NCHS Data Presentation Standards for Rates: Vital Statistics

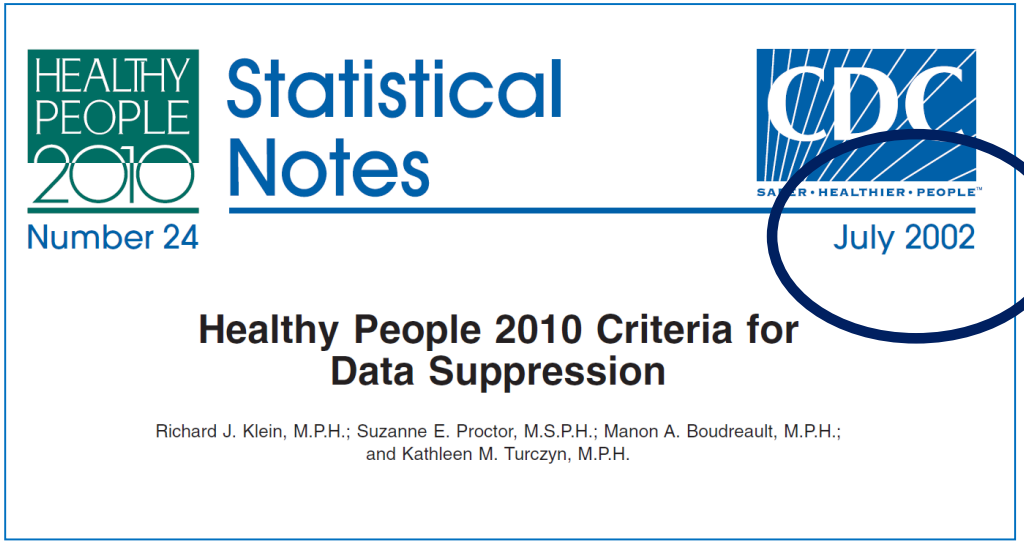
Jennifer D. Parker
Director, Division of Research and Methodology
National Center for Health Statistics

Meeting of the NCHS Board of Scientific Counselors
January 9, 2020

NCHS data products

- **General health data products** with a large set of health outcomes by many population subgroups, often from multiple data sources
  - Interest in many (often predetermined) subgroups, including small groups, multiply crossed categories and geographic areas
  - Some stand-alone estimates, although standard errors/intervals can be available
  - Minimal or no discussion in text

- **Topic specific data products:** Data Briefs; National Health Statistics Reports; journal articles.
  - Choice of subgroups often determined for specific report
    - Expected precision of estimates can be considered (e.g. coarsening age groups, collapsing race/eth or income groups)
  - Standard errors or confidence intervals can be provided and are discussed in the text.



- Listed criteria for about 20 data sources, including surveys and vital statistics.
  - At the time, most standards based on RSE and sample size.
- Did not make recommendations or compare criteria.

The first NCHS workgroup on data presentation standards

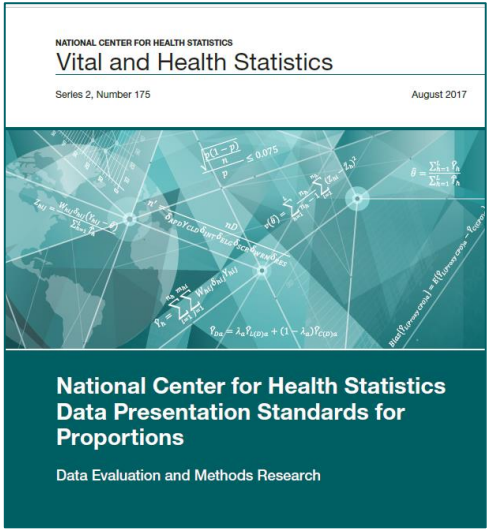
- NCHS workgroup formed in May 2013

- Task
  - Define scope
  - Assess of alternate methods of precision
  - Develop of proposed guidelines
  - Develop of software for implementation across different NCHS data systems
  - Compare former to proposed guidelines across data systems

- Focused on **proportions**, as most common statistic produced by NCHS

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, although not always, led to more estimates presented compared to prior criteria



The second NCHS workgroup on data presentation standard: Rates

- Rates are widely disseminated by NCHS, e.g.
  - Vital statistics
    - Death rates
    - Birth rates
  - Health care statistics
    - Visit rates
    - Hospitalization rates

The second NCHS workgroup on data presentation standard: Rates (2)

