



Department
of Health

Disclosure control protocol for abortion statistics

Please note: this document is out of date. It was withdrawn on 13 September 2019.

June 2015

Title: Disclosure control protocol for abortion statistics
Author: Public and International Health, Health Improvement Analytical Team (HIAT) Cost centre: 12420
Document Purpose: Consultation
Publication date: June 2015
Target audience: Users of abortion statistics
Contact details: Abortion Statistics 6 th Floor South Wellington House 133-155 Waterloo Road London SE1 8UG

You may re-use the text of this document (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence. To view this licence, visit www.nationalarchives.gov.uk/doc/open-government-licence/

© Crown copyright

Published to gov.uk, in PDF format only.

www.gov.uk/dh

Disclosure control protocol for abortion statistics

Prepared by Health Improvement Analytical Team, Department of Health

Contents

- Contents..... 4
- Summary..... 5
- Background..... 6
- Developing a Disclosure Protocol 8
- Annex A: Steps in Disclosure Control 9
 - Attribution..... 11
 - Identification..... 13

Summary

Registered medical practitioners are legally required to notify the Chief Medical Officer (CMO) of every abortion performed. The Department of Health receives these notifications on form HSA4 and undertakes the statistical processing and analysis. The Department of Health also process and publish abortion notifications on behalf of the Chief Medical Officer of Wales.

The HSA4 forms contain sensitive personal data. All releases of abortion statistics outputs are subject to the disclosure controls. The data are published and released in accordance with the Department's responsibilities for providing information under the Freedom of Information Act 2000 (FOIA), within the constraints of the Department's responsibilities for protecting personal data under the Data Protection Act.

The disclosure control protocol sets out the principles and process that the Department of Health (DH) follows to determine what abortion statistics can be released without disclosing personal information. It aims to provide users of the statistics with a clear explanation of why disclosure control is required in certain circumstances and how the Department makes as much information available as possible.

The protocol covers abortion statistics for England and Wales and replaces the previous guidance on disclosure control for abortion statistics that was released in July 2005. It aims to cover both the statistical reports that are released each year and information released in response to Parliamentary Questions (PQs) and ad hoc requests for information, including those explicitly made under the Freedom Of Information Act (FOIA). The protocol has been applied since July 2013. The annual statistical publication gives a brief account on the need for disclosure control; this document provides users with a comprehensive description of the steps taken to assess if there is a risk of releasing disclosive information.

The protocol has been developed by following the six steps set out in the GSS/GSRS Guidance on Disclosure Control for Tables Produced from Administrative Data Sources. It sets out the circumstances under which disclosure can occur through attribution and identification and presents a starting point for consideration of the controls required to prevent such disclosure.

Background

The Abortion Act 1967 permits the termination of a pregnancy by a registered medical practitioner subject to certain conditions. Under the Act and regulations made under it, the doctor taking responsibility for the procedure is required to notify the Chief Medical Officer (CMO) of the abortion within 14 days of the termination. The Department of Health provides form HSA4 for this purpose.

The Department of Health is responsible for disseminating abortion statistics for England and Wales.

The HSA4 forms contain personal data of the women who receive abortions and the medical practitioners who perform or certify the abortion. Statistics derived from the forms should only be published if they do not contain personal data. This is necessary for both practical and legal reasons. Practically, a failure to protect such personal data might discourage women from seeking treatment or discourage doctors from carrying out abortions. It could also affect the quality of the statistics if individuals and organisations refuse to submit complete or accurate information. Legally, such disclosure could breach human rights and data protection legislation, along with common law duties relating to confidentiality, require us to protect information that could identify an individual.

For many statistics, the risk of disclosing personal information is negligible and no disclosure control is necessary. Statistics would only contain “personal data” if they could lead to either:

- “Identification” – where the statistics taken together with additional information enables a third party to identify an individual; or
- “Attribution” – where the statistics taken together with additional information causes a third party to learn something new about an individual. For example, depending on how the information is broken down, a third party may be able to identify an individual and discover new details about that person that was previously unknown to them.

There are a number of benefits for individuals and organisations in accessing abortion statistics. As emphasised by the Code of Practice for Official Statistics, a balance needs to be struck in disclosure control and producers must “ensure that arrangements for confidentiality are sufficient to protect the privacy of individual information, but not so restrictive as to limit unduly the practical utility of official statistics.”

