Health and care 2006

Review of the Dissemination of Health Statistics: Confidentiality Guidance

Working Paper 4: Glossary

Contact points

For enquiries about this publication contact:

Statistical disclosure centre on 0845 601 3034

Email: info@statistics.gsi.gov.uk

For general enquiries, contact the National Statistics Customer Contact Centre:

Tel: 0845 601 3034

Minicom: 01633 812399

Email: info@statistics.gsi.gov.uk

Fax: 01633 652747

Post: Room 1015, Government Buildings, Cardiff Road, Newport NP10 8XG

About the Office for National Statistics

The Office for National Statistics (ONS) is the government agency responsible for compiling, analysing and disseminating economic, social and demographic statistics about the United Kingdom. It also administers the statutory registration of births, marriages and deaths in England and Wales. The Director of ONS is also the National Statistician and the Registrar General for England and Wales.

A National Statistics publication

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

Glossary

This working paper provides a glossary of terms and acronyms to support the guidance for the review of the dissemination of health statistics. A more detailed glossary on statistical disclosure control is available at http://neon.vb.cbs.nl/casc/default.htm

Term	Definition
Anonymised Data	Data that have had information removed from them that would have allowed an individual to be identified. Not only are direct identifiers removed, eg names and addresses but other steps are taken to ensure that the identification of an individual is highly unlikely.
Attribute Disclosure	This occurs when data reveals information about a statistical unit that is not in the public domain.
Barnardisation	A disclosure control method where table values are randomly adjusted by −1, 0 or +1.
CT	Care Trust.
Caldicott Guardian	Caldicott Guardians are senior staff in the NHS and Social Services appointed to protect patient information.
Caldicott Report	A review of the transfer of patient identifiable data published December 1997.
Complementary Suppression	Also called secondary suppression. These are additional suppressions designed to ensure that primary suppressions cannot be revealed using column or row totals.
Confidentiality Agreements	Confidentiality agreements can be made when data are transferred and in doing so control of what is done with the data is lost by the data provider. A confidentiality agreement can be used to oblige the recipient of the data to protect the confidentiality of statistical units.
Controlled Rounding	Controlled rounding involves rounding all cells in a table to a multiple of a given base whilst ensuring that the rounded cells sum to the rounded totals.
Coterminous	Of equal extent. It is used in this instance to describe when two geographies cover exactly the same area.
Count Data	Frequency data.
Denominator	The denominator is the lower part of a fraction used to calculate a rate or ratio.
Direct Identifier	A direct identifier is a piece of information that makes it possible to identify a unit as unique. For example, a national insurance number should uniquely identify an individual.

Disclosure	Disclosure occurs when information released reveals the identity or information about a respondent not already in the public domain.
Disclosure by Differencing	Disclosure by differencing occurs when it is possible to identify or reveal the confidential particulars of a respondent by subtracting one dataset from another.
Disclosure Risk	Occurring when information is published which could allow an intruder to deduce confidential information that may indicate the identity or particulars of an individual, business or other statistical unit.
DH	Department of Health.
FoI	Freedom of Information.
Frequency Table	A frequency table is a table that presents the number of responses for a category. For example, a table that shows the number of people registered with a dentist.
GMC	General Medical Council.
GOR	Government Office Region.
GP	General Practitioner.
GRO	General Register Office.
HES	Health Episode Statistics.
Identity Disclosure	Identity disclosure occurs when data can be associated with a particular statistical unit, thus revealing their identity.
LAD	Local Authority District.
LHB	Local Health Board.
Marginal Total	A marginal total is a total within a table, ie a row or column total.
Microdata	Microdata are the individual response data from statistical units.
NHS	National Health Service.
Non-coterminous	Of unequal extent. It is used in this instance to describe when two geographies do not cover exactly the same area.
Non-response	A non-response in a survey is the failure to obtain the information requested from a statistical unit.
Numerator	The numerator is the upper part of a fraction used to calculate a rate or ratio.
NWCS	NHS-Wide Clearing Service .
PCO	Primary Care Organisation, includes PCTs and CTs.
PCT	Primary Care Trust.
Perturbation	Perturbation is the introduction of error into data in such a way that the risk of disclosure is reduced whilst trying to retain the utility of the data. For example, the rounding of values in a table.

PHBF	Public Health Birth File.
PHMF	Public Health Mortality File.
PHO	Public Health Observatory.
PIAG	Patient Information Advisory Group.
Population	Population in this instance is an entire group with a set of particular characteristics. For example, the population of GP's in the UK is all GP's in the UK.
Pre-tabulation	Pre-tabulation refers to actions carried out on data before it is put into a tabular format.
Primary Suppression	Suppression of a cell or cells in a table identified as potentially disclosive.
The Protocol	Part of the National Statistics Code of Practice: The Protocol on Data Access and Confidentiality. See www.statistics.gov.uk/about_ns/cop/default. asp
Pseudo-Anonymised Data	Pseudo-anonymised data are data that have had direct identifiers removed and a random code applied so that data can be associated with the code but statistical units cannot be directly identified.
Random Rounding	Cell values are rounded, but instead of using standard rounding conventions a random decision is made as to whether they will rounded up or down. The rounding mechanism can be set up to produce unbiased rounding results.
Record Swapping	Record swapping involves swapping pairs of records in the microdata that are partially matched to alter the geographic locations attached to the records, but leaves all other aspects unchanged.
Rounding	Rounding in this instance is a method for protecting unsafe cells in a table. The basic idea is to round each count up or down to the nearest integer multiple of a rounding base. Common choices for a base are 3 or 5.
Safe cell/table	A cell within a table is described as safe when the cell cannot lead to a disclosure risk. A table is described as safe when each cell within that table is a safe cell.
S&CAG	Security & Confidentiality Advisory Group.
Secondary Suppression	See complementary suppressions
Self-Identification	Self-identification occurs when a statistical unit can identify themselves in a statistical output.
SHA	Strategic Health Authority.
Statistical Disclosure Control	Assessing the risk of disclosure from a potential release and taking measures, if appropriate, to lower that risk by modifying the data, changing the design of the statistic, or a combination of these.
Statistical Unit	A statistical unit is an individual unit, such as an individual, a household, or a business.

Suppression	Suppression is a method for protecting unsafe cells in a table where cells are not published but replaced by a special character such as '..' or 'X' to indicate a suppressed value.
Tabular Data	Tabular data are data presented in a table.
Unsafe Cells/Tables	A cell within a table is described as unsafe when the cell leads to a disclosure risk. A table is described as unsafe when one or more cells within that table are unsafe cells.