OHIO DEPARTMENT OF HEALTH
DATA METHODOLOGY STANDARDS FOR PUBLIC HEALTH PRACTICE

Disclosure Limitation Standard:
Tabulations of confidential Ohio Department of Health data shall be suppressed when the table denominator value minus the table numerator value is less than 10.

A. Overview/Summary of Standard

The standard is adopted to limit disclosure of confidential personal information when tabulating confidential information for the public.   A table generally includes the following components: a numerator, a denominator and a rate calculated from these two quantities. The numerator is usually a count of persons with some trait or condition. The denominator represents the population of persons from which the numerator was drawn and may or may not be shown in the table. The table rate allows for comparison across denominator populations. The key feature of the standard that allows public release of tables is the existence of a critical minimum number of persons (10) without the trait or condition among the population of interest. If the denominator minus numerator is at least 10, then we judge the likelihood of identity disclosure to be sufficiently small so as to allow for publication of the table. Non-confidential information need not hold to the standard. The standard is not a test of statistical reliability.

B. Rationale/Description of Problem

This standard has been developed to protect the confidentiality of personal health information released by ODH. As public health workers we have an ethical and legal obligation to provide such protection. This protection will help to ensure that providers of these data continue to participate in these data collection activities.

The disclosure limitation issue is one of numerators and denominators, or of cells in a table. Numerators are typically the cases in a public health statistic and denominators are the population from which the cases arise. In tabular data, one can think of a specific cell as the numerator and the row total as the denominator. The characteristic defining the frequency cells or defining the case is often confidential. The risk of disclosure is greatest when the denominator is small and the ratio of numerator to denominator is high. Small denominators are common in tabulations for smaller geographic areas and for subpopulations (e.g., narrow age ranges, race groups, ethnic groups, small geographic areas).   In situations with many cases drawn from a large pool of potential cases the risk of disclosure is small.

We usually report data for fairly large populations (e.g., County). Sometimes we need to report data for smaller areas such as census tracts or neighborhoods or for subpopulations (e.g., race groups). These data for small populations are often exactly what data users need to do their public health work of preventing disease and injury. With this standard, ODH has balanced disclosure limitation objectives against a responsibility to disseminate public health information to a wide variety of users and at a geographic and subpopulation level that supports public health work. In developing a disclosure limitation strategy, ODH has balanced the benefits and risks of cautious vs. liberal approaches to data release. On one hand, a cautious approach would suppress more tables based on small numbers and prevent misuse of the data. On the other hand, a liberal approach would disseminate more tables for the widest possible use at greater risk.   The standard ODH has chosen for disclosure limitation is a result of how it weighs the relative benefits of (i) preventing misuse of data and (ii) disseminating data to users.

Local health departments, as the principal public health practitioners in the field, have greater access to detailed confidential information than other users. Tabulations compiled for Local Health Departments and for other program-approved users need not abide by the data suppression methodology outlined in this standard. Those approved users must, however, abide by the ODH standard when they re-release

ODH tabulations to the public. Granting greater access to these users presents an added concern of preserving disclosure limitation at a level removed from ODH, and over which ODH has limited control.

The standard has been extensively discussed in the Data and Research Policy Committee of ODH. The standard applies to all tabulations of departmental confidential data, including those produced automatically over the internet in the Information Warehouse.   The standard does not apply to the release of observation-level datasets to approved users, except that those users may be expected to adhere to the tabulation standard when producing public reports.

C. Guidelines for Implementation of the Standard

Understand what is confidential   A complete and up-to-date listing of confidential datasets and data elements is an important component of this disclosure limitation standard. Research staff must understand which data elements in each dataset are protected by this standard. The standard does not apply to non-confidential datasets, although ODH may at times wish to prevent disclosure of sensitive information from the non-confidential datasets.

Define the numerator(s) and denominator(s)   Clear understanding and definition of the numerators and denominators in a table is critical for correct application of this standard. Program research staff must determine in advance which elements of a table represent numerators and which represent denominators. For example, a county low birth weight rates table is a series of 88 low birth weight numerators and total birth denominators (one set for each county). A table of pre-term babies by age is a series of age-specific counts of pre-term birth numerators and age-specific total birth denominators (one set for each county). Some indicator tables don't have clear single numerators. For example, in a county table of mother's marital status one must specify whether the married count or the unmarried count or both counts represent numerators.   A different sort of numerator/denominator pair occurs when the very existence of a person in a database is confidential. An example of this is the induced terminations registry. A table of abortion rates by county would have abortion counts as the numerator and the population of women as the denominator.

Assess the impact of applying the standard to a table   Once the numerator and denominator are defined, researchers should determine which rows in a table will be suppressed based on the standard. If no suppression is dictated then a single table will meet all needs for the table. If suppression is required the researcher may need to maintain a public version of the table as well as a confidential version for approved users. Researchers should also assess whether suppressed numbers in a table can be calculated from unsuppressed numbers in the same table. Also, researchers need to be aware of other tables already published that may be used to determine suppressed values by subtraction.

Consider changes to tables to increase dissemination of public information   There are several strategies available to reduce data suppression in tables.   Since disclosure risk is highest when tables include small denominators, researchers should consider aggregating smaller denominators into fewer and larger denominators. For example, researchers might combine multiple years of data together to increase the counts in table cells. A similar strategy would be to group geographic areas together. For example, if a census tract table is overly suppressed perhaps a zip code table would be adequate to represent the geographic variation in a health indicator.   Another example in an age-specific table would be to re-define age into broader categories.

Release data in multiple customized formats when necessary   Some population groups are important to public health but inherently small in size. For example, teen mothers or Hispanic mothers as denominator groups often lack sufficient observations to pass the standard for County level or City level tables. Researchers should consider special reports to allow for release of important public health statistics for smaller groups that are overly suppressed in automated tabulation systems.

Reference: The "denominator – numerator at least 10" rule was originated by Garland Land, Missouri Dept. Health. He presented the rule at the NAPHSIS/CDC Assessment Initiative Conference in January, 2002 at Minneapolis.