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

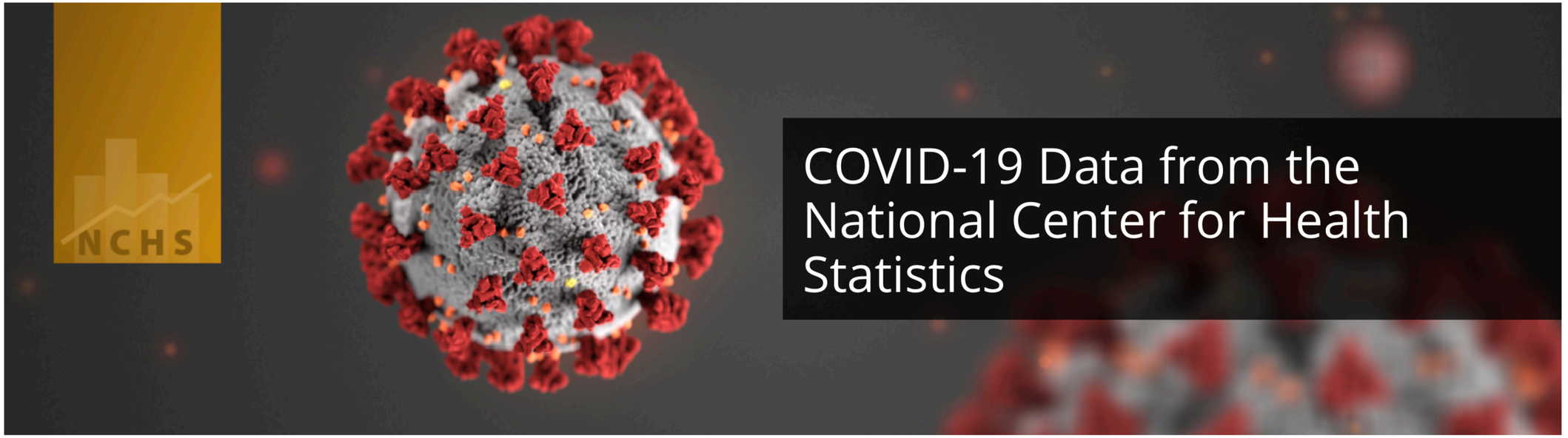
What do we produce?

ANNUALLY

~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



https://www.cdc.gov/nchs/covid19/index.htm

NCHS' COVID-19 index page has received 14.2 million hits since March

COVID-19 mortality pages have received 8 million hits

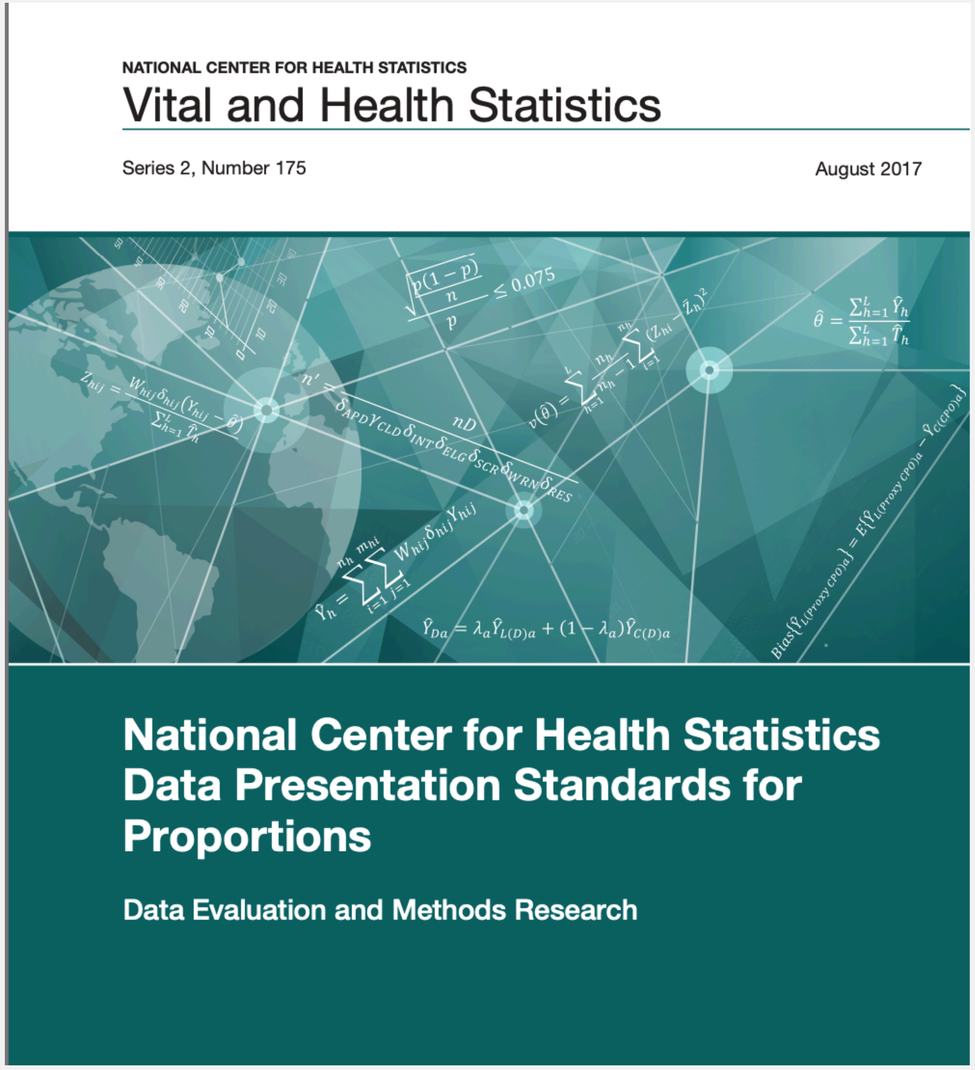
Q **What is reliable? What is “fit”? Why did we need to revisit this?**

