

NWX-US DEPT OF COMMERCE

Moderator: Kristina Barrett

August 5, 2021

4:40 pm CT

Coordinator: Welcome and thank you for standing by. At this time all participants are in a listen-only mode until the question-and-answer session of today's conference. At that time you may press Star 1 on your phone to ask a question. I would like to inform all parties at today's conference is being recorded. If you have any objections you may disconnect at this time. I will now turn the call over to Mr. Michael Cook. Thank you sir. You may begin.

Michael Cook: Good afternoon everyone and thank you for joining us virtually today. I'm Michael Cook, Chief of the Public Information Officer at the US Census Bureau and I will be moderator for today's event. We're hosting this Webinar ahead of the release of the 2020 Census P.L. 94-171 redistricting data.

This informational Webinar will include instructions on how to access redistricting data, information on improvements to the race and ethnicity questions design, processing and coding. Along with the presentation on how the Census Bureau is measuring diversity in the United States.

This will be a longer media briefing than we normally do because we have a lot of complex material to cover. We want to ensure you have the information you need to use it in your reporting. These are all important topics for members of the media and the public to use as you prepare to view and analyze the first local level results from the 2020 Census.

A live Q&A session for credentialed media with Census Bureau subject matter experts will immediately follow today's Webinar. Also for media members who subscribe to our emails, please check your inboxes after this Webinar for two important announcements from us.

Today we will hear from subject matter experts across the Census Bureau who will provide insights to help you prepare for the redistricting data release. I will first turn it over to Mr. James Whitehorne, Chief of the Redistricting and Voting Rights Data Office, to provide an overview of how we will release the 2020 Census Redistricting data. James?

James Whitehorne: Thank you Michael. I'm very happy to be here today to be able to provide everyone with some background on the 2020 Census redistricting data and the Redistricting Data program. I want to start by talking a little bit about the program itself. The Redistricting Data program at the Census is our response to Public Law 94-171, which was passed back in 1975.

The umbrella requirement for the program is that it's conducted in a nonpartisan manner. Now we accomplished that by having each state assign us a liaison to the program that can act in a nonpartisan manner on behalf of the state for the duration of the program.

The law dictates two primary activities. First, it requires that the Census Bureau allow states to help define the small area geography that they need for

redistricting. Over the decade these geographies have been census blocks which is the smallest geography for which census publishes data, voting districts which is our generic term for precincts and wards and then state legislative and congressional districts.

The second activity is to deliver tabulations for that geography in a timely manner which for a typical census would be one year from Census Day. However, as we all know the delivery has been delayed but we will soon be able to provide the high quality data that states need for redistricting.

So now I'd like to speak briefly about the data content that will be coming with the 2020 Census Redistricting Data release. I'm going to start by making a distinction between our already released apportionment count and the upcoming redistricting data.

As you know the Census Bureau released the apportionment counts on April 26 of this year. Those are numbers used for determining how many seats each state is entitled to in Congress. However, the apportionment counts are only provided at the state level and only for the total population.

The redistricting data that will soon be released is available for multiple characteristics and for a range of geography from the state all the way down to those individual census blocks. This provides redistricting officials with the data they need to do their work, but it also provides the public with the detailed information they need to know about their county, their township, their city, their neighborhood, their community.

This redistricting data will be protected using our modernized technique of differential privacy through our application of the top down algorithm. In order to make this data available as early as possible we have split our

delivery of the data by providing different formats of the data as they're ready for distribution. And I'm going to talk about that more in a moment.

The actual content of the 2020 redistricting data should look very familiar to anyone who worked with the 2010 data. The first five tables are exactly the same as what was reported in 2010. We have two tables one that contains information on race and another that identifies ethnicity as Hispanic or Latino first and then reports race for non-Hispanic or Latino populations.

We repeat those same two tables for the voting age population. We also have a table that shows housing unit counts by their occupied and vacant status. However we've added a table this decade.

So new this decade, and to support states that have a state statutory obligation to reallocate population from group quarters for the purposes of redistricting, we're including in the redistricting data a total population count by group quarters type. All the tables that I just mentioned are going to be provided down to that individual census block level.

In January and February of this year we provided the supporting geographic materials from the redistricting data program. So I want to touch briefly on what's - what we have available in that arena. We provide several types of geographic support. One of the primary products that we provide are the shapefiles.

For those of you who don't know shapefiles are the digital geographic boundary files which are used by Geographic Information Systems for mapping. Most redistricting officials will use GIS to be able to draw and adjust their plans by associating data with these shapefiles.

However, we also provide several other items. We have tabular listings showing how blocks relate to other geography and how blocks relate to each other between decades. But we also produce a series of maps and map types that can be especially useful to smaller jurisdictions that need to redistrict but don't have GIS resources.

We provide state legislative and voting district maps, school district maps, census tract maps and county block maps. The county block maps provide coverage of every county in the entire country with enough detail to identify the individual blocks. These can be used alongside the data to perform an old fashioned pen and paper redistricting if needed.

