

Small Numbers Reporting Guidelines: Overview

Office of Health Analytics

Version 2.0 – Updated September 2023

If you have questions about this document or if you would like this document in other languages, large print, braille, or another format you prefer, please email HPA.IDEA.Team@oha.oregon.gov

Table of Contents

I. Overview	3
Why do staff in the Office of Health Analytics use these guidelines?	3
II. Principles and Considerations	4
III. Small Numbers Reporting Criteria	6
Full count data	6
Survey data	8
IV. Important Concepts to Know and Understand.....	10
Confidentiality and Protected Health Information.....	10
V. Resources and References	11

I. Overview

This document describes the guidelines for reporting small numbers that staff in the Office of Health Analytics use for external facing reporting and when filling data requests. **Note: These are minimum guidelines, and some individual programs may require more conservative reporting standards.**

Why do staff in the Office of Health Analytics use these guidelines?

Staff in the Office of Health Analytics have a responsibility to protect the confidentiality of people in Oregon while reporting reliable data both internally and externally.

Our organizational integrity is what allows us to gather and use the data we collect/manage. If we are not systematically careful and consistent with our methods for de-identifying data, we risk not only violating a person's confidentiality, but losing the trust of the Legislature and the people of Oregon, in turn impacting our ability to improve health care in the state.

These guidelines serve as **minimum recommendations** for reporting small numbers for all data sets and systems within Health Analytics to ensure that we protect the confidentiality of people in Oregon while reporting reliable data.

Note: Federal and state laws and statutes supersede this document's standards.

For more information on federal guidance for de-identification of protected health information, please see the US Department of Health & Human Services "[Guidance Regarding Methods for De-Identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act \(HIPAA\) Privacy Rule.](#)"

II. Principles and Considerations

Health Analytics staff work with full count data (e.g., claims, enrollment, and hospital discharge data), as well as survey data (e.g., OHIS and CAHPS data). The type of data determines how the data must be treated to **protect confidentiality** and **ensure reliability**.

At a minimum, data must be released in a way that protects confidentiality, regardless of data type. Confidentiality is of greatest concern for full count data, as the risk for breach of confidentiality decreases as response rate decreases. Likewise, reliability is of greatest concern for survey data, as reliability decreases as the response rate decreases.

In general, all data releases should consider the need for suppression (or aggregation) to address confidentiality and reliability (in that order). In addition, after any initial suppression, the need for secondary suppression must always be considered.

Principles

1. Suppression for confidentiality

Values suppressed to protect confidentiality should include the following language: “Value suppressed to protect confidentiality.”

2. Suppression for reliability

Values suppressed to ensure reliability should include the following language: “Estimate suppressed due to small numbers; statistically unreliable.”

3. Secondary or complementary suppression

Following any needed suppression for confidentiality and reliability, non-suppressed values should be reviewed to ensure they do not allow backward calculation of suppressed values. In general, if backward calculation is possible, the next smallest value should be suppressed. If there is a tie between the next two smallest values, one of these should be suppressed at random. Alternatively, row or column totals may be suppressed to prevent backward calculation. Values suppressed due to complementary suppression should include the following language: “Value suppressed to prevent backward calculation of other suppressed value(s).”

Additional considerations

Criteria for treating survey data as ‘full count’ data

For survey data, the potential for breaches of confidentiality decreases as the proportion of the population in the sample (i.e., your response rate) decreases. **Surveys that include 80% or more of the eligible population should be treated in the same way as full count data.**

Aggregating groups

In some cases, aggregation across categories may be possible to ensure numbers are large enough not to require suppression. The Equity and Inclusion Division of OHA encourages aggregating race and ethnicity groups and using intermediate aggregation whenever possible instead of suppression.

Masking exact values

When it is necessary to suppress values, consider masking the exact values, rather than hiding the value. For example, you may be able to report suppressed values as a range (E.g., “1-4”) to provide some data without releasing exact values.