- Workgroup was formed in 2018 and includes representatives from DRM, DAE, DVS, DHCS, OCD and a DRM contractor (mathematical statistician)

- Approach
  - Identify current practice for data suppression of rates
  - Define scope
  - Review and examine statistical principles for rates, particularly properties of intervals calculated for rates

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

- Crude rates and age-specific rates

- Age-adjusted rates: linear combination of age-specific rates standardized to standard US population

Examples of rates

- Diabetes Death Rate for 2017
  - the number of deaths attributed to diabetes in 2017, divided by the US resident mid-year population for 2017, multiplied by 100,000

- Outpatient visit rate for preventive care, by region, 2017
  - the number of visits for preventive care divided by the US civilian non-institutionalized mid-year population by region for 2017, multiplied by 1000

Types of rates

- Rates with ‘random’ numerators and ‘fixed’ denominators from vital statistics, including:
  - national and state death rates, birth rates, rates for some subgroups

- Rates with ‘random’ numerators and ‘fixed’ denominators from surveys, including:
  - national and some subnational visit rates

Types of rates (2)

- Rates with ‘random’ numerators and ‘random’ denominators, including
  - some subnational and subgroup rates from vital statistics where population denominator estimated from American Community Survey
  - condition specific rates where population denominator is estimated from survey, such as the National Health Interview Survey
- Infant mortality rates (unlinked, period linked, cohort linked)

First step

- Rates with ‘random’ numerators and ‘fixed’ denominators from vital statistics, including:
  - national and state death rates, birth rates, rates for some subgroups

- Strategy
  - Identify and compare current presentation guidelines, methods of interval estimation, relative standard errors

Intervals

- Crude and age-specific rates
  - Approximate intervals calculated using normal approximation (Wald method: mean ±1.96×SE) for rates and counts with 100 or more events
  - Exact intervals calculated using Poisson distribution for rates with fewer than 100 events – operationalized using “Gamma method”

- Age-adjusted death rates are more complicated
  - Linear combination of Poisson random variables is not Poisson
  - Approximate intervals calculated using normal approximation for rates and counts with 100 or more events
  - Approximate intervals calculated using Poisson distribution for rates with fewer than 100 events – operationalized using “Gamma method” with “DVS” modification

Current presentation guidelines for crude and age-specific vital rates

- If **x < 20** (i.e., x ≤ 19), rate is deemed unreliable and is suppressed
  - Corresponds to relative standard error of ~23%
  - Relative width of gamma interval ~ 96%

Proposed change to presentation standard for age-specific and crude rates

FROM: If **x < 20** (i.e., x ≤ 19), rate is deemed unreliable and is suppressed

TO: If **x < 10*** (i.e., x ≤ 9), rate is deemed unreliable and is suppressed

*Note that x < 10 is the confidentiality cutoff for subnational data

Proposed change to presentation standard for age-specific and crude death rate (2)

- Rationale
  - $x < 10$ offers a single cutoff for suppression to ensure both confidentiality of subnational data and reliability
  - Whereas $x < 20$ is equivalent to RSE ≥ 23%, $x < 10$ is equivalent to RSE ≥ 33%
  - Relative width (RW) of gamma CI corresponding to $x < 10$ is $RW_{gamma} > 140\%$ (up from $RW_{gamma} \geq 96\%$ for $x < 20$)
    - This goes in the same direction as the RW criterion for proportions ($RW_{exact} > 130\%$)

Current presentation guidelines age-adjusted death rates (AADR)

- If $\sum x_i < 20$ (i.e., $\sum x_i \leq 19$), age-adjusted death rate is deemed unreliable and is suppressed

Proposed change to presentation standard for age-adjusted death rate

FROM: If $\sum x_i < 20$ (i.e., $\sum x_i \leq 19$), AADR is deemed unreliable and is suppressed

