

Department of Health and Human Services

Board of Scientific Counselors

January 22-23, 2015

NCHS Auditorium
3311 Toledo Road
Hyattsville, MD 20782

Meeting Minutes

The Board of Scientific Counselors was convened on January 22-23, 2015 at the National Center for Health Statistics in Hyattsville, MD. The meeting was open to the public.

Committee Members

Present

Raynard S. Kington, M.D., Ph.D., Chair BSC
Wendy Baldwin, Ph.D.
Virginia S. Cain, Ph.D., Executive Secretary
Michael Davern, Ph.D. (via phone)
Mark Flotow, M.A.
Hermann Habermann, Ph.D.
Christine L. Himes, Ph.D.
Genevieve M. Kenney, Ph.D. (via phone on January 23, 2015)
Virginia M. Lesser, D.Ph.
Wendy D. Manning, Ph.D.
Robert E. McKeown, Ph.D., FACE
F. Javier Nieto, M.D., M.P.H., Ph.D.
Ana Diez Roux, M.D., Ph.D., M.P.H. (January 22 only)
Margo Schwab
Linette T. Scott, M.D., M.P.H.
Katherine K. Wallman, Ex-Officio, OMB (via phone)

Not Present

Thomas A. LaVeist, Ph.D.
Trivellore E. Raghunathan, Ph.D.

Staff and Liaisons

Irma Arispe, Ph.D., OAE, NCHS Staff
Clarice Brown, M.S., Director, DHCS, NCHS Staff
Debbie Jackson, NCHS (January 23, 2015)
Jennifer Madans, Ph.D.
Charles Rothwell, NCHS
Nathaniel Schenker, ORM, NCHS Staff

Presenters

January 22, 2015

Robin A. Cohen, Ph.D.
Alan Dorfman, Ph.D.
Kristen Miller, Ph.D.
Jennifer Parker, Ph.D.
Charles Rothwell, NCHS
Joe Woodring, D.O., M.P.H., M.T.M.H.

January 23, 2015

Clarice Brown, M.S.
Yinong Chong, Ph.D.
Carol DeFrances, Ph.D.
Katherine Wallman, OMB

Others

January 22, 2015

Brenda Baker, DHANES
Stephen Blumberg, DHIS
Clarisse Brown, DHCS
Verita Buie, OPBL
Mary Ann Bush, OAE
Juanita Chinn, OCD
Jim Craven, OAE
Giovannina DiPietro, SSS
Nazck Elgadda, OAE
Cordell Golden, OAE
Jill Hensley
Rebecca Hines, OAE
Meena Khare, ORM
Ellen Kramarow, OAE
Xianfen Li, CDC/NCHS/OAE
Patsy Lloyd, OAE
Althelia Harris, OAE
Negasi Beyene, ORM
Hanyu Ni, CDC/NCHS/DVS
Colleen Nugent, DHIS
Van Parsons, ORM
O. Richard, CDC/OPHSS
Yahtyng Sheu, OAE
Sandy Smith, OCD
J. Tran, CDC/NCHS/ORM
Anjel Vahratian, Ph.D., DHIS
Brian Ward, DHIS
Kassi Webster, OPBL
Julie Weeks, OAE
R. Zach, OAE

January 23, 2015

Negasi Beyene, ORM
Karishma Chari, DHCS
Juanita Chinn, OCD
Giovannina DiPietro, SSS
David Huang, OAE
Penys Lan, DHCS
Xianfen Li, CDC/NCHS/OAE
Alena Maze, ORM
Pathaia Pastor, OAE
Shalean Levant, DHCS
Iris Shimizu, ORM
Makram Talih, OAE
Kassi Webster, OPBL
Sonja Williams, DCHS

MEETING SUMMARY **January 22-23, 2015**

ACTION STEPS

- Future BSC meetings will be held on May 21-22, 2015 and September 24-25, 2015.
- BSC members interested in participating in a small workgroup about data suppression should submit their names to Dr. Madans or Dr. Cain (Dr. Davern and Dr. Kenney have already volunteered).
- Requested information about various data systems from the Data Visualization for Discovery and Insight presentation will be distributed to BSC members.
- BSC members were asked to review Statistical Policy Directives #1 and #4 (and accompanying preambles) in order to provide feedback to Dr. Cain.
- In collaboration with Dr. Kington, Dr. Cain will draft a letter of support to the Department for review by BSC members about Statistical Policy Directives #1 and #4.
- Consider contacting the VA to address similar systems issues.

