Utah Department of Health and Human Services

# Data suppression/data aggregation guidelines

The Utah Department of Health and Human Services (DHHS) is committed to providing useful data to external parties, including the public, to inform policy and guide interventions while also protecting the privacy of individuals and presenting accurate and reliable results. Individuals may include patients, clients, employees, providers, or any other group who has or may have rights of privacy or confidentiality. The following are data suppression/data aggregation guidelines developed to inform data release.

Programs need to consider laws which govern their data. Some programs may have laws or regulations that allow data release using different standards. In the absence of other standards, DHHS follows the criteria below. The data stewards for each data source, in conjunction with the chief privacy officer, are the decision makers to determine which criteria to apply. Data stewards should consider using the stricter criteria of all those that may apply.

## Considerations

Data suppression/aggregation rules have 2 main purposes: (1) to protect against the release of identifying information; and (2) to release data with a maximum amount of precision so it can be used in decision making. The data steward should consider all data elements which are being released, the size of the population from which the data comes, and the relative standard of error. For example, releasing an incidence of 2 cases with age, gender, race, or ethnicity information in Salt Lake County may not be identifying, while 2 cases in Blanding may be.

## Criteria

**Reporting survey data**: Only report when > 10 cases in the numerator. Cells with fewer than 11 should be listed as <11.

**Reporting population data**: Typically you would not report counts less than 11. The base population should be at least 100. The base population should be limited to the portion of the population with the characteristics that are being released, such as the number of people of the specific gender, age, and/or race. In all cases, the ratio of numerator to denominator must also be considered. When the base population is low and the prevalence is high (the numerator and denominator are close to the same) then the information could potentially be identifying. Avoid reporting where the count and the base population are close enough in size to risk loss of privacy. Examples: 5 cases of X in the state of Utah may be okay to report if there is no other information being released with it that is sensitive. 85 cases of X in a town with only 100 people may be problematic.

A statistician or epidemiologist should review all types of data to provide due diligence to make sure the probability of identification is negligible given the numerator, denominator, and population characteristics.

If you can calculate relative standard error (RSE), also use criteria below:

Only report when RSE < 50%. If 30% < RSE < 50%, include an asterisk with a footnote that says: *Use caution in interpretation. The estimate has a relative standard error greater than 30% and does not meet DHHS standards for reliability. Age-adjusting RSE should be calculated based on crude rate and then applied to age adjusted rates.

If one data element in a table has to be suppressed, then enough other elements must be suppressed to prevent data users from being able to calculate the suppressed field (e.g., if the data value in a cell is suppressed, also suppress the total count or rate).

## Calculating relative standard error

Survey data:
If the estimated percentage is <50% then the RSE=100X(SE(R))/R
If the estimated percentage is >50% then the RSE=100X(SE(R))/(1-R)

Count data:
RSE =SQRT(100,000/P*R)
For counts <20 where the Poisson distribution is used to calculate CI
100 x [(UCL-LCL)/(2 x 1.96 x rate)]
When age-adjusting, the RSE should be calculated based on crude rate and then applied to age-adjusted rates.