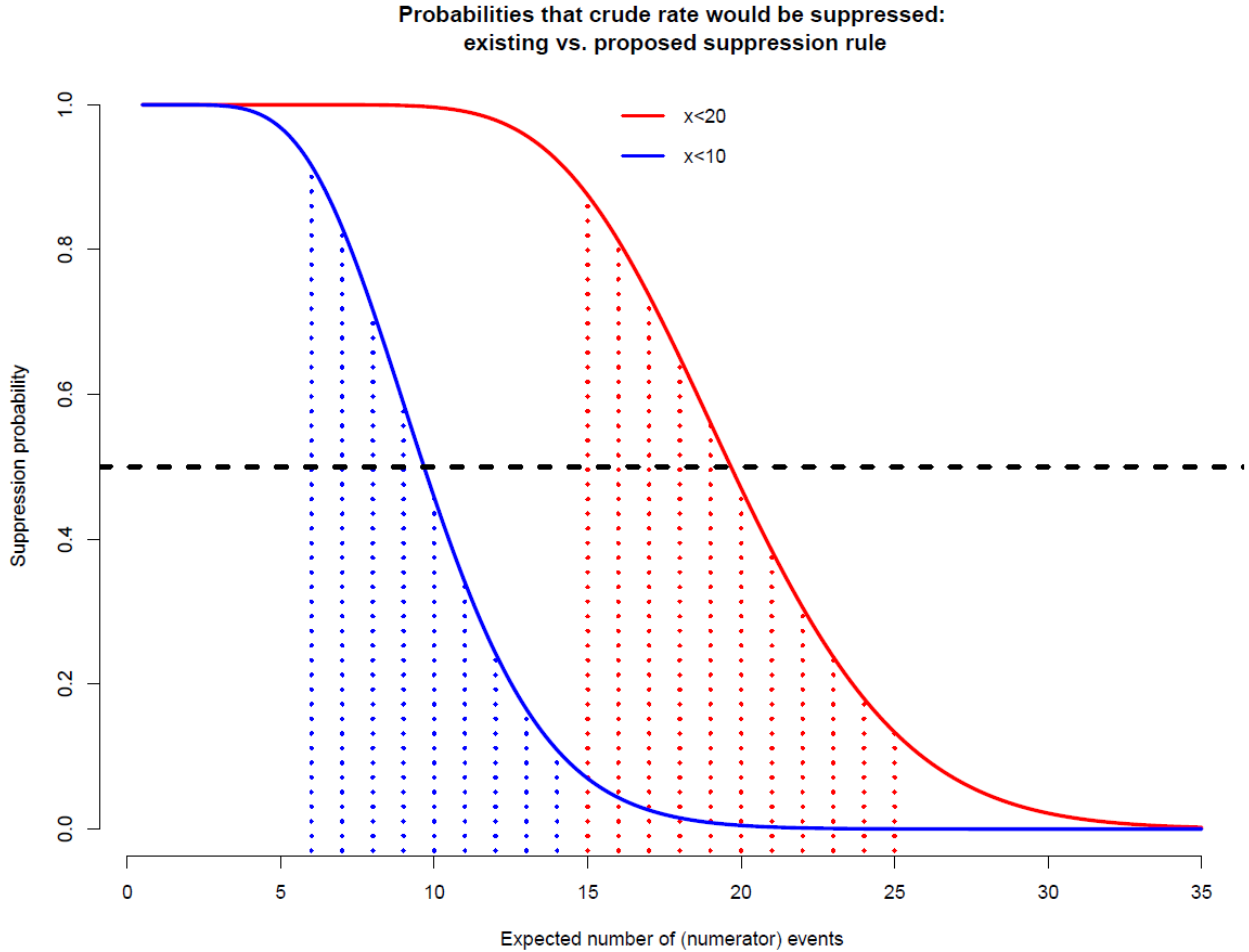
TO: If $\sum x_i < 10$ (i.e., $\sum x_i \leq 9$), AADR is deemed unreliable and is suppressed

