# NCHS Data Presentation Standards for Rates and Counts

Jennifer D. Parker
Director, Division of Research and Methodology
National Center for Health Statistics

FCSM Research and Policy Conference
October 25, 2022
Session 18: Data Linkage, Privacy, and Modernization

# The National Center for Health Statistics

- Principal statistical agency for health in the United States
- 1960 - National Office of Vital Statistics and the National Health Survey merged to form National Center for Health Statistics
- 1987 - NCHS became part of the Centers for Disease Control and Prevention
- Remains the official Federal Health Statistics Agency and part of the Federal Statistical System

# Suppress unreliable estimates?

## Yes, in some publications.

- Some publications do not have the format for standard errors or other indicators of precision alongside estimates to inform user.
  - Infographics, social media, some visualizations
- Do not fully trust the data user.
  - Misinterpretation of statistics can reflect badly on the agency and federal statistics more broadly.
- Nearly all NCHS estimates can be obtained using public use data sources and all can be obtained in the Research Data Center, subject to disclosure review.
  - Users who calculate estimates may be better able to interpret the results.

# Suppress estimates?

## Yes, in some publications.

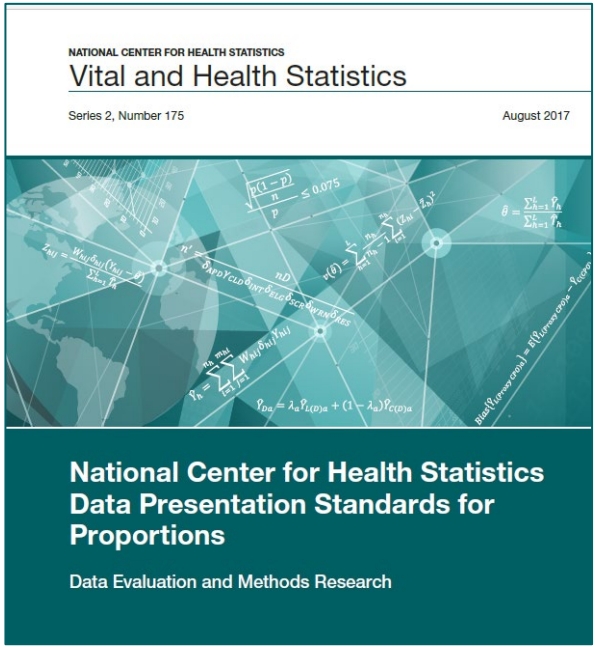
- Some publications do not have the format for standard errors or other indicators of precision alongside estimates to inform user.
  - Infographics, social media, some visualizations
- Do not fully trust the data user.
  - Misinterpretation of statistics can reflect badly on the agency and federal statistics more broadly.
- Nearly all NCHS estimates can be obtained using public use data sources and all can be obtained in the Research Data Center, subject to disclosure review.
  - Users who calculate estimates may be better able to interpret the results.

## No.

- Estimates can be informative even if they're not precise.
- Data for small subgroups may never meet conventional standards of 'good-enough' but information for these groups is needed for decisions.
- Trust the data user.

# NCHS Data Presentation Standards for Proportions released August 2017

- Standard
  - Criteria based on minimum sample size and absolute and relative width of exact confidence interval.
  - Estimated proportions with few degrees of freedom or with suppressed complementary proportions are to be evaluated individually.
- Implementation generally led to more estimates presented compared to prior presentation criteria.



