All official NCHS BSC documents are posted on the BSC website
(https://www.cdc.gov/nchs/about/bsc/bsc_meetings.htm)

Department of Health and Human Services (HHS)

Board of Scientific Counselors
National Center for Health Statistics
Centers for Disease Control and Prevention
May 19, 2021

Meeting Summary

The Board of Scientific Counselors (BSC) convened via Zoom on May 19, 2021. The virtual meeting was open to the public (via Zoom).

Board Members Present

Linette T. Scott, M.D., M.P.H., Chair, BSC
Kennon R. Copeland, Ph.D.
Robert M. Hauser, Ph.D.
Scott H. Holan, Ph.D.
Helen G. Levy, Ph.D.
R. John Lumpkin, M.D., M.P.H.
Bradly A. Malin, Ph.D.
Sally C. Morton, Ph.D.
Lucila Ohno-Machado, M.D., Ph.D.
Kristen M. Olson, Ph.D.
Andy Peytchev, Ph.D.
C. Matthew Snipp, Ph.D.
Gretchen Van Wye, Ph.D., M.A.

CDC/NCHS Panelists

Brian Moyer, Ph.D., Director, National Center for Health Statistics (NCHS)
Sayeedha Uddin, M.D., M.P.H., Designated Federal Officer, NCHS, BSC
Dan Jernigan, M.P.H., Acting Deputy Director for Public Health Science and Surveillance, CDC
Irma Arispe
Stephen Blumberg
Amy Branum
James Craver
Carol DeFrances
Lisa Mirel
Kiana Morris
Gwen Mustaf
Michelle Osterman
Lauren Rossen
Paul Sutton

Other Attendees

Dzifa Adjaye-Gbewonyo
Viviana Aguila
Farida Ahmad
Lara Akinbami
Jonathan Aram
Elizabeth Arias
Irma Arispe
Jill Ashman
Rihem Badwe
Brenda Baker
Bryan Bassig
Veronica Benson
Vladislav Beresovsky
Andres Berruti
Greg Binzer
Jonaki Bose
Matthew Bramlett
Amy Branum
Debra Brody
Sherry Brown-Scoggins
Verita Buie
Anjani Chandra
Te-Ching Chen
Jodi Cisewski
Jay Clark
Robin Cohen
Catina Conner
Traci Cook
Jenny Couse
Lauren Creamer
James Dahlhamer
Jerry Del Rosso
Anne Driscoll
Morgan Earp
Danielle Ely
Dorothy Farrell
Steven Fink
Sheila Franco
Alicia Frasier
Chris Freedman
Matthew Garnett
Debra Gilliam
Renee Gindi
Cordell Golden
Heidi Guyer
Brady Hamilton
Jon Hannings
Althelia Harris
Camillia Harris

Yulei He
Holly Hedegaard
Elizabeth Heitz
Kevin Heslin
Alston Hildreth
Rebecca Hines
Jacquie Hogan
Travis Hoppe
Isabelle Horon
John Hough
Rebecca Hu
David Huang
Ed Hunter
Katherine Irimata
Olga Joos
David Justice
Jenny Justice
Jessica Keralis
Kelly Keyes
Karen Knight
Ellen Kramarow
Nataliya Kravets
Peter Kyriacopoulos
Florence Lee
Jessica Lendon
Xianfen Li
Anthony Lipphardt
Christine Lucas
Aaron Maitland
Donald Malec
Crescent Martin
Joyce Martin
Gladys Martinez
Meredith Massey
Jody McLean
Justin Mezetin
Donna Miller
Kristen Miller
Leyla Mohadjer
Mary Moien
Jennifer Moore
Chris Moriarity
Alanna Moshfegh
Zakia Nelson
Jim Nowicki
Tatiana Nwankwo
Kathleen O'Connor
Anna Oganyan
Dagny Olivares

Jennifer Parker
Van Parsons
Kavita Patel
Ryne Paulose
Sue Pedrazzani
Steve Pierson
John Pleis
Paul Pulliam
Anthony Quintana
Daniela Relf
Cynthia Reuben
Askari Rizvi
Vincent Rome
Dorothy Roper
Charles Rothwell
Valerie Ryan
Ioredana Santo
Susan Schappert
Jeannine Schiller
Paul Schroeder
Steven Schwartz
Iris Shimizu
Sandra Smith
Zachary Smith
Merianne Spencer
Marietta Squire
Pamela Stephenson
Tammy Stewart-Prather
Bryan Stierman
Yu Sun
Makram Talih
Chally Tate
Tim Tilert
Anjel Vahratian
Claudia Valenzuela
Rajesh Virkar
Lisa Wagner
Meagan Walters
Brian Ward
Margaret Warner
Julie Weeks
Jean Williams
Sonja Williams
Stephanie Willson
David Woodwell
Jianmin Xu
Alana Yick
Carla Zelaya
Cindy Zhang

Qiming He	Vera Osidach	
		Call-in Participants (*n*=12)
Closed Captioner	Sabira Mohammed (Meeting Host, Rose Li & Associates (RLA))	Mike Kavounis (Presentation Tech, RLA)
		Dana Glei (Minutes, RLA)