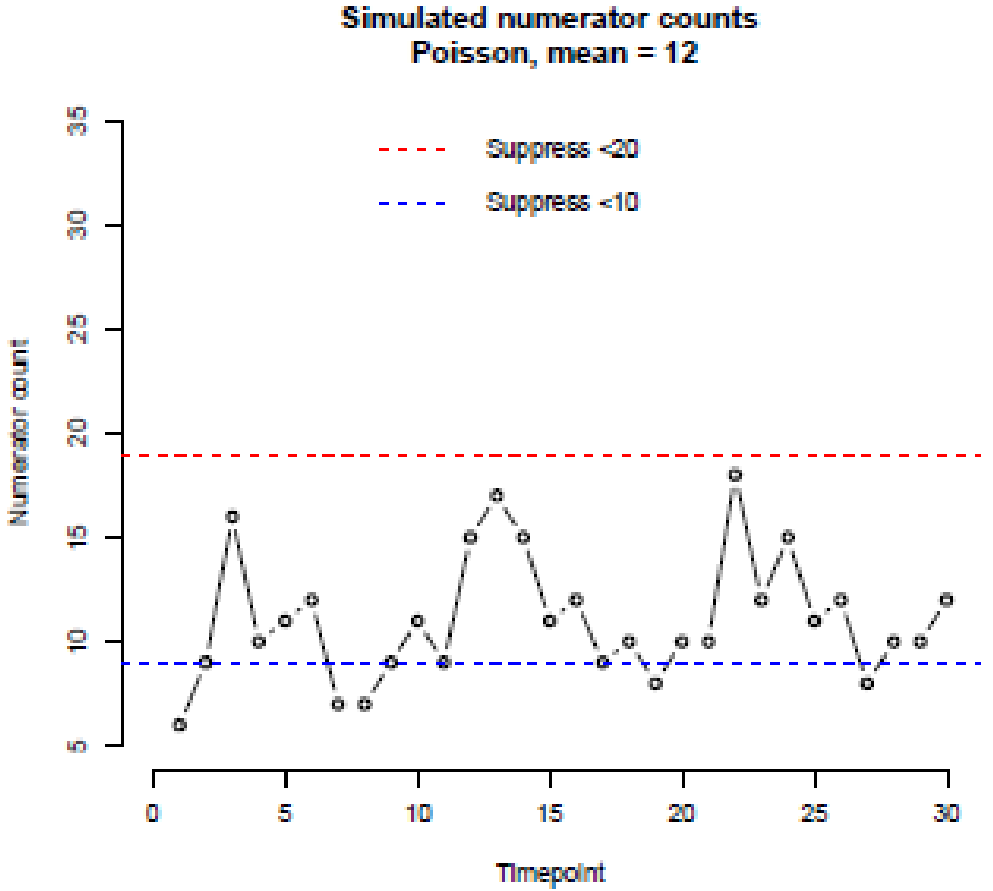
The AADR should continue to be suppressed when the crude rate $\sum x_i / \sum p_i$ needs to be suppressed for reliability or confidentiality reasons

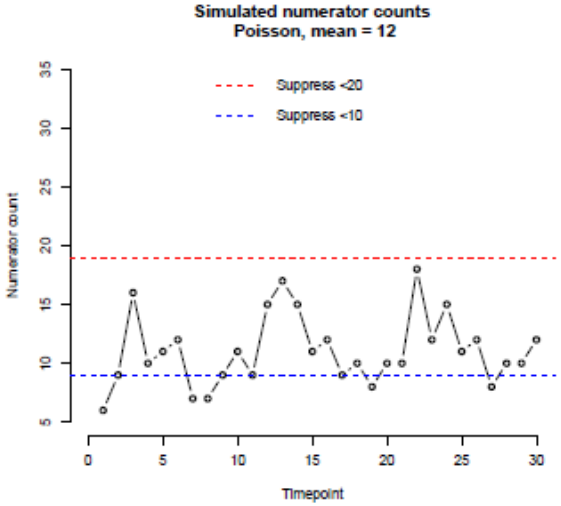
Proposed change to presentation standard for age-adjusted death rate (2)

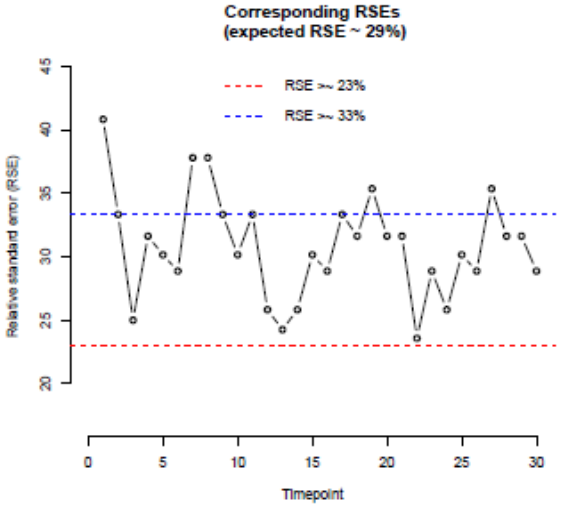
- Rationale
  - The AADR can be computed and presented even if the age-specific death rates are all suppressed (as is done in CDC WONDER and NCHS publications)
  - These cutoffs are consistent with those for the crude rate, and correspond to $\sum x_i < 10$ and $x_{adj} < 10$
  - In simulations, $RW_{DVS\ gamma} > 140\%$ corresponds to $RSE_{AADR} \geq 33\%$