Revised guidance has recently been produced by the Government Statistical Service (GSS) and Government Social Research Service (GSRS) on Disclosure Control for Tables Produced from Administrative Data Sources such as abortion notifications¹. The document provides more

¹ <http://www.ons.gov.uk/ons/guide-method/best-practice/disclosure-control-policy-for-tables/index.html>

detail about the legal and policy framework affecting the disclosure control of official statistics. It sets out a process that producers of official statistics should follow to assess and protect against the risk of disclosure. This protocol has been developed using that document. The document is also informed by the code of practice released by the Information Commissioner's Office relating to the anonymisation of personal data². It also takes into account the High Court judgment in the abortion statistics case³. This related to a request made under the FOIA for information about the principal medical condition for abortions performed over 24 weeks gestation.

The protocol has been applied since July 2013. The annual statistical publication gives a brief account on the need for disclosure control; this consultation provides users with a comprehensive description of the steps taken to assess if there is a risk of releasing disclosive information.

The protocol covers abortion statistics for England and Wales. It aims to cover the two main types of output that are produced using information from the abortion notification forms:

- The reports that are released each year by the Department of Health containing a set of tables and charts. The report covering abortions in 2012 contained 26 main tables. Disclosure control has been applied to some of the tables in the reports from 2003.
- Information released in response to Parliamentary Questions (PQs) and ad hoc requests for information from the public made under the FOIA.

The focus of the protocol is on setting broad principles to prevent the unlawful disclosure of personal data in statistics. It does not provide comprehensive guidance for dealing with FOIA requests. FOIA requests should still be assessed on a case by case basis, taking into account the principles in this protocol, where relevant.

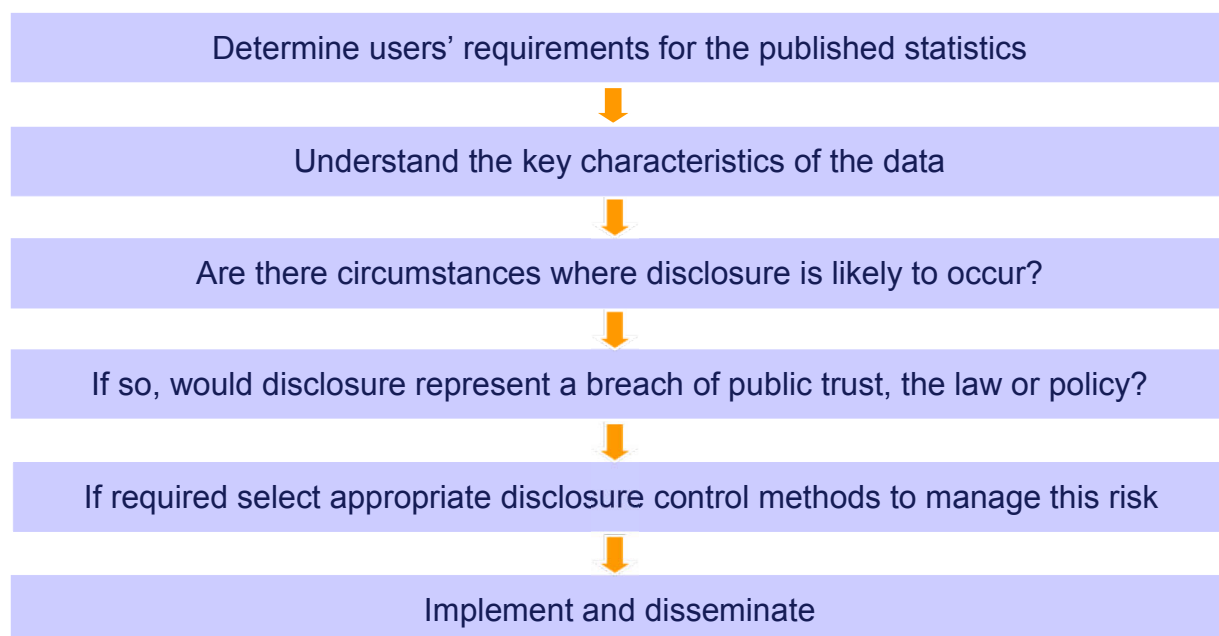
The protocol does not cover the sharing of identifiable data within the health community for administrative purposes, with ONS for the purpose of producing conception statistics, or with organisations for the purpose of bona fide research.