List of Abbreviations

API	Application Programming Interface
BSC	Board of Scientific Counselors
CDC	Centers for Disease Control and Prevention
CHC	Community Health Centers
COVID-19	Coronavirus Disease 2019
DAE	Division of Analysis and Epidemiology
DHANES	Division of Health and Nutrition Examination Surveys
DHCS	Division of Health Care Statistics
DMI	Data Modernization Initiative
EHR	Electronic Health Records
FTP	File Transfer Protocol
FY	Fiscal Year
HHS	Health and Human Services
NAMCS	National Ambulatory Medical Care Survey
NCHS	National Center for Health Statistics
NHANES	National Health and Nutrition Examination Survey
NHCS	National Hospital Care Survey
NHIS	National Health Interview Survey
NVSS	National Vital Statistics System
P20	Production 2020
PCORTF	Patient-Centered Outcomes Research Trust Fund
PII	Personally Identifiable Information
PPRL	Privacy Preserving Record Linkage
PSU	Primary Sampling Units
RDC	Research Data Center
RLA	Rose Li Associates
VDE	Virtual Data Enclave

Action Steps

- Dr. Moyer welcomes input from the BSC regarding the collection of demographic data in the context of promoting health equity.
- Dr. Moyer encouraged the BSC to contribute feedback regarding the Strategic Plan.
- The BSC voted unanimously to endorse the opinions submitted by the BSC National Hospital Ambulatory Medical Care Survey (NAMCS) Workgroup and for the BSC to formally issue those recommendations.
- Dr. Moyer requested feedback from the BSC regarding the best use for Workgroups in the future.
- Future BSC meeting dates for the remainder of 2021: October 22.

Presenters

Brian Moyer, Ph.D., Director, NCHS
Kiana C. Morris, M.B.A., C.M.C.P., Acting Director, Office of Planning, Budget and Legislation, NCHS
Daniel Jernigan, M.D., M.P.H., Acting Deputy Director for Public Health Science and Surveillance, CDC
James Craver, M.A.A., Special Assistant to the Center Director, NCHS
Lisa Mirel, M.S., Chief, Data Linkage Methodology and Analysis Branch, Division of Analysis and Epidemiology (DAE)
Stephen Blumberg, Ph.D., Director, Division of Health Interview Statistics
Andy Peytchev, Ph.D., BSC Member, Chair, PHSPMDP Workgroup
Lauren Rossen, Ph.D., Division of Research and Methodology
Michelle Osterman, M.H.S., Division of Vital Statistics
Paul Sutton, Ph.D., Deputy Director, Division of Vital Statistics
Carol DeFrances, Ph.D., Acting Director, Division of Health Care Statistics (DHCS)
John Lumpkin, M.D., Chair, NAMCS Workgroup

Welcome, Introductions, and Call to Order

Linette T. Scott, M.D., M.P.H., Chair, BSC
Sayeedha Uddin, M.D., M.P.H., Designated Federal Officer, NCHS, BSC

Dr. Uddin conducted roll call, asking members to introduce themselves and state conflicts of interest. None of the BSC members stated a conflict of interest.

Dr. Scott welcomed new BSC members.

NCHS Director's Update

Brian Moyer, Ph.D., Director, NCHS

Departing BSC Members and New Members

Dr. Moyer thanked the departing members of the BSC (Drs. Hauser, Scott, and Van Wye) for their service and valuable contributions. He also welcomed incoming Board members (Drs. Malin, Ohno-Machado, and Snipp). Next, he introduced the new Acting Deputy Director for Public Health Science and Surveillance, Dr. Daniel Jernigan, who replaced Dr. Chesley Richards, who retired in 2020. One of Dr. Jernigan's areas of focus is leading the Data Modernization Initiative (DMI). Dr. Moyer bid farewell to Dr. Uddin, who is assuming a new position as Chief Health Information Officer for the Assistant Secretary for Preparedness and Response.

New Administration Priorities

Key priorities for the new Biden administration are issues related to COVID-19, health equity, climate change, and the economy.

NCHS has placed a high priority on producing COVID-19-related data more quickly. Recently, NCHS released new COVID-19-related data from the National Hospital Care Survey (NHCS). Although those data are not nationally representative, they indicate that inpatient encounters with COVID-19 shifted from urban areas during March-May 2020 to rural areas after November 2020.

NCHS has been having internal discussions about its role with respect to health equity. Dr. Moyer welcomes input from the BSC regarding the collection of demographic data. The 2021 HHS Data Council, which is co-chaired by NCHS and the Assistant Secretary for Planning and Evaluation, has formed new subcommittees to focus on new priorities. One of those priorities is data sharing and linkage across NCHS, HHS, and the broader statistical community. NCHS has a representative on each of the new subcommittees.

As part of the Evidence Act, the federal government is developing a standard application process that will allow researchers to submit one application to gain access to any confidential datasets across the Federal Statistical System. NCHS remains fully committed to protecting privacy, but views this effort as important for encouraging greater use of NCHS data.

NCHS Budget Update

Dr. Moyer expects the Fiscal Year (FY) 2022 budget to be released next week. Until then, the budget remains the same as he reported in January: the enacted FY2021 budget ($175M) represented a $1M increase relative to the annual budget for FY2020. There was a realignment such that the $14M NCHS received annually via the Centers for Disease Control and Prevention (CDC) is now directly and explicitly represented in the NCHS budget. In FY2021, NCHS also received DMI funding from the Core Modernization Fund and from various acts of Congress (e.g., CARES Act).

NCHS Publications

During April/May, NCHS released various publications (e.g., Maternal Mortality Rates in 2019; Provisional Birth Data for 2020). Upcoming reports in May/June include reports regarding declines in infant mortality in Appalachia and the Delta (May 26), drug-involved infant deaths (June 3), and declines in monthly births (June 23).