(Please refer to PowerPoint presentations for further specifics)

Thursday, January 22, 2015

Welcome, Introductions and Call to Order

Raynard Kington, M.D., Ph.D., Chair, BSC and Charles Rothwell, Director, NCHS

NCHS Update

Charles Rothwell

New BSC members were welcomed and staff updates were announced. Budget update highlights included: FY 2015 level funding with FY 2014 for health statistics appropriation (\$155,397,000). No funding from the Prevention and Public Health Fund is forthcoming so in 2016 NCHS will discontinue NHIS and NAMCS survey enhancements which expand available state estimates. Because part of the budget previously funded through evaluation dollars will now come through direct appropriations, funding could be lost through taps or sequesters. Additional administrative updates were announced relative to space, travel ceilings and a possible National Conference on Health Statistics. The NCHS and CDC response to Ebola was shared.

Selected NCHS accomplishments were described to include: timely access to data that monitor the ACA; timeliness improvements for vital statistics; innovations in data collection as well as measure and data development; movement toward electronic data collection; and an exploration of data visualization methods. Priorities for 2015 were delineated relative to survey and data collection; timely access to data; timely and expanded use of vital statistics reporting for CDC surveillance activities; continued work with EHR vendors to develop an interface that allows hospitals to forward data to NCHS; recruitment and staff retention; mentoring and cross-training; continued investigation of data visualization methods; ensuring adequate IT infrastructure; and addressing the increasing health measures at state, regional and national levels.

Discussion DHCS is developing a standardization system for electronic health records. Decisions must be made about how to address so much data. There will be opportunities to mine the data for specific purposes (e.g., to identify drug-related cases). It is not clear what subsamples will be taken from such large data (or under what principles and practices). A request was made for the BSC to provide guidance to the electronic data collection project, noting that its success at NCHS requires funding and healthcare sector buy-in. It was suggested that buy-in would occur more quickly and robustly if those in the field understand how the project will improve metrics. With regard to looking at this quantity of data, a suggestion was made to examine anomalies and abnormalities in the data as is done for fraud investigations, noting that the data have the same analytics for fraud, data quality and disease outbreak. As with epidemiology principles, the “why” can be based on those variations (example given).

Mr. Rothwell expressed his view that NCHS’s job is to collect information for specific but broad purposes; and then to make that information available to the public with measures that impact health policy. Vital statistics have changed due to a variety of ICD format coding. After cleaning, NCHS could provide a file of natural language for the purpose of data mining or examining very rare cases.

Samples sizes and content of the 2016 NAMCS and NHIS will be affected by the budget cuts. A NHIS redesign will cut content material to allow for a 45-50 minute survey. Decisions about what to cut will likely be made by July 2015.

2014 Half-Year Insurance Estimates from the NHIS

Robin A. Cohen, Ph.D., DHIS

Early release health insurance estimates, quarterly health insurance estimates and preliminary microdata file were disseminated on December 16, 2014. Three different measures of “uninsurance” are produced through the preliminary microdata file. Specifics include numbers and percentage of uninsured in the first six months of 2014; changes in the percentage of persons of all ages who lacked health insurance coverage from 2013 to 2014 (January - June); percentage lacking health insurance coverage at the time of interview by age group; estimates of those without health insurance coverage at the time of interview by age group; persons under 65 covered by private health plans at the time of interview; and estimates by selected demographics including poverty status and race/ethnicity.

The decrease in uninsurance in the first six months of 2014 (6.5 million for those age 18-64; 2 million for age 19 - 25 but no decrease for children) is mostly due to more recent increases in coverage. Despite declines, uninsurance rates were greater for poor or near-poor adults. Significant declines in uninsurance for Hispanic and non-Hispanic Black adults were attributed to increases in private coverage although more than half of the decline for non-Hispanic Black adults was related to an increase in public plans. Uninsured decreases for White non-Hispanics were primarily due to private coverage increases. Within quarterly estimates, there was a decrease of uninsured for those under 65 from Quarter 4 in 2013 (16.2%) to 12.9% in Quarter 2 of 2014, noting that the surge was due to people joining private Exchange Plans. Quarterly decline estimates by age groups were also presented. There has been no statistical difference in the decrease of poor people who are uninsured, However, one can speculate that this decrease was probably due to an increase in public coverage (and private coverage for the near-poor).