² http://ico.org.uk/for_organisations/data_protection/topic_guides/anonymisation

³ R (on the application of the Department of Health) v Information Commissioner [2011] EWHC 1430 (Admin): <http://www.bailii.org/ew/cases/EWHC/Admin/2011/1430.html>

Developing a Disclosure Protocol

In line with the GSS/GSRS *Guidance on Disclosure Control for Tables Produced from Administrative Data Sources*, the following six steps have been taken in developing this disclosure control protocol for abortion statistics.



Detail of the six steps is presented in Annex A.

The need for disclosure control will be assessed on a case by case basis and will only be applied where a risk of disclosing personal information is identified. Through this protocol users are advised of how disclosure control is applied and give clarity of the detail that users can expect to see in the statistical reports and ad hoc data requests.

Users are asked to review and evaluate the disclosure control protocol and give feedback on the methodology.

Annex A: Steps in Disclosure Control

Step 1: Determine users' requirements

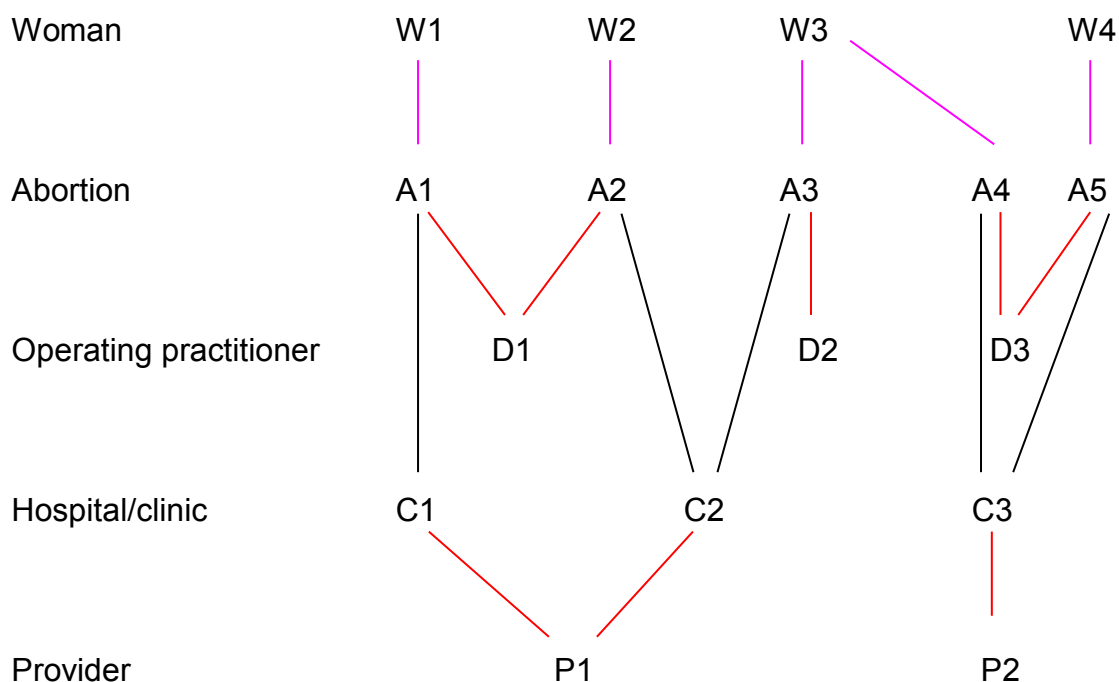
1. Providing one set of information can constrain what other information is released at a later date, because of the risk of making inferences about data subjects by combining information together. It is therefore important to understand user requirements in order to provide information that is of greatest value to users collectively.
2. There are a number of users and uses of abortion statistics, a summary of which can be found at <https://www.gov.uk/government/publications/the-use-of-abortion-statistics>.
3. The principal users are commissioners and providers of abortion and contraception services, professional bodies such as the Royal College of Obstetricians and Gynaecologists (RCOG), Parliament, special interest groups, the media, policy makers within the Department of Health, the Care Quality Commission, the Office for National Statistics (ONS) and researchers.
4. One important use is the monitoring of headline series at national level, such as the trend in the overall number and rate of abortion, the rates for different groups within the population, the breakdown by type of provider, grounds for abortion and gestation, and the frequency of repeat abortions. Such information typically does not present a risk of disclosure and no disclosure control is required.
5. A second important use is to support Parliament, special interest groups and others in checking compliance with abortion legislation. This may require, for example, detailed information at national level about the principal medical condition involved in abortions conducted under Ground E⁴. In some cases, where information is broken down two or more ways, there is a risk that a third party could discover details that were previously unknown to them (the situation referred to earlier as **attribution**).
6. A third important use is by commissioners to plan, monitor and evaluate services. In order to achieve this they will need statistics such as the number of abortions that are performed in their area by type of abortion (such as medical or surgical, gestation), characteristic of the patient (such as age and number of previous abortions) and source of funding. As above, where information is broken down two or more ways, there is a risk of attribution. In addition, in some cases, the number of women to whom such information might relate will be few enough for release of the information to pose an unacceptable risk of **identification**.
7. ONS use abortion data in the production of statistics on conceptions and publish figures for conceptions leading to abortion.
8. Abortion information attracts a lot of attention from users and the impact of disclosure from these statistics is potentially extremely high.