NCHS Strategic Planning

Brian Moyer, Ph.D., Director, NCHS
Kiana C. Morris, M.B.A., C.M.C.P., Acting Director, Office of Planning, Budget and Legislation, NCHS

Dr. Moyer emphasized that he welcomes feedback from the BSC regarding the Strategic Plan. At the last BSC meeting, NCHS presented information about five case. Since then, the case study approach has transitioned into a more formal process. A first step in this process is to identify key performance metrics. He then introduced Ms. Morris, who presented more detailed information regarding the Strategic Planning Framework.

Ms. Morris explained that the NCHS senior leadership team is engaging all staff to help revise the mission statement, vision, and core values. NCHS did not previously have a vision statement.

Strengths, Weaknesses, Opportunities, Threats (SWOT) Analysis
Ms. Morris reviewed key results from the SWOT Analysis (e.g., improve timeliness and data access; expand use of data science; increase diversity and collaboration). Based on that analysis, NCHS outlined four goals, two of which are more internally focused and two of which are more externally focused. She also summarized a set of objectives that will be used to achieve those goals. As part of this process, NCHS is using multiple tools to engage the staff (e.g., Employee Viewpoint Survey, Staff Surveys, feedback sessions with the Director).

Questions for the Board
Ms. Morris closed by posing six discussion questions for the Board:
1. What strategies would you suggest be added to support the objectives?
2. Are there gaps or opportunities that should be addressed to ensure success?
3. How can NCHS better engage partners and data users?
4. What additional goals should be added to better serve NCHS's external audiences?
5. How can NCHS address other priorities (e.g., health equity) in the plan?
6. How should NCHS balance introducing new data sources while maintaining with NCHS's reputation for high data quality?

Discussion/Reaction by the Board
As a whole, the Board endorsed the importance of Strategic Planning. The discussion centered on the importance of consulting with and engaging external users, the value of maintaining existing data sources, issues related to health equity, and suggested additions to core values.

Several Board members emphasized the value of consulting with and engaging external users, including not only academics but also other federal agencies. Although the data quality of external sources may be a concern, NCHS could enhance data quality by partnering with the producers of those data. Another BSC member suggested including K-12 education as part of external engagement (e.g., smaller datasets that could be used to help children learn about health statistics) in an effort to increase the relevance of NCHS data.

Multiple BSC members endorsed the importance of maintaining existing data sources and praised NCHS population surveys as a role model. One member questioned whether NCHS actually has the necessary discretion to make better use of existing resources or to expand staff and resources.

Regarding health equity, more consistency in the collection of data on race/ethnicity is needed; currently, these data are collected by many methods (e.g., voluntary vs. required, self-reported vs. proxy). One participant noted that the Blue Cross Association is prioritizing the collection of race/ethnicity data and would welcome advice from NCHS for how to do that in a consistent way. Are there models for imputing race/ethnicity? How racial/ethnic categories are aggregated is also an issue; the level of detail that can be reported at the state level differs from the national level. Another Board member stressed the value of contextualizing the data at the local level to ensure that NCHS does not reinforce biases. Engaging local support can improve equity.

Suggested additions to the state Core Values included: Accessibility, Usefulness, and Privacy. Protecting privacy is key to enhancing trust. Requiring some information may inhibit health equity because it deters some people from participating.

Data Modernization Initiative Update

Daniel Jernigan, M.D., M.P.H., Acting Deputy Director for Public Health Science and Surveillance, CDC
James Craver, M.A.A., Special Assistant to the Director, NCHS

Public Health Data and IT Modernization at CDC
Dr. Jernigan described the DMI as a national effort to create modern, integrated, and real-time public health surveillance. The DMI has base funding from CDC of $50M a year for two years. Additional funding consists of $500M from the CARES Act and $500M from the American Rescue Plan (still in progress). Each of the 59 jurisdictions has a DMI coordinator. Partner support is enhanced by support from Congress and the public.

Data modernization offers many benefits (e.g., better data, save time). DMI priorities include getting data to partners, to CDC, and to the U.S. government; building a public health workforce (which is a top priority for the states); and fostering ongoing data modernization/innovation. The CDC Roadmap outlines activities and the short-, intermediate-, and long-term outcomes.

Theme 1: Data Sharing Across the Public Health Ecosystem
Regarding data sharing, NCHS's key objectives are to develop and implement cloud-based approaches for automated data collection; reduce burden on data providers and public health agencies; and ensure that the systems are scalable, interoperable, and adaptable. CDC already has several important accomplishments (e.g., enabled gene sequencing for more than 800,000 unique SARS-CoV-2 genomes, facilitated vaccine response, increased real-time emergency department reporting, and released vital statistics more quickly through NCHS).

Theme 2: CDC Systems and Service Enhancements for Ongoing Data Modernization
Key objectives under Theme 2 are developing a cloud-based, enterprise-wide infrastructure, enabling data-linking across the agency, and integrating new nontraditional data sources that complement existing systems. To date, accomplishments include greater data access (i.e., 164 new datasets added in 2020, which represents a 32% increase compared with 2019), improved visualization, and numerous other advances.

Theme 3: New Standards and Approaches for Public Health Reporting
Theme 3 comprises the following key objectives: real-world testing of new standards, expansion of standardized application programming interfaces (APIs) and electronic health record (EHR) data elements, and hypothesis-driven discovery for fast-tracked implementation.

Unmet Needs and Priorities
CDC is also focused on addressing the following unmet needs: enhancing partner's capabilities to conduct effective surveillance and response, as well as its own capability for effective surveillance, and fostering an environment of continuous learning and innovation.

Data Modernization at NCHS
Mr. Craver reviewed two groups of DMI projects at NCHS. The first focuses on data sourcing/production, whereas the second involves data access.

