How many and what proportion of births were sole registered births in each (a) lower layer super output area, (b) medium layer super output area and (c) ward in each of the last five years?
5. How many and what proportion of children were born outside marriage in each (a) lower layer super output area, (b) medium layer super output area and (c) ward in each of the last five years?

---

<sup>16</sup> In 2013, the disproportionate cost threshold was set at £850. If it is estimated that it would cost an organisation more than this to answer a question, it can be refused.

6. How many and what proportion of births were sole registered live births in each (a) lower layer super output area, (b) medium layer super output area and (c) ward in each of the last five years?
7. How many and what proportion of births outside marriage were joint registrations in each (a) lower layer super output area, (b) medium layer super output area and (c) ward in each of the last five years?
8. How many and what proportion of birth registrations outside marriage were from parents living at the same address in each (a) lower layer super output area, (b) medium layer super output area and (c) ward in each of the last five years?

In the published answer, explanations are presented about which of the statistics requested can be published and which cannot, because of the possible risk of disclosure. Copies of the tables provided in the Answer are not published in the Hansard record as they were too large, although they are available on request from ONS, or via the [House of Commons Library](#).

## 11. Designing an open dataset

**An open dataset is one that is designed and released in a way that allows maximum re-use of information (so is in a machine readable form, based on open standards), while not compromising legal obligations to protect the confidentiality of individuals. Further information about open datasets produced by ONS is available from the [ONS website](#).**

This is exactly the same as if answering a customer request, with the exception that the person designing the dataset must consider what the most appropriate level of detail is to present the data, to ensure that the resulting table is designed to maximise the utility for potential customers.

In summary, as with the rest of the Government Statistical Service (GSS), ONS is committed to increasing the amount of data released in open and re-usable formats in the production of its statistical outputs. This is in line with the UK Government's policy, as set out in the Open Data White Paper.<sup>17</sup>

Principle 8 of the Code of Practice states that producers of official statistics should ensure that they are disseminated in forms that enable and encourage analysis and reuse. Producers should release datasets and reference databases, supported by documentation, in formats that are convenient to users.

Principle 5 of the Code of Practice states that official statistics should not reveal the identity of an individual or organisation, or any private information relating to them, taking into account other relevant sources of information. Arrangements for confidentiality protection must be sufficient to protect the privacy of individual information, but not so restrictive as to unduly limit the practical utility of official statistics.

Further information about [the position of the GSS on the release of open datasets](#) (published in 2012) can be found on the United Kingdom Statistics Authority's website, which also presents a summary of the Cabinet Office draft Public Data Principles (published in 2010), which underpins the release of open datasets.

---

<sup>17</sup> Link to White Paper <http://www.cabinetoffice.gov.uk/resource-library/open-data-white-paper-unleashing-potential>

## 12. Sensitive topics

**ONS has no evidence that more harm or distress has ever been (or would be) caused to a living relative of a person who has died through publication of detailed (but not disclosive) statistics. Arguably, more harm or distress may actually be caused if a relative felt suppression of detailed figures prevented those working in public health (or health services generally) from taking action to prevent further deaths. Therefore, ONS does not have a separate set of disclosure guidance for sensitive causes.**

Other countries and organisations (for example, ISD Scotland<sup>18</sup>) have defined lists of sensitive topics, for which there are different disclosure rules covering release in tabular form. The judgement to define a topic as 'sensitive' is based around the potential distress or embarrassment caused to an individual, if detailed statistics were published. By implication, such a list would refer to morbidity data rather than mortality data, although some of the topics listed are relevant to mortality statistics (such as suicide, self harm, and alcohol or drug misuse).

In the previous published briefing note on the ONS policy on protecting confidentiality within birth and death statistics, there was the following statement:

“Under the Data Protection Act 1998, ONS must consider any potential harm or distress to an individual mentioned in the register but also, in the case of death records, to the living relatives of those who have died.... Therefore for the information in the above lists, the only restrictions for statistics on registrations are where harm or distress to an individual could be caused by releasing small counts at low geographies (local authority and below).”

This statement was understood to mean that more stringent disclosure control was required for detailed death registration statistics based on potentially sensitive causes of death. The Data Protection Act does list some topic areas that should be considered more carefully, but none are specific to births and deaths data (the broad 'health' being the only related one). Otherwise, what is considered sensitive is very much down to individual judgement and might be difficult to defend (one way or the other) in a court of law. Moreover, in this revised guidance, ONS has decided to remove the concept of sensitive causes as these may be the very statistics that are of most use from a public health perspective.