Bias in survey data

This document primarily covers suppression to ensure confidentiality and reporting of reliable estimates. For survey data, bias must also be considered. Bias differs from the issue of precision of estimates in that bias is non-random error. For example, selection bias occurs

Small Numbers Reporting Guidelines

Office of Health Analytics

when the selection of respondents does not result in proper randomization and the resulting sample is not representative of the population you wish to generalize. **When the analyst is aware of likely bias in the sample, this should be noted. In some cases, this may necessitate data suppression with appropriate messaging (e.g., “not available due to selection bias”).**

III. Small Numbers Reporting Criteria

Full count data

Standards for reporting counts and rates from full count data (e.g., administrative, full count survey, enrollment, and claims data).

- 1) Protect Confidentiality (This may be skipped if data is not protected, see *Section V. Confidentiality and Protected Health Information* on page 10).

Report without any label or warning if:

- a. The denominator ≥ 50 .

Suppress data if:

- a. The denominator < 50 .
- b. The value for a rate is 100% (see note below for further explanation).
- c. Suppressed values should be noted with "Value suppressed to protect confidentiality."

- 2) Ensure reliability

Report:

- a. Values and rates if the numerator is ≥ 12 .

Report with reliability warning label if:

- a. The numerator of a value or rate is ≥ 5 and < 12 .*
- b. A warning should appear on the same page and say, "May be statistically unreliable due to small numbers; interpret with caution."

Suppress data if:

- a. The numerator of a value or rate is < 5 .*
- b. **It is up to the discretion of analysts and their manager whether they suppress zero values.**
- c. Suppress the estimate with the warning, "Estimate suppressed due to small numbers; statistically unreliable." Do not show the estimate and instead show a symbol referring to a warning, which should appear on the same page.

* For event or count data, any result of a rate calculation where the count of events is less than 12 is not reliable, because rate calculations where the count of events is < 12 result in Relative Standard Error (RSE) $\geq 30\%$ and calculations where the count of events is < 5 result in RSE $\geq 50\%$.

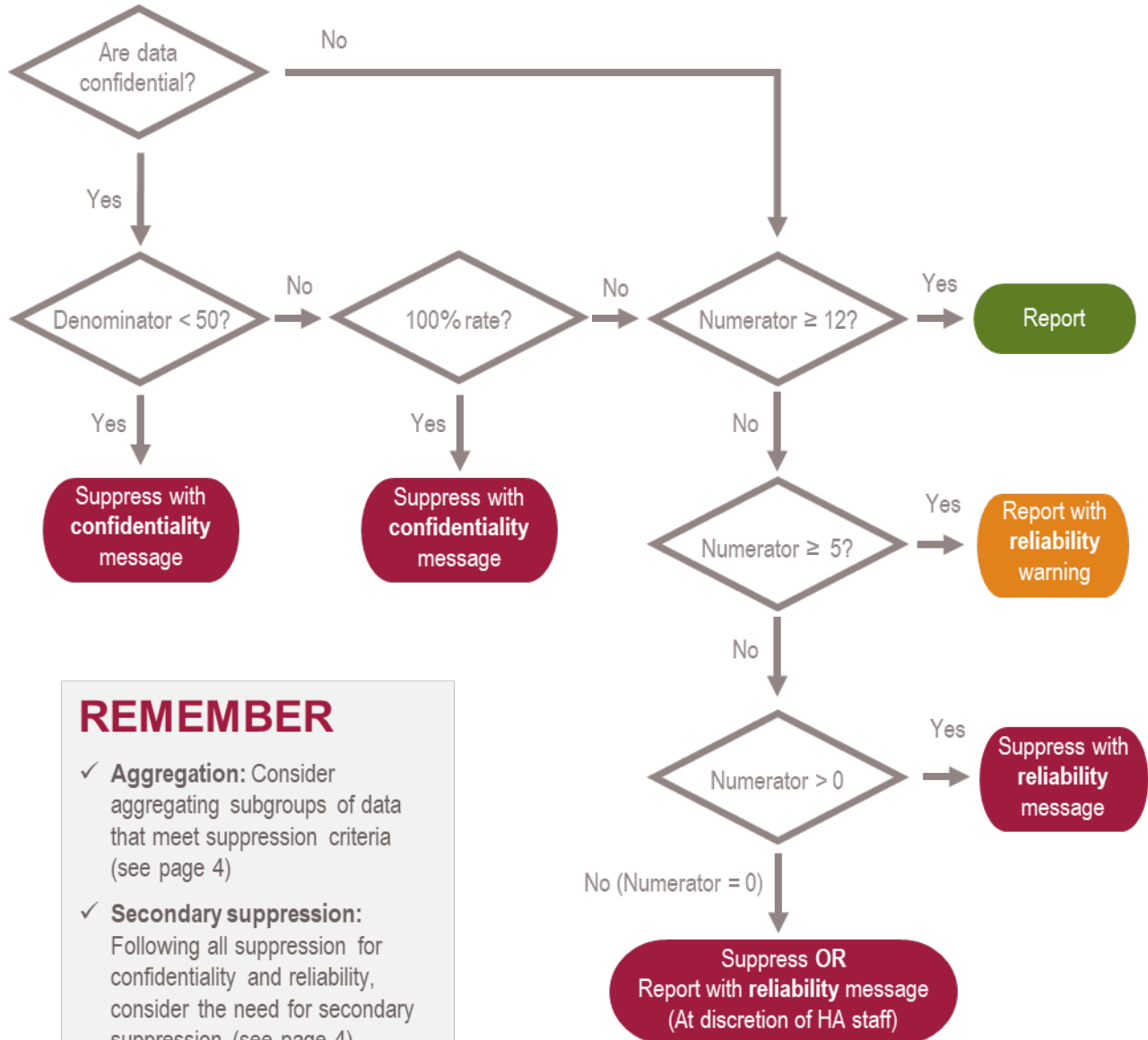
- 3) Use secondary suppression as needed (see page 4)