Discussion Within a year of the start of ACA, three million young adults were covered by their parents’ plans. Many young adults not covered by parents are eligible for public coverage. Policyholder data (not part of the preliminary microdata file) will be available in June 2015. Only large families experience an advantage with family coverage on the Exchange.

Historically, Hispanic and non-Hispanic Black people have been disproportionately uninsured (and therefore, a greater percentage have been covered relative to other ethnic groups). Nevertheless, non-Hispanic people have the highest number of people covered under the Exchange Plan. Many providers of public coverage are private health insurers who have benefitted. It is important to illustrate numbers versus proportions graphically to clarify the data.

A question was posed about how to reconcile administrative data on Medicaid enrollment with survey data on public coverage. There has been a public coverage increase within Medicaid expansion states, some of which disappears when considering national data. In response to a question about how to handle states (e.g., Michigan, Pennsylvania) that switch into the expanded Medicaid category, 2014 data buckets will remain until 2015 when the estimates change. The switch is at the end of the year for trend-reporting purposes.

Draft Suppression/Presentation Guidelines for Proportions

Jennifer Parker, Ph.D., OAE

Dr. Madans noted that the Divisions and Offices meet monthly to discuss clearance and to address cross-cutting scientific issues in Center. Two workgroups were formed, one on data suppression and the other on how to present and test for trends. Dr. Parker presented on the progress and recommendations of the workgroup on policies for data release and suppression. (Dr. Parker) The purpose of the workgroup is to update guidelines for data suppression in routinely published estimates. Workgroup background, scope, membership, recommendations and proposed guidelines were delineated (specifics and examples given). The workgroup focused on developing suppression/presentation criteria to be applied to proportions from survey data that will appear in standard data products with multiple tables and stand-alone estimates, such as *Health United States*, *Healthy People 2020*, or in other data products where estimates require readily applied and transparent suppression/presentation standards. Preliminary guidelines for data release were presented (see PowerPoint).

Discussion A decision about what to present to the public is complex. Should unreliable data be shared? Are the criteria reasonable? In response to a question about statistical justification for the guidelines, it was noted that the confidence interval is considered to be just that. It is hard to come up with criteria for a cut point. The guidelines have been designed for computer programs that create mass-produced tables, for which decisions must be made upfront. In contrast, determinations that present confidence intervals in addition to subject matter expertise are recommended for data briefs or health statistic reports. Despite a certain amount of arbitrariness, it was recommended that every table have a confidence interval.

One discussant emphasized a responsibility to present a viewpoint about data quality and to present all material in lieu of suppression. Another did not think there was a strong enough justification to suppress and opted for showing questionable data with some explanation or guidance (even if somewhat arbitrary). A third participant suggested communicating the notion that public health decisions should not rely on specific unreliable numbers except under unique circumstances. This issue is especially relevant to smaller data sets.

One must consider data timeliness verses accuracy. How data can or cannot be used must be refined. Within the federal government, there has been a culture shift toward making more data sets available although this creates greater inability to control how people use them. The trend is toward making data available while at the same time noting the limitations of the data. In response to a question about confidentiality within very small groups, it was noted that much of the data in question are public use data that use aggregate numbers. One big challenge is to determine criteria for growing amounts of data; and how decisions get made about cut-points or subgroups within the data. The hope is to develop a JSM Panel in 2015 for feedback from the statistical community about the issues and concerns raised. Online documentation and a comprehensive report will include a literature review and evaluation work.

Some issues transcend the scope of NCHS. A question was posed about involvement of the rest of the system in the beginning (e.g., Federal Interagency Council on Statistical Policy or the Federal Committee on Statistical Methodology). The Center is trying to bring together a variety of approaches. There is less concern about consistency across the federal government when the numbers are accompanied by footnotes about those that should not be used for policy-making.