## 13. Types of disclosure

**This section gives details about how disclosure may occur in birth and death registration data.**

When determining the possibility of disclosure a number of factors need to be considered, which are linked to the types of variables and topics listed in Section 3 above. The question that ultimately requires an answer is: 'what would be the impact on the individual if the data under discussion were in the public domain?' This is linked to whether the variables are considered to be potential identifiers (for example, postcode and/or occupation, in combination with other variables), or confidential variables (for example, marital or civil partnership status of the deceased).

---

<sup>18</sup> See page 12-13 of the ISD Scotland Statistical Disclosure Control Protocol available [here](#) for list complete list of sensitive diseases.

The use of variables that are considered to be strong identifiers (such as full postcode or specific occupation), whether they are on the public register or not, increase the risk that particular individuals may be identified. These factors determine whether or not there is a high risk of being able to identify an individual, whether the risk of identification includes the risk of disclosing confidential information, and how much time and effort would have to be taken by an individual to discover anything they did not already know.

The most likely types of disclosure are:

### 13.1 Attribute disclosure

**Attribute disclosure is a type of disclosure which occurs when an intruder is able to deduce something new that they did not already know about an individual in a dataset.**

This form of disclosure may occur when empty cells are present in a table or where a number of tables can be linked to produce sparse tables by subtraction. The possibility of attribute disclosure is less affected by the population at risk value than identity disclosure (see Section 13.2). Attributes can be disclosed due to the distribution of the information in a table. A table with a skewed distribution where almost all the members fall into a small number of categories could lead to attributes of those people in the least populous cells being discovered.

Attribute disclosure may be possible for a wide range of outputs but is of greatest concern if the variables are potential identifiers and/or are confidential (and so not in the public domain). In summary, the more identifiable the data are, the greater the requirement for protection.

In a table, attribute disclosure will occur if any row or column contains only one non-zero value – that is, if all individuals are in the same category in a row or column in a table. This shows how non-structural zeros can be problematic. These are zeros that occur because no members of the population fall into the relevant category (or categories) although it would be possible to do so. Structural zeros occur when a category must be zero because of practical reasons. For example, there are upper and lower limits on the ages at which a woman can give birth.

A table containing counts of stillbirths by low level geography for a particular year will contain many zero cells and a small number of low frequencies. This could encourage an intruder who may have known a woman in this area had been pregnant but now did not appear to have a child to see if they were present in this table. A similar case could occur if a table of deaths from very specific causes and at a very low geographic level (for example LSOA) were published. This may enable the data to be linked with a reasonable level of accuracy to that in another table held by an individual, potentially leading to attribute disclosure.

### 13.2 Identification and self identification

**Identification disclosure occurs when an individual can be discovered in a published table.**

Identification is closely linked to population at risk. Tables based on populations at risk are more likely to experience identification disclosure problems particularly when combined with low cell counts. A cell with a count of 1 or 2 in a table based on a small population at risk could encourage an intruder to attempt to identify these individuals in the population. Self identification could also be an issue with a particular individual noticing that they were unique or rare in the population.

Identification and self identification will be more of a concern when the variables building the table are confidential and there would be greater impact if the table was published. A cell indicating the mere fact of a death is less of an issue than a cell presenting cause of death by marital or civil partnership status, as the latter is information that could potentially be used to attempt to identify an individual.

### 13.3 Disclosure by Differencing (Residual Disclosure)

**Disclosure by differencing can occur when two or more similar tables are compared.**

In many 'differencing' scenarios the variable to compare is geography. Tables can be produced at clinical commissioning group level, local authority level and parliamentary constituency level among others. In some places these differ slightly and it is possible that if tables are published at these different geographical levels they could be compared to produce slivers with low counts thus leading to disclosure problems.

Tables with non-geographical variables could also be differenced. One example is age group. If tables of 'cause of death' are produced with marginally different age categories (e.g. Table 1 age groups are 16-18, 19-25 etc and in Table 2 they are 16-19, 20-25...) it could be possible to obtain the exact value in a cell by subtracting one table from another thereby revealing the cause of death of the 19 year olds in the table. If this was a unique cell and the table was presenting information about confidential variables this would be a disclosure problem.

In summary, the importance of disclosure by differencing is closely related to whether the information presented in a table is public record information only, or whether it is combined with confidential variables.