Data Sourcing and Production Improvements
The National Vital Statistics System (NVSS) uses automated, bi-directional data sharing between NCHS and jurisdictions. Over time, data quality and timeliness has dramatically improved (e.g., the percentage of death records obtained within 10 days increased from 11% in 2011 to 67% in 2020, thus rapidly approaching the goal of 80%).

NCHS is planning to move the Healthcare Electronic Health Record (HEHR) platform to a cloud-based infrastructure. To supplement the data collection, NCHS has been purchasing EHR data for the NHCS.

The National Health Interview (NHIS) has two pilot projects: Piloting In-Home Health Examinations and NHIS-Teen project (i.e., to assess the feasibility of data collection from adolescents). NCHS is also using model-based estimates to produce faster, more granular estimates and continues to expand the data linkage program.

Data Accessibility
Efforts to improve data access include developing a Data Query System to improve discoverability and integration of NCHS data assets, redesigning the NCHS website to improve accessibility, standardizing metadata and APIs, and planning to create a Virtual Data Enclave (VDE). The VDE will allow remote access to an encrypted, secure server. The primary advantages are reduced cost to researchers and greater accessibility.

Discussion/Reaction by the Board
The discussion focused on concerns regarding the costs involved with cloud-based storage and purchasing EHR data, the problem of implicit bias in AI, the need to hire more data scientists, and changes in governance to facilitate collaboration across jurisdictions and different levels.

One BSC member asked who would be the cloud-based provider (internal at NCHS or external?) and how NCHS plans to integrate purchased EHR data. Dr. Jernigan confirmed that CDC will use external providers of cloud services. The BSC member advised careful attention to the fixed and variable costs (e.g., analytics can be very expensive). Dr. Jernigan explained that CDC is carefully reviewing what EHR data are being collected and who uses them within CDC and among their partners. Are there additional data sources CDC should collect? Should CDC purchase it? Can it replace some of the information currently being collected via surveys?

Another BSC member noted that, in the interest of health equity, NCHS should consider how implicit bias works its way into artificial intelligence. Some commercial firms are starting to provide consulting services in this vein. Dr. Jernigan noted that CDC is doing research on forecasting and predictive analytics to gain a better understanding of introduced biases that can become incorporated into machine learning.

In terms of workforce transformation, the BSC discussed the need for data scientists, the difficulty in hiring them when the health department does not even have a job classification for data scientists, and how to pay for them. Dr. Jernigan acknowledged that the transformation from distributed systems to an enterprise-wide approach is complex.

Finally, a BSC member highlighted that the desire for more real-time data may require changes in governance and better strategies for collaborating across jurisdictional levels to solve national and state public health issues. A couple organizations that may be helpful were mentioned (i.e., National Academy of State Health Policy; National Association of Medicaid Directors).

Privacy Preserving Techniques: Case Studies from the Data Linkage Program

Lisa Mirel, M.S., Chief, Data Linkage Methodology and Analysis Branch, DAE

NCHS links survey data with administrative data, but because of privacy concerns, many of these linked datasets are available only through the Research Data Center (RDC). NCHS has been unable to link with some other data sources because of privacy issues. Today, Ms. Mirel will present two case studies that explore new methods that might enable such data linkage. Both case studies use patient-level encounter data from the NHCS linked with the National Death Index.

Case Study I: Synthetic Data

NCHS already releases a partially-synthetic public-use linked mortality file, but wants to explore a fully-synthetic dataset. The key question is: How do the estimates compare with those based on the original restricted-use file? NCHS computed predicted probabilities of death and synthesized the date of death using predictors (e.g., race/ethnicity, diagnoses). A classification tree was used to predict cause of death. The results for all-cause mortality showed that parameter estimates based on synthetic data closely matched those based on the original data. There were more differences for the cause-specific estimates.

NCHS received Patient-Centered Outcomes Research Trust Fund (PCORTF) FY21 funding to create fully-synthetic, linked data files. The next steps include convening a stakeholders meeting to determine which variables should be included, developing and implementing the methodology, and creating a method by which researchers can validate their results. Synthetic data would increase data accessibility, potentially expanding the user base. The challenges include communicating with users who are skeptical of synthetic data and ensuring that the resulting datasets return valid results.

Case Study II: Privacy Preserving Record Linkage (PPRL)

PPRL is a method that can be used to link de-identified data using hashing algorithms and tokens. The critical question is: How do the results compare with the standard linkage algorithm? NCHS compared sensitivity and specificity from both the initial and refined (e.g., removed tokens with high false positivity rates) versions of PPRL with standard linkage. The results demonstrated very high sensitivity (97.8-98.7%) and specificity (99.7-99.9%). Furthermore, the effect of PPRL linkage on secondary data analysis was minimal (e.g., compared with standard linkage, the death rates differed by less than 5% for the initial PPRL and by less than 1% for the refined PPRL). In conclusion, PPRL can be an effective record linkage technique.

As next steps, NCHS will investigate the robustness of the results when personally identifiable information (PII) is less complete (e.g., only the last 4 digits of the social security number, as collected by NHIS since 2007). The Center also plans to evaluate the impact of PPRL for health outcomes other than mortality. Benefits of PPRL are that it would allow linkage beyond federal data sources without sharing PII, and it could help address emerging public health threats.

Challenges include ensuring that the PPRL methodology is safe and secure when sharing hashed tokens. NCHS will continue to evaluate and calibrate linkage accuracy.

Thoughts for Discussion
Ms. Mirel posed several questions to focus the discussion. First, does synthetic data entail unique considerations (e.g., user communication, validation, privacy)? Second, what are the priority sources (e.g., private sector, emerging public health data) for which NCHS might use this method?

