Protocol for the Sharing of Population Level QChild Data on the Ear Health of Aboriginal and Torres Strait Islander Children

Developed by the Healthy Hearing Unit and the Deadly Ears Program

July 2017

Queensland Government

Protocol for the Sharing of Population Level QChild Data on the Ear Health of Aboriginal and Torres Strait Islander Children

Published by the State of Queensland (Queensland Health), July 2017.

For more information contact:

Healthy Hearing, Children's Health Queensland Hospital and Health Service, 10 Chapel Street, Nundah 4012. Phone 07 3310 6222.

An electronic version of this document is available at https://www.childrens.health.qld.gov.au/chq/our-services/community-health-services/deadly-ears/resources/

Table of contents

Appendix

Appendix 1: NHISSC Guidelines for the Disclosure of Secondary Use Health Information for Statistical Reporting, Research and Analysis (2015)

Background

Context

In 2016 the Queensland Government launched the policy framework Deadly Kids Deadly Futures (DKDF) and the associated DKDF 2016-17 Action Plan. The Action Plan included six initiatives to improve service delivery, one of which was the commitment to, *"Share population-level newborn hearing screening data with primary healthcare providers to improve the planning and delivery of services for children at higher risk of acquiring hearing loss."*

To enable this, Children's Health Queensland Hospital and Health Service has developed this Protocol to support the secure and confidential sharing of information from the QChild database.

Objective

The objective is to ensure that primary healthcare providers have accurate and robust population health information to inform their planning of health services for children – especially those Indigenous children identified by newborn hearing screening as having a higher risk of acquiring hearing loss

Scope

This Protocol relates solely to the sharing of population-level data at the community level or above, disaggregated geographically to areas no smaller than Statistical Area 2. It does not encompass the sharing of QChild data which are identifiable and relate to specific children and/or their families.

The data which may be included in the release comprises:

a) demographic information:
    a. year of birth or multiple years of birth
    b. community (or suburb or postcode as it appears in QChild) and
    c. Indigenous status.

b) data associated with the newborn hearing screening process:
    a. % children referred (on newborn hearing screen)
    b. % children with risk factors (ie family history, bacterial meningitis, prolonged ventilation, severe asphyxia, syndromes associated with hearing loss, craniofacial anomalies, hyperbilirubinemia, infection, professional concern)
    c. % of children who are regarded as "lost contact" (comprising incomplete newborn hearing screening data, appointment series at audiology, or DNA audiology)
    d. % of children with a confirmed hearing loss (ie where children have completed the audiology process and are confirmed to have hearing loss – either sensorineural, permanent conductive, mixed or ANSD)
    e. % of children found with a transient conductive hearing.

Disclosure Risks

There are two key types of disclosure risk which this Protocol seeks to avoid. The first is identity disclosure, which is when an individual's identity can be identified from the information disclosed, either directly or by cross-referring the disclosed information with other information available to the user. The risk of this increases significantly in the case of information relating to small communities, but the NSW Centre for Epidemiology and Evidence (CEE) states that the most common cause of identify disclosure is the release (either deliberately or by mistake) of a person's name, address, photograph or geocoded location (CEE, 2015).

The second type of disclosure risk is attribute disclosure. This occurs when information about an individual's attributes is released that could result in the person themselves being identified. This is, of course, the key concern from a privacy perspective, but the CEE regards attribute disclosure as a cause for concern for three other reasons:

"Information about an individual has been released (even if we cannot identify the individual); the public interest arguments (in relation to the epidemiological utility of the data) associated with publishing any information about an individual (or a small number of individuals) are questionable; finally, the statistical inferences drawn from such a small sample are likely to be very compromised."[1] (CEE, 2015, p 7).

Both of the above disclosure risks may be managed by anonymising the data that is to be released. To ensure that this is done effectively, the sharing of QChild ear health data will be managed in accordance with the protocols set out in the National Health Information Standards and Statistics Committee's (NHISSC) *Guidelinesfor the Disclosure of Secondary Use Health Information for Statistical Reporting, Research and Analysis (2015*). For convenience, the full NHISSC Guidelines are attached as Appendix 1.[2]

Informed Consent

A third type of disclosure risk identified in the literature is community disclosure. However, in the case of the current initiative the data to be shared may often be specific to particular communities so community disclosure is an inherent part of the process. The clearest example of this may be where localised data is requested by a provider to plan or improve the resourcing of health services for one or more identified communities. In such cases the data will still be anonymised such that identity or attribute disclosure does not occur, but the ethical question becomes one of informed community consent to the disclosure.

