Data Suppression and Working with Small Numbers

NCDHHS Division of Public Health, Injury and Violence Prevention Branch (IVPB), Epidemiology, Surveillance, and Informatics (ESI) Unit

What Is Data Suppression?

Data suppression is when data on very small numbers are not shared publicly.

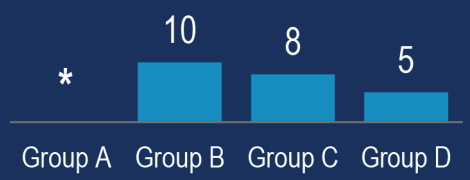
The number of events that are okay to share publicly and what needs to be suppressed or hidden depends on the data and the program or agency.

- Some types of data do not need to be suppressed at all, while others do.
- Different programs and agencies have different policies on what is okay to share publicly and what is not.
    - Some programs suppress counts less than 50, while others suppress counts less than 10, or less than five.
    - A count of zero is usually not suppressed.
    - Some programs consider both the total population size and the event count when deciding if data should be suppressed.

WHAT ARE SMALL NUMBERS?

In injury surveillance, when the number used to describe how many injuries happened within a population or group is very low (usually fewer than 10), it is considered a "small number" or a "small count."

An **asterisk (*)** or another marker are often used to indicate data are suppressed in charts and figures.



Why Suppress Data?

Protecting Privacy

Some data are hidden or suppressed to keep people's information private.

When numbers are very small it can be possible to figure out who the data refer to.

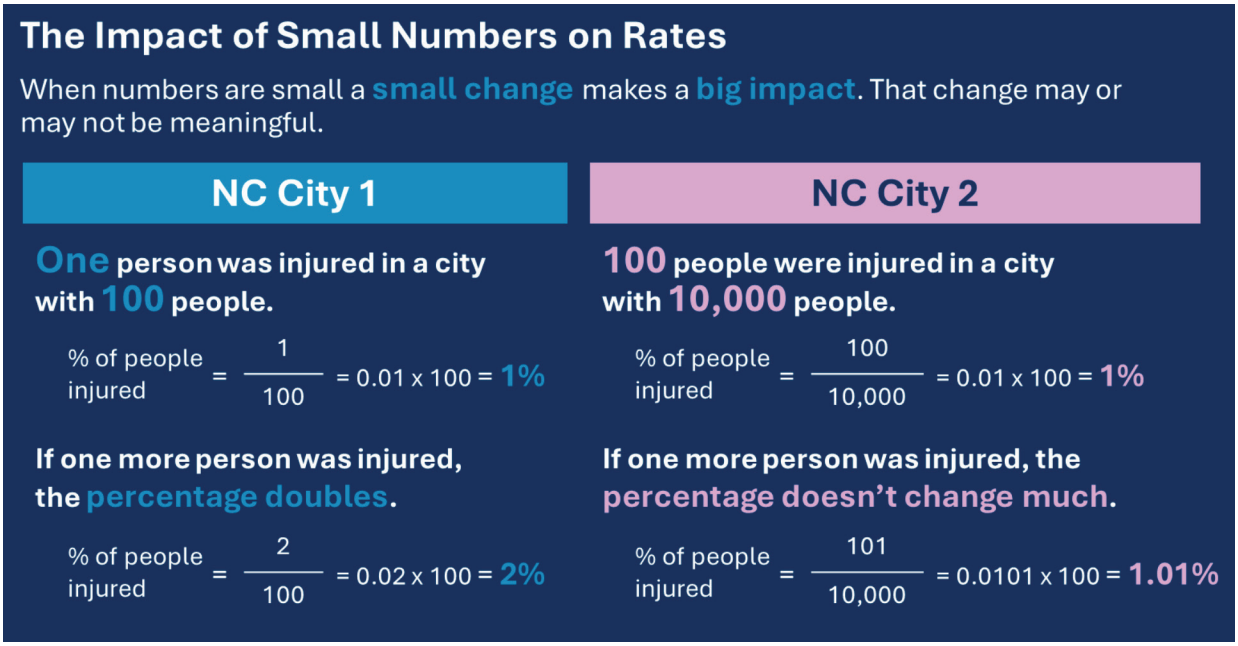
- This is especially true when small numbers are combined with other information, like sex, age, race, or place (city, county, etc.).
- For example, imagine a report shares that one person between the ages of 10-17 died by suicide in Tyrrell County last year.
    - Tyrrell County is a rural county in NC and there is only a small number of people living there, including a small number of teenagers.
    - Even though the report does not list the person by name, someone from Tyrrell County might be able to figure out who the person was who died by suicide. This could cause people to judge or treat the surviving family members differently.