Discussion/Reaction by the Board
The discussion highlighted questions regarding the ethics of linkage and privacy issues, the challenges of validation, and the tradeoffs of allocating scarce resources to developing a VDE vs. synthetic data. Ms. Mirel noted that all of the linkage conducted through the NCHS Data Linkage Program are reviewed and approved by the NCHS Ethics Review Board.

One Board member noted that NIH is having a workshop (June 29-30) on policy and ethics of record linkage. Ms. Mirel plans to attend the workshop. This BSC member questioned whether NCHS has the right to make such linkages. Participants noted that death data have different levels of protection depending on the law under which they were collected (e.g., under HIPAA, death data are protected for 50 years). Another Board member expressed concern about ensuring that private companies do not unmask the data.

A BSC member noted that the use of synthetic data could extend beyond training datasets to potentially include fully-synthetic analytic datasets. The critical question is: How can NCHS validate the data sufficiently so that researchers will be willing to use them analytically? Ms. Mirel acknowledged that NCHS recognizes the need to provide service whereby researchers can submit their results from complex analysis of the synthetic data and, in return, NCHS will inform the users how close those estimates are to the results based on the original data. The main benefit of synthetic data is that it would expand access to a big potential group of researchers who might not otherwise use the data (e.g., would not have access even via a VDE). Ms. Mirel confirmed that the restricted-use data would still be available from the RDC, whereas the synthetic data would be public-use.

There was an active discussion of the tradeoffs between investing scarce resources in synthetic data versus development of a VDE. One Board member argued that VDE was a higher priority. Others countered that NCHS does not need to choose one over the other; both can help expand access. Someone else pointed out that there is already a model of using synthetic data to encourage use of the virtual RDC and comparing the results to validate synthetic data (e.g., John Abowd at the Census Bureau). An unresolved, but difficult to answer question is how to handle informative sampling mechanisms in synthetic data (i.e., How do you assign the weights? How do you deal with sampling by stratum and primary sampling unit?).

Changes in NHIS Data Collection During the COVID-19 Pandemic

Stephen Blumberg, Ph.D., Director, Division of Health Interview Statistics
Andy Peytchev, Ph.D., BSC Member, Chair, PHSPMDP Workgroup

Dr. Blumberg reviewed the consequences of the COVID-19 pandemic for NHIS data collection during 2020. For January through March 18, data collection was in-person as usual. After that, NHIS had to stop personal visits; between March 19 and June 30, all interviews were conducted by telephone only. In July, NHIS restarted some in-person interviews, but with a telephone-first

policy (i.e., in-person visits were used to follow up on non-response, drop-off recruitment materials, or if the respondent expressed a preference for an in-person visit). During August through December, NHIS used a split sample method. Half of the sample was selected as usual and interviewed on a telephone-first basis with in-person visits as needed in all areas. The other half of the sample comprised re-interviews of respondents who were first interviewed in the 2019 NHIS. This follow-back sample provided increased coverage and permitted NCHS to investigate intra-individual differences between the pre- and post-pandemic periods. An important question is: How should NCHS combine and weight these four different sampling designs? NCHS staff members, Drs. Bramlett and Dalhamer, developed a proposal, and NCHS requested feedback from the BSC Workgroup.

Dr. Peytchev explained that the Population Health Survey Planning, Methodology and Data Presentation (PHSPMDP) Workgroup was convened to obtain external input on NHIS weighting options for data collected during 2020. The goal was to review materials and convene a meeting in time for NCHS to prepare weights for the September 2021 release of 2020 data. In April 2021, the workgroup held a virtual meeting to discuss NCHS's proposed approach. The 2020 NHIS data include five different files for weighting: Q1 (in-person), Q2 (telephone only), Q3 (normal sample for July, split-sample for Aug-Sept; both telephone and in-person), Q4 (half sample that was drawn as usual), and the half sample that comprises a follow-back survey of adults originally sampled in 2019.

Needed Weights
NHIS needs at least two weights. First, a longitudinal weight is needed for the follow-back survey of 2019 respondents who were re-interviewed in 2020. In this case, NCHS could take advantage of substantive variables from 2019, which could be used to inform adjustments. NCHS used a tree-based method (i.e., recursive partitioning model) to identify variables and combinations of variables to form adjustment cells. This method allows adjustment on multiple levels, although it is important to be mindful of node size. NCHS also tested alternative raking to combinations of demographics and the 24 resulting nodes. The results were evaluated using the 2019 sample adult file.

Second, a combined annual weight is needed for single-year or multi-year analysis. The regular production 2020 (P20) samples—which exclude the re-interviews of 2019 respondents—follow a different nonresponse adjustment stream. Key questions include (1) When do you post-stratify the samples? (2) How should the follow-back sample be combined with the P20 samples? (3) Should you control for the relative contribution of the follow-back sample (i.e., proportional adjustment reflecting the relative weight assigned to one sample vs. another)? and (4) If so, is it better to perform proportional adjustment before or after raking to the 2020 population control totals?

Options with Which the Workgroup was Presented
The Workgroup was presented with four options to evaluate. The first two options do not include a proportional adjustment, whereas the other two options include such adjustment. For that reason, the Workgroup preferred the latter two options, but slightly preferred the third option: (1) rake the follow-back sample to the internal 2019 control totals; (2) rake the P20 samples to the external 2020 control totals; (3) combine them using proportional adjustment; and (4) finally, rake the combined samples—follow-back plus P20—to the external 2020 control totals again. Nonetheless, the Workgroup suspected that the results from the fourth option (i.e., similar to option 3, but re-raking the follow-back sample to the external 2020 control totals *before*

combining the follow-back with the P20 samples using proportional adjustment) would yield very similar results.