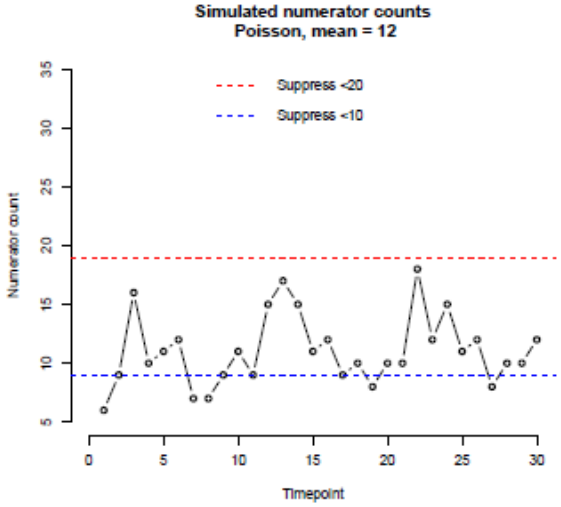
*x_{adj} is defined as $[100/RSE(y)]^2$; $y=\sum w_i(x_i/p_i)$=AADR

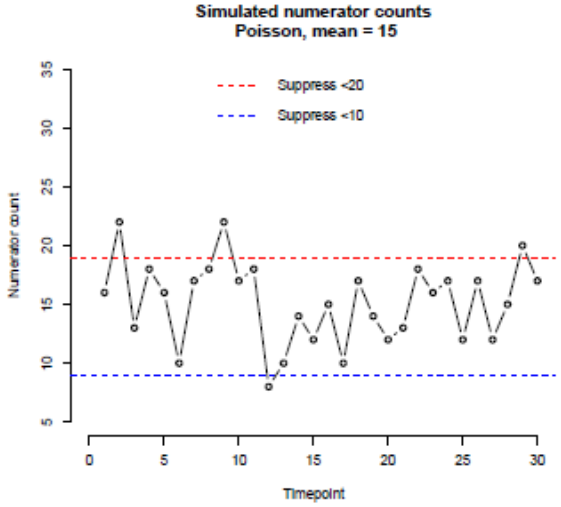

Probabilities that crude rate would be suppressed:
existing vs. proposed suppression rule
x<20
x<10
Suppression probability
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
30
35
Expected number of (numerator) events

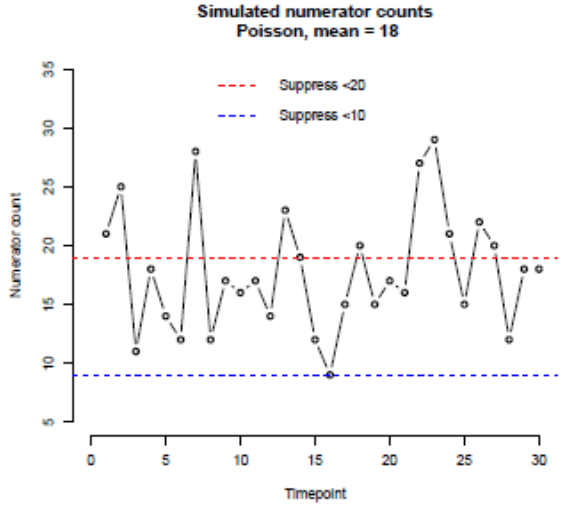

Simulated numerator counts
Poisson, mean = 12
Suppress <20
Suppress <10
Numerator count
5
10
15
20
25
30
35
0
5
10
15
20
25
30
Timepoint

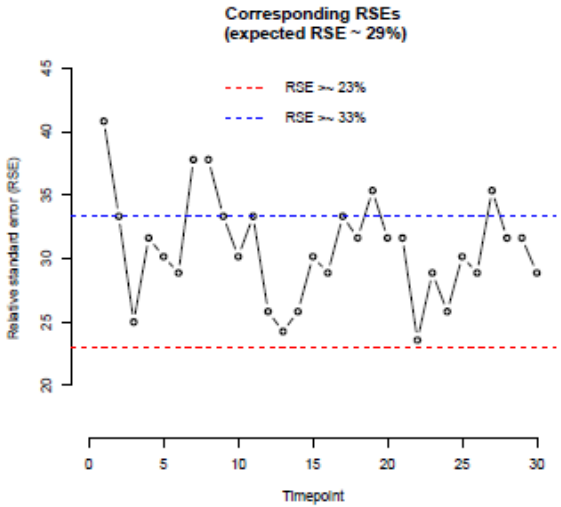

Simulated numerator counts
Poisson, mean = 12
Suppress <20
Suppress <10
Numerator count
Timepoint