Notes:

- When data meet the criteria for suppression, you may consider aggregating categories to report more data.
- Further explanation for suppression of rates of 100%: For most situations in which we talk about diseases, zero cases are not a concern for confidentiality breach. An exception is double negative descriptions, e.g., "None of these four teens are non-drinkers." The guidelines need to apply regardless of how the description is structured. For example, if we flip the sentence structure, we will then apply the thresholds on "all four teens are drinkers" rather than "none of these four teens are non-drinkers." Be aware of implications if you choose to report 100% rates.

Suppression Decision Flow Chart: Full count data

Use this flow chart to determine whether you should report a number and, if so, how. See the minimum reporting criteria on page 6.



REMEMBER

- ✓ **Aggregation:** Consider aggregating subgroups of data that meet suppression criteria (see page 4)
- ✓ **Secondary suppression:** Following all suppression for confidentiality and reliability, consider the need for secondary suppression (see page 4)

Survey data

Standards for reporting survey data

- 1) Report data if:
 - a. Denominator (unweighted) ≥ 50 for full population,* or ≥ 30 for a subpopulation;* and
 - b. Numerator (unweighted) ≥ 3 ; and
 - c. Relative Standard Error (RSE) $< 30\%$; and
 - d. Reporting the estimate cannot lead to backward calculation of other suppressed value(s). See Secondary Suppression on page 4.

* You are working with a full population if your denominator is aligned with the full sample frame, or highest level of your sample (such as all people in Oregon or all Medicaid members). You are working with a subpopulation if your denominator is a subset of the sample frame or of respondents who share a common trait (such as a county or a race group).

- 2) Report data with a warning label if:
 - a. Denominator (unweighted) ≥ 50 for full population, or ≥ 30 for a subpopulation; and
 - b. Numerator (unweighted) ≥ 3 ; and
 - c. $30\% \leq$ Relative Standard Error (RSE) $< 50\%$.

Warning label: Put a symbol next to the estimate referring to a warning on the same page that says: “May be statistically unreliable due to small numbers; interpret with caution.”

- 3) Suppress data with a suppression label if:
 - a. Denominator (unweighted) < 50 for full population, or < 30 for a subpopulation; or
 - b. Numerator (unweighted) < 3 ; or
 - c. Relative Standard Error (RSE) $\geq 50\%$; or
 - d. If value is 0% or 100%.