The parents or guardians of children whose ear health has been assessed and/or treated by the Deadly Ears Program, and whose health information is held on QChild, have already consented to that information being shared with a range of other agencies (see Text Box 1). However, where the release of information may result in the disclosure of one or more specific communities, consent should be sought from the community or communities concerned before the release occurs.

[1] CEE (2015), p 7.

[2] The National Health Information Standards and Statistics Committee (NHISSC) was formed in 2008 as a sub-committee of the National Health Information and Performance Principal Committee (NHIPPC). NHIPPC is one of several principal committees that report to the Australian Health Ministers' Advisory Council (AHMAC). AHMAC provides support to the Standing Council on Health under arrangements for the Council of Australian Governments (COAG). The NHISSC assumed roles previously undertaken by the Statistical Information Management Committee (SIMC), the Health Data Standards Committee (HDSC), and some of the roles of the National Health Performance Committee (NHPC)

Text Box 1: Existing Consent

All parents or guardians of children whose ear health has been assessed and/or treated by the Deadly Ears Program, and whose health information is held on QChild have been asked to consent to:

- the Program providing and or obtaining the following information relevant to their assessment or treatment:
  - discharge summaries
  - operation/ anaesthetic reports
  - medical notes
  - results/tests
  - letters
  - special instructions (where relevant),
- from and/or to:
  - their child's school and related school-based support services
  - Australia Hearing
  - the local Primary Health Network
  - other community health providers
  - where relevant, their child's childcare and support teams that work in the childcare.

Such consent should be documented and obtained freely and transparently, allowing reasonable timeframes for community discussion. It should include (AHMRC, 1998):

- an explanation of why the information is being collected and how it will be used;
- assurances that additional consent will be sought if there are any changes in the specified
- use of the information;
- identification of who will, or is likely to, have access to the information; and
- advice regarding the right to withdraw consent.

In some situations it may be problematic identifying the person or persons in any given community with the mandate to provide consent on behalf of that community. Some Community Controlled Health organisations have formed local health committees to provide community consultation and liaison (such as the Health Action Teams created by the Apunipima Cape York Health Council). In other communities agencies may need to seek consent from community elders, the Mayor, or some other group or community representatives.

Data Sharing Principles

Given the disclosure risks and issues discussed above, any sharing of QChild data on the ear health of Aboriginal and Torres Strait Islander children should follow the principles set out below. These have been adapted from the QAIHC Data Governance Protocols (2014) and the National Aboriginal and Torres Strait Islander Health Data Principles (as endorsed by AHMAC, 2006).

1. The sharing of population health information about Aboriginal and Torres Strait Islander children must support improved health outcomes for Aboriginal and Torres Strait Islander peoples and/or better planning and delivery of health services.
2. The privacy and confidentiality of Aboriginal and Torres Strait Islander people will be protected in accordance with any relevant legislation and privacy codes.

3. All QChild data to be shared will be anonymised so that the identity of individual children or families is not apparent, and cannot reasonably be ascertained from the dataset either on its own or in combination with any other information to which the user may reasonably be considered to have access. As discussed above, this anonymisation will be in accordance with the principles and techniques set out in the National Health Information Standards and Statistics Committee's *Guidelines for the Disclosure of Secondary Use Health Information for Statistical Reporting, Research and Analysis (2015*) (see Appendix 1).
4. Consistent with principle 3, information released will be aggregated to the maximum level consistent with achieving the user's epidemiological or service-planning requirements.
5. The specifics of each data request will be reviewed against the risk of identity disclosure due to low incidence. Data requests will be refused if the confidentiality of individuals cannot be protected.
6. When information is to be shared on specific Aboriginal and/or Torres Strait Islander communities, the user of that information must first obtain free and informed consent from the key representatives of those communities, except where mandatory reporting or legislative provisions apply. Otherwise, where there is a proposal to share a community's information without that community's consent, the user must clearly demonstrate both that the information-sharing will advance the interests of Aboriginal and Torres Strait Islander people and that it is impractical and infeasible to obtain specific consent.

   Note: community consent is not required if the information shared relates solely to demographics (eg population numbers, age profiles etc). Consent is only required if the data describes health status or socio-economic risk factors (or enables those risk factors to be inferred).

7. When information is to be shared on specific Aboriginal and/or Torres Strait Islander communities, access for those communities should be provided to the results of subsequent data analysis. This should include access to products such as data reports, research results and plans for health services to be provided to the communities in question.

