OREGON **HEALTH** AUTHORITY

Small Numbers Reporting Guidelines - Summary

This document is a plain language summary of the Small Numbers Reporting Guidelines that the Office of Health Analytics uses when we report aggregate or summary data that sometimes includes small numbers.[1] For more detailed technical information, please read the complete guidelines.

Sometimes we must suppress data

The Office of Health Analytics is committed to providing accurate data while protecting the privacy of the people these data represent. To achieve these goals, sometimes we must suppress data. **Suppressing data means not sharing it.** We only suppress data when we have to and share as much data as we can.

We suppress data for two reasons

1. Confidentiality: we must protect the people who data represents.
   When data represents people, it's our job to protect people and their privacy. So, when there's a chance that releasing data could share information about individual people, we do not release that data.

2. Reliability: our data must be trustworthy and consistent
   We have a responsibility to share data that are reliable. Reliability means that we are relatively confident the numbers are correct and are not simply due to chance. This is important because data are used to inform decisions that impact

---

[1] Aggregate or summary data often describes the health care system or people's experiences with health care. Examples include how many people in Oregon have health insurance, or how much money is spent on primary care across the state.

people in Oregon. When we do not believe the data are reliable, we do not release the data.

Confidentiality is a concern only when the data are about people. Reliability is a concern when the data are about people or things. Confidentiality and reliability are most at risk when the numbers in the data are small.

We are more likely to suppress some of the data when numbers are small

When we suppress data, it's because showing the small numbers would risk sharing information about people (confidentiality) or when we can't be sure the data are accurate (reliability). How small a number must be to suppress it depends on the type of data. You can read all the rules we use in the Small Numbers Reporting Guidelines: Overview.

If we are looking at the number of people in Oregon who had a common medical procedure, the numbers are probably large, and we would share the data because it would be hard to identify a single person. An example of this is the number of people who received a common vaccine.

But if we are looking at the number of people who speak a certain language and had a less common medical procedure, then the numbers will be smaller. An example of this might be Urdu speakers who had a rare diagnosis. If the numbers are too small, then we would suppress the data.

There are many ways to suppress data

There are different ways HPA may suppress data when numbers are too small to share. When someone requests data, it can be helpful for us understand what they are trying to learn, so we can choose suppression methods that give them the most information.

Let's use the imaginary dataset on the next page and say that the number five (for the age group 0-12) is too small to share.

Age group	#
0-12	5
13-18	20
19-64	100
65+	15

Simple suppression

In some cases, we could simply hide (suppress) the number five:

Age group	#
0-12	-
13-18	20
19-64	100
65+	15

Secondary suppression

Sometimes suppressing one number isn't enough to protect privacy or reliability because someone could do the math to figure out the number. This is often true when the dataset includes a total. Using the same example, if we know the total is 140, then a person could subtract the other numbers from the total (140) to figure out the suppressed value (five).

If only one age groups is suppressed, then a person could use math to figure out the missing number (as shown at right).

Age group	#
0-12	-
13-18	20
19-64	100
65+	15
Total	**140**

140 (Total)
-20 (Ages 13-18)
-100 (Ages 19-64)
-15 (Ages 65+)
_____
5 The suppressed number

Therefore, we need to suppress a second value, called 'secondary suppression.' For example, we could suppress the total, or we could suppress another age group. As described earlier, it's helpful to know the goals of a request so we can decide which second number to suppress. For example, if the total is important, then we will choose to suppress another age group instead.

Age group	#
0-12	-
13-18	20
19-64	100
65+	15
Total	**-**

or

Age group	#
0-12	-
13-18	20
19-64	100
65+	-
Total	**140**

Aggregating

Aggregating means combining groups. For example, we can combine the age groups 0-12 and 13-18 into one larger age group of 0-18. Then, none of the numbers will be too small to share.

Age group	#
0-18	25
19-64	100
65+	15
Total	**140**

Another example of aggregating data might be combining several counties into one geographical region.

Blurring

Blurring means we don't provide the exact number, but instead we provide a range. For example, instead of sharing the number five, we could share that the value is "less than 10" or "between one and five."

Age group	#
0-12	<10
13-18	20
19-64	100
65+	15

Questions?

We hope this was a helpful summary of how and why we suppress data. For more detail, read the Small Numbers Reporting Guidelines: Overview. If you have questions, please email HPA.IDEA.Team@odhsoha.oregon.gov.

You can get this document in other languages, large print, braille or a format you prefer free of charge. Contact HPA.IDEA.Team@odhsoha.oregon.gov

Health Policy and Analytics Division
Office of Health Analytics

OREGON
HEALTH
AUTHORITY