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

WWW.CDC.GOV/NCHS/DATA/SERIES/SR_02/SR02_175.PDF

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:

where, in this case, the design effect is:

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 v
As a result, the estimated variance of the proportion will be 0, and the effective sample si
undefined. In these cases, the sample size should be used to determine whether the minin
size criterion is met, and it should also
be used for CIs and other computations that include the effective sample size. Because ob
events or events for everyone in a category can provide important information (e.g., in th
rare health outcomes or conditions), estimates based on 0 events (or the complement) tha
absolute CI and degrees of freedom criteria should be flagged and considered for presenta
statistical review by a clearance official to confirm the validity of the point and interval e

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 v
minimum sample size criterion is met, and it should also be used for CIs and other compu
include the effective sample size.

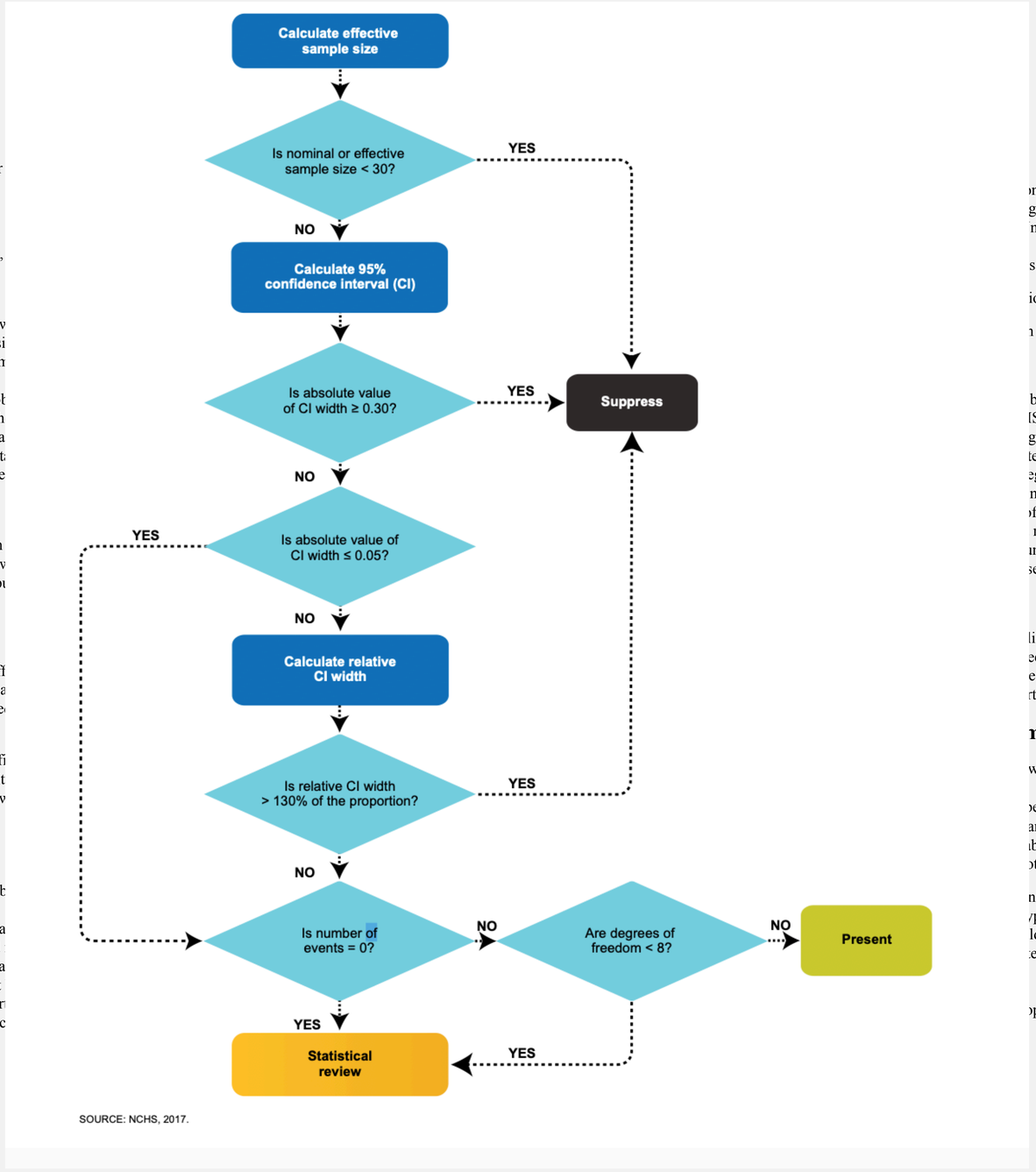
Standard

● Estimated proportions should be based on a minimum denominator sample size and eff
denominator sample size (when applicable) of 30. Estimates with either a denominator sa
an effective denominator sample size (when applicable) less than 30 should be suppresse

● If the number of numerator events
is 0 (or its complement), then the denominator sample size should be used to obtain confi
intervals. If all other criteria are met for presentation, an estimate based on 0 events (or it
complement) should be flagged for statistical review by the clearance official. The reviev
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 ab
CI widths. CIs provide a way to assess
an estimate's precision, and technical definitions are available in many standard statistica
and Casela 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 calcula
calculate CIs for proportions are available and the expectation of 95% coverage may not
some conditions. Methods used to calculate a CI lead to undercoverage if the true propor
expected number of intervals (e.g., less than 95%). Conversely, methods are considered c



n of the estimated variance is approximately related to the square root
g SEs with low precision to assess estimated proportions may lead to poor measures
nder certain conditions, the variance estimate is approximately proportional to a chi-

sample survey can be approximated as

ions based