## 14. Selecting methods of statistical disclosure control

**If a table is deemed to be disclosive, a method of disclosure control must be applied before it is published or released, to reduce the risk of information about identifiable individuals being disclosed. The choice of method must balance how the information is to be used with the simplicity of approach.**

The methods are divided into three categories: those that determine the design of the table, those that modify the values in the table and those that adjust the data before tables are designed. Descriptions of each method with advantages and disadvantages are provided below. In addition, examples where each method has been implemented are outlined. Each example dataset (other than the 1991 Census) can be found on the ONS Neighbourhood Statistics website ([www.neighbourhood.statistics.gov.uk](http://www.neighbourhood.statistics.gov.uk)).

### 14.1 Table design

Table redesign is the most recommended method to use when designing a table that includes birth and death registration data. It is simple in concept, as the aim is to minimise the number of unsafe cells and preserve original counts. However, the use of this method should be balanced against consistency in table design and publication plans.

**Table 1: Statistical disclosure control methods - design the table**

Method	Description	Advantages	Disadvantages	Examples
Table redesign	<p>Disguise unsafe cells by:</p> <ul style="list-style-type: none"> <li>- grouping categories within a table (for example, age groups)</li> <li>- aggregating to a higher level geography or for a larger population sub-group</li> <li>- aggregating tables across a number of years, months or quarters</li> </ul>	<p>Original counts in the data are not damaged.</p> <p>Easy to implement</p>	<p>Detail in the table will be reduced.</p> <p>May be policy or practical reasons for requiring a particular table design</p>	<p>Teenage conception statistics are published for Local Authority or higher level. City of London is combined with Hackney, Rutland UA is combined with Leicester UA and Isles of Scilly UA are combined with Cornwall UA<sup>10</sup></p> <p>To avoid suppression of small cell counts, data across a number of reference years are routinely aggregated if answering a particularly detailed Parliamentary Question or customer request.</p>

## 14.2 Cell modification

If unsafe cells remain in the output tabulation, further protection methods should be considered in order to disguise them. If table redesign is not a feasible solution, the recommended method for post-tabular protection is cell suppression. However, this is generally only a useful method if the number of unsafe cells is low. If the number of unsafe cells is high, then the utility of the table is substantially reduced.

An alternative method of post-tabular protection is controlled rounding. However this method requires specialist software and therefore will not always be practical.

Controlled rounding and cell suppression can be implemented in the Tau-Argus software (version 3.5.0 available at <http://neon.vb.cbs.nl/casc>)

Table 2 shows different methods which can be used to modify cell values.

**Table 2: Statistical disclosure control methods - modify cell values**

Method	Description	Advantages	Disadvantages	Examples
Cell suppression	Unsafe cells are not published. They are suppressed and replaced by a special character, such as '.' or 'X', to indicate a suppressed value. Such suppressions are called primary suppressions. To make sure that the primary suppressions cannot be derived by subtraction from totals, it may be necessary to select additional cells for secondary suppression	<p>Original counts in the data that are not suppressed are not adjusted.</p> <p>Can provide protection for zeros</p>	<p>Most of the information about suppressed cells will be lost.</p> <p>Secondary suppressions will hide information in safe cells.</p> <p>Information loss will be high if more than a few suppressions are required.</p> <p>In order to protect any disclosive zeros these will need to be suppressed.</p> <p>Does not protect against disclosure by differencing.</p> <p>Complex to implement optimally if more than a few suppressions are required and particularly complex for linked tables.</p>	Cell suppression is used in detailed characteristics of birth tables <sup>19</sup>
Rounding	Rounding involves adjusting the values in all cells in a table to a specified base. This creates uncertainty about the real value for any cell while adding a small but acceptable amount of distortion to the data	<p>Counts are provided for all cells.</p> <p>Provides protection for zeros.</p> <p>Protects against disclosure by differencing and across linked tables.</p> <p>Controlled rounding preserves the additivity of the table and can be applied to hierarchical data</p>	<p>Cannot be used to protect cells that are determined unsafe by a rule based on the number of statistical units contributing to a cell.</p> <p>For example, if a cell had an original count of 17 events all associated with one practitioner, then rounding this to 15 means that the count still relates to only one practitioner, the unsafe cell is not disguised.</p> <p>Random rounding requires auditing; controlled rounding requires specialist software, which is readily available.</p>	Counts from the New Zealand Census are rounded to base 3 <sup>20</sup>

<sup>19</sup> <http://www.ons.gov.uk/ons/publications/re-reference-tables.html?edition=tcm%3A77-230095>

<sup>20</sup> <http://www.stats.govt.nz/Census.aspx>

### 14.3 Database modification

If a data provider has access to the individual record level data then disclosure control methods can be implemented that adjust the data before tables are designed. This is not a method that is used routinely for release of birth and death registration microdata, as other methods exist to allow researchers access microdata (see the information on becoming an Approved Researcher, available from the [ONS website](#)). However, in certain rare circumstances, this method may be appropriate.