Data Reliability

Statistics, like rates and percentages, that are calculated using small numbers can be unstable, or less reliable.

With small numbers a very small change can make a big impact on a count or a rate.

- Rates can go up or down a lot, even if only one case is added or taken away.
    - This makes it hard to tell if a change was caused by something real happening or if it was random.
- It is important to be careful when interpreting trends or when making decisions based on small numbers.
    - You want to be sure a change is caused by a shift in overall trends and not just caused by chance.



The figure above gives an example of an injury rate based on small numbers and how that rate can change even with only a small change in count. The rate calculated using small numbers (1/100) doubled when another case was added (2/100), representing a 100% increase in the number of injuries in NC City 1. The rate with larger numbers (100/10,000) did not change much when another case was added (101/10,000), representing a 1% increase in the number of injuries in NC City 2.

How Does IVPB Suppress Data?

IVPB believes in being open and honest about the data it collects and uses (data transparency). The Branch shares as much data as it can, while also keeping people's private information safe.

Counts

IVPB suppresses the number of injuries differently for different data sources.

Death Data

In NC, Vital Statistics death certificate data are available publicly. Because of this, IVPB does not suppress the number of injury deaths, even when counts are very small.

- Small counts are not suppressed when using data from the NC State Unintentional Drug Overdose Reporting System (NC-SUDORS) and the NC Violent Death Reporting System (NC-VDRS), since both systems use death certificate data.

Non-Fatal Data

IVPB follows the suppression guidelines outlined by North Carolina Disease Event Tracking and Epidemiologic Collection Tool (NC DETECT) when using emergency department (ED) visit and hospital discharge data.

- In addition to the number of injuries, these guidelines also consider the size of the population and therefore the chance that someone could be identified from the data.

Non-fatal data are suppressed when:

The number of injuries is greater than zero and less than five. (1 - 4 Injuries)		
	AND	
• Data are for a county or region • Time frame of data is less than one year • There were fewer than 500 total ED visits in that same county or region	**OR**	• Data are broken out by demographic group (sex, race, etc.) • Any timeframe of data • There were fewer than 500 total ED visits among that same demographic group

For example (these examples use fictional data):

- There were 6 hospitalizations in Watauga County in May 2022 – **Not suppressed**
- There were 2 ED visits for drowning-related injuries among adults ages 65-69 between 2023 and 2025, and 4,400 total ED visits for those ages 65-69 – **Not suppressed**
- There were 3 ED visits for motor vehicle traffic injuries in Hyde County in January 2024, and 450 total ED visits for the county that month – **Suppressed**
- There were 4 hospitalizations for pedestrian injuries in NC in February 2025 – **Not Suppressed**

Rates

IVPB does not usually calculate rates of injury when the number of injuries is greater than zero and fewer than five (1-4). This is because these numbers are too small to give reliable results.

- Rates are usually calculated when the number of observed events is between five and nine but should be interpreted with caution.
- If the number of events observed is zero, a rate is calculated (which also equals zero) and shared.

