

# Data disaggregation as a means to improved health research and policy-making: Event summary

On September 26, 2019, the Brookings Institution hosted a symposium titled [“Data disaggregation as a means to improved health research and policy-making.”](#) The event convened experts in data collection, research and health policy. The event was sponsored by the Robert Wood Johnson Foundation.

## Introduction and welcome: William H. Frey

The keynote speaker, William H. Frey (Senior Fellow, Metropolitan Policy Program, Brookings Institution) discussed the increasing racial and ethnic diversity of the US in coming decades. His comments framed the important policy concerns that will be influenced by this demographic change in the realm of health policy. In coming decades, the number of individuals who list two or more races will eventually eclipse the existing standard race categories. This poses an issue for categorizing and anticipating health and wellness needs or programs that are based on single-race categories.

## Panel 1. “Critical Needs for Data Disaggregation for Race and Ethnic Groups” – Why do we need to disaggregate?

The first panel was moderated by Richard Reeves (John C. and Nancy D. Whitehead Chair, Senior Fellow, Economic Studies; Director, Future of the Middle Class Initiative, Co-director, Center on Children and Families) and focused on the need for data disaggregation in formulating policy analysis and interventions. The first panelist, John Yang (President and Executive Director, Asian Americans Advancing Justice [AAJC]) described issues in the Asian American community with aggregated data and the model minority myth. In fact, he notes on measures such as income and educational attainment the data may be much more bimodal for Asians. He also noted that there are differential growth rates by different subgroups of Asians in the US. The second panelist, Ana Gonzalez (Senior researcher, Pew Research Center), discussed the relative size and growth rates of Hispanics in the US. Hispanics are a relatively large group at approximately 60 million people in the US, but the median age is young at 30 years old. Many lack health insurance and have relatively high poverty rates. Additionally, few have paid attention to the differences within the Hispanic group by country of origin, income, or educational attainment. The third panelist, Yvette Roubideaux (Vice President for Research and Director, Policy Research Center, National Congress of American Indians), emphasized that American Indians and Alaska Natives are a small proportion of the US population (around 2 %) and are quite different from one another. Dr. Roubideaux emphasized that in many data sets the AIAN population doesn't appear due to small sample sizes. She said that “being invisible is a policy problem.”

Following this initial panel, the symposium continued with the presentation of five research papers.

### Paper 1. “Asians and Hispanics Earnings Inequality in Administrative Data”

Randall Akee (UCLA and Brookings Institution) presented earnings data on Asians and Hispanics disaggregated into their subgroups such as Chinese, Japanese, Korean, Vietnamese, Asian Indian, Filipino, Mexican, Puerto Rican, Cuban, Central American and Latin American. The disaggregated data shows considerable variation across the subgroups of Asian and Hispanic that are otherwise obscured when looking at only the single aggregated category. The analysis shows the benefits of disaggregated data for identifying social determinants of health outcomes, in this case earnings and inequality.

### Paper 2. “Disaggregating Data to Advance Health Equity: The Case of Native Hawaiians and Pacific Islanders”

Sela Panapasa (University of Michigan) presented the second paper on “Disaggregating Data to Advance Health Equity: The Case of Native Hawaiians and Pacific Islanders” which examined the diversity within the Native Hawaiian and Pacific Islander population. She noted the fact that the NHPI group has the highest percentage of more than one race in national data. One important issue that may affect health outcomes is that new procedures and health interventions aimed at certain groups may be based on self-identified race and ethnic categories as opposed to genetic make-up. If these different categorizations do not align, this may increase misdiagnosis or incorrect treatment. Finding ways to self-identify and engage new and evolving categories of race or ethnicity are important for healthcare delivery and diagnoses.

### Paper 3. “Analyzing the Impact of Differential Privacy on the Accuracy of Decennial Census Data”

David van Riper (University of Minnesota) discussed how the new differential privacy policy may affect US Census data in the future. He emphasized that this new privacy will contain an explicit trade-off in privacy loss and data accuracy. There is concern that this may have an effect on data for small populations or small, remote geographic locations. Using public-use 1940 US Census data, he implemented the differential privacy procedure with varying levels of privacy loss. He provided several figures showing, as suspected, that lower levels of privacy loss are associated with less accuracy in tables (relative to the true counts) and that this differs by the unit of geographic analysis. He said that it appears that the problem is the most acute for the smallest race groups who will be most affected and have the least accurate data.

### Paper 4. “Welfare Reform and Adolescents: Are the Boys Alright?”

Hope Corman (Rider University) presented her research on the effect of welfare reform in the US on subsequent generations. Her analysis shows that aggregated results for children shows little to no effect on child outcomes. However, upon disaggregating by gender, she finds that there is an increase in delinquency behavior for boys and substance abuse.

### Paper 5. “Disaggregating the Data for Bisexual People”

Shabab Ahmed Mirza (Center for American Progress) presented research from a new survey on bisexuals in the US. Specifically, the analysis focuses on Bi+ groups which are individuals that identify as bisexual, queer, pansexual, or attraction to more than one gender. The new data shows that Bi + women differ significantly from lesbian women. Bi + women tend to report more

mental health problems and are more likely to use public benefits, SNAP and have higher poverty levels. In disaggregating the data by race between white and non-white populations, the differences between Bi+ and monosexual peers disappears for measures of physical health. These results by race and Bi + identity seem to suggest that race may play a larger role in these health outcomes

## Panel 2: “Data collection and data access: current issues and future challenges” – Where disaggregation is going and obstacles

The second panel focused on data collection and the challenges faced in these efforts. Emilia Simeonova (Johns Hopkins University Carey School of Business) moderated the panel. The first panelist, Ninez Ponce (Director, UCLA Center for Health Policy Research, UCLA), described the challenges and successes in administering the California Health Interview Survey. The data contains oversamples for race and minority groups in California and it is one of the first to query gender expression and gender identity as well. An indicator of the success is that after CHIS data became widely available, the number publications focusing on Asian subgroups increased dramatically as measured in PubMed. This underscores the demand for more data at the disaggregated level in health and other disciplines. The second panelist, Megan T. Khau (Director, Data and Policy Analytics Group, Centers for Medicare & Medicaid Services), discussed the methods for collecting health data in federal agencies. She emphasized the changes that are occurring at the federal level, for instance, the Affordable Care Act mandates that data now be collected on race for patients for 5 subcategories for Hispanics, 6 subcategories for Asians, and 4 subcategories for Native Hawaiian and Pacific Islanders. The third panelist, Makada Henry-Nickie (David M. Rubenstein Fellow, Governance Studies; Race, Prosperity, and Inclusion Initiative), discussed issues around identifying race and ethnic group in mortgage data such as the HMDA dataset. She also mentioned the issue of race imputation that is commonly done which may lead to discrepancies or errors in analysis.

## Concluding discussion and summary of next steps

In the final session, the participants engaged in a group discussion of topics and issues to consider for the future of data collection and how to disaggregate data. One point raised was that guaranteed privacy of data may not exist in the future, however, perhaps there should be more advocacy for laws that govern how the data will be used. A second point raised emphasized that data disaggregation is informed by and assisted by qualitative interviews. These efforts should be seen as complementary, as they inform and enrich existing or new race or ethnic categories. Third, there were concerns that there should be group protections regarding data in addition to individual-level protections. Finally, there is continued concern for how data relating to undocumented individuals in the US will be collected and whether this will be representative of that population.