8. Requests by third parties for access to, and sharing of, QChild information will be determined with reference to the principles and terms of this Protocol. Third parties making such requests will be provided with a copy of this Protocol, and if access is granted, must formally acknowledge and agree in writing to adhere to the principles.

9. Applications for population level data held on QChild will be considered on a case by case basis, and Healthy Hearing reserves the right to place limitations on access to and use of QChild information.

10. Data will be released no more than quarterly over the course of any financial or calendar year.

Applications for Data

Applications may be made by emailing the custodian of QChild data: the Director of the Healthy Hearing program (Rachael.Beswick@health.qld.gov.au).

Appendix 1: NHISSC Guidelines for the Disclosure of Secondary Use Health Information for Statistical Reporting, Research and Analysis (2015)

National Health Information Standards and Statistics Committee

Endorsed 4 September 2015

Guidelines for the Disclosure of Secondary Use Health Information for Statistical Reporting, Research and Analysis 2015

National Health Information Standards and Statistics Committee

Endorsed 4 September 2015

Introduction

Development

These guidelines were developed by the National Health Information Standards and Statistics Committee (NHISSC) in 2015 to set out principles and techniques regarding the disclosure of ***secondary use health information*** for statistical reporting, research and analysis. They replace the Guidelines for the Use and Disclosure of Health Data for Statistical Purposes developed by the Statistical Information Management Committee (SIMC) in 2007. The name of the Guidelines has been changed to better reflect their content. Terms that are defined in Appendix A: Glossary are ***italicised*** and ***bolded*** the first time they appear in the body of the Guidelines

Scope

These guidelines are appropriate for health information held by National Health Information Agreement (NHIA) signatories such as National Minimum Datasets (NMDSs) and Dataset Specifications (DSSs) disclosed for purposes of statistical reporting, research and analysis. It applies to ***unit record*** and ***aggregate data*** and to individuals and health services and includes the output of data linkage.

Out of scope for these guidelines are data sharing arrangements between governments including the specification and provision of NMDSs & DSSs; specified agreements or data outside the NHIA scope; data security and data reliability. Also excluded are data linkage protocols and data for payment or linkage purposes, which by necessity contain ***re-identifiable*** or ***identifiable*** information.

Context

This document provides general guidance to assist in the management of risks regarding the identification of individual patients/clients and health service providers, where legislative provisions do not provide sufficient detail about the release of data (Although the case studies included in this document use hospital admitted patient data to illustrate the principles and techniques, those principles and techniques are intended to be broadly applicable to other health information).

These guidelines are intended to be used in conjunction with other more specific agreements or arrangements, including both existing agreements between parties to the NHIA and agreements with regard to the subsequent release to third parties of data owned by jurisdictions which are parties to the NHIA. For example, the National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Research Involving Humans (2007, updated March 2014, p70) states that institutions may choose to exempt from ethical review research that (a) is negligible risk research; and (b) involves the use of existing collections of data or records that contain only ***non-identifiable data*** about human beings.

Although health privacy legislation and policies vary between Australian jurisdictions, their common purpose is to govern the collection, use and disclosure of ***personal information*** about the health of, or health services provided to, individuals whose identity is apparent or can be reasonably ascertained.

This means that, before providing a health dataset (in either unit record or aggregate form) to other agencies or individuals, the providing agency must satisfy itself that:

- if the dataset is identifiable or re-identifiable data, it will only be disclosed where there is patient consent or for purposes for which the use or disclosure of personal information is permitted by its policies and legislation, or
- the dataset is non-identifiable data, in the sense that the identity of individual patients is not apparent, and cannot reasonably be ascertained from the dataset either on its own or in combination with any other information to which the user may reasonably be considered to have access.

The focus of these guidelines is on the second requirement. The Australian Institute of Health and Welfare's (AIHW) *Policy on reporting to manage confidentiality and reliability, Nov. 2013* contends that a cell in a table is identifiable if, as well as being able to identify the entity, other details are revealed (i.e. 'attribute disclosure'). Similarly, NHIA Clause 26f states that aggregate data suppression rules should be case specific and only enacted where there is a risk that an individual could have information disclosed that was previously unknown to the recipients of the data. NHISSC supports the AIHW and NHIA approaches but also recognises that minimising the risk of identification is the safest way to prevent attribute disclosure.