Workgroup Tentative Opinions
The Workgroup formed the tentative opinion that the general weighting approach to the follow-back sample (i.e., tree-based methodology to identify the nodes and make adjustments) is appropriate and that proceeding with the third option is preferable.

Discussion/Reaction by the Board
There was little discussion except to express admiration for the capability of NHIS staff and to applaud the fact that despite all the challenges and different designs implemented in 2020, Dr. Blumberg plans to release the NHIS data on the same schedule as the previous year. Dr. Blumberg cautioned that it remains to be seen whether that goal can be attained.

Innovation in Vital Statistics: Nowcasting Techniques, Maternal COVID Data, and Provisional Data Dissemination

Early Provisional Estimates of Drug Overdose, Suicide, and Transportation-related Deaths: Nowcasting Methods to Account for Reporting Lags
Lauren Rossen, Ph.D., Division of Research and Methodology

Dr. Rossen briefly reviewed a recent report on drug overdose, suicide, and traffic-related accidents deaths. Injury-related deaths often undergo lengthy investigations and testing to determine the cause and manner of death. Although 60% of death certificates are obtained within 10 days post-death, the delay can be substantially longer for external causes of death; in some cases, it can be nine months or longer before the data are complete. To ensure that reporting is reasonably complete and to avoid potentially misleading declines that may be a result of incomplete reporting for the most recent months, NCHS imposes a six-month lag on the release of monthly provisional drug overdose death counts. The timeliness of drug overdose deaths has drastically improved over recent years (e.g., in 2016, about one-third were reported within 3 months, but by 2018 that percentage had increased to more than 70%).

Impact of COVID-19 Pandemic on Injury-Related Causes of Death
The pandemic has shifted expectations for how quickly data can and should be released. NCHS realized that it needed to release more timely provisional estimates, but still needed to account for reporting lags. Thus, NCHS explored "nowcasting" (i.e., method for predicting the number of deaths that have already occurred, but have not yet been reported to NCHS). Using its archive of provisional data from the past several years, NCHS modeled the completeness of provisional data within specific intervals (e.g., two weeks post-death, four weeks, …39 weeks). Based on the model-based estimates, NCHS created a set of weights based on the inverse of predicted completeness (e.g., if provisional data were 50% complete within 12 weeks of the death, then the weight would be 2 for deaths that occurred 12 weeks ago in that jurisdiction).

Next, NCHS used a multi-stage hierarchical modeling process to generate the predicted number of deaths using the weighted counts. The approach was validated by comparing the estimates based on the older provisional data with the reported counts 6-10 months later when reporting was assumed to be virtually complete. Staff evaluated different scenarios for how the reporting lags may have changed in 2020 relative to 2019. The estimates that assumed timeliness would be the same as the median timeliness in 2019 tended to greatly over-predict the number of overdose deaths, which implies that the reporting lags were shorter in 2020 than in 2019. The best estimates (i.e., predicted death counts were closest to eventual reported counts) were

those that allowed the improvement in timeliness to vary by jurisdiction. Using that approach, NCHS produced provisional weekly counts of drug overdoses, suicides, and transportation-related deaths for each jurisdiction and the United States as a whole. Dr. Rossen cautioned that NCHS will not be able to determine how reporting lags actually changed over the course of the pandemic until much later. NCHS plans to refine the methods and expand them to additional outcomes, data sources, and smaller geographic levels (e.g., county).

Discussion/Reaction by the Board
The discussion focused on whether users have questioned the veracity of provisional estimates and suggestions from the BSC for alternative approaches.

One Board member asked whether users have questioned the validity of the provisional estimates. Dr. Rossen replied that over the past several years, people have become more accustomed to NCHS releasing provisional estimates, but those types of questions still arise, especially for model-based estimates.

In terms of alternative methods, one BSC member noted that NCHS is using R-INLA (Integrated Nested Laplace Approximation), which limits the flexibility of the model to deviate from the random effects specifications available in INLA (e.g., this approach does not allow you to jointly model outcomes, space, time, and delays in a single multivariate model). Instead, the BSC member suggested building a multivariate space-time model over outcomes, jurisdictions and over time lags to handle potential non-separable space-time-delay interactions. Dr. Rossen explained that, for the purposes of modeling completeness, she purposefully did not include spatially structured random effects because she did not want to assume that reporting lags are geographically-clustered. However, in the second stage, when they modeled the death counts, they did use spatially structured random effects and included space-time interactions. Dr. Rossen acknowledged that a key limitation is computational efficiency (i.e., they cannot run complicated multivariate space-time-delay models with the existing computational resources at NCHS). Dr. Holan agreed that some dimension reduction would be necessary. Dr. Rossen added that they also want to be able to manually adjust for sudden changes in reporting in a particular jurisdiction (i.e., in cases where they have external information). Another Board member focused on the change-point issues in 2020 (i.e., there was relative consistency in under-reporting in 2015-19, but then everything suddenly changed in 2020). NCHS could consider modeling this problem to determine where the changes actually occurred and ensure the model is not relying outdated estimates of reporting lags. Dr. Rossen replied that NCHS hopes to eventually incorporate such methods and will add it as a question for the DMI to explore.

Maternal and Infant Characteristics Among Women with Coronavirus Disease (COVID-19) During Pregnancy for Selected Reporting Areas
Michelle Osterman, M.H.S., Division of Vital Statistics

What impact does having COVID-19 during pregnancy have on the mother and newborn?
In March 2020, the Natality team began to encourage states to collect maternal COVID-19 status because the U.S. Standard Birth Certificate cannot be modified quickly. NCHS used the national Birth Data Quality Workgroup to develop recommended questions, alternative approaches for data collection, and standards/processes for reporting.

