RCPCH Audits

RCPCH Guidance for Managing Data Disclosure Risks in Public Reports and Dashboards – Examples

Example I: Documentation of the risk of disclosure management in the National Neonatal Audit Programme (NNAP) clinical audit.

Table 1 displays an assessment of the potential risks of disclosure associated with the indicator 'Parents on ward rounds,' as reported by the NNAP, delivered by the RCPCH. The table also outlines the agreed actions for disclosure control.

Table 1: Documentation of the risk of disclosure management in the NNAP clinical audit.

	Could an intruder discover any protected information from these figures?						
NNAP measure	Can any individual be identified from the data or table, with any degree of certainty?	If yes, is any new information revealed about them?	Is any information about any other living person connected with them?	Steps already taken to reduce deductive disclosure	Risk to utility of data of applying further statistical disclosure control	Further action required following consultation with Methodology and Dataset Group.	Application of masking to unit level data.
Parents on ward rounds	If the proportion for the measure is 0% or 100% this would indicate something about any parents with babies at that unit in that year, or where the denominator for main measure is less than 2.	Information on whether a parent was present on a ward round or not.	Yes, parents are implicated in the measure.	No gestational age breakdowns provided.	tests on 2021 data indicated only x% units affected.	measure reviewed with MDG on xx/xx/xx and it was agreed to adopt the recommended suppression	Mask units where denominator for main measure is <2, or where sub-measure proportion is <1% or >99%

Example II: Small numbers suppression and secondary suppression within a table

Table 2 shows a hypothetical group of children living with Type 1 diabetes and using the most advanced technology for glucose monitoring, by different ethnic groups and NHS region.[1]

The table shows two examples of small numbers and secondary suppression.

(I) The number of children from Asian ethnicity was below the small number threshold set by the audit, in the North-West region.

Decision-making process:

[1] All the numbers included in the table were randomly generated.

a. Should the denominator 'N' be suppressed, and the percentages be displayed? Or should only the percentage related to the small number be suppressed?

Considering that the same denominator will probably be used in several tables throughout the report, the best alternative is to suppress the percentages.

b. After the suppression of just one cell, can the small number be deduced by differencing with respect to the total by column (i.e. the total number of Asian children reported)?

Yes, if only one cell is suppressed within a column, then there is a high risk of disclosure by differencing. Secondary suppression will be necessary.

c. How to select another cell for secondary suppression when no other region meets the small number criteria?

A common approach is to suppress the next smallest number, which, in this case, is the South-East with 9 children.

d. Can the suppressed percentages be deduced by differencing with respect to the totals by row (i.e. the total number of children by region)?

Not in this case, as there are additional small numbers within the same row, specifically in the 'Not stated' column. No additional suppression will be needed.

(II) The number of children from 'Mixed' ethnicity is below the small numbers threshold set by the audit in two regions, the East of England and the Midlands. A third region, London, shows a small number of children from 'Other' ethnicity.

a. The choice remains to suppress the percentages rather than the denominators, for the same previously stated reason.

b. The column 'Mixed' doesn't need secondary suppression as more than one region were below the small number threshold.

c. The column 'Other' needs secondary suppression as only one region was below the small number threshold.

d. Considering that the columns 'Mixed' and 'Other' contain the smallest figures among the three regions requiring masking, an alternative method is to combine these categories into one, showing the combined percentage as shown in Table 3.

Table 2: Number of children by ethnic group and region.

Region	Total	Ethnic group					
		White	Asian	Black	Mixed	Other	Not stated
East of England	63	9	20	17	1	7	9
London	56	13	15	15	5	2	6
Midlands	60	10	16	19	2	5	8
North East and Yorkshire	64	8	23	12	3	10	8
North West	61	11	2	35	5	7	1
South East	51	12	9	14	5	9	2
South West	59	8	10	20	6	12	3
England	414	71	95	132	27	52	37

Table 3: Percentage of children by ethnic group and region.

Region/country	Total	Ethnic group					
		White	Asian	Black	Mixed	Other	Not stated
East of England	63	14.3%	31.7%	27.0%	12.7% †		14.3%
London	56	23.2%	26.8%	26.8%	12.5% †		10.7%
Midlands	60	16.7%	26.7%	31.7%	11.7% †		13.3%
North East and Yorkshire	64	12.5%	35.9%	18.8%	4.7%	15.6%	12.5%
North West	61	18.0%	*	57.4%	8.2%	11.5%	*
South East	51	23.5%	*	27.5%	9.8%	17.6%	*
South West	59	13.6%	16.9%	33.9%	10.2%	20.3%	5.1%
England	414	17.1%	22.9%	31.9%	6.5%	12.6%	8.9%
* number <3							
† Results merged to mask number <3							

Example III: using rolling proportions to mask time series data

In cases where it is desirable to release granular time series data, one method of masking small numbers can be to use rolling totals.

This is done by summing the numerator and denominator of the time period that will be displayed with a set of preceding time periods. Typically, rolling periods cover a year, so each month or quarter of the year is summed with the preceding 3 (quarter) or 11 (month) periods to arrive at an annual total, displayed for each month.

As these periods cover a longer time range their numbers are larger and therefore less likely to present small numbers, while still preserving the ability to present up to date data on a much shorter time period. Masking can be further enhanced by converting this data to a proportion, which makes it harder (though not impossible) to identify the underlying numerator and denominator.

The table below shows how small numbers can be masked by using this process:

Table 4: Example of masking using rolling proportions.

Data Type	Q1 2020	Q2 2020	Q3 2020	Q4 2020	Q1 2021	Q2 2021	Q3 2021	Q4 2021
Raw data	4/16	1/18	6/12	9/32	3/15	2/14	5/14	12/28
Rolling total	n/a	n/a	n/a	20/78	19/77	20/73	19/75	22/71
Rolling %	n/a	n/a	n/a	25.6%	24.7%	27.4%	25.3%	31.0%

Although this process is effective at removing small numbers from most time series data, it does come with some drawbacks:

- The initial time periods that do not have complete data need to be hidden for two reasons:
    1) Leaving the initial totals would allow a reader to uncover the remainder of the results by differencing from the initial results, where totals are known.
    2) It is bad practice to present data representing different time periods as the same – seasonal or other variations within the time period may make them inappropriate for comparison.
- For very low number datasets, the time period selected may still not be enough to provide adequate masking, therefore checks should be made on the data before this process is used.
- Rolling totals can hide temporal variations within a period, for example monthly rolling annual totals will hide seasonal variations in the data, because each data point contains all months of the year.
- Improvements in processes can be difficult to fully assess until the period of review has completed a full cycle. For this reason, rolling totals are generally more appropriate for assurance than for improvement.