The Health Statistics NSW paper: *Privacy Issues and the Reporting of Small Numbers* also refers to community disclosure where a data release has the potential to disclose information about small communities. It is NHISSC's view that managing the risk of identification and attribute disclosure for individuals will generally also ensure that small communities are similarly protected. However additional action may be necessary, e.g. the Northern Territory department does not always release information for certain Statistical Area 2s because those SA2s, in combination with Indigenous Status, would immediately identify specific Aboriginal communities.

Purpose

The aim of these guidelines is to assist data custodians to manage the risk of identification of individual patients/clients (the simple fact that a person attended a specific health service may raise privacy issues for the person concerned) and the disclosure of previously unknown information about that patient (e.g. their diagnosis or procedures) by defining a number of underpinning principles and recommending some of the available techniques.

Principles

There are a number of principles that apply to disclosure of health information for statistical reporting, research and analysis:

1. Information is a strategic national asset and agencies should provide as much public access as possible whilst minimising the risk of personal information disclosure.
2. Data disclosure must comply with legislation and interagency agreements, including but not limited to information privacy principles, secrecy, consent, commercial-in-confidence, freedom of information and Commonwealth data integration principles.
3. All ***data custodians*** should have, adopt, or develop their own guidelines which conform to legislation specifically applicable to them and utilise the principles and techniques outlined in these guidelines.
4. Original data custodians retain primary responsibility for their unit record data at all times. The original data custodian is defined as the primary source of that particular health information. For example, where states and territories provide national minimum datasets to the AIHW, Dept of Health, IHPA, NHPA, etc. the state and territory is the original data custodian and all other parties provided with that dataset are custodians of the data.
5. Where other data custodians seek to publish or disseminate health information for research or analysis purposes, they need to reach agreement with all of the original data custodians. The AIHW's policy guideline 6 says that where data suppliers require the application of additional suppression rules to an AIHW release in order to manage confidentiality – the supplier's proposed suppression rules should be applied.
6. In regard to unit record data for research and analysis purposes, custodians should aim to provide data that satisfies the purpose of the request whilst seeking to prevent individuals or an organisation's commercial interests being identified or having information disclosed about them that was previously unknown to the recipients. Further, only those data items essential to the user's purpose should be released. It is not good practice to provide more information than is needed for a specific project (data minimisation principle).
7. Aggregate data suppression rules should be case specific and only enacted where: (i) there is a risk that an individual could be identified possibly leading to information about them being disclosed that was previously unknown to any possible recipient of the data; or (ii) there is a risk of exposing an organisation's commercial operations. Custodians must protect the 'commercial in confidence' nature of private hospitals in both aggregate and unit record data releases, unless they have received explicit approval from all relevant private organisations. The provision of aggregate information about the performance of individual public hospitals is the prerogative of State/Territory health authorities.

Techniques

The ABS provides a comprehensive range of techniques, including alternative approaches for managing the risk of personal identification that could be applied to health information by data custodians. The ABS' Confidentiality Series can be accessed at:

http://www.nss.gov.au/nss/home.NSF/pages/Confidentiality+Information+Sheets

NHISSC has endorsed some specific techniques. In all cases professional judgement is required to assess the privacy implications of the request, to utilise one or more of the following techniques, and to assess whether an individual or an organisation's commercial operations can be identified and have previously unknown information about them disclosed:

1. In seeking to minimise the risk of identification and attribute disclosure in unit record data releases, custodians should anonymise data by:

    - removing and/or modifying personal identifiers such as a person's name, address, date of birth and unit record number. For example, if it is an essential requirement of the data request to know that multiple episodes relate to the same person in the same hospital, then the unit record numbers provided should be encrypted (see Case Study 1);
    - not providing other specific dates unless absolutely necessary (see Case Study 2), otherwise provide month and year of admission/separation, etc.;
    - aggregating variables wherever possible: e.g. provide 5 year age groups rather than date of birth; a metropolitan/rural indicator or statistical level area 2 (SA2) rather than postcode and locality of residence, diagnosis related group instead of individual diagnosis and procedure codes, etc.

      This technique, which can also be applied to aggregate data, is based on the data minimisation principle and addresses the concern that if the 'denominator' population (i.e. the population in the community) is too small it can provide a risk of individual identification and information disclosure.

      Custodians should ensure that the pool of people who could potentially have contributed to unit record data or to a cell in aggregate data is as large as possible while still enabling the user to do their job. This approach could be assisted by a numerical test, i.e. unit record data would not be provided for sub-groups where their estimated population is less than a value set by the custodian. (See Case Study 3).

