16

Technical report

Pathways of Care Longitudinal Study: Outcomes of Children and Young People in Out-of-Home Care

Guidelines for Publishing Results with Small Sample Sizes

Pathways of Care Longitudinal Study: Outcomes of Children and Young People in Out-of-Home Care in NSW

Technical Report No. 16

Guidelines for Publishing Results with Small Sample Sizes

Published by
New South Wales Department of Communities and Justice (DCJ)
Insights Analysis and Research
320 Liverpool Road
Ashfield NSW 2131

Phone + 61 2 9716 2222

March 2020

ISBN: 978-0-6485156-4-7

Recommended citation
NSW Department of Communities and Justice. (2020). Guidelines for Publishing Results with Small Sample Sizes. Pathways of Care Longitudinal Study: Outcomes of Children and Young People in Out-of-Home Care. Technical Report Number 16. Sydney. NSW Department of Communities and Justice.

Prepared by
Albert Zhou, NSW Department of Communities and Justice.

Pathways of Care Longitudinal Study Clearinghouse
All study publications including research reports, technical reports and evidence to action notes can be found on the study webpage www.facs.nsw.gov.au/resources/research/pathways-of-care

Study design by NSW Department of Communities and Justice (Insights, Analysis and Research); Australian Institute of Family Studies; Professor Judy Cashmore, University of Sydney; Professor Paul Delfabbro, University of Adelaide; Professor Ilan Katz, University of NSW; Dr Fred Wulczyn, Chapin Hall Center for Children University of Chicago.

Study data collection by I-view Social Research.

Advisors Expert advice and support has been provided by the CREATE Foundation; Aboriginal Child, Family and Community Care State Secretariat (AbSec); My Forever Family NSW; and program areas.

Acknowledgements We wish to extend our thanks to all the children, young people and caregivers who participated in interviews; childcare teachers, school teachers and caseworkers who participated in on-line surveys; and the data custodians in the relevant NSW and Commonwealth government departments. Ms Sammy Verma grew up in care and played a key role in the production of the study video for children and stakeholders. Ms Billy Black also grew up in care and designed the study artwork. Ms Sammy Verma and Mr Samuel Eyeson-Annan both did the voiceover for the audio computer-assisted self-interview (ACASI) for the child/young person interview.

Ethics approval by The University of NSW Human Research Ethics Committee (approval number HC10335 & HC16542); Aboriginal Health and Medical Research Council of NSW Ethics Committee (approval number 766/10); NSW Department of Education and Communities State Education Research Approval Process (SERAP, approval number 2012250); NSW Population & Health Services Research Ethics Committee (Ref: HREC/14/CIPHS/74 Cancer Institute NSW: 2014/12/570).

NSW
GOVERNMENT
Communities
& Justice

Contents

List of Tables

Preface

The Pathways of Care Longitudinal Study (POCLS) is funded and managed by the New South Wales Department of Communities and Justice (DCJ). It is the first large-scale prospective longitudinal study of children and young people in out-of-home care (OOHC) in Australia. Information on safety, permanency and wellbeing is being collected from various sources. The child developmental domains of interest are physical health, socio-emotional wellbeing and cognitive/learning ability.

The overall aim of this study is to collect detailed information about the life course development of children who enter OOHC for the first time and the factors that influence their development. The POCLS objectives are to:

- Describe the characteristics, child protection history, development and wellbeing of children and young people at the time they enter OOHC for the first time.
- Describe the services, interventions and pathways for children and young people in OOHC, post restoration, post adoption and on leaving care at 18 years.
- Describe children's and young people's experiences while growing up in OOHC, post restoration, post adoption and on leaving care at 18 years.
- Understand the factors that influence the outcomes for children and young people who grow up in OOHC, are restored home, are adopted or leave care at 18 years.
- Inform policy and practice to strengthen the OOHC service system in NSW to improve the outcomes for children and young people in OOHC.

