HOST: NHANES – the National Health and Nutrition Examination Survey - is designed to assess the health and nutritional status of adults and children in the United States. The survey uses complex sampling design to ensure that the data collected are nationally representative. And it also combines information collected across several different survey components. During a home interview, participants provide information on demographic characteristics, health conditions, and risk factors and behaviors, such as which dietary supplements and prescription medications they are using. Then, participants are invited to travel to a nearby mobile examination center to participate in a health exam, in which they undergo tests, provide lab specimens, and take part in additional interviews. After the exam, participants may be contacted again to participate in post-exam content. This could include activities such as dietary recall, interviews, or wearing a physical activity monitor. The survey content varies over time and covers a wide variety of health conditions and public-health topics. Conducting examinations, along with health interviews, provides data that is invaluable for public health. But it also poses operational and statistical challenges.

Dr. Lara Akinbami, a pediatrician and medical officer with the Division of Health and Nutrition Examination Surveys at NCHS, discussed some of those challenges in a webinar last year:

AKINBAMI: NHANES is conducted in 15 sites per year due to the intricate field operations of the survey. The mobile exam centers, or MECs, must be driven to each new site and set up and maintained according to exact specifications. Teams includes interviewers, clinicians, technicians, and engineers live in the field full time as they travel among the different survey locations over the year. Each MEC contains a mobile laboratory with all the equipment needed for specimen processing and storage until specimens can be shipped to laboratories for testing. On-site testing is also performed for some health measures to provide immediate results to participants. There's also other equipment for medical testing in the MECs. For example, a spirometer has been used to measure lung function and a sound-isolating room is used to test hearing. Depending on which health exams and measures are being performed, equipment can be swapped in and out of the MECs. The range of pre-pandemic activities that occurred in the MECs was broad. These included body measures, such as weight and height, Blood-pressure measurements, DEXA scans to assess bone density, oral health exams, and phlebotomy and urine collection to collect specimens for a wide array of lab tests. In addition, participants responded to additional interviews, such as audio computer-assisted self-interviews for more sensitive topics that included reproductive health and alcohol and substance use. The MECs provide a way to standardize protocols, equipment, and exam environments across different locations and across time, so that results are comparable. This allows for more accurate interpretation of health differences between groups and of trends over time. NHANES began continuous field operations in 1999.

And, although data collection continued from year to year, data were released in two-year cycles. Each two-year cycle is drawn from a multiyear sample design. These sample designs have changed over time to keep up with changes in the U.S. population. For instance, the 2015-to-

2018 sample design selected 60 locations to be visited over four years. The 2015-to-2016 data-collection cycle visited the other 30. Although data are available for two-year cycles, NHANES advises combining cycles together into four-year data sets to calculate reliable estimates for subgroups. For example, estimating the prevalence of a health condition by age group separately for men and women, or for race and Hispanic-origin groups among children, is best done with a four-year data set.

HOST: Dr. Akinbami also discussed in the webinar how the pandemic impacted NHANES data collection.

AKINBAMI: So, like almost everything else, NHANES was affected by the COVID-19 pandemic. The 2019-to-2022 sample design also chose 60 locations to be sampled over four years. NHANES entered the field in 2019 with a plan of visiting the first 30 locations in the 2019-to-2020 data collection cycle. In March 2020, a growing number of cases of COVID-19 disease were being reported to CDC. This suggested that community spread was occurring. Widespread shutdowns had not yet occurred, but the environment was starting to change. For example, mobility data show that, during March, normal patterns and movements started to decline. The NHANES program needed to decide whether continuing field operations posed a risk of coronavirus transmission to participants and staff and their close contacts. On March 16, field operations for NHANES were suspended. And, although it wasn't clear at the time, this meant that the 30 locations planned for the 2019-to-2020 cycle would not all be visited. When field operations were stopped in March of 2020, the survey had been to 18 of 30 planned locations. And, as 2020 progressed, it was clear that there was no feasible way to resume in-person exams. The potentially long pause before field operations could be resumed raised questions about how a break in data collection would affect estimates of health conditions. Resuming data collection when it was safe to do so would mix pre-pandemic data and pandemic data together and potentially introduce bias into the estimates, especially for a two-year cycle that would have to be extended.

Therefore, it was decided not to collect more data for this cycle. Because no additional data would be collected, the 2019-to-March 2020 sample was not nationally representative. There was no method to create sample weights using the 2019-to-2022 sample design. Additionally, publicly releasing the data for fewer than 30 locations could pose disclosure risks for participants. However, the data that were collected represented a significant investment by survey participants, the federal government, and collaborators; and simply not using the data wasn't an option. So, a solution was found in the creation of a pre-pandemic data file. The 2017-to-2018 two-year cycle contained a complete sample and was nationally representative. It could be used to build a larger data set. And methodology to combine a probability sample with a nonprobability sample was used but adapted to this situation. The probability sample in this case was the 2017-to-2018 sample. And, rather than a nonprobability sample, the 2019-to-March 2020 sample was a partial probability sample, because it was selected based on the 2019-to-2022 sample design.