# Why presentation standards for rates?

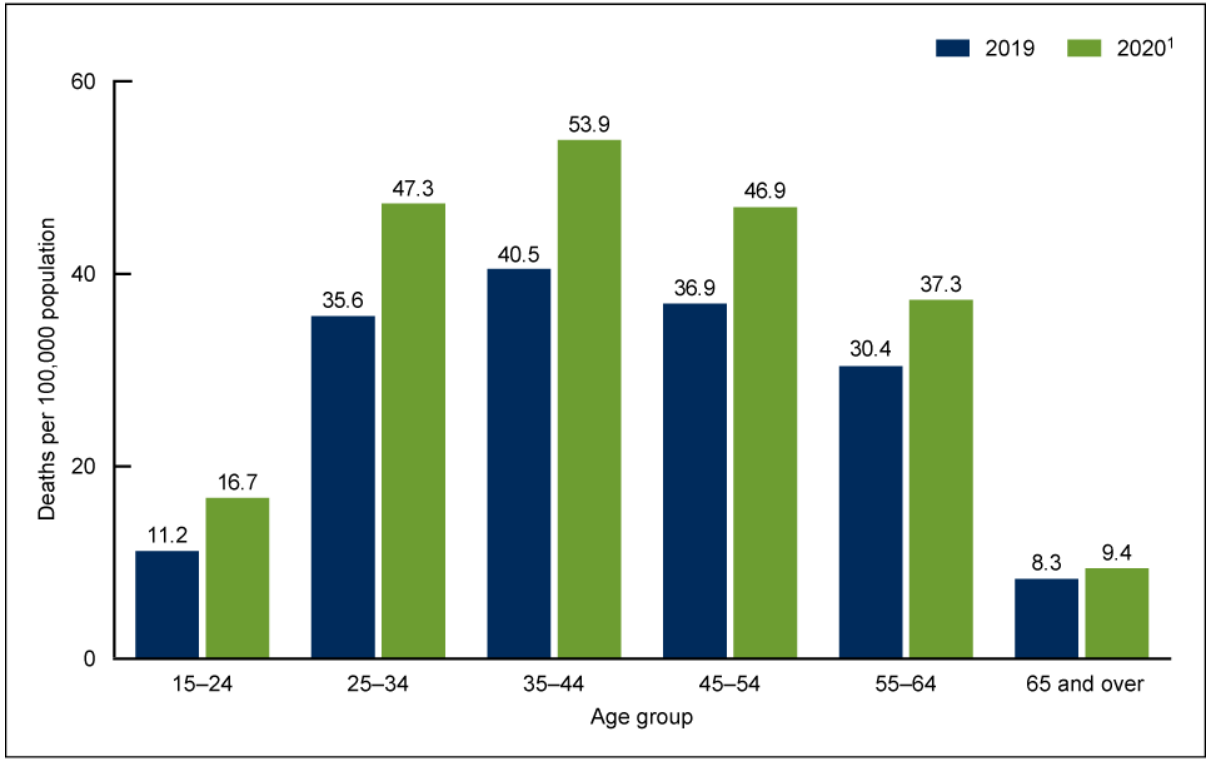
- Crude, age-specific, and age-adjusted rates are widely disseminated by NCHS
  - Vital statistics
    - Death rates
    - Birth rates, fertility rates
  - Health care statistics
    - Visit rates
    - Hospitalization rates
- Counts follow as the ‘numerator’ of the rate

# What are rates?

- The working definition of a rate is the estimated number of events for a given time period (typically year) divided by a population count for the number at risk during that year (typically a Census mid-year population count) and often multiplied by a constant.
- Age-adjusted rates:  linear combination of age-specific rates standardized to a standard US population.

# Drug overdose death rates among those aged 15 and over, by selected age group. US 2019-2020

Figure 2. Drug overdose death rates among those aged 15 and over, by selected age group: United States, 2019 and 2020




[1]Rates in 2020 were significantly higher than in 2019 for all age groups, $p < 0.05$.
NOTES: Drug overdose deaths are identified using the *International Classification of Diseases, 10th Revision* (ICD–10) underlying cause-of-death codes X40–X44, X60–X64, X85, and Y10–Y14. Access data table for Figure 2 at: https://www.cdc.gov/nchs/data/databriefs/db428-tables.pdf#2.
SOURCE: National Center for Health Statistics, National Vital Statistics System, Mortality.

Hedegaard H, Miniño AM, Spencer MR, Warner M. Drug overdose deaths in the United States, 1999–2020. NCHS Data Brief, no 428. Hyattsville, MD: NCHS. 2021.