The POCLS is the first study to link data on children's child protection backgrounds, OOHC placements, health, education and offending held by multiple government agencies; and match it to first-hand accounts from children, caregivers, caseworkers and teachers. The POCLS database will allow researchers to track children's trajectories and experiences from birth.

The population cohort is a census of all children and young people who entered OOHC for the first time in NSW over the 18 month period between May 2010 and October 2011 (n=4,126). A subset of those children and young people who went on to receive final Children's Court care and protection orders by April 2013 (2,828) were eligible to participate in the study. For more information about the study please visit the study webpage www.facs.nsw.gov.au/resources/research/pathways-of-care.

The POCLS acknowledges and honours Aboriginal people as our First Peoples of NSW and is committed to working with the DCJ Aboriginal Outcomes team to ensure that Aboriginal children, young people, families and communities are supported and empowered to improve their life outcomes. The POCLS data asset will be used to improve how services and supports are designed and delivered in partnership with Aboriginal people and communities.

DCJ recognises the importance of Indigenous Data Sovereignty (IDS) and Indigenous Data Governance (IDG) in the design, collection, analysis, dissemination and management of all data related to Aboriginal Australians. The POCLS is subject to ethics approval, including from the Aboriginal Health & Medical Research Council of NSW. DCJ is currently in the process of scoping the development of IDS and IDG principles that will apply to future Aboriginal data creation, development, stewardship, analysis, dissemination and infrastructure. The DCJ will continue to collaborate with Aboriginal Peoples and will apply the DCJ research governance principles once developed.

1 The POCLS guidelines for reporting results with small sample sizes

This section sets up the guidelines for reporting findings from analysis of the POCLS survey and DCJ administrative data involving small sample sizes. The purpose of these guidelines is to help users understand the issue of confidentiality in the context of the POCLS data and how it can be managed thus reducing the risk of disclosure (i.e., identification of children, carers or caseworkers) to a minimum. These guidelines apply to results presented in any form (e.g., tables and graphs), to both internal and external reports and publications, and outputs (including syntax) taken out of SURE.

In the process of developing these guidelines, the POCLS team has reviewed two major international and national longitudinal studies in the areas of child welfare and/or child developmental outcomes (NSCAW and LSAC). We haven’t been able to find any published guidelines for reporting small numbers on their websites or elsewhere. There are reasons why small sample sizes may not be a concern for these two studies. Firstly, both NSCAW and LSAC have much larger sample sizes than POCLS does. For example, LSAC has around n=5,000 children in each of its two cohorts while the sample size for POCLS is around n=1,300 only; Secondly, both studies are usually analysed with sample weights. The use of sample weights add a layer of ambiguity to the data that makes re-identification of individuals extremely difficult. In a personal email communication, the Data Manger at the Australian Institute of Family Studies (AIFS) advised not to report on cells with less than five cases when reporting on results from LSAC.

Several factors that are of particular relevance to the POCLS have been considered:

- The POCLS is a sample survey of children who entered OOHC between May 2010 and October 2011. Some children might be still in OOHC. Others might have left OOHC
- The POCLS data are stored in SURE, which is a secured data environment with remote access
- Researchers who access the POCLS data in SURE are appropriately authorised
- The POCLS data in SURE have already been confidentialised to an extent with identifying information about individuals having been removed. For example, individual names and the suburb they live in have been removed. Information about the height and weight of each child has also been removed and replaced with the Body Mass Index (BMI)

- No unit record data are allowed to be taken out of SURE
- Outputs/results are subjected to review and approval before they are taken out of SURE. This includes review of previous outputs taken out of SURE to ensure that when outputs are combined there is no risk of disclosure.

Given the above factors, it is judged that the risk of an individual in the POCLS data being identified is small. However, presentation/reporting of small counts may still pose some disclosure risk.