**Table 3: Statistical disclosure control methods - adjust the data**

Method	Description	Advantages	Disadvantages	Examples
Record swapping <sup>21</sup>	Swap pairs of records within a micro-dataset that are partially matched to alter the geographic locations attached to the records but leave all other aspects unchanged.	<p>Protects against disclosure by differencing.</p> <p>Once modified all tables are produced, this can be useful when protecting online databases.</p> <p>Can target risky records.</p> <p>Gives consistent and additive tables.</p> <p>Counts at high geographies are unaffected.</p>	<p>High level of swapping may be required in order to disguise unsafe cells.</p> <p>Will distort distributions in the data.</p> <p>Method not transparent to users.</p> <p>May be a perceived risk as cells with low counts will be published.</p> <p>Understanding the theory and practicality of this method may not be easy. Considerable communication and education will be required.</p> <p>Calculations relating to loss of data utility and doubt may need to be calculated before all output tables are produced.</p>	Used in combination with table design to protect the 2011 Census for England, Wales, Scotland and Northern Ireland.
Removal of risky records	A small number of records may be unique in the data for a number of variables. Rather than protecting tables using these variables it would be simpler to remove the record	Less protection is required in the published tables without having to allow for an outlying record	A subjective decision has been made to remove information from the dataset. Users of the data may not know this has taken place or the methodology behind the removal of certain records	<p>This is more likely to be a technique used prior to releasing microdata.</p> <p>There are no examples of the use of this method for ONS tabular data that have been made public.</p>

<sup>21</sup> Details of record swapping methodology can be found at <http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-prospectus/new-developments-for-2011-census-results/statistical-disclosure-control/index.html>

There are many other methods of disclosure control not in the list of recommended options in the above tables. For further information, see the document entitled '[Evaluating a statistical disclosure control \(SDC\) strategy for 2011 Census outputs](#)' available from the ONS website. This covers a modified version of the cell perturbation method developed by the Australian Bureau of Statistics, over-imputation as well as record swapping. Another option is creating a synthetic dataset which maintains all the properties of and relationships in the true dataset. From these data, non-disclosive tables can be created.

Alternative methods for presenting data can be considered as an approach for providing users access to information without disclosing the underlying data. In many cases this will provide a more robust analysis than reliance on the accuracy of small cell counts. These could include presenting data using smoothed maps and/or funnel plots. Smoothed maps provide a strategic overview of geographic patterns within a detailed dataset, without revealing the (potentially disclosive) data on which the map is based. Funnel plots can be used to show how data points differ from the national average, without revealing any disclosive data. [Examples of these methods of data presentation](#) published by ONS, are available on the ONS website.

## 15. Definitions

**A list of definitions is presented to assist people in reading and understanding this document.**

**Anonymisation** involves removing the direct identifiers from a microdata record. This term is used frequently when microdata are being protected. *Direct Identifiers* are variables which will enable an individual to be identified with a high degree of confidence. *Direct identifiers* include name, NHS number and date of birth (if presented in combination with other variables).

**Attribute Disclosure** If an intruder knows something about a person, for example which row the person will be counted in, then they will deduce something new about the person if all the cases in that row are in one column and all other columns in the row contain zero. This is attribute disclosure. Such situations are certain to arise when there is only one case in the row, and more likely to occur when the number of cases in the row is small.

**Data Utility** is a measure of the value of a table for researchers or other users. The lower the level of *disclosure control*, the greater is the utility of the data. However *disclosure risk* is also greater. A balance needs to be found between utility and risk.

**Differencing** may occur if similar tables are produced at slightly different geographies (for example local authority and middle super output area). It may be that by differencing the tables, slivers are found which allow individuals to be identified. Differencing may also occur when tabulations are created where non geographical variables (such as Age group) have slightly different categories.

**Direct Identifiers** are variables in a dataset will help an intruder easily identify an individual. These include name, NHS number and date of birth (if presented in combination with other variables).

**Disclosive data** are data (tables or microdata) containing no personal identifiers from which it would be possible to gain more information about an individual or other statistical unit.

**Disclosure** has two components: *identification* and *attribution* and occurs when an intruder makes a correct claim regarding a member of the data (microdata or table). The claim could be about an individual, household business or other statistical unit.