## Table C. Deaths and percentage of total deaths for the 10 leading causes of death: United States, 2018 and 2019

[An asterisk (*) preceding a cause-of-death code indicates that the code is not included in the *International Classification of Diseases, 10th Revision* (ICD–10)]

| Cause of death (based on ICD–10) | Rank[1] | 2019 | | 2018 | |
|---|---|---|---|---|---|
| | | Deaths | Percent of total deaths | Deaths | Percent of total deaths |
| All causes | ... | 2,854,838 | 100.0 | 2,839,205 | 100.0 |
| Diseases of heart (I00–I09,I11,I13,I20–I51) | 1 | 659,041 | 23.1 | 655,381 | 23.1 |
| Malignant neoplasms (C00–C97) | 2 | 599,601 | 21.0 | 599,274 | 21.1 |
| Accidents (unintentional injuries) (V01–X59,Y85–Y86) | 3 | 173,040 | 6.1 | 167,127 | 5.9 |
| Chronic lower respiratory diseases (J40–J47) | 4 | 156,979 | 5.5 | 159,486 | 5.6 |
| Cerebrovascular diseases (I60–I69) | 5 | 150,005 | 5.3 | 147,810 | 5.2 |
| Alzheimer disease (G30) | 6 | 121,499 | 4.3 | 122,019 | 4.3 |
| Diabetes mellitus (E10–E14) | 7 | 87,647 | 3.1 | 84,946 | 3.0 |
| Nephritis, nephrotic syndrome and nephrosis (N00–N07,N17–N19,N25–N27) | 8 | 51,565 | 1.8 | 51,386 | 1.8 |
| Influenza and pneumonia (J09–J18) | 9 | 49,783 | 1.7 | 59,120 | 2.1 |
| Intentional self-harm (suicide) (*U03,X60–X84,Y87.0) | 10 | 47,511 | 1.7 | 48,344 | 1.7 |

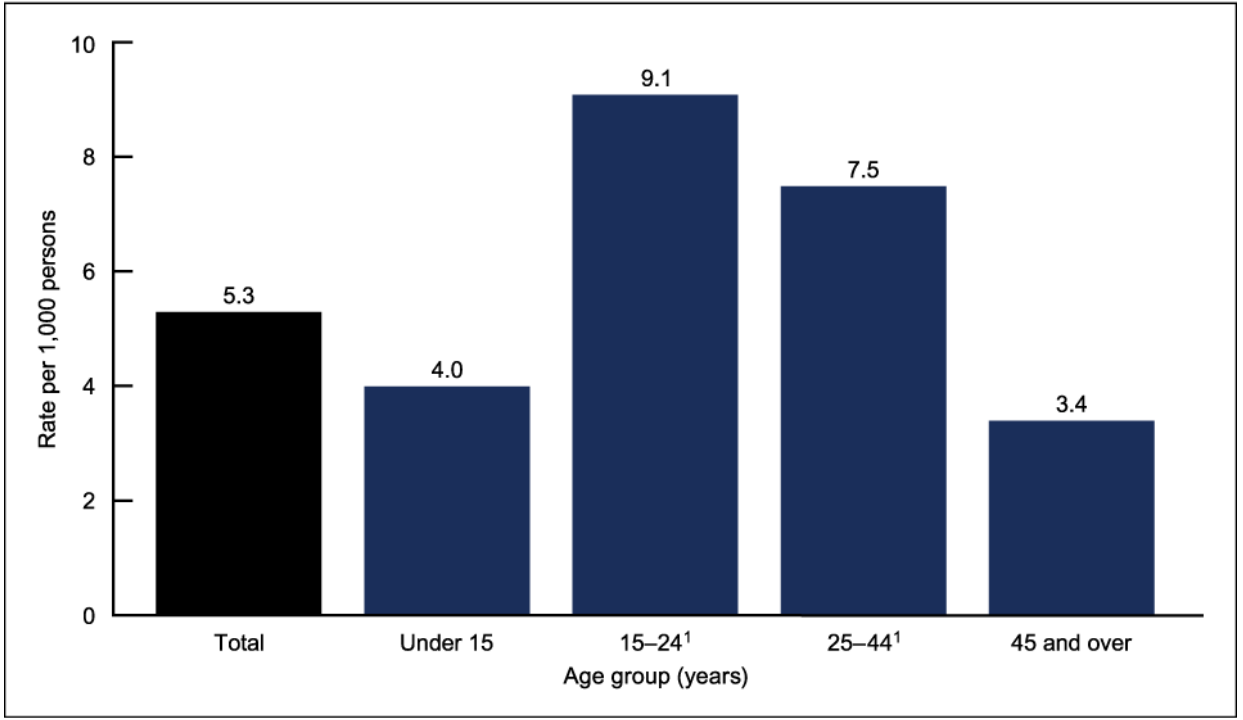
... Category not applicable.
[1]Based on number of deaths.