⁴ That there is a substantial risk that if the child were born it would suffer from such physical or mental abnormalities as to be seriously handicapped.

9. The issues of attribution and identification in relation to abortion statistics are examined in more detail in Step 3.

Step 2: Understand the key characteristics of the data

10. The HSA4 forms contain personal data of the women who receive abortions and the medical practitioners who perform or certify the abortion. The links between abortions, women, practitioners and abortion providers are outlined in the diagram below. An abortion is linked to one woman. However, some women may have more than one abortion within the period concerned, e.g. A3 and A4 are linked to W3. Each abortion is linked to a practitioner. Each practitioner typically performs more than one abortion, e.g. A1 and A2 are both linked to D1. Each abortion is also linked to one hospital or clinic, which in turn is linked to one healthcare provider (P1 and P2 in figure 2). Practitioners may operate within one hospital or clinic, but are commonly linked to more than one clinic or more than one provider, e.g. D1 works in both C1 and C2.



11. Some information on the HSA4 relates to the woman, such as age, area of residence and ethnicity. Some information relates to the abortion, such as gestation, method, grounds and complications. Other information relates to the practitioners, such as their name, address and nature of involvement.
12. Information is typically published as tables containing numbers of abortions, rather than tables giving the number of women, practitioners, clinics or providers.

Step 3: Are there circumstances where disclosure is likely to occur?

13. As stated above, the release of statistics can lead to either attribution or identification. In assessing the risk of each, thought needs to be given to both the likelihood of disclosure (related to the design of the table) and the impact of disclosure (related to the subject matter covered by the table).

Attribution

14. Attribution occurs when a third party is able to deduce something new about a data subject that is not in the public domain. For example, someone who already knows that a woman has had an abortion and who knows some of the information about her that is in a table, might be able to find her in the table and disclose further information. If a woman's partner or estranged partner knows her age, where she lives and that she had an abortion, he might be able to identify her in a table and discover the gestation period or the statutory grounds for the abortion.
15. In assessing the risk of attribution, consideration needs to be given to other information that might reasonably be available to third parties. The test isn't that identification is a remote possibility. It must be reasonably likely that a person would have other information enabling them to make the deduction.
16. 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. For example, a table where all the data subjects in a row or column fall into one or two categories could lead to attributes of those people being discovered. It is more likely to occur when tables contain multiple and/or detailed breakdowns, such as a table containing figures broken down by individual year of age and week of gestation.
17. Two examples of tables resulting in attribution are given below.

Example 1

This example might occur if a request is made for a breakdown by grounds and gestation for a particular locality, such as a Local Authority. Anybody who knew one of the 67 women in the area who had an abortion at 20 weeks or over and knew that was the case would be able to deduce that it was carried out under Ground C. One approach to deal with this would be to combine the columns for 13 to 19 weeks and 20 weeks and over.