There is no universal consensus on how low a count is too low and would therefore pose a disclosure risk. Data custodians decide on a threshold value that is deemed appropriate for them. This is often a trade-off between maintaining the usefulness of the data (setting a high threshold may decrease the usefulness of the data) and maintaining the confidentiality of individual children, carers and caseworkers involved. According to the National Statistical Service, the frequency threshold values that are commonly used are 3, 5 and 10.

It is recommended the following guidelines to be adopted for reporting on the POCLS findings:

- A frequency threshold value of 5 is adopted for reporting on analyses involving geographic location and/or demographic variables (i.e., age, sex, Aboriginality, primary cultural identity and language spoken). This rule applies to frequency tables involving a single geographic/ demographic variable or cross-tabulation tables involving at least two geographic/demographic variables (e.g., age by sex, Aboriginality by district). It means cells with counts less than 5 are deemed to pose an unacceptable disclosure risk and need to be protected, and cells with counts greater than or equal to 5 are deemed as acceptable. In some cases, zero cells or 100% cells can also pose a disclosure risk. This has to be looked at in context and will be assessed on a case by case basis. Apart from issues with low frequency counts, there are also situations where a few units/ individuals may contribute a large percentage of a cell's total value although it is believed that this is less likely to happen to the POCLS data
- For analyses involving response variables (i.e., variables other than the geographic and demographic variables) or their interaction with geographic/demographic variables, the above rule can be relaxed. For example, the rule would not be automatically required when a response variable is cross-tabulated with a demographic variable. Researchers should consult with DCJ in instances where they would like to relax this rule. The end goal is that there should be no re-identification possible

- Use weighted data for analysis and reporting purposes where available unless there are specific reasons not to use weights in the analysis
- Report on percentages and column or row totals (rather than the actual frequency value for each cell) in the tables and/or graphs if possible. Although this is not a data confidentialisation technique, presenting data this way helps keeping readers' attention on overall distribution of the data and the relative proportions of each sub-group.

If a cell is deemed as unacceptable (i.e., cell size <5), the researcher(s) (and ultimately DCJ) have a responsibility to ensure the data is protected. There are a number of data treatment techniques that one can employ to address this including data reduction and data modification techniques. Data reduction involves combining variable categories or cell suppression while data modification involves rounding and perturbation. While there are pros and cons with each method, data reduction methods are briefly described below as they are considered more relevant for the POCLS. See ABS (2017) for details about other methods.

It is advisable that researcher(s) run frequency tables on the key variables of interest and check the counts before starting an analysis. A small sample size not only poses a disclosure risk, but may also lack the statistical power to detect significant differences. If the sample sizes are small, the researchers may need to re-think their research questions and/or strategies to dealing with small counts.

For example, research into young people (aged 12-17 years) in OOHC involves a sample of 124 young people at wave 1. While this count appears reasonable, a break-down by age and other variables such as district, may result in small counts, and pose a disclosure risk. A further break-down by another variable, such as Aboriginal status would make the counts even smaller.

A hypothetical example[1] of the low frequency count and its treatment are illustrated in Tables 1-8 below. Applying the threshold value of 5 to Table 1, Far West, South Eastern Sydney and Sydney are deemed as unacceptable.

---

[1] All data used in this document are made up for illustration only.

Table 1: Example of low frequency counts in a frequency table

District	Young people (12-17 years) at interview
Hunter/Central Coast	10
Far West	2
Hunter New England	42
Illawarra Shoalhaven	7
Mid North Coast	6
Murrumbidgee	12
Nepean Blue Mountains	10
Northern NSW	12
Northern Sydney	5
South Eastern Sydney	1
South Western Sydney	16
Southern NSW	0
Sydney	2
Western NSW	18
Western Sydney	7
Total	150

Combining categories is one way to increase cell sizes. Using the district classification scheme as shown in Table 2, we can combine small areas to larger ones. Table 3 combines the fifteen districts (geographic level 3) into the seven larger areas (geographic level 2) while Table 4 makes a distinction between metro vs. non-metro areas only (geographic level 1).