SOURCE: National Center for Health Statistics, National Vital Statistics System, Mortality.

Suggested citation Heron M. Deaths: Leading causes for 2019. National Vital Statistics Reports; vol 70 no 9. Hyattsville, MD: National Center for Health Statistics. 2021. DOI: https://dx.doi.org/10.15620/cdc:107021

# Emergency department visit rates for motor vehicle crashes, for total and by age group: US 2017-2018

Figure 1. Emergency department visit rates for motor vehicle crashes, by age group: United States, 2017–2018



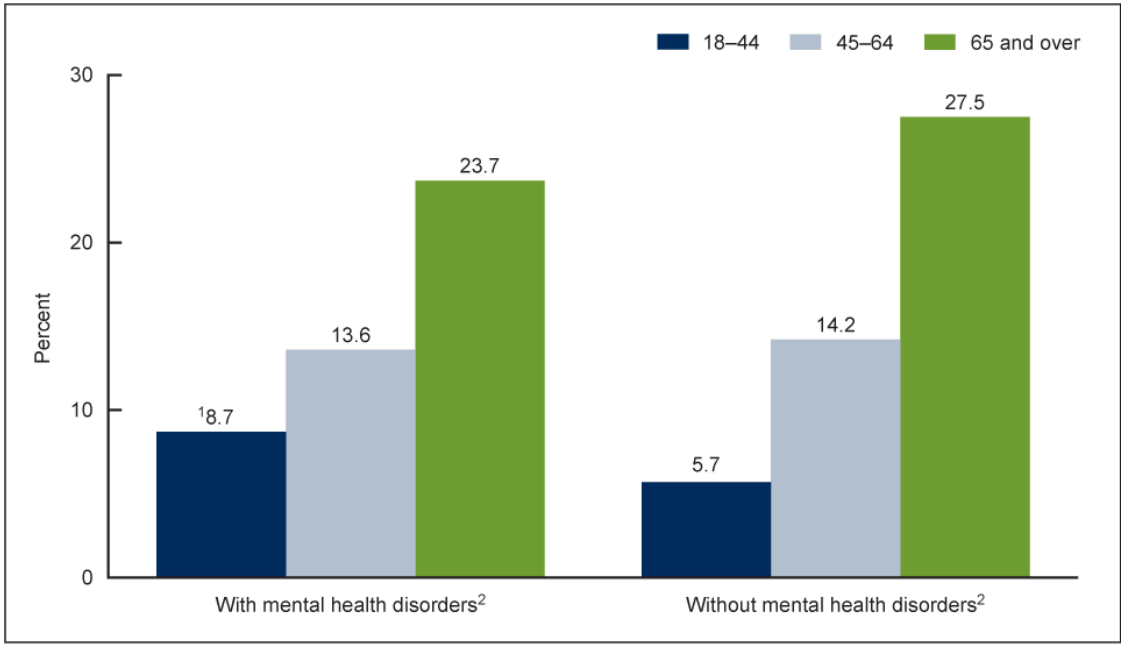

[1]Significantly different from persons aged under 15 and 45 and over.
NOTES: Based on a sample of 917 emergency department (ED) visits for motor vehicle crashes, representing an annual average of 3.4 million ED visits. Visit rates are based on the July 1, 2017, and July 1, 2018, sets of estimates of the civilian noninstitutionalized population of the United States, as developed by the U.S. Census Bureau, Population Division. Motor vehicle crashes defined as a visit with International Classification of Diseases, 10th Revision, Clinical Modification codes: V02–V04 (.1, .9), V09.2, V09.3, V12–V14 (.3–.5, .9), V20–V28 (.3–.5, .9), V19.4–V19.6, V19.9, V29.4–V29.9, V30–V79 (.4–.9), V83–V86 (.0–.3), V80.3–V80.5, V81.1, V82.1, V87.0–V87.8, V89.2, X81.0, X82, Y02.0, Y03, and Y32. Injured persons included motor vehicle occupants, motorcyclists, pedal cyclists, and pedestrians. Access data table for Figure 1 at: https://www.cdc.gov/nchs/data/databriefs/db410-tables-508.pdf#1.
SOURCE: National Center for Health Statistics, National Hospital Ambulatory Medical Care Survey, 2017–2018.