- Custodians should require external users of unit record data to sign 'conditions of release' covering specific confidentiality requirements such as the purpose for which data may be used, requirements for the secure storage and retention of data, restrictions on the publication of data, the provision of data to a third party and any attempt to re-identify individuals. The Conditions should indicate the applicable laws covering release and the penalties that apply for a breach of the conditions.

2. To maintain the anonymity of individual private hospitals or private hospital owners in aggregate tables derived from hospital morbidity data, cells should be suppressed if:
   - there are fewer than three (3) separately owned private hospitals; or
   - there are three or more separately owned private hospitals and one private hospital owner contributed more than 85% of the total separations; or
   - there are three or more separately owned private hospitals and two private hospital owners contributed more than 90% of the total separations. (see Case Study 4)
3. Small cell suppression in aggregate data is commonly used in national statistical reporting and by many of the individual states/territories. It is an easy test to apply and detects cells with potential identification problems possibly leading to the release of previously unknown information, e.g. age, diagnosis or procedure (see Case Study 5).

   Small cells (e.g. containing values between 1 and 4) may be avoided by aggregating variables, e.g. age group ranges 65-74, 75-84, 85+ are replaced with 65+, data from small areas or communities are aggregated over a number of years, etc. If this is not possible, then the small cells may be suppressed.
4. Cells in aggregate data where the value of the cell is the same as a row/column total should be suppressed if it is considered that it could lead to disclosure of an additional attribute.
5. The application of the private hospital, small cell, and cell = row/column total techniques may require the suppression or amalgamation of several cells in a table, possibly including some with values of zero or greater than 4, in order that a cell not be derivable by subtraction. In these circumstances, it is advisable that the compiler of the table choose a method of confidentialisation that maintains the column and row totals and results in the loss of the least amount of useful information (see Case Study 6).
6. As the original data custodians, if individual states/territories want to apply particular rules for a national statistical publication or other specific aggregate data releases involving their data (e.g. some requiring no restrictions, some requiring small cell suppressions with different threshold values e.g. <3, <5, <10, etc.) the most conservative option will apply in line with the principles espoused in this document.

Case Studies

These Case Studies are provided to illustrate the techniques

Unit Record Data

Case Study 1: A contractor working for a Government Department on a service planning project requests unit record data, including amongst other data items:

- Campus
- Patient ID
- Admission date
- Separation date
- Date of Birth
- All diagnoses codes
- All procedure codes
- Postcode
- Locality

After discussion with the contractor, who confirms that knowing the campus, and that multiple episodes relate to the same patient in that campus, are essential to the work, a dataset is negotiated that seeks to prevent individual identification, but still enables the contractor to complete the required job. Any data items that are not essential to the user's purpose are also not provided. The dataset released after conditions of release were signed included:

- Campus
- Encrypted Patient ID
- Admission Month and Year
- Length of Stay (capped at 30+ days)
- Age in 5 year groups
- Diagnosis Related Group
- Statistical Local Area of Usual Residence

Case Study 2: A University researcher investigating the link between air pollution and admission for respiratory conditions requests unit record data, including amongst other data items:

- Campus
- Patient ID
- Admission date
- Separation date
- Date of Birth
- All diagnoses codes
- Postcode
- Locality

After discussion with the researcher, who confirms that knowing the admission date and a more detailed usual residence than SLA is essential to the work, a dataset is negotiated that prevents individual and health service identification, but still enables them to complete the required job. Any data items that are not essential to the user's purpose are also not provided. The dataset released after conditions of release were signed included:

- Encrypted Campus code
- Encrypted Patient ID
- Admission Date
- Length of Stay (capped at 30+ days)
- Age in 5 year groups
- Diagnosis codes (Only those episodes with respiratory condition as principal or additional diagnosis)
- Locality of Usual Residence

Case Study 3: A consultant working for a community action group and looking at self-sufficiency issues for a specific hospital and its catchment (as defined by a list of 8 postcodes), requests unit record data, including amongst other data items:

- Campus
- Admission date
- Separation date
- Age
- Sex
- Diagnosis Related Group
- Postcode
- Discharge Status
- Insurance Status
- Weighted Separations

Examination of underlying population data grouped by five year age groups, sex and postcode reveals that a significant number of cells are less than 1000 (a pre-set minimum value chosen by this particular custodian). Indeed for some of the postcodes within the prescribed catchment, the total for all ages for males and females is under 1000.