Some jurisdictions were able to modify their electronic birth certificate reporting systems (i.e., added a maternal COVID-19 item). Others linked COVID-19 status from their infectious disease surveillance systems to the birth certificates or collected data via supplementary forms. NCHS recommended some options for collecting this information (e.g., insert an additional check box, add a question at the end). Data collection indicated only whether the mother had COVID-19 at any time during pregnancy; NCHS has no information about when infection occurred. Fifteen jurisdictions are currently providing data, but the start of reporting varied from spring to mid-June 2020. All 15 jurisdictions cover the period from June 19, 2020 through February 2021. Maternal COVID-19 cases are reported to NCHS where they are linked to the standard birth record.

Results

The results are published every two months on https://www.cdc.gov/nchs/covid19/technical linkage.htm (e.g., data for April-October 2020 were reported on January 12, 2021; data for April 2020 through February 2021 should be published by the end of May). Compared with all U.S. births, the maternal COVID-19 reporting area has a slightly larger proportion of births to Hispanic women, but lower percentage of births to non-Hispanic Black and non-Hispanic White women. NCHS was able to link 99.6% of the cases to the full birth record. Over the period from April 2020 to February 2021, there were 23,000 births to moms with presumed/confirmed COVID-19 and approximately 750,000 births to moms without COVID-19 during pregnancy. Mothers with COVID-19 during pregnancy tended to be younger, less educated, and more likely to be admitted to the ICU or have a preterm birth than those without COVID-19.

Limitations and Next Steps

These data are not nationally representative. There is no information about the timing of COVID-19 during pregnancy. Reporting methods and coverage periods varied by jurisdiction. Nonetheless, it is the largest, most complete national dataset available for maternal COVID-19.

Innovation in Data Dissemination: Provisional Data
Paul Sutton, Ph.D., Deputy Director, Division of Vital Statistics

Improvements in Timeliness of Mortality Reporting

As noted earlier, the percentage of deaths reported to NCHS within 10 days has dramatically improved, but researchers are primarily concerned with how quickly NCHS releases the data. In the fourth quarter of 2018, NCHS began providing quarterly provisional estimates on mortality. In 2020, the timeline for releasing provisional data shortened such that they began issuing releases at least five days a week. To the extent possible, NCHS also tried to accommodate requests for ad-hoc datasets given limited resources and capacity.

NCHS's long-term vision is to make provisional data available through CDC WONDER, where researchers can create their own tabulations. The Center is using some of the DMI funds to modernize dissemination via WONDER. In particular, NCHS is considering a new query screen for provisional data that would be similar to the interface by which users tabulate the final mortality data. NCHS also wants to add week of death, the state in which the death occurred (currently, the data only provide the state of residence of the decedent), and an enhanced API.

NCHS imposes a lag on the release of provisional data, which varies by cause (e.g., six months for injury-related deaths, three months for most natural causes). Data from a recent report (Ahmad et al., 2020, www.cdc.gov/nchs/data/vsrr/vsrr009-508.pdf) show how the reporting lags

varied by cause. For example, nearly 95% of all-cause deaths were reported within 13 weeks, but the percentage declined to 89% for suicide, 84% for all injuries, and 70% for drug overdoses. By 39 weeks after death, more than 97% of deaths were reported for all those cause groups, although the percentage was still lowest for drug overdoses (97.7%) and highest for all-causes combined (99.8%). The reporting lags also varied by state: by 13 weeks, some states reported nearly 100% of deaths (e.g., IL, CO, OH), while others reported less than 70% (i.e., CT, PA).

Future Plans
NCHS is considering whether provisional data released via WONDER should use the same lag (i.e., six months) for all causes or whether to incorporate lags that vary by cause and/or jurisdiction. NCHS is also debating the frequency at which provisional data on WONDER should be updated. The current practice of updating provisional data related to COVID-19 on a daily basis seems excessive; NCHS is considering updates on a quarterly basis, concurrent with the current quarterly provisional releases. Of course, provisional data are subject to error and incompleteness. NCHS needs to inform users of those limitations to prevent them from making serious errors. Finally, Dr. Sutton asked: How much descriptive analysis is needed to accompany provisional data releases? In the past, major new data releases have always been accompanied by an analytic, descriptive analysis. However, if provisional data will be updated frequently, it may become prohibitive to continue that practice.

Discussion/Reaction by the Board
Questions from the BSC related to the possibility of suppressing data based on expected completeness (e.g., could NCHS suppress estimates for states that are below some threshold level of completeness?) and whether there are key factors that explain variation in reporting delays. Dr. Sutton explained that there are many reasons for variation in reporting delays, but the biggest factor is whether or not the state has an electronic system. Also, sometimes there are state-specific rules that cause delays. The map that Dr. Sutton presented was for 2017, but he noted it would not be very different for 2020. Surprisingly, the states maintained their levels of timeliness remarkably well in 2020 despite all the challenges.

Update on Division of Health Care Statistics EHR Activities

Carol DeFrances, Ph.D., Acting Director, DHCS

Division EHR Activities
DHCS will be moving the health care EHR system to the cloud. A key issue is ensuring privacy and security standards. It will be the first Confidential Information Protection and Statistical Efficiency Act (CIPSEA)-compliant cloud at NCHS. DHCS is part of a PCORTF project titled, "Making EHR Data More Available to Research and Public Health (MedMorph)." Dr. DeFrances also sits on a stakeholder committee for the Pandemic-Ready Interoperability Modernization Effort (PRIME), which is a multi-year effort with CDC and the U.S. Digital Service to strengthen data quality and information technology.