A National Academy of Sciences panel entitled "Addressing Priority Technical Issues for the Next Decade of the American Community Survey" recommended that the Census Bureau for avoid data suppression due to quality in the American Community Survey (ACS) (although suppression must occur relative to confidentiality). Because numbers are aggregated in one-

three- and five-year estimates, it was noted that the Census Bureau should release the data in tandem with a program to educate data users about quality issues. Is NCHS different from the ACS or the Census Bureau? NCHS must decide whether it should enforce a data quality standard. An alternative to full suppression would be to mark certain estimates as unstable.

OAE would like feedback from BSC members as to whether all data should be presented: and if so, what should accompany the information relative to use and stability? In addition, what educational component should be attached? Feedback is sought about priorities relative to these issues. The BSC is in general agreement that the data ought to be released and that OAE's direction with regard to these issues seems reasonable. More discussion is warranted about whether to move away from the relative standard error to a more nuanced approach. Dr. Davern (who supports moving away from RSEs) volunteered to consider the issue further as did Dr. Kenney (other volunteers may be forthcoming) in a small workgroup. Concerns were raised about whether the new approach would lead to some frustrating anomalous suppressions and about the criteria for suppression. Reasons why the standard error is less effective were presented. Consensus about suppression will be very difficult to achieve within the federal statistical system. If all estimates are shown, it is important to responsibly warn people about what is and is not stable.

ORM Research and Development Project on Question Testing and Use of the Web **Nathaniel Schenker, Ph.D., ORM**

Big data are "out there to stay" as are web panels and well-designed surveys. ORM has proposed a research project (still in early stages) of methodological surveys known as *The ORM Research and Development Survey (RANDS)* that will examine web panel methodology and develop new mixed-method question/evaluation techniques. Research goals were described as were commercial web panel survey methodology; new question evaluation methods; and an example of a potential Question Set. If valid data can be obtained from commercial web panels, one project motivation is to see whether such panels can supplement the HIS.

Declining response rates (which affect data quality and underlying estimate assumptions) are identified as a problem with survey statistics. Simultaneously, use of the web to collect data has grown, particularly by surveys (including panel surveys) in that it is fast and inexpensive. It is not yet known whether web panel surveys can be used for official government statistics. This project compares what is obtained from a well-established survey to that obtained by Survey Monkey or other panel surveys to determine if the data are comparable.

Discussion Questions arose about whether demographics would be the same for the web and non-web-based respondent groups. Rather than ask the same questions on the web, a suggestion was made to enhance the quality and accuracy of reporting. Internet panel surveys have been used for the past three to four years during flu season, targeting specific subgroups with policy concerns about immunization rates. While no web survey evaluations exist at present, imperfect data are better than no data to help policymakers make difficult decisions during flu season (further conversation is warranted about this concept).

Comparing in-person versus commercial vendors on the internet excludes other considerations such as use of phone surveys or people who would not participate in-person surveys. What are the differences between internet-based surveys versus web panels? A suggestion was made to consider doing on-line surveys that have nothing to do with commercial panels. It was reiterated that commercial panels are cost-effective and quick; and that the project is a first step to see

what can be learned. The University of Michigan project and Danny Pfeffermann of Hebrew University were suggested as project resources.

Additional study questions were suggested: will there be reduced variability on responses to the simple questions asked on the web because they target a more selected (homogeneous) sample? Could this affect the extent to which responses correlate? Are there factors related to web panel participation that could modify the relationship (e.g., SES)? One must be careful about inferring a constant relationship across categories within a relatively homogeneous web panel because participants have varying attributes.

It is possible to specify some desirable variability within panels although there is no transparency about how people are included. The project aims for as much panel variability as possible. It was noted that there are different ways to request data from panel companies. It might be useful to examine how people who use the web for the NHIS differ from those who don't (which might be a variable for model development). In the end, a statistical adjustment is needed to match the web panel to the NHIS as closely as possible (although there are times when adjustments do not work). Core questions can supplement socioeconomic variables.

This type of research must be done despite the complexity and certain disadvantages because panel surveys are increasingly used. Probability samples must be examined with coverage and response rates in mind. More research is needed to understand numeracy and literacy issues among included panel populations. Post-stratification probably means having to go further than the usual age, sex, race, income and education categories. Even if the research yields a negative result, some good information will have been obtained. ORM might eventually make its own panels. Moving forward with panels is essential, noting that places like NORC are quickly moving into the panel world.