**Disclosure Control** refers to a number of techniques which can be applied to the data to limit *disclosure risk*. The most common techniques include *recoding*, *suppression* and *rounding*.

**Disclosure risk** occurs if a respondent's confidential information can be ascertained either exactly or to within a defined narrow bound by an *intruder* with a high level of confidence. This risk can be mitigated by applying *disclosure control*. The level of disclosure risk is often not zero but is manageable.

**Final cause of death** Occasionally a medical practitioner or coroner may submit extra information about a death directly to ONS, after the original cause of death information has been submitted by the Registrar General. This does not match the information available via the public register (the original cause of death), and so requires a higher level of disclosure control than original cause of death information. ONS publishes statistics based on final cause of death at national level (England & Wales, or England and Wales separately) and for regions within England. Sub-regional outputs are routinely based on original cause of death information.

**Frequency tables** are tables of counts. They are usually in an additive format with individual cell summing to marginal totals and then to an overall total.

**Hierarchical Recoding** is a type of *rounding* possible with variables which have a defined layered structure. The hierarchical recode can be applied to display the data at a higher level of a variable such as Geography. For example data can be recoded from Local Authority level to Region level.

**Identifying variable** is a variable which may identify an individual. These might be *direct identifiers* such as name or NHS number or *indirect identifiers* which are variables of varying degrees of visibility such as sex, age group, occupation, etc.

**Identity Disclosure** occurs when at least one person in the released data is identified. Identification may be by oneself or by a third party *intruder*.

**Indirect Identifiers** are variables in a dataset that assist with identification of an individual without directly referring to them. For example combinations of age and occupation in a table could allow an individual to be identified with a great degree of confidence.

**Intruder** refers to a group or individual who wishes to identify people in the table or attributes relating to these people. Also known as an attacker, although there may or may be malicious intent.

**Key variable** is a variable which is commonly used in tabulations. If a large number of tables are produced they are likely to be linked via one or more key variable. By combining these tables an *intruder* may be able to identify an individual or associated attributes.

**Marginal Totals** are the totals for each row or column in the table. The overall total is the sum of the relevant marginal totals.

**Microdata** are individual records, either relating to individual people (in the case of birth or death registrations) or events (in the case of cancer registrations). Microdata may contain direct identifiers (such as names, NHS number), or indirect identifiers (such as occupation, postcodes, in combination with other

variables), and so be disclosive. Where an individual is not identifiable, data are said to be anonymous, or non-disclosive (see below).

**Non disclosive data** are data (table or microdata) containing no personal identifiers from which it is not possible to gain more information about an individual or other statistical unit.

**Open data** are data (tables or microdata) that are released in a way that allows maximum re-use of information (so is in a machine readable form, based on open standards), while not compromising legal obligations to protect the confidentiality of individuals.

**Original cause of death** is the original cause of death information submitted to ONS by the Registrar General, via a medical practitioner or coroner, which ONS then codes. In many cases this matches the un-coded information available via the public register. Coded original underlying cause of death may differ from the un-coded original underlying cause of death because of the application of internationally agreed cause of death coding rules. Occasionally, additional information is provided by medical practitioners or coroners after the death has been registered. This is termed final cause of death (see above) and does not match the public register. ONS publishes statistics based on final cause of death at national level (England & Wales, or England and Wales separately) and for regions within England. Original underlying cause is used if sub-regional statistics are also presented.

**Personal data** are information relating to an identified or identifiable person. An identifiable person is one who can be identified, directly or indirectly. Where an individual is not identifiable, data are said to be anonymous.

**Personal information** (as defined by the Statistics & Registration Service Act, 2007) is information that relates to and identifies a particular person (or corporate body). Information identifies a particular person if the identity of that person is either specified in the information, can be deduced from the information, or can be deduced from the information taken together with any other published information. Under the Statistics and Registration Service Act (2007), personal information is deemed confidential.

**Population, resident** refers to estimated number of people living in a specified geographic area. For a tabulation based on local authorities, the resident population size to consider would be that of the smallest local authority for which data were presented. This is used throughout the majority of this guidance, as it is much easier to define than *population at risk*.

**Population at risk** is the maximum number of people who could feature in a released table. This could be estimated at a tabular or cell level. For example for birth the population at risk is usually defined as the female population aged 15-44 (Note: this may be modified in the near future to take account of the increasing numbers of older mothers). A table with a small population at risk is more likely to have disclosure issues than one with a large population at risk.