n 8 degrees of freedom have

be calculated as the number of PSUs minus the number of strata. This calculation is used
lS surveys and implemented in survey software, although specific calculations can vary
ges. However, default calculations of degrees of freedom from survey software may not
te for subgroups represented in only a subset of PSUs (e.g., some racial and ethnic
egion-specific estimates) and when calculating annual or survey cycle estimates using a
multicycle data file. In these instances, the relevant information should be extracted and
f freedom directly calculated to assess estimate precision. The calculation of degrees of
measure of precision for the SE may not be applicable for all surveys (see survey-
mentation) and does not apply to vital statistics. For additional information on degrees
see Korn and Graubard (13) and Valliant and Rust (14).

licable for complex surveys, if the sample size and CI criteria are met for presentation
ees of freedom are fewer than 8, then the proportion should be flagged for statistical
e clearance official. This review may result in either the presentation or the suppression
rtion.

nentary Proportions

width of the CI for the complement of a proportion $(1 - p)$ are the same as those for the

ed in previous sections, relative measures for the smaller proportion are much larger
arger complement. Consequently, there is a range of proportions where the CI criteria
bility. For these proportions, the relative CI width may indicate that a small proportion
ot.

n, the larger proportion may be the most salient measure, while for others, the smaller
ypically, both proportions are not shown (e.g., only
ld be shown, not both the proportion with and the proportion without health insurance).
ed proportion can be determined by subtraction, consideration of the precision of the

portions if one of the proportions is identified as unreliable. However, this practice may

Stata J. Author manuscript; available in PMC 2019 Dec 6.
Published in final edited form as:
Stata J. 2019 Sep; 19(3): 510–522.
Published online 2019 Sep 20. doi: 10.1177/1536867X19874221

PMCID: PMC6896998
NIHMSID: NIHMS1043397
PMID: 31814807

kg_nchs: A command for Korn-Graubard confidence intervals and National Center for Health Statistics' *Data Presentation Standards for Proportions*

Brian W. Ward

Standards can be implemented with a number of software packages

NCHS highly recommends that data users familiarize themselves with the data standards for proportions and implement them when using NCHS data

STANDARDS FOR RATES IN PROGRESS AND COMING SOON!