Suppression label: Suppress the estimate with the warning, “Estimate suppressed due to small numbers; statistically unreliable.” Do not show the estimate. Instead, show a symbol referring to the warning, which should appear on the same page.

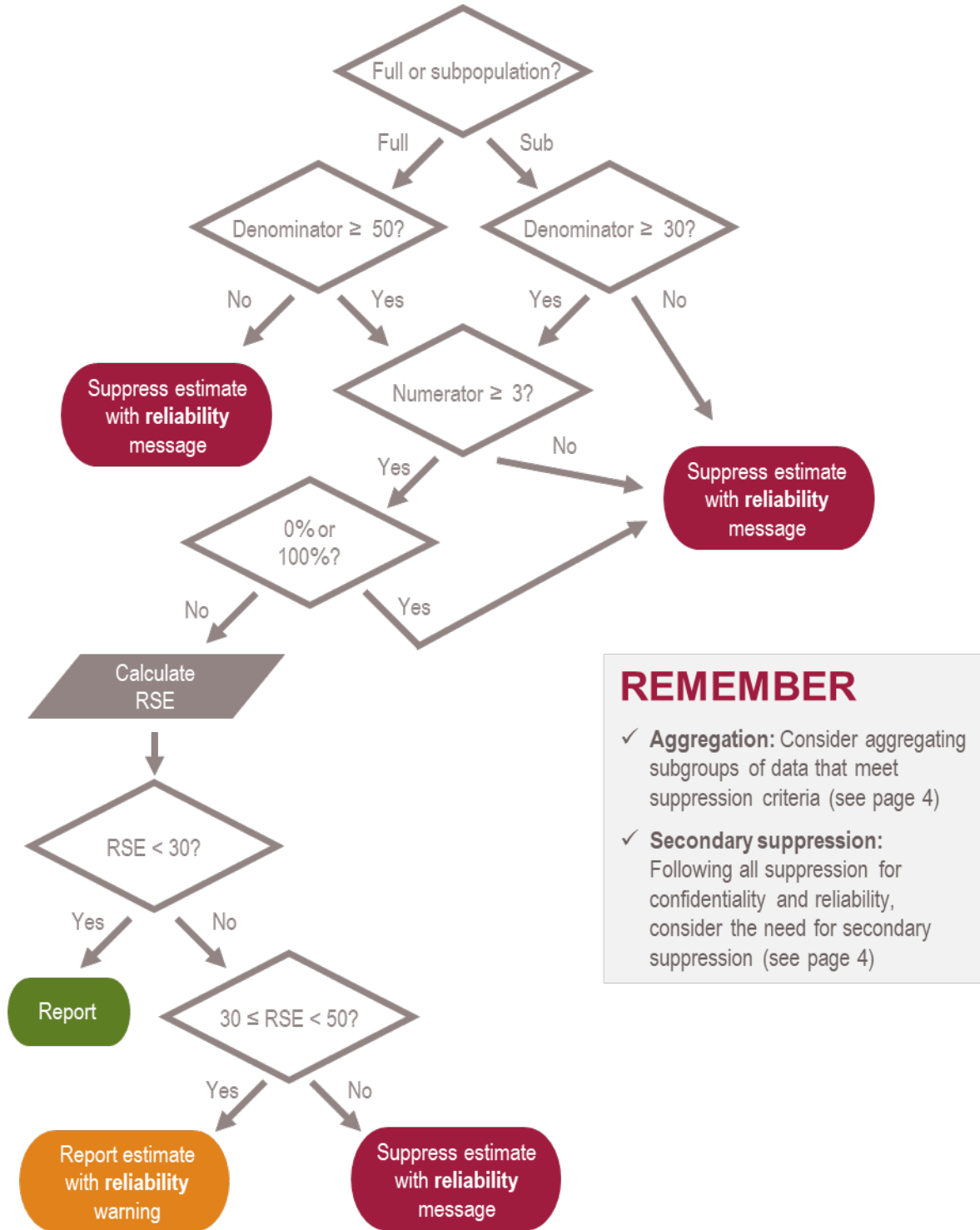
Use secondary suppression, if necessary, to prevent backward calculation of other estimates (Secondary Suppression, see page 4).