**Public register** some personal information collected at birth/death registration is available (on application and for a charge) from the General Register Office (GRO), via the supply of individual certificate copies. This information is considered to be on the public register. For more information, see the [GRO website](#).

**Pseudonymised data** where a 'real' identifier – such as a name or NHS number – is replaced by a 'false' identifier such as a 'hashed' code number. This is a privacy-enhancing technique used in contexts such as medical research. It allows individuals to be 'tracked' longitudinally – for example – without revealing their identities to the researcher.

**Recoding** is a *disclosure control* technique. A variable in a table can be recoded so that the information in a particular row or column is combined with neighbouring rows or columns. An example is individual age which could be recoded into 5 year age groups. *Hierarchical recoding* is a specific type of recoding.

**Rounding** is a *disclosure control* technique based on perturbation of the output. It is used to protect small counts in tabular data against disclosure. Each cell count can be rounded up or down either deterministically or probabilistically to the nearest integer multiple of a rounding base (typically 3, 5 or 10). Table additivity is generally lost when rounding is applied unless the operational research technique known as controlled rounding is used.

**Suppression** is a *disclosure control* technique where information is lost rather than perturbed. Low values and other sensitive cells are replaced a symbol showing suppression has taken place. A row or column with a suppressed cell will require at least one additional cell to be suppressed to avoid disclosure by subtraction from the marginal total. These other cell suppressions are referred to as secondary suppressions. Selecting cells for secondary suppression can be a time consuming process.

**Table** in the context of this guidance, a table is defined as a cross tabulation of two or more variables, created to summarise categorical data. This type of table is also known as a contingency table. It is assumed that in aggregating the data to carry out the cross tabulation, the number of rows will be less than the number of events being summarised.

**Table dimension** defines the level of detail present in the table. A two dimensional table will have two defining variables such as age group and cause of death while a three dimensional table will have an additional variable such as geography. Increasing dimensions lead to a smaller number of average counts per cell.

**Underlying cause of death** is defined by the World Health Organisation as:

- a) the disease or injury that initiated the train of events directly leading to death; or
- b) the circumstances of the accident or violence that produced the fatal injury.

**Visible variable** is a variable which enables identification of an individual or other statistical unit by placing them in a certain category for particular key variables.

## Annex 1A: Summary guidance on disclosure control – birth registrations

		Fact of live birth			Fact of stillbirth		
		Year	Agegroup	All combined	Year	Agegroup	All combined
	<b>Year and agegroup variables not used in combination with any below</b>	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
<b>Identified / confidential variables</b>	Marital / civil partnership status	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
	Exact date of birth / marriage / civil partnership	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
	Occupation of parent (mother / father / 2nd female parent)	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
	Number of previous live born children	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
	Birthweight & gestational age (categorised: stillbirths only)	N/A	N/A	N/A	Nat/Reg	Nat/Reg	Nat/Reg
		N/A	N/A	N/A	Sub-nat 1	Sub-nat 1	Sub-nat 1
		N/A	N/A	N/A	Sub-nat 2	Sub-nat 2	Sub-nat 2
	Number of previous stillborn children	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
<b>Other routinely used variables</b>	Type of registration (sole / joint etc.)	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
	Whether part of a multiple birth	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
	Sex of the child	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
	Place of birth (categorised: home, hospital, etc.)	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
National Statistics Socio-Economic Classification (NS-SEC)	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	
	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	
	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	
Month of birth	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	
	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	
	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	
Country of birth of the parent (individual or grouped)	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	
	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	
	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	
Nat/Reg	Tables at national (England & Wales, England or Wales) or regional level only.						
Sub-nat 1	Table at sub-national level (local authority, merging Isles of Scilly with Cornwall, clinical commissioning group, & local health board only).						
Sub-nat 2	Table at sub-national level: all other standard and non-standard areas (incl. parliamentary constituencies, electoral wards, middle and lower super output areas).						
	No disclosure control or table redesign required, even if cell counts are low (<3).						
	Disclosure control or table redesign may be required, if cell counts are low (<3), depending on what the risk of disclosure is.						
	Disclosure control or table redesign will be required, if cell counts are low (<3). Alternatively, a full risk assessment should be carried out. If this demonstrates the risk of disclosure is small, the table can be published.						
N/A	Not applicable (see footnote 5 below).						
	<sup>1</sup> Disclosure guidance for stillbirths also covers that for perinatal deaths, which are stillbirths plus early neonatal deaths (or deaths under 7 days).						
	<sup>2</sup> Year of birth is the standard variable used to present birth registration statistics.						
	<sup>3</sup> In general, month of birth would be used in combination with year of birth.						
	<sup>4</sup> For a list of the standard groupings used for the 'country of birth' variables, see the 'Parent's country of birth, England and Wales' publication here: <a href="http://www.ons.gov.uk/ons/rel/vsob1/parents-country-of-birth-england-and-wales/index.html">http://www.ons.gov.uk/ons/rel/vsob1/parents-country-of-birth-england-and-wales/index.html</a>						
	<sup>5</sup> For stillbirths, details of the gestational age and weight of the foetus are supplied on a certificate by a doctor or midwife. The certificate is then taken by the informant to the registrar. If information is missing at registration, but the registration is linked to the NHS birth notification, then the information from the notification is taken. Gestational age and birthweight for live births are only available via linkage with birth notification (NHS) data.						