Grounds	Gestation weeks				
	Total	3 -9	10- 12	13-19	20 & over
A (alone or with B, C, D) or F or G	0	0	0	0	0
B (alone or with C or D)	4	2	2	0	0
C (alone)	1333	934	201	131	67
D (alone or with C)	17	11	6	0	0
E (alone or with A, B C or D)	13	0	4	9	0
Total	1367	947	213	140	67

Example 2

This example might occur if a request is made for a breakdown by age and gestation for a particular locality, such as a Local Authority. The figures for under 18s present a risk of attribution. Attribution would occur if the female aged under 18 who underwent an abortion between 10 to 12 weeks gestation knows another of the females aged under 18 in that area and knows they underwent an abortion, as she would be able to deduce that the abortion occurred between 3 to 9 weeks gestation. One approach to deal with this would be to suppress the four figures for 3 to 9 weeks and 10 to 12 weeks for under 18 and 18 to 19 year olds.

Age	Gestation weeks				
	Total	3 -9	10- 12	13-19	20 & over
Under 18	10	9	1	0	0
18-19	14	8	4	2	0
20-24	42	26	11	3	2
25-29	33	17	9	2	4
30-34	24	16	6	0	2
35+	21	14	7	0	0
Total	144	90	38	7	8

18. Attribution may also occur in relation to practitioners. A third party may know some information about a practitioner that leads to identification from the table and deduction of further information, eg the third party discovers that the practitioner performs abortions on certain grounds. For example, if a third party knew that a practitioner carried out all abortions at a particular clinic during a certain period, then the release of information about the number and type of abortions carried out at that clinic could result in attribution.
19. A table resulting in attribution may be released where it is judged that deduction of the attribute would be of negligible consequence. This might occur, eg, if a third party would need to know a range of detailed information about the abortion and would only be able to deduce the broad age group to which the woman belonged. The situation might also occur in Example 1 above, where anyone knowing a woman in the area had undergone an abortion would be able to deduce it had not taken place under Ground A.
20. A zero in a table also enables attribution, as it can be deduced that all members of the population do not have that attribute. For example, the zeros in the first row of Example 1 enable people to deduce that no one underwent an abortion under Ground A (alone or with B, C, D) or F or G. Similarly, the zeros in the second row enable people to deduce that anyone undergoing an abortion under B (alone or with C or D) did so within the first twelve weeks of pregnancy. The GSS guidance states that “The risk from many zeros within tables may not be significant, but, in some cases, they may need to be protected” and that “care

should also be taken where a row or column is dominated by zeros or in some cases has a single zero”, “in particular where a marginal total is a 1 or 2.”

Identification

21. Identification occurs when a third party is able to identify an individual in a released table, either through determined effort or inadvertently, typically in combination with other information or knowledge. Third parties might include partners, family, friends, neighbours, work colleagues, pupils or staff at a school, staff at a primary care practice or pharmacy and journalists.
22. The risk of identification has to be more than remote. In assessing the risk of identification, consideration needs to be given to other information that might reasonably be available to third parties. This includes informal information conveyed verbally or via social networking sites, as well as official sources. It includes knowledge held by pupils or staff at a school about recent absences or sickness, similar knowledge held by colleagues at a workplace, knowledge held by employees or customers at a pharmacist about individuals who have bought a pregnancy testing kit, knowledge held by family, friends or partners that a pregnancy had occurred (but who might have been told that it had ended through miscarriage). The ability to obtain small pieces of informal information should not be underestimated as the summation of these could result in sufficient detail when combined with released tables to help identify an individual.

23. The risk of identification is affected by a number of factors:

- The Population at Risk.

The population at risk is the number of women in the underlying population who share the same characteristics as those in the cell of the table. For example, if the table contains figures for the number of abortions broken down by age, the population at risk associated with the figure for a particular age group would be the number of women in the underlying population of that age. For tables broken down by local authority, the population at risk would be women of child bearing age living in the local authority area.

The likelihood of identification will generally increase as the population at risk decreases. There would be a greater risk of disclosure from releasing counts of abortions for an area that only contained a few hundred women of child bearing age, compared to releasing counts at national level.

- Visibility of the variables.

The likelihood of identification will be higher if the information in the table is of a highly visible nature – that is, it is a characteristic that can be determined or estimated by observing the individual or is something that is likely to be known to friends, family or associates (eg age, ethnicity, area of residence). Other variables are likely to be of lower visibility, such as grounds and the procedure used. Others may be somewhere between the two, such as gestation and the name or location of the provider. This issue is closely linked to population at risk, because highly visible variables can be treated as resulting in smaller populations at risk.