Notes:

- The relative standard error (RSE) is a measure of the variability of the estimate compared with the magnitude of the estimate. It may be thought of as the percentage of the magnitude of the estimate that is subject to random error. **Estimates with large RSE are considered unreliable.**
- When data meet the criteria for suppression, you may consider aggregating categories to report more data.
- Unweighted denominators and numerators are the raw or actual counts from the survey, not weighted estimates of the entire population or subpopulation represented by the group in the survey.

Suppression Decision Flow Chart: **Survey data**

Use this flow chart to determine whether you should report an estimate obtained from a survey, and if so, how. See the minimum reporting criteria on page 8.



REMEMBER

- ✓ **Aggregation:** Consider aggregating subgroups of data that meet suppression criteria (see page 4)
- ✓ **Secondary suppression:** Following all suppression for confidentiality and reliability, consider the need for secondary suppression (see page 4)

IV. Important Concepts to Know and Understand

Confidentiality and Protected Health Information

Confidentiality involves a set of rules or a promise, usually executed through confidentiality agreements, that limits access or places restrictions on certain types of information.

Confidentiality in the Health Insurance Portability and Accountability Act of 1996 (HIPAA) refers to the duty of *covered entities* entrusted with health information to keep that information private. Covered entities include health care providers who electronically submit health information, health plans, and health care clearing houses ([Learn more](#))

Oregon statutes and administrative rules that impact OHA programs may also require confidentiality of data that is not necessarily protected health information. An example is the Health Care Workforce Reporting Program database where Oregon Revised Statute (ORS) 676.410 and Oregon Administrative Rule (OAR) 409-026 requires the program keep health care professional demographics and practice characteristic information de-identified.

Protected health information (PHI) in HIPAA, is any health information that can be tied to an individual and includes one or more of the following 18 identifiers. Information is only considered PHI when an individual could be identified from the information. If all identifiers are stripped from health data, the data is no longer PHI, and the HIPAA Privacy Rule's restrictions on uses and disclosures no longer apply. Stripping out the following 18 items is known as the "HIPAA Safe Harbor method."

1. Names (Full or last name and initial).
2. All geographical identifiers smaller than a state, except for the initial three digits of a zip code if, according to the currently publicly available data from the U.S. Bureau of the Census: the geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people; and the initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000
3. Dates (other than year) directly related to an individual
4. Phone numbers
5. Fax numbers
6. Email addresses
7. Social Security numbers
8. Medical record numbers
9. Health insurance beneficiary numbers
10. Account numbers
11. Certificate/license numbers
12. Vehicle identifiers (including serial numbers and license plate numbers)
13. Device identifiers and serial numbers
14. Web Uniform Resource Locators (URLs)
15. Internet Protocol (IP) address numbers
16. Biometric identifiers, including finger, retinal and voice prints
17. Full face photographic images and any comparable images
18. Any other unique identifying number, characteristic, or code except the unique code assigned by the investigator to code the data

V. Resources and References

These resources are intended to provide an opportunity for those would like to know more about small number suppression to read more discussion of federal standards and other agency's standards.

Oregon Health Authority Public Health Division: [Health Promotion and Chronic Disease Prevention Guidelines for Reporting Reliable Numbers](#)

Oregon Health Authority Public Health Division: [Guidelines for Reporting Small Numbers to Protect Confidentiality](#)

Utah Department of Health Data Suppression Decision Rules Work Group: [Report of Guidelines for Data Result Suppression](#)

U.S. Department of Health and Human Services, Centers for Medicare & Medicaid Services: [Instructions for Completing the Data Use Agreement \(DUA\) Form CMS-R-0235](#)

U.S. Centers for Disease Control and Prevention, United States Cancer Statistics (USCS): [Suppression of Rates and Counts](#)

U.S. Department of Health & Human Services: [Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act \(HIPAA\) Privacy Rule](#)

U.S. Department of Health & Human Services: [Disclosures for Public Health Activities](#)

Washington State Department of Health: [Department of Health Agency Standards for Reporting Data with Small Numbers](#)