## Annex 1B: Summary guidance on disclosure control – death registrations

		Fact of death				Underlying cause of death			
		Year	Agegroup	Sex	All combined	Year	Agegroup	Sex	All combined
	<b>Year, agegroup &amp; sex variables not used in combination with any below</b>	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
<b>Identified / confidential variables</b>	Marital / civil partnership status	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
	Exact date of birth / date of death / age	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
	Occupation	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
<b>Other routinely used variables</b>	Place of death (categorised: home, hospital, hospice, etc.)	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
	National Statistics Socio-Economic Classification (NS-SEC)	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
<b>Other variables available for statistical analysis</b>	Month of registration	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
	Coded 'mentions' of conditions that contributed to the death	N/A	N/A	N/A	N/A	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		N/A	N/A	N/A	N/A	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		N/A	N/A	N/A	N/A	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
	Country of birth (individual or grouped)	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
	If the death may have been related to employment	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
	If the death was referred to a coroner	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg
		Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1
		Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2
Coded text information, for example mentions of specific drugs	N/A	N/A	N/A	N/A	Nat/Reg	Nat/Reg	Nat/Reg	Nat/Reg	
	N/A	N/A	N/A	N/A	Sub-nat 1	Sub-nat 1	Sub-nat 1	Sub-nat 1	
	N/A	N/A	N/A	N/A	Sub-nat 2	Sub-nat 2	Sub-nat 2	Sub-nat 2	

Nat/Reg	Table at national (England & Wales, England or Wales) or regional level only.
Sub-nat 1	Table at sub-national level where underlying population is ≥5,000 (incl. local authority, merging Isles of Scilly with Cornwall, clinical commissioning group, local health board, middle super output area & parliamentary constituency).
Sub-nat 2	Table at sub-national level where underlying population is <5,000 (incl. some electoral wards, lower super output areas).
	No disclosure control or table redesign required, even if cell counts are low (<3).
	Disclosure control or table redesign may be required, if cell counts are low (<3), depending on what the risk of disclosure is.
	Disclosure control or table redesign will be required, if cell counts are low (<3). Alternatively, a full risk assessment should be carried out. If this demonstrates the risk of disclosure is small, the table can be published.
N/A	Not applicable to present variables just by 'fact of death' as these include information related to underlying cause.

<sup>1</sup> Final underlying cause is the standard variable used to present death registration statistics at national and region level. Original underlying cause is used if sub-regional statistics are also presented.

<sup>2</sup> Year of registration is the standard variable used for death registration tables. In some circumstances this may be substituted for year of occurrence.

<sup>3</sup> In general, month of registration would be used in combination with year of registration, and month of occurrence with year of occurrence.

<sup>4</sup> For a list of the standard groupings used for the 'country of birth' variables, see the 'Parent's country of birth, England and Wales' publication here: <http://www.ons.gov.uk/ons/rel/vsob1/parents--country-of-birth--england-and-wales/index.html>

<sup>5</sup> Care must be taken when presenting figures by communal establishment groupings, to prevent private organisations from being identified. This impacts on the presentation of figures below the level of country or region, as some local authorities may only have one (privately owned) hospice, for example.

<sup>6</sup> Previously, standard age groups for death registration data were: <1, 1-4, 5-9 etc. up to 85+. With the adoption of the new European Standard Population in 2014, the upper age group range will be extended to 85-89, 90-94 and 95+.

## Annex 2: Risk assessment

### 1. Introduction

The process of risk assessment carried out by the ONS statistical disclosure control group (SDC) is based on:

- The nature of the data,
- Possible intruder scenarios,
- Key variables.