Disaggregating Data

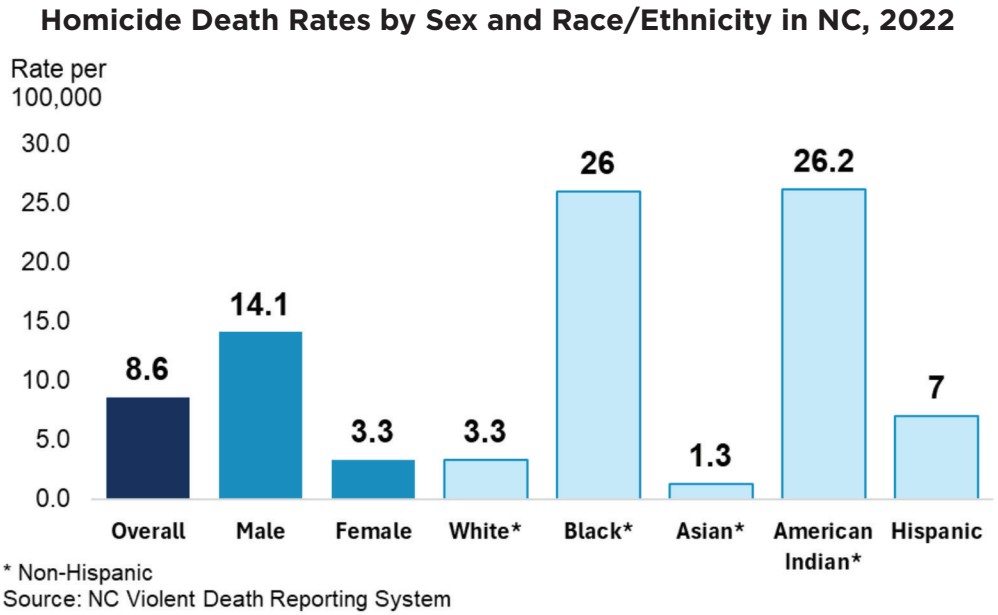
Disaggregating data is when data are separated into smaller parts to better understand differences between groups within the data.

IVPB regularly breaks out data by things like sex, age, race/ethnicity, and county of residence.

- This helps identify patterns that might be missed when looking at the data all together.
- IVPB shares disaggregated data to inform injury prevention planning.
    - These data can be used to ensure prevention strategies align with what injuries are happening, where they are happening, and to whom.

For example:

- The homicide rate in NC in 2022 was 8.6 per 100,000.
- After disaggregating data, you can see some groups have higher rates than others.
- The homicide rates for males, non-Hispanic (NH) Black, and NH American Indian residents are all higher than the state rate.



This figure shows the rate of homicides in NC by sex and by race/ethnicity in 2022.

Disaggregating data can cause small numbers.

The more you break data into specific groups, like by both race and sex together, the smaller the numbers you are working with can get.

- To publicly share data for these specific groups, even when there are small numbers, IVPB often combines, or aggregates, data over multiple years, calculating multi-year rates.
- Combining data over several years helps increase the number of injuries included so the data can be shared for a more specific group or population.
- Multi-year rates that are calculated over time (rolling rates) can be used to help interpret trends when there is a small number of injury events.
    - For more information on multi-year rates and rolling rates, visit Understanding Counts and Rates.

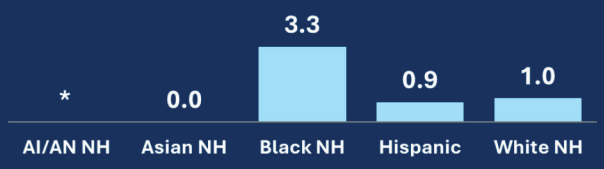
Combining data over time allows IVPB to share data on more specific groups.

The example below looks at **female homicide deaths by race/ethnicity in NC.**

1 Year of Data: 2022

In 2022, **3** non-Hispanic (NH) American Indian/ Alaskan Native (AI/AN) women died by homicide in NC.

The homicide rate for NH AI/AN women is unstable and is not shared because of the small count.

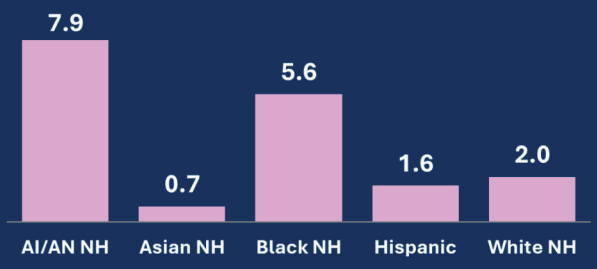


5 Years of Data: 2018-2022

Combining data from 2018-2022, there were a total of **23** NH AI/AN women who died by homicide in NC.

NH AI/AN women had the highest homicide rate compared to women of other races/ethnicities.

IVPB could not share these data with communities if they were not combined over multiple years.



* Data are suppressed due to count 1-4.

Questions

If you have any questions about injury rates shared by IVPB, contact:

- SubstanceUseData@dhhs.nc.gov for questions related to overdose, substance use, and alcohol
- InjuryData@dhhs.nc.gov for all other injury-related questions

FOR MORE RESOURCES:

Visit our Injury Data Users Toolkit