- Degree of concentration.

The likelihood of identification will also increase if the relevant population is concentrated within an organisation or area. For example, the risk of identification is likely to be greater for figures relating to females aged under 16, because those concerned are likely to be enrolled at school and school networks would be expected to aid identification.

- Sensitivity of the information.

Information that is of a more sensitive nature is likely to increase both the motivation to, and therefore likelihood of, identification and the impact of identification. Sensitivity can be affected by the subject matter, such as young ages (under 16), older ages (50 or over), late gestation (over 24 weeks), medical condition, and procedure by gestation. Sensitivity can also be affected by rarity or uniqueness – whether the case is the only, or one of a very few, of a particular type. This guidance does not attempt to define all combinations of information that would be highly sensitive. The level of sensitivity should instead be considered on a case by case basis.

24. It follows that tables with more dimensions are likely to carry a greater risk of identification. The population at risk for each cell will decrease as the number of dimensions increases. The more dimensions there are in a table, the more information there is available to third parties to identify an individual. Tables with more dimensions will typically include cells with smaller counts.
25. In response to the above issues, the approach taken for abortion statistics will be different for:
- information relating to females aged 16 or under and information relating to broader age ranges
 - information that involves highly sensitive subject matter and information that does not
 - different sizes of population at risk
26. It will be necessary to consider information on a case by case basis. However, the following table sets out a starting point for those deliberations.

Table 2: Approach to identity disclosure control

Figures relating to females aged under 16	Sensitivity (additional to information on young females)	
	High	Other
Population at risk (female population aged 15):		
<p>Fewer than 400</p> <p>Equates to less than five average-sized secondary schools. This would apply to the smallest local authorities. Even taking into account the fact that pupils may attend schools in neighbouring authorities and some pupils move from one authority to another during a year, the risk of identification associated with this limited number of schools may necessitate the suppression of all figures.</p>	Suppress all figures	Suppress all figures
<p>More than 399 but fewer than 800</p> <p>Equates to more than five, but fewer than ten average-sized secondary schools. This would apply to medium-sized local authorities. The risk of identification associated with this limited number of schools may necessitate the suppression of figures where additional sensitive subject matter or small counts (<3) are</p>	Suppress all figures	Suppress counts under 3

involved.		
More than 799	No suppression	No suppression
Equates to more than ten average-sized secondary schools. This would apply to large local authorities. There is likely to be negligible risk of identification even where highly sensitive subject matter and very small counts.		
Figures relating to broader age ranges	Sensitivity	
	High	Other
Population at risk (female population aged 11-49):		
Fewer than 1,500	Suppress all figures	Suppress all figures
Equates to a small town with a total population of about 6,000. In more remote areas in particular, most women may use the same medical practice, pharmacy and other facilities. Risk of identification may necessitate the suppression of all figures.		
More than 1,499 but fewer than 12,500	Suppress all figures	Suppress counts under 3
Equates to smaller local authorities such as Rutland and Eden. The population will use more than one doctor's surgery but in more remote areas most residents obtaining NHS hospital treatment are likely to use a single provider. Risk of identification may necessitate suppression of all figures for highly sensitive subject matter and of very small counts (<3) for other items.		
More than 12,499 but fewer than 25,000	Suppress counts under 3	No suppression
Equates to medium-sized local authorities such as Darlington. The population will tend to be treated by more than one major hospital. Risk of identification may necessitate suppression of highly sensitive subject matter where there are very small counts (<3).		
More than 24,999	No suppression	No suppression
Equates to large local authorities. There is likely to be negligible risk of identification even where highly sensitive subject matter and very small counts.		

27. This is illustrated through three specific examples:

Example 1: Figures for the number of abortions to females aged under 16 by local authority.

This example does not involve highly sensitive variables in addition to young females. Therefore, the figures would be suppressed if either:

- There were fewer than 400 females aged 15 resident in the local authority

- There were fewer than 800 females aged 15 resident in the local authority and there were fewer than three abortions

Example 2: Figures for the number of abortions broken down by ethnicity by local authority.

This example does not involve highly sensitive variables. Figures would be suppressed if either:

- There were fewer than 1,500 females aged 11-49 within the ethnic group resident in the local authority
- There were more than 1,499 but fewer than 12,500 females aged 11-49 within the ethnic group resident in the local authority and there were fewer than three abortions

Example 3: Figures for the number of abortions carried out after 24 weeks broken down by local authority.