Table 2: Geographic classification

Geographic level 1	Geographic level 2	Geographic level 3
Metro	South Eastern, Northern & Sydney Districts	South Eastern Sydney
		Northern Sydney
		Sydney
	South Western Sydney Districts	South Western Sydney
	Western Sydney & Nepean Blue Mountains Districts	Nepean Blue Mountains
		Western Sydney
Regional/remote	Illawarra Shoalhaven & Southern NSW Districts	Illawarra Shoalhaven
		Southern NSW
	Mid North Coast & Northern NSW Districts	Mid North Coast
		Northern NSW
	Murrumbidgee, Far West & Western Districts	Far West
		Murrumbidgee
		Western NSW
	Hunter New England & Central Coast Districts	Hunter New England
		Central Coast

Table 3: Example of treatment – combing categories (1)

District	Young people (12-17 years) at interview
South Eastern, Northern & Sydney	8
South Western Sydney Districts	16
Western Sydney & Nepean Blue Mountains Districts	17
Illawarra Shoalhaven & Southern NSW Districts	7
Mid North Coast & Northern NSW Districts	18
Murrumbidgee, Far West & Western NSW Districts	32
Hunter New England & Central Coast Districts	52
Total	150

Table 4: Example of treatment – combing categories (2)

District	Young people (12-17 years) at interview
Metro	41
Regional/ remote	109
Total	150

Another option for data reduction is to suppress small counts in a table as per the frequency threshold value. That is, to suppress any cells with a count less than 3 and to replace with 'not publishable' or 'np'. This is often referred to as primary suppression. See the example in Table 5.

Table 5: Example of primary suppression of cells

District	Young people (12-17 years) at interview
Hunter/Central Coast	10
Far West	np
Hunter New England	42
Illawarra Shoalhaven	7
Mid North Coast	6
Murrumbidgee	12
Nepean Blue Mountains	10
Northern NSW	12
Northern Sydney	5
South Eastern Sydney	np
South Western Sydney	16
Southern NSW	0
Sydney	np
Western NSW	18
Western Sydney	7
Total	150

The suppression protocol used means that secondary or consequential suppression may also be required to suppress cells with a value of 5 or greater to prevent the primary suppressed cell from being calculated. Table 7 provides an example of the use of both primary and consequential suppression to treat the hypothetical data including small cells shown in Table 6. Other cell suppression combinations, such as the suppression of the totals could also be used to protect the data, as shown in Table 8. However, care needs to be taken to ensure that the totals are not available in other tables and/or elsewhere in the report. The same approach and technique are applicable to data involving either Aboriginal or non-Aboriginal children or both.

These guidelines for reporting are designed to apply to the POCLS survey data (including carer, child, caseworker and teacher survey data) and DCJ administrative data only. For the record linkage data, other data custodians may have different threshold values/requirements that will need to be complied with when reporting their data. See the Appendix for a summary of different publishing rules across various data custodians.

As pointed out by ABS, the process of implementing data treatment and/or checking data for possible disclosure is not a trivial process and may take considerable time/resources. We encourage researchers to discuss/consult with the DCJ POCLS research team if they are not sure about the most appropriate technique to use and how to best apply it. In some cases, the researcher may seek an exemption to the rules, for example, if the researcher is reporting results from a qualitative study (e.g., case studies). In that case, the researcher should provide a rationale for the exemption together with the results/report to be reviewed by the DCJ POCLS research team. Protecting the POCLS data confidentiality is a shared responsibility between researchers and DCJ.

