



DPH Guidelines for Releasing Aggregate Population Data

Background

During 2016, the DPH Epidemiology Advisory Team created a workgroup to assess data suppression rules across the Division and to develop standard guidelines for the release of aggregate statistics from population data. This document outlines guidelines for population-based statistics released to the public on the DHS website or during meetings or presentations. Data suppression guidelines for ad-hoc aggregate data requests, media requests, and open records requests are assessed on an individual basis by the Data Governance Board. The DPH's online query system (WISH) applies a different set of rules because the query is interactive and users can do some data manipulation themselves. There are separate guidelines for sample survey estimates.

During 2000, the Bureau of Health Information reviewed its practices with the Non-confidential Sensitive Information (NoSI) report. Over time, information delivery systems have advanced and more data are available online. Nonetheless, two key issues persist:

1. Confidentiality and privacy: even absent *explicitly* identifying data elements, when information is released from a data system, information regarding small groups may *implicitly* identify an individual *and* reveal new and sensitive information about that individual, unless precautions are taken.
2. Statistical reliability: rates based on small numerators or small denominators may be deceptively unstable from year to year. While small numerators present a challenge for the interpretation of short-term variation, small denominators are a more fundamental statistical challenge.

The following recommendations are intended to provide practical guidelines for addressing confidentiality and statistical reliability issues of presenting population data. As recommendations, they are not inflexible rules; every presentation requires professional judgment about the methods and the audience.

Recommendations

The workgroup believes that these guidelines will accommodate the needs and requirements of most DPH programs. Some programs have established national standards as grant requirements (e.g., the North American Association of Central Cancer Registries standards used by the Wisconsin Cancer Reporting System). Other programs, such as AIDS/HIV, may have more stringent standards because of the sensitive and/or chronic nature of the conditions on which they report.

The guidelines are detailed in the table 'DPH Guidelines for Releasing Aggregate Population Data (5-20-100)', in a checklist of techniques and considerations, and in supplementary notes on age adjusting and maps. The guidelines are premised on the observation that both confidentiality and statistical reliability are enhanced when multiple data years are combined, when base populations are larger, and when analytic categories are broader. More detail on the underlying concerns raised by small numbers is presented in a final section.



5-20-100 Table: DPH Guidelines for Releasing Aggregate Population Data

Geographic Level	Data Years	Number of Breakout Dimensions	OK to Publish Cell Counts?	OK to Publish Cell Counts?	OK to Publish Crude Rates?**	Examples
			Count = 5+	Count = 1 - 4		
Statewide	Any period	None	Yes	Yes	Yes	Total number, statewide
		One	Yes	Yes, if Condition A*	Yes, if Condition B*	Number by age group OR gender, statewide
		Two or more	Yes, if Condition A	Yes, if Condition A	Yes, if Condition B	Number by age group AND gender, statewide
County	One or less	None	Yes	Yes	Yes	Total number by county
		One	Yes, if Condition A	Yes, if Condition A	Yes, if Condition B	Number by age group OR gender by county
		Two or more	Yes, if Condition A	Yes, if Condition A	Yes, if Condition B	Number by age group AND gender by county
	Two or more	None	Yes	Yes	Yes	Total number by county
		One	Yes	Yes	Yes, if Condition B	Number by age group OR gender by county
		Two or more	Yes	Yes	Yes, if Condition B	Number by age group AND gender by county
Sub-County (MCD, Zip, census tract, etc.)	One or less	None	Yes	Yes	Yes, if Condition B	Total number by sub-county area
		One	Yes, if Condition A	Yes, if Condition A	Yes, if Condition B	Number by age group OR gender by area
		Two or more	Yes, if Condition A	Yes, if Condition A	Yes, if Condition B	Number by age group AND gender by area
	Two or more	None	Yes	Yes	Yes, if Condition B	Total number by area
		One	Yes	Yes	Yes, if Condition B	Number by age group OR gender by area
		Two or more	Yes	Yes	Yes, if Condition B	Number by age group AND gender by area

Notes: **Condition A*:** The population underlying the cell count is **100+**. For example, if there are at least 100 persons age 18-44 in a county, then single-year counts of any number for this age group may be published.

Condition B*: Denominator for rate calculation is **100+**.

Aggregate statistics are counts and rates.

****Age-adjusted Rates:** When calculating age-adjusted rates, the underlying age-specific rates should each have a **denominator of at least 20**.

Breakout dimensions include demographic dimensions (age, gender, race, ethnicity) and other attributes. Age groups may be broader or narrower and racial categories may be restricted to larger populations (e.g., whites and non-whites).

Counts of zero may always be published.

Population data are complete censuses and not sample-based.

See checklist below for techniques and situations that may require notes or caveats.



Age-adjusted rates

- Reliable age-specific rates require that each age subgroup have a denominator of at least 20.
- Age-adjusted rates typically use direct standardization techniques by adjusting an all-age rate using observed age-specific rates and the population distribution of a reference population. An alternative that avoids issues of small denominators is to use indirect standardization techniques, in which the observed number of cases is compared to an expected or predicted number of cases.¹

Maps

- Grouping rates using quintile or quartile distributions is generally recommended, although some analyses may appropriately use natural breaks.
- Data points should be aggregated to, at minimum, census block groups, using the same guidelines for sub-county counts and rates noted above.