CDC's Ebola Response in Liberian Counties: October 15 – November 14, 2014 **Joe Woodring, D.O., M.P.H., M.T.M.H., DHANES**

Dr. Woodring described CDC's county-level response to the Ebola crisis in Liberia, where he was stationed for several months in the Fall of 2014, focusing on epidemiology, surveillance and country reporting and operational studies. His work took him to Nimba County (bordering Guinea and the Ivory Coast) and the sparsely populated Sinoe County, noting terrible road conditions. Training was provided to 25 health clinics, especially with regard to dead body management and ways that healthcare workers could prevent infection.

The difficult circumstances and repercussions of the spread of the disease were described, including school closures; faulty and limited communication systems; and highly inadequate treatment and protective gear options. A retrospective case of a 45-year old man who denied Ebola symptoms and exposure history was traced, noting that 65 cases emanated from the initial patient within six weeks in three villages with 72% case fatality. The high-risk of traditional funerals and the importance of involving tribal elders early on in outbreak areas were stressed. A confirmed index case was followed at the UNHCR Bahn Refugee Center (3500 mostly Ivory Coast refugees), identifying 58 contacts in four homes but thankfully, no secondary cases.

The focus in Sinoe County was to encourage triage and establish community Clinical Care Centers with viable septic systems. The difficulties of tracing the spread of the disease in an overpopulated government camp with substandard housing (populated by Liberian civil war

rebels, now illicit gold miners) was described; as were fears about exposure of American health workers returning from places like Liberia.

Discussion Despite a surveillance chart implying a decrease in the number of Ebola cases, much work remains relative to the trajectory of cases. Inaccurate data in Nimba County made it hard to keep an accurate count of confirmed, suspected and probable cases. With regard to the challenges of data in the field, it was noted that a national surveillance system generally worked while development of a separate data management center for Ebola by counties did not. As a result, enhancement of the national surveillance system with Ebola information was recommended. NCHS has an important role to play.

There is an emerging consensus that data collection under such circumstances is overcomplicated; and that collected data are not being used. The Global Health Security Agenda is “real”; and many countries will set up Emergency Operation Centers with CDC’s help. Distilling recommendations from all NCHS staff with hands-on experience in high-risk Ebola communities is recommended. Training and background for this kind of work were described, noting that risks to health professionals in the field also pose risks to CDC’s role in these countries. The need for coordination between various agencies (e.g., WHO; UNICEF; Africare; the African Union) was stressed.

The meeting was adjourned at 4:30 p.m.

Friday, January 23, 2015

Welcome and Call to Order

Raynard Kington, M.D., Ph.D., Chair, BSC

Hospital Participation Rates and EHRs

Clarice Brown, Director, DHCS; and Carol DeFrances, Ph.D., DHCS

Overview: National Hospital Care Survey

Ms. Brown provided an overview of the National Hospital Care Survey (NHCS), noting various data gathering sites. NHCS’s background, purpose, goals and objectives, opportunities for data collection; linkage; health services research; and healthcare organizations were described as were information about sample design; collected data; and protected health information (PHI). A data collection status summary was given with regard to hospital recruitment and key participation challenges (i.e., competing demands; timing; and resources), noting use of a focus group to understand how to engage resistant hospitals. Activities to move NHCS forward were presented including development of the HL7 Implementation Guide for clinical Department Architecture (CDA) release to National Healthcare Surveys; Census pilot and EHR vendor studies.

Discussion Participating hospitals have been randomly chosen although the type of EHR system and meaningful use payment are known via the HMS analytic database. It was noted that hospitals have been paid millions of dollars to use certified EHRs. NCHS’s request is consistent with this requirement so long as survey requests fall within the meaningful use package.

Surveillance data on heart failure and stroke rates and patient-level data for hospitalized people will be used. Population rates will also be used when there is a full sample. Noting that NIH is

partnering with the VA Hospital in examining the Million Person cohort, a suggestion was made for NCHS to consider partnering with VA hospitals as a way to tap into those systems. Survey design questions were discussed with regard to samples, noting the low 20% response rate.