Table 6: Example of low counts in a cross-tabulation table

Do you think Study Child will want to continue living here after [He/She] leaves care	**Child’s cultural background**			
	Aboriginal	CALD	Other Australian	Total
Definitely will	11	1	6	18
Probably will	5	7	6	18
Unsure	2	0	10	12
Probably won’t	1	2	5	8
Definitely won’t	6	0	14	20
Total	25	10	41	76

Table 7: Example of primary and secondary suppression of cells

Do you think Study Child will want to continue living here after [He/She] leaves care	**Child’s cultural background**			
	Aboriginal	CALD	Other Australian	Total
Definitely will	11	np	np	18
Probably will	5	7	6	18
Unsure	np	np	10	12
Probably won’t	np	np	np	8
Definitely won’t	6	0	14	20
Total	25	10	41	76

Table 8: Example of suppression of totals

Do you think Study Child will want to continue living here after [He/She] leaves care	**Child’s cultural background**			
	Aboriginal	CALD	Other Australian	Total
Definitely will	11	np	np	>17
Probably will	5	7	6	18
Unsure	np	0	10	>10
Probably won’t	np	np	np	>5
Definitely won’t	6	0	14	20
Total	>22	>7	41	>70

Appendix 1: Publishing Rules for Linked Administrative Data Sets

Custodians of the various administrative data sets linked to the POCLS data have applied different publishing rules depending on the agency. These are summarised below.

Each agency website publishes their own comprehensive and detailed guidelines, which should be consulted if in any doubt.

1.1. AEDC

(See AEDC guidelines page 32-40 for detail https://www.aedc.gov.au/resources/detail/aedc-data-guidelines-2017).

Confidentiality rules

- No data should be published which discloses adverse information about all members of a group

- 90% vulnerability rule – if more than 90% of a population group are considered developmentally vulnerable in any domain or subdomain the number and percentage of children should be reported as '90 per cent and over'. The converse is also the case such that 'Where the number or proportion of children not developmentally vulnerable is also reported (e.g. developmentally on track and developmentally at risk) the number and proportion of children should be grouped to less than 10 per cent to ensure the vulnerable value cannot be derived' (see page 34)

- Rule of three – lowest number that can be published is 4

- Two teachers rule – data can only be published for a community where data was reported by two or more teachers.

Interpretability rules

- 15 valid AEDC children rule - the total population represented in an AEDC indicator should be at last 15 with valid AEDC data – if less than 15 this should not be reported

- 80% coverage rule – if publishing AEDC results for a group, 80% of the surveyed population must have valid instruments (excluding special needs children)

- Estimated resident population (ERP) rule. If AEDC data collection is (a) less than 60% of the ABS resident four year old population or (b) between 60 and 80% of the ABS resident four year old population special proviso clauses re interpreting with caution need to be added in as a footnote (see page 35 for respective wording).

Publishing rules

- Data Users to send a draft embargoed copy to the DMA (Data Management Agency) at AEDC and allow 10 business days from date of receipt to provide feedback
- All queries must be answered and then sent again to the DMA until agreed revised version (each time 10 days)
- The two points above apply to a government agencies, or material produced under contract for a government agency (as has to go to the AEDC National Committee), except 20 business days should be allowed for instead of 10
- All publications to have following disclaimer:

  “This [paper/project/research/report etc.] uses data from the Australian Early Development Census (AEDC). The AEDC is funded by the Australian Government Department of Education and Training. The findings and views reported are those of the author and should not be attributed to the Department *or the Australian Government.”*

Further suggestions from AEDC (not rules)–

- Suppressed cells should be highlighted in yellow
- Make sure that there is no differencing – e.g. where two tables can be combined in some way to produce a third table which contravenes publishing rules
- If primary statistics have been suppressed averages should be suppressed
- Check it is not necessary to apply secondary suppression (see page 39 for a geographic example and page 40 for a categorical variable example).

1.2. NAPLAN

- Cannot publish numbers less than 5
- All reporting should be related to groups and not individual children.