In discussion with the consultant the need for an identified campus was established and agreed and it was confirmed that a patient identifier of any kind was not required. Also agreed was separation month and year as opposed to admission and separation dates. After discussion about compacting either the geographic area, the age dimensions or both, it was agreed that the eight postcodes would be grouped together as the "catchment" and that the area outside the catchment would be Statistical Area Level 2. Additionally the oldest age category would be "70+".

The dataset released after conditions of release were signed included:

- Campus
- Separation month and Year
- 5 Year Age Groups ending with 70+
- Sex
- Diagnosis Related Group
- Catchment as a whole/Non-catchment by SA2
- Discharge Status
- Insurance Status
- Weighted Separations

Aggregate Data

Case Study 4: Private Hospital data

An external requestor asks for the following aggregate table where Parameter X could be a sub-group of hospitals, defined age group, geographical location, etc. or a combination thereof.

For the purposes of illustration only a subset of diagnosis are shown.

Separations by Parameter X by Principal Diagnoses, Private Hospitals	
Principal Diagnosis	**Seps**
G47 Sleep Disorders	491
I20 Angina Pectoris	159
J35 Chronic Diseases of Tonsils and Adenoids	188
K40 Inguinal Hernia	196
K80 Cholelithiasis	175
M17 Gonarthrosis (Arthrosis of knee)	253
M75 Shoulder Lesions	234
O80 Single Spontaneous Delivery	179
O82 Single Delivery by Caesarian Section	185
R07 Pain in Throat and Chest	148
Other diagnoses with less than 10 separations	1124
Total	**3332**

Step 1 - Suppress numbers where there are less than 3 separately owned private hospitals or where 1 private hospital owner has more than 85% of separations or two have more than 90%. In this example, there were only two private hospitals providing maternity services.

Separations by Parameter X by Principal Diagnoses, Private Hospitals	
Principal Diagnosis	**Seps**
G47 Sleep Disorders	491
I20 Angina Pectoris	159
J35 Chronic Diseases of Tonsils and Adenoids	188
K40 Inguinal Hernia	196
K80 Cholelithiasis	175
M17 Gonarthrosis (Arthrosis of knee)	253
M75 Shoulder Lesions	234
O80 Single Spontaneous Delivery	np
O82 Single Delivery by Caesarian Section	np
R07 Pain in Throat and Chest	148
Other diagnoses with less than 10 separations	1124
Total	**3332**

np Commercial-in-confidence

Step 2 - Further action to prevent deduction. From the table above you can deduce that the "np"s total 364. Whilst it is not possible to accurately split between O80 and

O82, this number provides a good indication of the level of obstetric activity. By including the np numbers in with "Other" this information is protected as follows:

Separations by Parameter X by Principal Diagnoses, Private Hospitals	
Principal Diagnosis	**Seps**
G47 Sleep Disorders	491
I20 Angina Pectoris	159
J35 Chronic Diseases of Tonsils and Adenoids	188
K40 Inguinal Hernia	196
K80 Cholelithiasis	175
M17 Gonarthrosis (Arthrosis of knee)	253
M75 Shoulder Lesions	234
O80 Single Spontaneous Delivery	np
O82 Single Delivery by Caesarian Section	np
R07 Pain in Throat and Chest	148
Other - commercial-in-confidence data and diagnoses w ith less than 10 separations	1488
Total	**3332**

np Commercial-in-confidence

Case Study 5: Attribute Disclosure

An external requestor specifies the following table which contains a small cell (a value between 1 and 4). In order to identify a person in that or any other cell, all of the following detail would need to be already known: that they were admitted to hospital in that time period, their indigenous status and their specific principal diagnosis. No additional information is provided to that already known; i.e. there is possible identification but no attribute disclosure, so the custodian may decide to release the table without alteration.

Hospital Separations for Cardio-Vascular Disease by sub-type and indigenous status

Principal Diagnosis	Indigenous status	
	Non-indigenous	Indigenous
Acute rheumatic fever I00–I02	22	1
Chronic rheumatic heart diseases I05–I09	474	13
Hypertensive diseases I10–I15	1,696	5
Ischaemic heart diseases I20–I25	40,204	212
Pulmonary heart disease and diseases of pulmonary circulation I26–I28	2,548	18
Other forms of heart disease I30–I52	35,337	73
Cerebrovascular diseases I60–I69	11,250	42
Diseases of arteries, arterioles and capillaries I70–I79	7,350	22
Diseases of veins, lymphatic vessels and lymph nodes, not elsewhere classified I80-I89	22,021	49
Other and unspecified disorders of the circulatory system I95–I99	2,491	12
Total	123,393	447

A second table specified contains small cells. In order to identify a person in those or any other cell, all of the following would need to be already known: that they were admitted to hospital in that financial year, their indigenous status and their specific principal diagnosis. It is possible that this would enable the person's age group to be determined; i.e. there is possible identification and attribute disclosure. Techniques outlined in this document and in Case Study 6 would need to be applied to prevent attribute disclosure.