Davis D, Cairns C. Emergency department visit rates for motor vehicle crashes by selected characteristics: United States, 2017–2018. NCHS Data Brief, no 410. Hyattsville, MD: National Center for Health Statistics. 2021

# Percentage of adult ED visits with and without mental health disorders that resulted in a hospital admission, by age

Santo L, Peters ZJ, DeFrances CJ. Emergency department visits among adults with mental health disorders: United States, 2017–2019. NCHS Data Brief, no 426. Hyattsville, MD: National Center for Health Statistics. 2021

Figure 3. Percentage of adult emergency department visits with and without mental health disorders that resulted in a hospital admission, by age group: United States, 2017–2019




[1]Significantly higher than for patients without mental health disorders in same age group ($p < 0.05$).
[2]Statistically significant increasing trend with increasing age ($p < 0.05$).
NOTES: Based on a sample of 5,978 emergency department (ED) visits made by adults with any listed diagnosis of mental health disorders (*International Classification of Diseases, 10th Revision, Clinical Modification*, codes F01–F99), representing an annual average of approximately 13.2 million ED visits (12.3% of all adult ED visits) and a sample of 37,541 ED visits made by adults without mental health disorders (87.7% of all adult ED visits). Access data table for Figure 3 at: https://www.cdc.gov/nchs/data/databriefs/db426-tables.pdf#3.
SOURCE: National Center for Health Statistics, National Hospital Ambulatory Medical Care Survey, 2017–2019.

# Types of rates

- Rates with numerators subject to random variability and denominators assumed relatively free from variability from vital statistics, including:
  - National and state death rates; age-specific deaths rates, including infant death rates (deaths for persons under 1 year of age); birth rates, fertility rates.

- Rates with numerators from surveys subject to sampling variability and denominators assumed relatively free from variability, including:
  - National and some subnational health care visit rates.

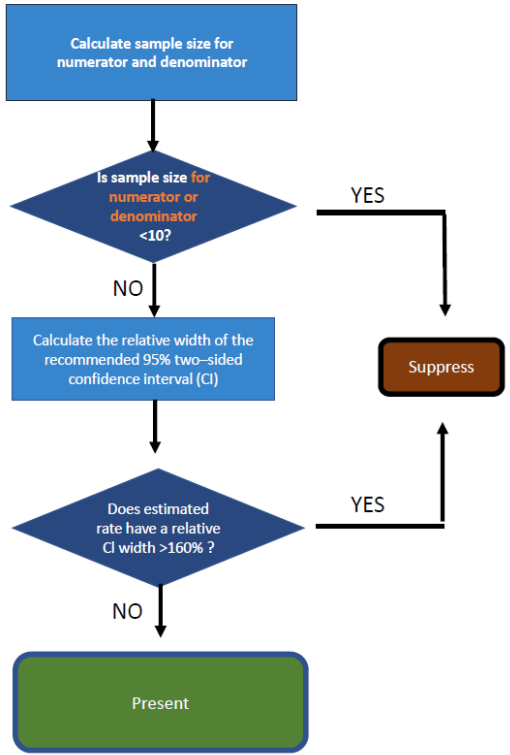
# Types of rates (2)

- Rates with numerators subject to sampling variability and denominators subject to sampling variability, including:
  - Condition specific rates where population denominator is estimated from a survey, such as the National Health Interview Survey.