1.3. BOCSAR

- BOCSAR themselves publish results down to 1 (e.g. one murder in 2016 in Campbelltown). Nevertheless the rule is that there should be no re-identification possible which means there would have to be at least three or four as readers can attach quite detailed data to an individual

- Acknowledge BOCSAR as the source of the data in any publication or presentation using the reoffending (ROD) data. The acknowledgement should take the following form: Source: NSW Bureau of Crime Statistics and Research: <insert BOCSAR reference code and year>
- Provide BOCSAR with an electronic copy of any and every research report, article or conference presentation at least a fortnight prior to publication/presentation.

1.4. DCJ data

- No findings that have an adverse impact on individuals and/or all members of a group
- Numbers less than five need to be suppressed.

1.5. Health data

There are detailed guidelines to the publication of health data developed by the National Health Standards and Statistical Committee (NHISSC) in 2015. These provide detailed examples and can be found appended at: https://www.childrens.health.qld.gov.au/wp-content/uploads/PDF/healthy-hearing/QChild-data-sharing-protocol.pdf

In NSW the Health publication rules are in line with those of other external data custodians. The disclosure risks and privacy concerns associated with specific attributes can be found in Privacy issues and the reporting of small numbers in HealthStats NSW (2015) which are cited and/or quoted here as follows (see page 11): (www.health.nsw.gov.au/hsnsw/Publications/privacy-small-numbers.pdf)

- Small area: small areas often report small numbers of people (lower denominators) that increase the risk of identity disclosure. They are also likely to contain small communities which may present additional community disclosure risks
- Aboriginality: There are additional policy obligations associated with reporting Aboriginal health information at both individual and community disclosure
- Infectious diseases: There is a social stigma attached to many infectious diseases
- Analysis type: Count data carries a high risk whereas calculated values such as life expectancy have small disclosure risks

- Small numbers: ‘As the number reported (n) increases the risk of attribute disclosure decreases. As a general rule, reporting cell counts n<5 is not recommended (in order to preserve the privacy of individuals) although this depends on the size of the denominator

- Census or survey: reports based on stratified random surveys have inherently less risk to privacy than those based on a census. In addition, the type of data reported may have an impact on privacy risks. For instance individual/case based (e.g. births or deaths) is associated with a higher risk of identification and services based (e.g. hospitalisations) carries a lower risk, as one hospitalisation does not necessarily reflect on individual as one individual can be admitted multiple times.

There is also a section on disclosure control methods which has been quoted here (page 12-14 of the document):

Disclosure control methods

There has been significant research on statistical disclosure and there exists a toolbox of strategies from the field of statistical disclosure control which have varying costs, risks and benefits. Two types of information are commonly presented in statistical tables: count (frequency) data or magnitude data. These two types of data present different challenges for statistical disclosure control and will be considered separately.

Rules for count data

Count data present the number of individuals who meet certain categorical criteria. These criteria are usually specified by the intersection of a row and column, and perhaps page, which define a particular cell in a table. In some cases, count data are reported as relative frequencies or percentages of a category or combination of categories.

The US Federal Committee on Statistical Methodology (2005) identified two general approaches for disclosure control of count data: special rules and the threshold rule.

- Special rules are simply agency-specific conditions used to constrain the resolution of a public report. For example, ages will never be presented at a finer scale than a 5-year age class. Tables are then constructed based upon these rules

- The threshold rule is more generic and is usually stated that table cells with less than k individuals (sometimes referred to as sensitive cells) must not be reported. Common values of k are 3, 5 or 10. If a table cell reports a number less than k, then there are a number of general strategies for managing the disclosure risk associated with that table including: table restructure; cell suppression; and changing table values.

These are discussed below:

Table restructure

The first, and simplest, is to restructure the table to increase the number of people reported in that cell by combining rows or columns until the threshold k is reached. Sometimes this will require human judgment about what groups should be sensibly combined and the process may be difficult to automate for some types of tables, particularly if there is not a natural hierarchy or order in the classification used.

Cell suppression

The second commonly used approach is cell suppression: replacing the contents of a table cell with a number below a threshold value with an identifiable character such as an asterisk (*).