National Hospital Care Survey: Electronic Health Record (EHR) Activities

Dr. DeFrances reviewed the electronic data process for NHCS. EHR data are requested first; Uniform Bill (UB)-04 administrative claims data come next; and state files are the third choice. Benefits and challenges of UB-04 data were presented as were the benefits, initial approach and challenges associated with EHR data. In contrast to UB-04, a data standard, EHR data constitute a system developed by commercial vendors that pulls information from multiple sources. In collaboration with the Census Bureau, an EHR pilot study and an EHR/UB-04 pilot study (both using hospitals from NHAMCS) were developed; and lessons learned from the pilots were shared (for example: a mechanism is needed to extract the data). An EHR vendor study was also described which aimed at identifying vendor interest in developing a module, template or interface to help get the data out. To date, NCHS has located one sampled hospital to work with Epic (Epic will support this interface, once built; and continue to allow hospitals to use it) as it continues to look for others to work with Cerner and Allscripts. Other EHR activities include the development of a Request for Information (RFI) published in October 2014; and exploration of natural language processing (NLP) to help identify substance-involved ED visits.

Discussion Quality data department staff members are pushing hospitals to develop and use such interfaces to help get the data out because doing so will improve quality management. CMS is very interested in national data; and they believe that the survey will act as a laboratory for developing quality measures they don't yet have. To date, mostly UB-04 data have been gathered with diagnosis, procedures and demographics being examined first. Gathering EHR and UB-04 data will allow for procedure comparisons. A pilot study was done for NAMCS that compared abstracted data to EHR data. The data were very comparable because they were abstracted from the EHR (i.e., the same source yields the same content). With the move to EHRs, it is important to recognize where required data can be found.

To date, work has not been done with the clearinghouses nor have there been discussions with the larger Health Information Exchanges (many hospitals go to the HIE). A non-profit of HIEs called K-Hi has emerged in California to facilitate public health reporting (data extraction, translation and delivery, single and batch), which is legally required but not funded. Part of Meaningful Use, Stage II stipulates that providers must transfer CDAs with 10% to systems unlike their own. Collection of NPIs and license numbers were discussed as were the 837-UB-04 cross-over and the work CMS is doing on the Transfer of Medicaid Statistics Information System (TMSIS). CMS will have mapped the 837 to the TMSIS files by the end of 2015. Universities are another possible resource for investigative development work. California's Office of Statewide Health Planning Development uses a model that feeds data back to hospitals for a fee. Although states have gone to new vital statistics systems with new web-based vendors, this has actually slowed down information exchange. In order to gather a composite picture of what is going on, the systems have been re-engineered to send batches of records at a time.

Data Visualization for Discovery and Insight – The NCHS Experience **Yinong Chong, Ph.D., DHANES**

The presentation explored tools capable of new kinds of reporting for data organization, based on current resources. A theoretical review and comparison of major data systems (vital

statistics; health surveys; longitudinal study; linked data) with government reports; scientific journals; chart books; electronic data/reports; online publications and interactive statistics was put forth in response to the question of whether data visualization tools improve data presentation and storytelling. Advanced technology was reviewed in the form of several user-friendly tools. Gartner's Magic Quadrant for BI Platforms was mentioned as a widely quoted review that rates business intelligence software in terms of vision execution and completeness.

The search for tools began with Tableau, noting constraints with the IT environment and the budget. The CDC Innovation Lab provided a virtual server onto which a range of tools representing different strengths and combination of strengths were downloaded. Different kinds of data mapped out the expedition (e.g., SAS Visual Analytics (VA), a statistical package for complex survey design; and Tableau for micro-data and its relative maturity in visual products). NHIS and NHANES data were used for R-SHINNY; and an attempt was made to mimic and repackage some quick statistics for publications (example given: influenza MMWR tables verses R-SHINNY charts). The new technology allows a small piece of the web to hold a great deal of information. An application has been developed to help individuals interact with the growth chart using R-SHINNY. This "one-stop" application allows for the entire complex survey to be taken care of in one suite.

HIS and NHANES data were tested with the SAS VA (video rather than live demo distributed). Such tools can be powerful for internal exploration in that they allow analysts to see them (as well as patterns and influential points) visually without complicated programming. SAS has also been used to repackage all NHANES cardiovascular risk factors (for example) into one place, allowing for different years and data as well as an ability to maneuver between survey cycles and use of different chart types. Tableau examples illustrating new interactive ways to organize data and graphics were presented as were new ways to publish such data on the web.