### 2. Possible intruder scenarios

Both tabular data and microdata may be at risk of disclosure. When assessing the disclosure risk, it is necessary to first consider the relevant intruder scenarios. When producing tables based on birth and death data, there are two scenarios which need to be considered: the use of published datasets and spontaneous recognition. In relation to microdata the scenarios depend on assumptions about what an intruder might know about individuals in the data, and what information will be available to match against the microdata and potentially make an identification. The consideration of scenarios indicates some of the variables which are likely to be used by an intruder.

#### 2.1 Scenario 1 – use of published datasets

This scenario is relevant to both tables and microdata. There are datasets in the public domain which contain demographic variables, examples are the Electoral Register and commercial datasets, such as consumer profile databases. A typical commercially available dataset may include the following variables:

- Name,
- Address and postcode,
- Age,
- Sex,
- Ethnicity,
- Number of cars,
- Number of children, given in five year age-groups (e.g. one child aged 0-4, two children aged 5-9),
- Size of household,
- Tenure, house type and number of rooms,
- Occupation (high-level),
- Income (banded),
- Qualifications.

Of these, sex, age, occupation and address are relevant to births and deaths data. An intruder could match these public datasets with microdata or very detailed tables of birth and death registration data (for example, presenting the data at a detailed level of geography) on the key demographic variables and potentially identify an individual. Therefore, the level of detail available needs to be restricted in order to

make it sufficiently difficult for an intruder to do this with any degree of certainty. This gives sufficient uncertainty to any supposed identification of an individual made by use of such data.

#### (i) Birth registration data

Mother's address, parents' occupation, marital status and age at birth of child could be used with published datasets to identify individuals in birth data. As a consequence birth information for named individuals that is not in the public domain may be discovered.

#### (ii) Death registration data

Basic death registration data, including fact of death, date and age at death are considered to be in the public domain. However, sex, date of birth, occupation and address, could be used to link with published datasets, potentially allowing the disclosure of information not in the public domain, or sensitive information that is discoverable, via the register may be made more readily accessible.

## 2.2 Scenario 2 – spontaneous recognition

An intruder may recognise an individual in birth or death data, for example because of the rarity of a particular combination of variables collected at birth registration, or because of a similarly rare combination collected at death registration. The individual might be an acquaintance of the intruder, or a well-known public figure about whom much information is already in the public domain. The key variables in this scenario include:

- Name,
- Age,
- Sex,
- Marital status,
- Address,
- Occupation – job title, which may be equivalent to 4-digit occupation and industry codings.

## 2.3 Scenario 3 – Self identification

Individuals who identify themselves in births data may conclude that others are also able to identify them, and learn confidential information about them.

## 2.4 Scenario 4 – Nosy neighbour

This scenario is more relevant to microdata. The intruder has considerable information which is their personal knowledge of the person rather than information obtained from a formal data source. Typically the intruder is a neighbour, colleague, or someone else the respondent knows. Key variables in this scenario are sex, date of birth, place of birth, location, occupation.

## 2.5 Scenario 5 – Commercial gain

Certain diseases have been linked to industrial or occupational causes. Firms connected with the management of claims against former employers may seek to contact relatives of those who have died from such a disease. Care should be taken when considering requests for bespoke microdata at sub-

regional level that information provided does not facilitate obtaining address information through death registration indexes. Key variables in this scenario are:

- Date of death,
- Age,
- Location,
- Occupation,
- Specific causes of death.

### 3. Key variables

Although the intruder scenarios indicate that the risk of identifying an individual in birth or death registration tables is low, we should consider the key variables which are generally requested with such data.

#### 3.1 Guidance on preparing non-disclosive tables of births and deaths data

See Annex 1A and 1B above for guidance on how to prepare a non-disclosive table of birth or death registration data respectively. Variables included are those that are most commonly requested in relation to births and deaths registration data, in standard groupings or classifications.

Ultimately, the requirements of the user should always be borne in mind; if one variable is required at a more detailed level, then another one should be provided at less detail. For example, if detailed occupation data are required, then broader age groups could be used compared to the standard five-year ones, or geography could be national level only. For a full list of the variables that are available for birth and death registration data, see Section 4 above.

### 4. Summary

Birth and death registration data are a valuable resource for research. If the guidance above is followed, the data can be published in tabular form in a way that protects confidentiality.

Although this guidance is aimed at allowing ONS staff and others to design their own tables (and to understand why ONS tables are designed in the way they are), there will inevitably be requests for data that do not fall neatly within the parameters described. In these cases, tables must be assessed individually by the SDC team, who use a combination of statistical skills, data knowledge and experience.