This example involves highly sensitive information on late gestation. Therefore, figures would be suppressed if either:

- There were fewer than 12,500 females aged 11-49 resident in the local authority
- There were more than 12,499 but fewer than 25,000 females aged 11-49 resident in the local authority and there were fewer than three abortions

28. The same approach of assessing the likelihood and impact of identification will be taken for practitioners. This would typically involve considering the population of practitioners at risk (eg the number that might perform a particular procedure), the degree of concentration (eg the number operating within a particular area or organisation), etc.
29. Care will need to be taken in publishing information for overlapping geographies, in particular Clinical Commissioning Groups (CCGs) and local authorities (LAs), to make sure that the two sets of information do not result in disclosure when combined together. The overlaps may enable a third party to derive information for a smaller geographical sliver and result in attribute or identity disclosure. In the light of responses received to the consultation conducted in summer 2013 from users, priority will be given to the release of information by CCG where such a conflict arises and where the flexibility to do so exists.

Step 4: Would disclosure represent a breach of public trust, the law or policy?

30. As stated in the introduction, human rights and data protection legislation, along with common law duties relating to confidentiality, require us to protect personal data.
31. The release of personal data relating to abortions could lead to a breach of Article 8 of the European Convention on Human Rights ("ECHR"), which provides that:
 - Everyone has the right to respect for his private and family life, his home and his correspondence.
 - There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others..

32. The release of such information may also breach the Data Protection Act 1998. The Act sets out what can be done with personal data. Additional controls apply to sensitive personal data, within which personal abortion data fall. For the release of such data to be lawful, at least one of the conditions in schedules 2 and 3 of the Act would need to be met - for example, if release was “necessary for the exercise of any functions of the Crown, a Minister of the Crown or a government department”.
33. The Abortion Act 1967 prohibits the disclosure of information provided from the abortion notification forms. Regulation 5 of the Abortion Regulations 1991 lists the exceptions to that general prohibition. However, statistics derived from that information can be published if they have been made sufficiently abstract, whereby it is possible to say that they are not the information originally furnished, but data derived from that information.
34. The FOIA creates a public right of access to information held by public authorities. However, the Act contains a variety of provisions that exempt disclosure of certain types of information. The exemptions include where:
- The requested information is “personal data” and its disclosure would contravene (i) any of the data protection principles, or (ii) section 10 of the Data Protection Act (covering the right to prevent processing likely to cause damage or distress) (section 40(2)).
 - The disclosure of information is prohibited by or under any enactment (section 44).
 - The disclosure would or would be likely to, prejudice the commercial interests of any person, (section 43), and the public interest in maintaining the exemption outweighs the public interest in disclosing the information (section 2).
35. In responding to requests for information made under FOIA, consideration will need to be given on a case by case basis whether these, or any other, exemptions apply.

Step 5: Select appropriate disclosure control methods to manage the risk

36. The method of disclosure control that is used to address the risk of attribution and identification aims to maximise the usefulness of released information to users, whilst meeting our obligations under the Freedom of Information Act and operating within resource constraints.
37. The first step is to consider how the required tables can be designed to minimise the number of cells that require disclosure control whilst maximising its usefulness. Design considerations centre around the breakdowns that are used for each of the variables (eg reducing or altering the age groups or gestation periods that are offered).
38. Where the level of table design needed to deal with all disclosure issues is such that it would significantly compromise the usefulness of information, further disclosure control is carried out in the form of cell suppression. This involves the suppression of ‘unsafe’ cells and their replacement by a special character, such as ‘..’. Such suppressions are called primary suppressions. To make sure that the primary suppressions cannot be derived by subtraction from totals, it is often necessary to select additional cells for secondary suppression. Secondary suppression may be necessary not only within a table but between tables.

39. Cell suppression has the advantages of simplicity, of not adjusting original counts in the data and of being practical to implement. The main disadvantage is that information loss can be high if more than a few suppressions are required.

Step 6: Implement and disseminate

40. This protocol is being made available on the GOV.UK web site and is already implemented. It will be reviewed in the light of its application, as further guidance on and experience of disclosure control develops, following any changes to the legal framework or case law, and in response to developments in the availability of information and information processing technology.