Hospital Separations for Hypertensive diseases (I10–I15)
by age group and indigenous status

	Indigenous status		
Age Group	**Non-indigenous**	**Indigenous**	**Total**
00-24	58		58
25-29	16		16
30-34	38	1	39
35-39	49		49
40-44	59	1	60
45-49	82		82
50-54	123	1	124
55-59	124	1	125
60-64	128		128
65-69	160		160
70-74	171	1	172
75-79	200		200
80-84	243		243
85+	245		245
Total	1,696	5	1,701

Case Study 6: Small Cell Suppression

An external requestor asks for the following aggregate table where Parameter X could be a group of hospitals, defined age group, geographical location, etc or a combination thereof.

Separations for Parameter X by principal diagnosis of mental health related conditions and sex			
Principal diagnosis	**Males**	**Females**	**Total**
Mental & behavioural disorders due to psychoactive substance use (F10–F19)	183	130	313
Schizophrenia, schizotypal and delusional disorders (F20–F29)	169	204	373
Mood disorders (F30–F39)	136	146	282
Neurotic, stress-related disorders (F40–F49)	59	114	173
Disorders of adult personality and behaviour (F60–F69)	14	34	48
Behavioural and emotional disorders (F90–F98)	6	9	15
Organic, including symptomatic, mental disorders (F00–F09)	6	7	13
Behavioural syndromes assoc. with physiological disturbances (F50–F59)	1	16	17
Unspecified mental disorder (F99)	1	7	8
Mental retardation (F70–F79)	1	2	3
Disorders of psych. Development (F80–F89)	1	4	5
Other[(a)]	37	39	76
Total	**614**	**712**	**1326**

(a) list of codes included in other

It is determined that action is required because the small cells may lead to identification and attribute disclosure

Step 1 - Suppress all numbers between 1 and 4

Principal diagnosis	**Males**	**Females**	**Total**
Mental & behavioural disorders due to psychoactive substance use (F10–F19)	183	130	313
Schizophrenia, schizotypal and delusional disorders (F20–F29)	169	204	373
Mood disorders (F30–F39)	136	146	282
Neurotic, stress-related disorders (F40–F49)	59	114	173
Disorders of adult personality and behaviour (F60–F69)	14	34	48
Behavioural and emotional disorders (F90–F98)	6	9	15
Organic, including symptomatic, mental disorders (F00–F09)	6	7	13
Behavioural syndromes assoc. with physiological disturbances (F50–F59)	np	16	17
Unspecified mental disorder (F99)	np	7	8
Mental retardation (F70–F79)	np	np	np
Disorders of psych. Development (F80–F89)	np	np	5
Other[(a)]	37	39	76
Total	**614**	**712**	**1326**

(a) list of codes included in other
np Numbers between 1 and 4 are not published.

Step 2 - Further suppression to prevent deduction

- It is very easy to deduce the value of each of "np"
- The male column adds to 610 therefore each male “np” must be 1, and the female numbers can be determined by deduction.
- The total column adds to 1323 without the "np", therefore the "np" must be 3
- To prevent identification by deduction you would also suppress the female numbers for F50-F59 and F99; plus the male and female numbers for F00-F09*; and the Total for F80-F89. *F00-F09 was chosen as it had the lowest numbers
- Note how Column totals and all Row totals bar two are preserved.

Principal diagnosis	Males	Females	Total
Mental & behavioural disorders due to psychoactive substance use (F10–F19)	183	130	313
Schizophrenia, schizotypal and delusional disorders (F20–F29)	169	204	373
Mood disorders (F30–F39)	136	146	282
Neurotic, stress-related disorders (F40–F49)	59	114	173
Disorders of adult personality and behaviour (F60–F69)	14	34	48
Behavioural and emotional disorders (F90–F98)	6	9	15
Organic, including symptomatic, mental disorders (F00–F09)	np	np	13
Behavioural syndromes assoc. with physiological disturbances (F50–F59)	np	np	17
Unspecified mental disorder (F99)	np	np	8
Mental retardation (F70–F79)	np	np	np
Disorders of psych. Development (F80–F89)	np	np	np
Other[(a)]	37	39	76
Total	**614**	**712**	**1326**

(a) list of codes included in other
np Not published to prevent disclosure of numbers between 1 and 4.