So here's an overview of how a 2017-to-March 2020 pre-pandemic data set was created and some analytic considerations when working with the data. The 2015-to-2018 sample design

specified the locations chosen for the 2017-to-2018 data collection cycle. And, as we mentioned previously, all 30 locations were visited in 2017 to 2018. The 2019-to-2022 sample design specified 30 locations that were supposed to have been visited in the 2019-to-2020 data collection cycle and only 18 were visited. Combining the 2017-to-2018 sample with the 2019-to-March 2020 sample posed a problem. The 2015-to-2018 and the 2019-to-2022 sample designs were different because the 2019-to-2022 sample design was updated to reflect the changing United States. So the chosen solution was to pick one of these sample designs. Because the 2017-to-2018 data collection cycle fully adhered to the 2015-to-2018 sample design, this design was chosen. The 18 sites that were visited in 2019 to March 2020 were reassigned to the 2015-to-2018 sample design. Now that a design was chosen, the sample weights could be calculated. However, there were still some issues that needed to be resolved. The 2019-to-March 2020 locations didn't line up exactly with the 2015-to-2018 sample design. The result was that some portions of the country were underrepresented in the data. An adjustment factor was used to equalize representation over the sites visited from 2017 to March 2020. And, once that was done, interview weights and exam weights were then calculated using previous methodology. Extensive assessments confirmed that the final sample was nationally representative by making demographic comparisons to the American Community Survey, which is a population survey administered by the U.S. Census.

HOST: Dr. Akinbami concludes by discussing some important analytic considerations for users of the data.

AKINBAMI: The resulting 2017 to March 2020 pre-pandemic data can be used to calculate nationally representative estimates of health conditions and behaviors. It can be used like the previously released data sets for two-year cycles. However, the data from the partial 2019-to-March 2020 cycle by themselves are not nationally representative.

Therefore 2017-to-2018 data cannot be compared to the 2019-to-March 2020 data. And remember that, because the 2019-to-March 2020 data did not conform to the 2019-to-2020 survey design, no separate survey weights could be constructed for this cycle. It is not appropriate to use the 2017-to-March 2020 pre-pandemic weights for the partial sample collected in 2019 to March 2020. The weight adjustment that was applied to the 2017-to-March 2020 data was designed for overall estimates but not necessarily for subgroups. So, therefore, when 2017-to-March 2020 estimates for subgroups are compared to earlier estimates, trends should be interpreted with caution. For example, when the adjustment factor and other measures were applied to the survey weights, national representation by sex

was achieved and so is representation by age. But some sex-specific age

groups, for example, may have larger variation in estimates depending on how the participants are distributed across survey locations.

MUSICAL BRIDGE

HOST: In part two of this feature on pre and post pandemic NHANES, Dr. Bryan Steirman discusses a published report on health estimates from this NHANES data set. This webinar is accessible on the NCHS web site.

HOST: On April 12, NCHS released its [quarterly mortality](#) data on several leading causes of death, with disease-related mortality rates featured through the third quarter of 2021. The web feature ["Stats of the States"](#) was also updated the same day. "Stats of the States" features key vital statistics on topics such as Births, Deaths, and Marriages & Divorces by state. Users can rank States according to rates, either highest to lowest or lowest to highest. This data visualization was updated with final 2020 data for all these measures. Included in each State fact sheet are the 10 leading causes of death for each state, which always presents some interesting variation from state to state. 2020, of course, features the introduction of COVID-19 as the 3rd leading cause of death in the U.S. And at the state level, COVID-19 indeed was the 3rd leading cause of death in 44 states and DC. As for the other states, COVID was 4th among leading killers in 3 states: Alaska, New Hampshire, and Utah. The virus was the 5th leading cause of death in 2 states: Washington and West Virginia., as well as the 7th leading cause of death in 2 states: Hawaii and Oregon, and the 8thth leading cause of death in 2 states: Maine and Vermont. Provisional data for 2021 suggests some changes to those rankings, based on regional outbreaks of the virus.

Finally, on April 20th, NCHS released a new [report](#) comparing dental utilization rates among adults in 2019 with the arrival of the pandemic in 2020.

Next month promises to be a more active month of data releases for NCHS, including full-year 2021 drug overdose death data, full year 2021 birth data, and a new report on sexual orientation differences in access to care and health status, behaviors and beliefs.

MUSIC OUT