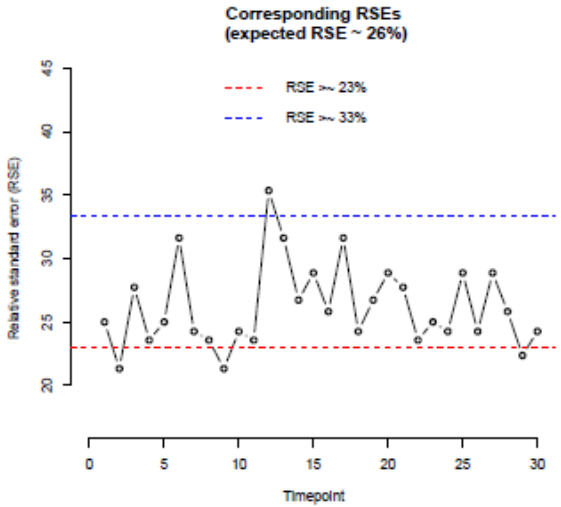

Corresponding RSEs
(expected RSE ~ 29%)
RSE >~ 23%
RSE >~ 33%
Relative standard error (RSE)
Timepoint

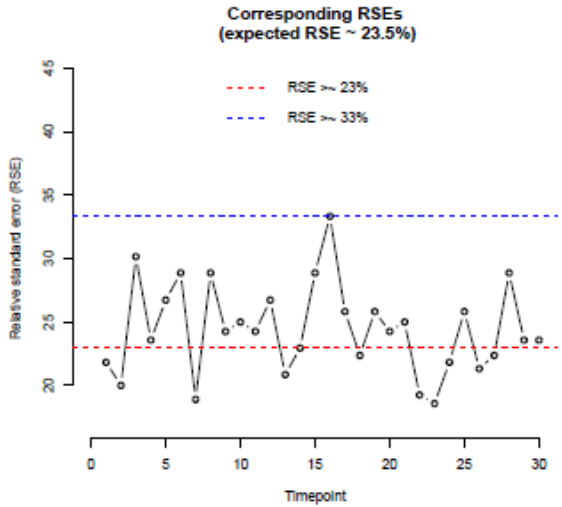

Simulated numerator counts
Poisson, mean = 12
Suppress <20
Suppress <10
Numerator count
Timepoint


Simulated numerator counts
Poisson, mean = 15
Suppress <20
Suppress <10
Numerator count
Timepoint


Simulated numerator counts
Poisson, mean = 18
Suppress <20
Suppress <10
Numerator count
Timepoint


Corresponding RSEs
(expected RSE ~ 29%)
RSE >~ 23%
RSE >~ 33%
Relative standard error (RSE)
Timepoint


Corresponding RSEs
(expected RSE ~ 26%)
RSE >~ 23%
RSE >~ 33%
Relative standard error (RSE)
Timepoint


Corresponding RSEs
(expected RSE ~ 23.5%)
RSE >~ 23%
RSE >~ 33%
Relative standard error (RSE)
Timepoint

Summary

- Recommendation to lower presentation threshold from 20 events to 10 events for crude, age-specific and age-adjusted rates for vital rates with random numerator and fixed denominator
  - Corresponds to RSE increase from 23% to 33%
  - Relative confidence interval width ~ 140%

Current and future activities

- Health care surveys with ‘random’ numerator and fixed denominator
  - Differences from vitals
    - Survey design (e.g. degrees of freedom, sample weights, etc)
    - Multiple visits per person

- Future
  - ‘Random’ denominators

Data Presentation Standards Workgroup

- **Division of Research and Methodology**
  - Jennifer Parker
  - Katherine Irimata
  - Don Malec
  - Guangyu Zhang
  - Makram Talih (SIS)
- **Office of Center Director**
  - Amy Branum
- **Division of Vital Statistics**
  - Brady Hamilton
  - Ken Kochanek
- **Division of Health Care Statistics**
  - Alex Strashny
- **Division of Analysis and Epidemiology**
  - Barnali Das

Thanks to DVS

- Division of Vital Statistics
  - Robert Anderson
  - Elizabeth Arias
  - Danielle Ely
  - Isabelle Horan
  - Hanyu Ni

Jennifer Parker
jdparker@cdc.gov