What is NCHS doing to get “fit”? 

Amy M. Branum, PhD
Acting Associate Director for Science
National Center for Health Statistics
NCHS - who are we?

The US' federal statistical agency for health

Home of NHANES, NHIS, NSFG, health care surveys, and vital statistics
What do we produce?

| ~100 scientific reports and analyses | Dozens of data files | Reams of technical documentation | Hours of technical expertise and guidance |
What are we known for?

Production of high-quality, **reliable**, transparent statistics on the health of the US that educate, inform, and shape health policy.
NCHS’ COVID-19 index page has received 14.2 million hits since March

COVID-19 mortality pages have received 8 million hits
What is reliable? What is “fit”? Why did we need to revisit this?

Q

A

Needed a more consistent approach across data divisions

A

Needed better guidance for staff and data users

A

Narrowly focused on standard errors as measures of variation

A

Statements from ASA and others regarding p-values
Internal workgroup formed

Recognized a need to have clear and transparent presentation criteria that can be broadly and efficiently implemented

Published this report as a culmination of that effort
estimate from a complex survey, the effective sample size, \( n_e \), is defined as the sample size, \( n \), divided by the design effect (7). One approach used to calculate \( n_e \) for sample survey is:

\[ n_e = \frac{n}{D^2} \]

where, in this case, the design effect is:

\[ D = \sqrt{\frac{N-n}{n}} \]

Documentation for specific surveys should be consulted when calculating design effects, and for specific analytic purposes.

If the number of numerator events is 0 or equal to the denominator (the complement of 0 events), the estimated proportion is undefined. In these cases, the sample size should be used to determine whether the mininum sample size criterion is met, and it should also be used for CIs and other computations that include the effective sample size. Because of events or events for everyone in a category can provide important information (e.g., in health outcomes or conditions), estimates based on 0 events (or the complement) the absolute CI and degrees of freedom criteria should be flagged and considered for present statistical review by a clearance official to confirm the validity of the point and interval estimates.

For complex sample surveys, due to sampling design and variability, there may be cases where the effective sample size is greater than the sample size. When sample size is greater than the sample size, the sample size should be used to determine minimum sample size criterion is met, and it should also be used for CIs and other computations that include the effective sample size.

**Standard**

- Estimated proportions should be based on a minimum denominator sample size and an effective denominator sample size (when applicable) less than 30 should be suppressed.
- If the number of numerator events is 0 (or its complement), then the denominator sample size should be used to obtain confidence intervals. If all other criteria are met for presentation, an estimate based on 0 events (or its complement) should be flagged for statistical review by the clearance official. The review result in either the presentation or the suppression of the proportion.

**Confidence Intervals**

The NCHS Data Presentation Standards for Proportions are based on the evaluation of CI widths. CIs provide a way to assess an estimate’s precision, and technical definitions are available in many standard statistics textbooks and Casella and Berger (9). More generally, under repeated sampling, if a proportion and sample, the true value of the proportion is expected to be contained in 95% of the calculated CIs. CIs for proportions are available and the expectation of 95% coverage may not be met, and the effective sample size is greater than the sample size, the sample size should be used to determine whether the minimum sample size criterion is met, and it should also be used for CIs and other computations that include the effective sample size.

\[ p \pm 1.96 \sqrt{\frac{p(1-p)}{n}} \]

\[ p \pm 1.96 \sqrt{\frac{p(1-p)}{n_e}} \]

The SE and width of the CI for the complement of a proportion (1 – \( p \)) are the same as those for the proportion. Consequently, there is a range of proportions where the CI criteria are not met for presentation of the proportion. For these proportions, the relative CI width may indicate that a small proportion of the estimated variance is approximately related to the square root of \( p(1-p) \) for the proportion, the relative CI width may be calculated as the number of PSUs minus the number of strata. This calculation is used in surveys and implemented in survey software, although specific calculations can vary. Generally, default calculations of degrees of freedom from survey software may not be appropriate for subgroups represented in only a subset of PSUs (e.g., some racial and ethnic group-specific estimates) and when calculating annual or survey cycle estimates using a multiyear data file. In these instances, the relevant information should be extracted and the freedom directly calculated to assess estimate precision. The calculation of degrees of freedom is approximately proportional to a chi-square distribution, and the RSE is undefined. In these cases, the sample size should be used to determine whether the minimum sample size criterion is met, and it should also be used for CIs and other computations that include the effective sample size.