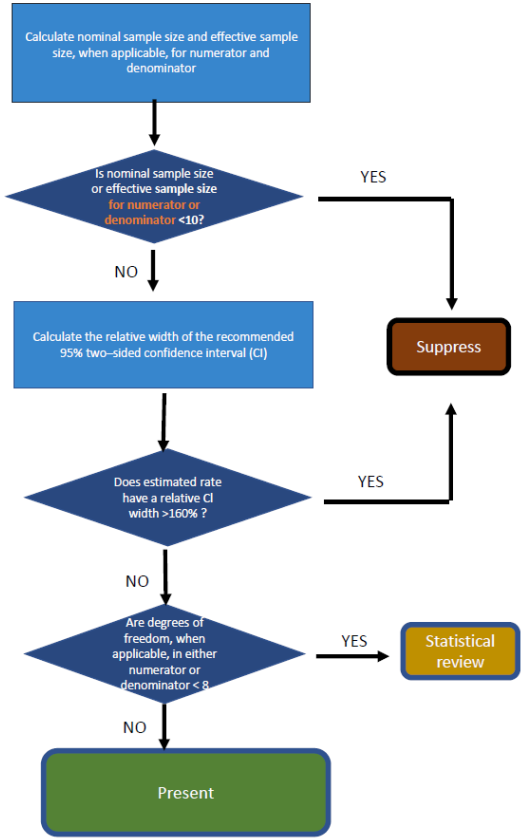
# Presentation Criteria for Rates and Counts

- Minimum sample size or effective sample size
  - Minimum sample size and effective sample size (when applicable) of 10 in the numerator (rates and counts) and in the denominator (rates only).
- Maximum relative width of confidence interval
  - If the sample size criteria are met, calculate a 95% two-sided confidence interval using the appropriate method and obtain its relative width.
  - Appropriate confidence interval method used ensures 95% coverage.
  - Estimated rates should have a relative confidence interval width of 160% or lower.
- Minimum degrees of freedom
  - When applicable, if the degrees of freedom are fewer than 8 then flag for statistical review.

## NCHS Presentation Standards for Rates without Sampling Variability




## NCHS Presentation Standards for Rates with Sampling Variability



# Variability

- Numerators or denominators estimated from a survey are subject to **sampling variability**.
- Even when the actual number of events in the numerator or the size of the population at risk in the denominator is recorded and is free from sampling variability, there is **natural** (or **“stochastic”**) **variability** in the realized values.
  - For numerators or denominators enumerated from vital statistics and free of sampling variability, the number of events (that is, deaths or births) will be assumed to arise from a Poisson distribution.
  - For a denominator that is enumerated from a decennial Census or postcensal/intercensal population estimate there may be some natural variability in the realized value, however such random variation will be considered negligible and will not be considered in calculations.

# Presentation Criteria – Confidence Intervals for Rates without sampling variability (1)

- Numerator subject to random variation and denominator assumed free from variability
  - Calculate gamma interval* where the lower limit is the 0.025 quantile of the standard gamma with $\alpha=x$ and $\beta=1$, where x=number of vital events, and the upper limit is the 0.975 quantile of the standard gamma with $\alpha=x+1$ and $\beta=1$.
  - For age-specific and crude rates, this confidence interval is considered exact.
  - The adjustment used by Fay-Feuer should be applied for age-adjusted vital rates.

* In Excel, the function GAMMA.INV (probability, alpha, beta), with beta set to 1, returns the quantile of the standard gamma distribution for a given probability between 0 and 1. For 95% confidence limits, the probability associated with the lower limit is 0.05/2 = 0.025, and with the upper limit 1 - (0.05/2) = 0.975.

# Presentation Criteria – Confidence Intervals for Rates without sampling variability (2)

- Numerator and denominator subject to random variation (infant mortality rates)
  - Calculate a Student's t confidence interval for the logarithm of the rate and transform to get CI for the rate.

$$\exp\left\{\ln(r) \pm t_{\alpha/2,\ \mathrm{df}} \sqrt{\frac{v}{r^2}}\right\}$$

$$v = \left(\frac{x}{y}\right)^2 \left[\frac{1}{x} + \frac{1}{y}\right]$$

The rate r = x/y, x is number of infant deaths, y is number of live births, v is the variance of r. df=min(x,y)-1

# Presentation Criteria – Confidence Intervals for Rates with sampling variability (1)

- Numerator from survey (e.g., National Health Care Survey) with sampling variability, and denominator assumed free from variability
  - Calculate a Student's t confidence interval for the logarithm of the rate and transform to get CI for the rate.

$$\frac{1}{y}\exp\left\{\ln\left(x\right)\pm t_{\alpha/2,df}\sqrt{\frac{s_x^2}{x^2}}\right\}$$

The rate r = x/y, x is the estimated number of events, y is the population count assumed free of variation, and $s_x^2$ is the standard error of x. df=min($n_x$, $n_{x\_eff}$)-1