Visual communication is dynamic, integrated, interactive and automated. It is possible to combine multiple data sources in one dashboard; allow users to create charts of interest and share easily using the support for existing security. Visual tools integrate mega-data in varying sizes and formats; and combine analyses and graphics while providing dynamic and interactive user interface, automated real-time reporting and instant publication and sharing.

Discussion There is a need for integration between computer scientists, statisticians, modelers, sociologists, demographers and others. An important question is how to begin understanding and expanding the expeditions and explorations on a larger scale. A concern was raised about a lack of nuance in data coming out of the visualization structures, noting that the provision of explanatory footnotes is a big commitment. Most visualization tools come with two kinds of security precautions. With the insertion of larger raw data, *who* has access can be specified in the design phase. No finished products published on the web are raw data. Suppression can also be built in (e.g., maps; data download). In addition, notes can be inserted. Of what is currently available, SAS and R-SHINNY are probably the best tools for survey data professionals in that they are easier to manipulate; they bring the data indirectly; and they can deal with large survey data in a complex survey. Tableau is very nimble and versatile although one must go to R integration to complete the survey data. A concern about R (an open source) was raised relative to quality control. SAS Jump is another user-friendly piece of data visualization.

The visualization strategy is to first use all the free and open resources and then to publish small pieces in order to gauge how users interact with them. Without commitment to a major tool deployment, further exploration could reveal how to get to clearance and publication. The

push around open data and transparency is to get data out so that they can be used. A question was posed about whether it is bad for the end products to look official. Sometimes, it is easier for non-governmental entities to develop useful applications (example given).

Data visualization is currently a hot topic in the statistics community, especially with the large number of datasets available. The field of data analytics is also growing. The federal government's current perspective is toward open access, which means an inability to control how people manipulate or characterize data. The current Administration is pushing for more open license as are universities. Who is responsible for addressing issues with open source software security (e.g., risks, patches, testing, updates and maintenance) and where are the resources? A suggestion was made to not only make the data source public but to add "created by" (or an electronic logo) to cite associated agencies.

Statistical Policy Directive #1: Fundamental Responsibilities of Federal Statistical Agencies and Recognized Statistical Units
Katherine Wallman, Chief Statistician, OMB

Ms. Wallman introduced the Trust Directive, otherwise known as the Fundamental Responsibilities of Federal Statistical Agencies and Recognized Statistical Units by outlining the legislative and executive initiatives to maintain confidence in the system. She described the function of the Chief Statistician's Office and OMB's role in establishing and enforcing standards. Examples of OMB Statistical Policy Standards were presented, followed by a discussion of the reason for this fundamental directive at the present time. This is a good time for the Executive Office of the President to articulate expectations for the core values of agencies producing official statistics. Implementation of this high-level Directive, with all its challenges, is primarily the responsibility of departments and agencies.

The Directive applies to 13 federal principal statistical agencies; three recognized statistical units and over 100 federal programs. Its four fundamental responsibilities were delineated. A 120-day reporting requirement from each department and agency hosting a federal statistical agency or unit to OMB is intended to ensure that the Directive's mandates are being met and to identify best practices. Due in March 2015, the reports are expected to propose solutions to implementation obstacles. They will become an ongoing activity to help address agency concerns while placing this fundamental Directive at the core of OMB's standards and guidance.

Discussion The report is the right thing to do at the right time, particularly in light of changing leadership in CDC, HHS and NCHS. It will serve as a mechanism to ensure that people carefully consider the Directive in addition to providing a training opportunity that encourages dialogue. One issue to discuss with the Department is the fact that there are no policies or guidelines (as mentioned about the Census Bureau) that address what is in this Directive. The BSC expressed support for the Directive and its implementation and developed plans to voice it to the Department. Board members were encouraged to examine the Directive to provide feedback after which, a letter of support to the Department will be drafted by Dr. Cain and Dr. Kington. The letter might show interest in how the Directive will be implemented; demonstrate support for the Directive; inquire about next steps; and even state the OMB has done a good job. It should state the Directive is important and should be implemented appropriately. A suggestion was made for BSC members to also examine Directive #4 about releases, bearing in mind an intention to link internal policies to it. A suggestion was made for BSC members to also review accompanying preambles to the Directives.