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

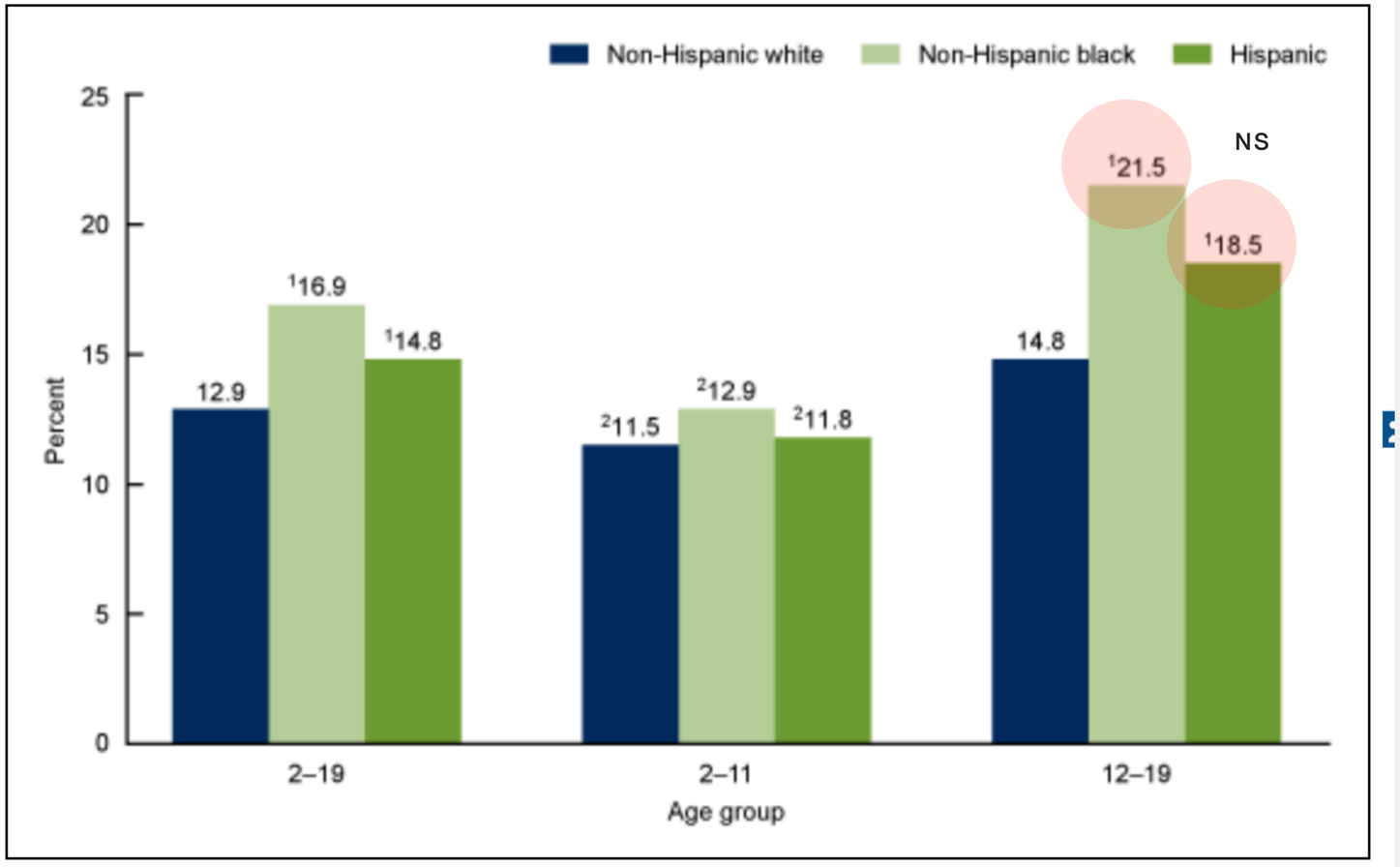
Some surveys have had decreasing sample sizes

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



[1]SIGNIFICANTLY DIFFERENT FROM NON-HISPANIC WHITE CHILDREN AND ADOLESCENTS OF THE SAME AGE GROUP.
[2]SIGNIFICANTLY DIFFERENT FROM THOSE AGED 12–19 YEARS.