# Presentation Criteria – Confidence Intervals for Rates with sampling variability (2)

- Numerator subject to random variation and denominator subject to sampling variability
  - Calculate a Student's t interval for the logarithm of the rate. Parameters for intervals for age-adjusted intervals can be formed using weighted combination of age-specific intervals.

$$\exp\left\{\ln(r) \pm t_{\alpha/2,\ \mathrm{df}}\sqrt{\frac{v}{r^2}}\right\} \qquad v = \left(\frac{x}{y}\right)^2\left(\frac{1}{x}+\frac{s_y^2}{y^2}\right)$$

The rate r = x/y and $s_y$ is the standard error of y. v is the variance of the rate. df=min(x, $n_y$ ,$n_{yeff}$)-1

NOTE: When the denominator is based on the ACS or CPS and subject to sampling variability, the CI uses GVF parameters provided by Census

# Presentation Criteria – Confidence Intervals for Rates with sampling variability (3)

- Numerator and denominator both subject to sampling variability
  - Calculate a Student's t confidence interval for the logarithm of the rate
  - The interval accounts for the sampling variability of the denominator

$$\exp\left\{\ln\left(\frac{x}{y}\right) \pm t_{\alpha/2,df}\sqrt{\frac{s_x^2}{x^2}+\frac{s_y^2}{y^2}}\right\}$$

The rate r = x/y, x and y are survey estimates of outcomes and population. $s_x$ and $s_y$ are the standard errors of x and y. Df=min($n_x$, $n_{x\_eff}$, $n_y$, $n_{y\_eff}$)-1

# Evaluations for Vital Statistics

- Distributional assumptions: comparison of ‘exact’ gamma interval width thresholds, relative standard errors (RSE), and counts
- Age-adjusted death rates: simulation-based comparison of Anderson-Rosenberg method, used at NCHS and in CDC WONDER, with approaches by Fay/Feuer and Tiwari et al, used in U.S. Cancer registries.
- Simulation to examine proposed and previous standards for county-level data by race/Hispanic origin and cause of death
- Simulations evaluating log Student’s t interval for non-constant denominator, including correlation assumptions for numerator and denominator

# Evaluations for National Health Care Surveys

- Simulations based on NAMCS data structure
  - Evaluation of confidence intervals when the denominator is constant
  - Evaluation of confidence intervals by nominal and effective sample sizes and survey years when the denominator is constant
  - Evaluation of the design effect using row proportions and using totals or counts when the denominator is constant
  - Evaluation of rates and age-adjusted rates when both the numerator and the denominator are non-constant

# Next Steps

- Reports
  - The National Center for Health Statistics Data Presentation Standards for Rates and Counts
  - Evaluation of the National Center for Health Statistics Data Presentation Standard for Rates from Vital Statistics and Sample Surveys
- Implementation with NCHS Data Systems with 2023 data
- Web page with worked examples

# Data Presentation Standards Workgroup

- Amy Branum
- Barnali Das
- Danielle Davis
- Brady Hamilton
- Katherine Irimata
- Ken Kochanek
- Don Malec (retired)
- Frances McCarty (SIS, Virginia Tech)
- Jennifer Parker
- Alex Strashny
- Makram Talih (SIS)
- Guangyu Zhang

# Questions?

Jennifer Parker
jdparker@cdc.gov

# Presentation Criteria – Confidence Intervals for Rates with sampling variability (4)

- Denominators subject to sampling variability from American Community Survey
  - Calculate a Student's t interval for the logarithm of the rates with variance estimated using generalized variance function (GVF) with parameters supplied by Census. Parameters for age-adjusted intervals can be formed using weighted combinations of age-specific estimates.

$$\exp\left\{\ln(r) \pm t_{\alpha/2,\ \mathrm{df}}\sqrt{\frac{v}{r^2}}\right\} \qquad v = \left(\frac{x}{y}\right)^2\left[\frac{1}{x} + f\left(a + \frac{b}{y}\right)\right] \text{ and } \frac{v}{r^2} = \frac{1}{x} + f\left(a + \frac{b}{y}\right)$$

Rate r =x/y. x is number of vital events; y is population estimate. v is the variance of r. Parameters for GVF, a and b, provided by Census. df=min(x,y)-1