Cell suppression is a commonly used approach but has associated complications. For example, if marginal totals are provided (i.e. total counts for a row or column) then complementary suppression of another table cell may be required. Otherwise readers will be able to calculate the count in the suppressed cell using subtraction. In large tables, the patterns of complementary suppression required may become exceedingly complex and specialised algorithms in linear programming need to be used to determine the effective patterns of cell suppression. This problem can become even less tractable if, for example, the marginal totals have been presented in another table in the same or a different report. Elliot (2001) discusses the complexities associated with these so-called table linkages and the complementary concept of table differences.

It should be noted that, while the cell suppression rule is easy to implement, it may be unnecessary in relation to disclosure risk if the small number is drawn from a large population (see the “Denominator rule” below). Conversely, a larger number of cases (above the specified threshold) for a sensitive issue drawn from a smaller population may represent an unacceptable risk of disclosure.

Changing table values

The third approach, changing table values, usually involves some type of rounding or randomisation of cells values. Rounding (say to the nearest five people) requires care as it usually results in marginal totals that are incorrect, which can erode public confidence in the table. More sophisticated algorithms such as controlled rounding or controlled tabular adjustment can be applied to maintain accurate marginal totals but these require the implementation of more complex algorithms to fully implement.

Randomisation involves replacing the contents of a table cell with a number which has been randomly selected below a threshold value. This approach is commonly used by the Australian Bureau of Statistics. For example, when reporting suicide deaths in children (5–17 years), cell counts with small values have been randomly assigned to protect the confidentiality of individuals. In this example, readers are informed in the table key that randomisation has been applied.

One aspect of some health indicators which does not seem to have gained much attention is that in some cases the health records in a cell will be based upon the same person. For example, if a person is re-admitted to hospital within a year there may be two hospital separations recorded, but they are for the same person. Unless a unique identifier is available for a person (such as will be available for a linked dataset), and a query which processes only unique cases is used, then many tables will overestimate the number of individuals being reported from administrative datasets. In such cases, more conservative threshold rules may need to be used.

Denominator rule

As discussed above, in relation to small areas or subpopulations, the size of the population denominator may be more relevant to disclosure than the size of the numerator. The Australian Statistical Information Management Committee (SIMC) guidelines (SIMC 2007) suggest that statistical results involving small numbers can be presented if the population from which they are drawn is more than 1000 people. The reasoning behind this is that, even for rare conditions, in populations or communities larger than 1000 people, the likelihood of identifying an individual would be very small.

It should be noted that an emphasis within the SIMC guidelines is on the denominator, rather than the numerator, associated with statistics. The SIMC argues that risks to privacy are more commonly associated with size and nature of the population that statistics are drawn from rather than the actual number of people reported.

Rules for magnitude data

Magnitude data are based upon a simple statistic of a numerical characteristic of individuals represented by a table cell. This simple statistic is usually the sum (such as the total income), but it could also be the average or a percentile. Magnitude data require additional considerations for statistical disclosure control beyond that required for count data (Federal Committee on Statistical Methodology 2005).

Tables of magnitude data are usually tested against the (n,k) dominance rule, which is a generalisation of the threshold rule. The (n,k) dominance rule checks if n unique entities contribute more than k% of the value of that cell. If a cell breaks the (n,k) dominance rule, then the same types of approaches as described above are used to modify the table until all cells are considered safe from disclosure.

HealthStats NSW indicators which are based on magnitude data are generally either calculated from large Australian Bureau of Statistics surveys or a census of births and deaths, therefore none of these tables represents any significant risk of disclosure.’

Reference

Australian Bureau of Statistics (ABS), 2017, 1160.0 - ABS Confidentiality Series, Aug 2017, http://www.abs.gov.au/ausstats/abs@.nsf/mf/1160.0, accessed on 6 March 2018.

NSW
GOVERNMENT