SOURCE: NCHS DATA BRIEF #375, FIGURE 2

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.”

Low-risk cesarean delivery—The low-risk cesarean delivery rate also decreased in 2018, from 26.0% to 25.9% for 2017–2018. The low-risk cesarean delivery rate is cesarean delivery among nulliparous (first birth), term (37 or more completed weeks based on the obstetric estimate), singleton (one fetus), cephalic (head first) births. See Tables 13, 14, 17, and 18 for details by age and race and Hispanic origin of the mother.

LOW-RISK CESAREAN RATES BY YEAR

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

Low-risk cesarean delivery—The low-risk cesarean delivery rate also decreased in 2018, from 26.0% to 25.9% for 2017–2018. The low-risk cesarean delivery rate is cesarean delivery among nulliparous (first birth), term (37 or more completed weeks based on the obstetric estimate), singleton (one fetus), cephalic (head first) births. See Tables 13, 14, 17, and 18 for details by age and race and Hispanic origin of the mother.

SOURCE: NCHS/NATIONAL VITAL STATISTICS REPORT NO. 68, VOL 13

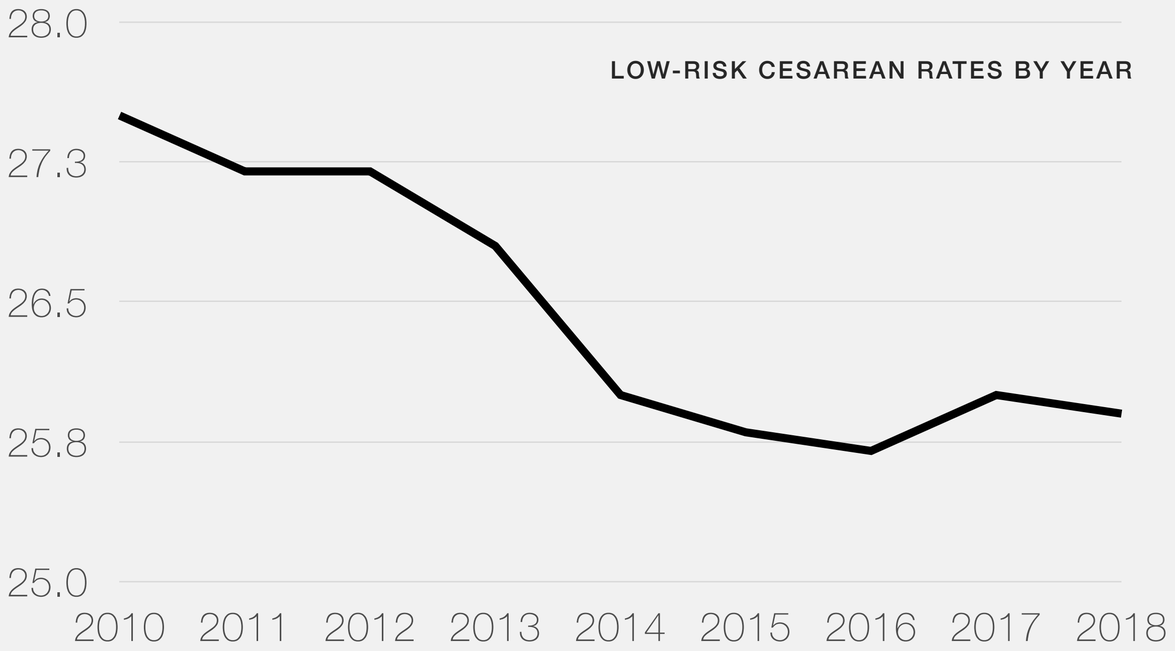
Very small change, but statistically significant

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?

Low-risk cesarean delivery—The low-risk cesarean delivery rate also decreased in 2018, from 26.0% to 25.9% for 2017–2018. The low-risk cesarean delivery rate is cesarean delivery among nulliparous (first birth), term (37 or more completed weeks based on the obstetric estimate), singleton (one fetus), cephalic (head first) births. See Tables 13, 14, 17, and 18 for details by age and race and Hispanic origin of the mother.



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?

Smoking status	
Never	1.00
Former	1.23 (0.99, 1.54)
Current	**1.20 (1.02, 1.40)**

SOURCE: UNNAMED MANUSCRIPT TO PROTECT THE INNOCENT!

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

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!