Survey-Specific EHR Activities
The NHCS closed out in March; 157 hospitals submitted data (among 198 who had agreed to participate), but the coverage remains insufficient to generate national estimates. NCHS is purchasing data to supplement the 2020 data collection. The purchased data do not include any PII and thus cannot be linked with other data sources. Some issues have arisen during the course of negotiation (e.g., some vendors would not allow NCHS to use the proprietary data in public-use or RDC files; some wanted to impose limits on the length of time the data can be

used—that is, NCHS would need to continue paying a licensing fee to continue to use the data). The greatest challenge is how to transfer the massive amounts of data to NCHS, given that NCHS does not have a cloud or an FTP site. Also, integrating and harmonizing data from five different sources, which include a mixture of EHR and claims data, will also pose a difficult challenge. In terms of methodological work, DHCS is working with the Division of Research and Methodology to explore the possibility of modeling national estimates for 2020. DHCS is also exploring the possible use of synthetic data for public-use and RDC files.

For NAMCS, the Health Center component has not yet begun (it is awaiting approval from the Office of Management and Budget), but DHCS will partner with the Health Resources and Services Administration to collect EHR data from 50 federally qualified health centers and similar centers. NCHS will provide health centers with a $10,000 set-up fee to activate the interface/module on their EHR system. DHCS is exploring whether it is possible to buy EHR data from physicians and other providers including large networks (e.g., Kaiser).

DHCS is also considering transitioning the long-term care survey (National Post-Acute and Long-Term Care Study) to collect data from EHRs.

Dr. DeFrances closed by posing several questions to the BSC:
1. Does anyone have knowledge of other CIPSEA-compliant data on the cloud?
2. Has anyone created privacy and security controls for CIPSEA?
3. Are you aware of potential EHR data sources for physician and other ambulatory providers? Residential care communities? Adult day service centers?

Discussion/Reaction by the Board
Discussion focused on the need to coordinate across agencies and ethical issues regarding the purchase of EHR data.

More coordination is needed across agencies. Many groups are frustrated with the government, particularly public health, because they impose reporting requirements and require the data in their own special format. Also, every registry has its own connection point. Dr. DeFrances acknowledged that more coordination within CDC is needed so that data are not being sent multiple times to different divisions. NCHS began purchasing data as an interim approach because, despite having been in the field for 10 years, it still cannot make any national estimates. Currently, it purchases only de-identified data. It would require negotiating a lot of additional contracts to obtain PII.

Two BSC members expressed concern regarding the ethical issues surrounding the purchase of EHR data (e.g., did patients know their data would be sold?). It may send the wrong message if NCHS purchases data about patients who do not know their data are being sold. It is important to emphasize that NCHS is purchasing only de-identified data (although the data NCHS collect itself often include PII); those data are not linked with anything else.

Update from NAMCS Workgroup
John Lumpkin, M.D.,M.P.H., Chair, NAMCS Workgroup

The NAMCS team posed a couple of additional questions to the Workgroup.
1. Should visit-level data be managed at the group or at the individual provider level?
2. Should supplemental data be used to enhance NAMCS visit data?

The Workgroup sought advice from participants of previous workgroup meetings (see slides for a list of the respondents), discussed the responses, and then reached a consensus.

Most of the Workgroup's conversations focused on the first question. In the Workgroup's opinion, it is valuable to sample both groups and individual providers, but there is a concern that data collection at the group level could risk over-sampling larger groups/communities. If so, NAMCS may not obtain a complete understanding of ambulatory care in rural communities.

Given their lack of expertise on the topic, the Workgroup members did not have any substantial recommendations regarding the use of supplemental data. However, based on comments provided by subject matter experts, the Workgroup encourages NAMCS to capture more of the increasingly diverse group of ambulatory care providers. As technology improves, the NAMCS could consider enhancing its point-in-time model to longitudinal sampling.

Actions
There was no discussion. Dr. Lumpkin moved to approve the report submitted by the Workgroup and for the BSC to formally submit those opinions to NCHS. The vote was unanimous in support.

BSC Wrap-up

Linette T. Scott, M.D., M.P.H., Chair, BSC
Brian Moyer, Ph.D., Director, NCHS

Dr. Scott apologized that the letter they discussed in September has been delayed. She circulated it to Dr. Uddin and will soon forward the letter to the BSC for comments and edits. As she nears the end of her term, Dr. Scott expressed her pleasure at serving on the BSC and acting as chair for several years.

Dr. Moyer thanked everyone for their input and extended a special thanks to those members rotating off the BSC. Soon, he expects to share the next iteration of the strategic plan with the BSC for review. He closed by asking the BSC for their feedback regarding the best use for Workgroups in the future.

Dr. Hauser expressed his admiration of the progress achieved by NCHS over the years. He also expressed concern regarding the lack of occupational data in many of the surveys and other data collection operations at NCHS. He noted that advances in machine-based coding methods have increased the efficiency of collecting data on occupation and occupational characteristics. Dr. Sutton noted that NCHS is working in partnership with NIOSH to re-establish industry and occupational data within the mortality dataset.

Public Comment

Dr. Scott noted that earlier this morning there were some public comments in the virtual Q&A. During the public comment session, there were some further comments regard validating synthetic data.

The meeting was adjourned at 5:00 pm ET.

To the best of my knowledge, the foregoing summary of minutes is accurate and complete.

_____/s/_____   _____9/28/2021_____
Linette T. Scott, M.D., M.P.H.   DATE
Chair, BSC