- **Complementary Proportions**

  The SE and width of the CI for the complement of a proportion (1 – \( p \)) are the same as those for the proportion. Consequently, there is a range of proportions where the CI criteria are not met for presentation of the proportion. For these proportions, the relative CI width may indicate that a small proportion of the estimated variance is approximately related to the square root of \( p(1-p) \) for the proportion, the relative CI width may be calculated as the number of PSUs minus the number of strata. This calculation is used in surveys and implemented in survey software, although specific calculations can vary. Generally, default calculations of degrees of freedom from survey software may not be appropriate for subgroups represented in only a subset of PSUs (e.g., some racial and ethnic group-specific estimates) and when calculating annual or survey cycle estimates using a multiyear data file. In these instances, the relevant information should be extracted and the freedom directly calculated to assess estimate precision. The calculation of degrees of freedom is approximately proportional to a chi-square distribution, and the RSE is undefined. In these cases, the sample size should be used to determine whether the minimum sample size criterion is met, and it should also be used for CIs and other computations that include the effective sample size.

- **Confidence Intervals**

  The SE and width of the CI for the complement of a proportion (1 – \( p \)) are the same as those for the proportion. Consequently, there is a range of proportions where the CI criteria are not met for presentation of the proportion. For these proportions, the relative CI width may indicate that a small proportion of the estimated variance is approximately related to the square root of \( p(1-p) \) for the proportion, the relative CI width may be calculated as the number of PSUs minus the number of strata. This calculation is used in surveys and implemented in survey software, although specific calculations can vary. Generally, default calculations of degrees of freedom from survey software may not be appropriate for subgroups represented in only a subset of PSUs (e.g., some racial and ethnic group-specific estimates) and when calculating annual or survey cycle estimates using a multiyear data file. In these instances, the relevant information should be extracted and the freedom directly calculated to assess estimate precision. The calculation of degrees of freedom is approximately proportional to a chi-square distribution, and the RSE is undefined. In these cases, the sample size should be used to determine whether the minimum sample size criterion is met, and it should also be used for CIs and other computations that include the effective sample size.
NCHS highly recommends that data users familiarize themselves with the data standards for proportions and implement them when using NCHS data.

Standards can be implemented with a number of software packages.
WHAT ABOUT THAT P-VALUE?

1. **ASSESSING SIGNIFICANCE**

   NCHS has a few challenges

2. **SIGNIFICANT BUT NOT MEANINGFUL CHANGES**

   In our “big data” systems, does a 0.1 percentage point change mean anything even if statistically significant at the p<0.05 level?

3. **BIG CHANGES, LIMITED STATISTICAL POWER**

   Some surveys have had decreasing sample sizes

   Harder to retain sufficient power to detect seemingly large differences

4. **NO CONCRETE GUIDANCE ON NAVIGATING SIGNIFICANCE**

   NCHS has not issued guidance to staff on how to assess statistical significance

   Work in progress
Some difference but lack of significance

These two groups look like they could be different but not statistically significant at \( p<0.05 \)

Instead of saying there is no difference…

“The observed difference between non-Hispanic black and Hispanic adolescents was not significant.”
Very small change, but statistically significant

Is this statistically significant change meaningful?

When longer-term trends are assessed, clearly a decreasing pattern emerges

Should we be assessing annual changes or describing patterns more broadly?

SOURCE: NCHS/NATIONAL VITAL STATISTICS REPORT NO. 68, VOL 13
Very small change, but statistically significant

Is this statistically significant change meaningful?

When longer-term trends are assessed, clearly a decreasing pattern emerges.

Should we be assessing annual changes or describing patterns more broadly?

SOURCE: NCHS/NATIONAL VITAL STATISTICS REPORT NO. 68, VOL 13
Very small change, but statistically significant

Is this statistically significant change meaningful?

When longer-term trends are assessed, clearly a decreasing pattern emerges

Should we be assessing annual changes or describing patterns more broadly?
Are these estimates alike or different?

Original text results only mentioned association with current smokers!

Clearly the association is consistent between former and current smokers but former smokers are just slightly underpowered.

Text revised to acknowledge that association with former smokers was similar although 95% CI did not include 1

<table>
<thead>
<tr>
<th>Smoking status</th>
<th>Estimate (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never</td>
<td>1.00</td>
</tr>
<tr>
<td>Former</td>
<td>1.23 (0.99, 1.54)</td>
</tr>
<tr>
<td>Current</td>
<td>1.20 (1.02, 1.40)</td>
</tr>
</tbody>
</table>

SOURCE: UNNAMED MANUSCRIPT TO PROTECT THE INNOCENT!
What are we doing moving forward?

1. Try and use a common sense approach

2. Advise staff to not live and die by the p-value: if 2 estimates are the same and one is significant but the other isn't, talk about them BOTH!

3. Try and get staff to think more critically, use context, and break away from previous training which relied on p-values

4. Convened a workgroup to provide guidance to staff (derailed by COVID)
Questions

ambranum@cdc.gov

Thank you!