Checklist of Techniques and Considerations

There are options for increasing the appropriate release of small numbers from population data. The appropriateness and effectiveness of each adjustment will depend on the data's robustness, the analysis and interpretation provided with the data, and the needs of the intended audience. Please consult with your Bureau or Office's Epi Advisory Team member if you are unsure whether or how to apply data suppression.

Data techniques: Aggregation

- ☐ Increase the time frame: present multiple aggregated years of data instead of a single year. Consider rolling multi-year averages as a way to track trends.
- ☐ Increase the base population size.
 - Use a larger geographic unit (e.g., county instead of Zip).
 - Increase the size of demographic categories (e.g., use broader age groups or combine racial/ethnic groups).

Data techniques: Suppression

- ☐ Mask small counts. Consider whether further suppression of row or column cells or totals is necessary (such as when only one cell is suppressed).
- ☐ Drop small categories (e.g., smaller race/ethnic groups) or dimensions with small populations.
- ☐ Decrease stratification (e.g., present data by age or by sex or both separately, but not both cross-tabulated).

Provide interpretation:

- ☐ Add specialized statistics (e.g., present standard deviations or confidence intervals for less stable estimates).
- ☐ Add caveats and warnings to discussion or table notes. For example, note that rates based on small numerators (e.g., <10) may appear to be more unstable over time.



Confidentiality check:

- ☐ Is there any plausible risk of identifying persons because of special characteristics of the topic or data?
- ☐ What additional information would be needed to identify a person in the tally? Is this information freely available elsewhere?
- ☐ Have you consulted any existing program or data source-specific suppression rules or consulted any national standards?

Functionality of the statistics:

- ☐ Is there a clear utility in the released statistics to the public health community or in general?
- ☐ If statistics are suppressed, will the table still meet public health information needs?
- ☐ Would indirect standardization be useful?

Statistical Reliability and Validity:

- ☐ Have you considered (and shared, where appropriate) the limitations and assumptions associated with the data source and the analyses?

Why are small numbers a concern in public health assessment?

There are at least four types of concerns raised regarding the publication of aggregate statistics from population data sources when those statistics involve relatively few cases:

1. Useful and reliable quantitative information is critical for the work of local health officials, program and policy developers and managers, health care providers, advocacy groups, and the media. Questions concerning health outcomes and related health behaviors and environmental factors often are studied within small subgroups of a population, because many activities to improve health affect relatively small populations.
2. Legal and ethical considerations require that the privacy and confidentiality of persons regarding their health and health care be respected. Violations may injure persons and their families and may also compromise the future reliability of population data sources.
3. Data owners and decision-makers have statutory or assigned responsibilities for protecting the privacy of persons in data systems. Absent clear, evidence-based rules, decisions often may rely on convention, custom, and personal intuition or comfort levels.
4. Statistics for rare cases are susceptible to misinterpretation and misunderstanding. Small changes in counts result in apparently large changes in rates. Some calculations, such as age-adjusting rates, may have their reliability compromised when denominators are small.

In the context of aggregate data from whole populations, the key question is whether someone would *learn something they did not already know* about a person who is included in an aggregate statistic, such as a cell count. Knowing the identity of a person who is included in a cell count because of prior personal knowledge of all the information provided by the tabulated dimensions (geography, time



period, health condition, and demographic group) is *not* a privacy violation, because nothing new is learned from the tabulation. A breach of confidentiality occurs when analysts release information in a way that allows an individual to be identified and reveals confidential information about that person. This document provides cues to situations that present some risk to privacy and suggestions on how to reduce this risk.

References and Additional Materials

Current Practices for Protecting Non-Confidential Sensitive Information In Bureau of Health Information Data Systems ([The NoSI Report](#) and its [NAHDO and NAPHSIS presentation](#))
[Guidelines for Working with Small Numbers](#) (Washington State Public Health)
AIDSvu, a leading resource for visualizing the HIV epidemic in the United States <http://aidsvu.org/data-methods-statecounty/>
[Wisconsin HIV Cell Suppression Rules](#)
[Guidance on De-identification of Protected Health Information](#) (US Office for Civil Rights)
[Suggested Guidelines in Public Health Disclosures to External Parties](#) (WI DHS guidelines)
[NCHS Staff Manual on Confidentiality](#)

Workgroup Members

Milda Aksamitauskas, OHI
Ashley Bergeron, OPEHC
Sarah Blackwell, BCHP
Paul Creswell, BEOH
Ousmane Diallo, OHI
Mary Foote, OHI
Ruth Koepke, BCD
Richard Miller, OHI
Christine Muganda, OHI
Karl Pearson, OHI
Casey Schumann, BCD
Lauren Stockman, BCD
Carrie Tomasallo, BEOH
Anne Ziege, OHI

Dates

Last report version:	September 8, 2016
Epi Advisory Team approval:	September 21, 2016
Data Governance Board approval:	September 28, 2016

¹Juliet Van Eenwyk. *Guidelines for Using and Developing Rates for Public Health Assessment*. Washington State Department of Health, May 2012.