Step 2 Alternative - Expanding "Other"
You could choose to expand "other" to include suppressed male and female numbers.

Principal diagnosis	Males	Females	Total
Mental & behavioural disorders due to psychoactive substance use (F10–F19)	183	130	313
Schizophrenia, schizotypal and delusional disorders (F20–F29)	169	204	373
Mood disorders (F30–F39)	136	146	282
Neurotic, stress-related disorders (F40–F49)	59	114	173
Disorders of adult personality and behaviour (F60–F69)	14	34	48
Behavioural and emotional disorders (F90–F98)	6	9	15
Other[(a)]	47	75	122
Total	**614**	**712**	**1326**

(a) original list of codes plus F00-F09, F50-F59, F70-F89 and F99

Appendix A: Glossary

Aggregate data

Aggregate data are produced by grouping information into categories and aggregating values within these categories. For example, a count of the number of people of a particular age (obtained from the question 'In what year were you born?'). Aggregate data is typically presented in tables. Aggregate data is also referred to as tabular data or macrodata. (ABS)

Anonymisation

The process of removing identifiers and/or other data items from a dataset with the intention that the dataset content changes from identifiable to re-identifiable or non-identifiable data or from re-identifiable data to non-identifiable data.

Data Custodian

The organisation or agency which is responsible for the collection, use and disclosure of information in a dataset. Data custodians have an obligation to keep the confidential information they are entrusted with secret. (ABS)

Health Information

Includes any data required to inform health research/health status. This includes health data sets as well as data sets linked to data where the primary issue is health (e.g. socio economic status, education, occupational health and safety)
(Health and Health related data - NHMRC's National Statement on Ethical Conduct in Human Research (2007, updated March 2014))

Identifiable Data

Data from which the identity of a specific individual can be reasonably ascertained (NHMRC's National Statement on Ethical Conduct in Human Research (2007, updated March 2014))

Non-identifiable data

Data that have never been labelled with individual identifiers or from which identifiers have been permanently removed, and by means of which no specific individual can be identified.
(NHMRC's National Statement on Ethical Conduct in Human Research (2007, updated March 2014))

Personal information

Information or an opinion (including information or an opinion forming part of a database), whether true or not, and whether recorded in a material form or not, about an individual whose identity is apparent, or can be reasonably ascertained, from the information or opinion (whether directly from the information or from the information when read in combination with other information held by or available to the organisation).

Re-identifiable Data

Data from which identifiers have been removed and replaced by a code, but it remains possible to re-identify a specific individual by, for example, using the code or linking to a different data set.
(NHMRC's National Statement on Ethical Conduct in Human Research (2007, updated March 2014))

It is also possible to re-identify a specific individual via a combination of variables.

Secondary Use

The use of data for any authorised purpose other than the use for which the data was originally collected (primary purpose)

Unit Record Data (Also known as Patient level data or Microdata)

Each record represents observations for an individual or organisation. Unit record data may contain individual responses to questions on a survey questionnaire or administrative form. For example, answers given to the question 'In what year were you born?'. (Microdata - ABS)

References

AHMRC (1998). Aboriginal Health Information Guidelines. Aboriginal Health and Medical Research Council of NSW. Retrieved Nov 2016 from http://www.ahmrc.org.au/media/resources/ethics/ethics-background-resources/281-nsw-health-information-guidelines.html

Centre for Epidemiology and Evidence (2015). HealthStats NSW: Privacy issues and the reporting of small numbers. Sydney: NSW Ministry of Health.

National Advisory Group on Aboriginal and Torres Strait Islander Health Information and Data. National Aboriginal and Torres Strait Islander Health Data Principles (endorsed by AHMAC 2006). Retrieved Nov 2016 from www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=6442472790

NHISSC (2015) Guidelines for the Disclosure of Secondary Use Health Information for Statistical Reporting, Research and Analysis. Retrieved Nov 2016 from http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=60129553404

NSW Health (1998) NSW Aboriginal Health Information Guidelines. Retrieved Nov 2016 from

Queensland Aboriginal and Islander Health Council (2014) Data Governance Protocols. Retrieved Nov 2016 from http://www.qaihc.com.au/resources/publications/