Now you keep hearing me mention census blocks. The reason why census blocks have been identified as being important is because they're the piece of geography from which all other geography is built. You can see by following the lines on this chart the relationships between geographies.

And you follow the line from blocks up the geographic spine you can see that blocks nest within block groups which nests within census tracts within counties and then state. But as you can see by the lines going up the side off the spine blocks also make up the primary building geography for voting district which nest within county. And blocks make the primary building geography for legislative and congressional districts which only nest within state.

Now I want to address the split data delivery that I mentioned earlier. The Census Bureau has been working hard to be able to provide the high quality data that states need for redistricting. Since the release of the apportionment counts we've made certain that all records are coded to their most detailed level of geography, we've conducted our characteristic editing and

implementation to ensure that all records have valid values for all the major characteristics, we've applied the top down algorithm to protect the confidentiality of respondents and where we find ourselves today is at the tail end of tabulating and reviewing the data and its multiple formats.

Once this is complete the legacy format summary files will be ready for release. It will take us more time to generate the easier to use formats which will be ready in September, and I'll touch on that in a moment as well.

The legacy format summary files that we're close to releasing were always part of our product plan. We're just releasing them earlier due to the needs of the states to get started with their redistricting. These are a familiar format that's been around since at least the 2000 Census and will be completely reviewed before we release them?

However they do require some additional handling to be able to pull out familiar tables for specific geography. We've consulted with major software vendors, redistricting software vendors. We've talked to the redistricting and elections staff at the National Conference of State Legislatures, we've talked with the nonprofit group Redistricting Data Hub, and we've had several conversations one on one with states all to ensure that they understand and are comfortable using this format.

To the ones that we've spoken with they have all indicated that they can use this format as many have done so over the last several decades. But we also know that despite the states being able to use this format for redistricting there will be many in the public that want access to this data. We've been working hard to make sure that they too will have tools to help them look at this data but we're still working to present it in September through our data.census.gov primary dissemination platform.

So with the August release we will have some mapping tools, data visualizations, America count stories and we'll include highlights and news releases. I'm going to single out two of the support tools and you'll hear about others in some of the later parts of the presentation.

The first of these is our QuickFacts application, we'll load the 2020 Census population counts for the nation, states, Puerto Rico and counties into this tool when we release the data or very soon after. This tool will also include the population counts for places, zonas or urbanas and townships that have populations of 5000 or more people.

The second tool I want to point out is our 2020 Census demographic data map application. This mapping allows data users to look at state, county and tract level data for most of the characteristics that are contained in the redistricting data. It will also show people population change and density. And by hovering over a specific piece of geography in the map data users will be able to see the actual number for that map type behind each specific piece of geography as it's being displayed.

For those data users that want to jump in and work directly with the legacy format file we have provided a large number of support materials that are already available online. We've got the technical documentation - one second we can go back - and we've got the technical documentation which is already posted and this has all the critical detailed information about the organization and the files.

We have prototype redistricting data summary files. These are prototype data that we've made from our 2018 end to end census test that was conducted in Providence County, Rhode Island that are in the same format as what's

planned for the 2020 release. These data allow users to build systems, test systems, practice making queries from the data so that they are prepared once these files hit the streets.

The link that you see there on the Redistricting Data Program Management page that'll take you to the page that hosts this prototype data. And we've got another easier way to find that as well coming up here in the second.

We have header files in Excel format. So if people are planning to bring these into their database and they want to have something that already has the layout or the header for each of the files that's already prepared. And they can pull that from an Excel table.

We have a Microsoft access database shell. This is an empty database with just shells for the data that makes it easy for people to import the data. We also have example queries in there so people can understand the logic of how you work with these files to extract data from them.

We have an illustrated guide that goes along with that in PDF format. And we've also added a video just recently that walks people through using that guide and gives them a visual representation of the steps to be able to pull data from that.

And then for those folks who are just going to work with this data in a statistical method we've got two import scripts one for SAS software and one for R that allows them to easily import this data and then they can begin their work of analysis.

So we're providing several consolidated locations where the redistricting data materials and support products can be accessed. The first is my office's site.

We have a very simple Web address, www.census.gov/rdo. Here you can find access to the prototype data.

And it's indicated on the slide where you would go for that. And support products for the redistricting data release which you can also see where that will be available. You'll be able to find links to the geography files, the data files and many of these other data access tools that I've also spoken of just in earlier slides.

We also have a 2020 Census Results page which will also work to direct - as a direct access point for many of these materials. In addition the press kit that's provided by the Census Bureau will continue to be updated with relevant materials including links to many of the materials that I'm discussing today and the actual data itself.

So to wrap things up I want to touch on the September data release. This will be the same data but in an easier to use format. One format is the DVDs and flash drives with an integrated data browsing software that we're going to send to official recipients which are the governors, the state legislative leaders and any existing commissions that are in the country.

The second will be the census primary data dissemination tool called data.census.gov. This tool will provide everyone with full easy to use access to everything contained in the redistricting data set.

So I want to thank you for your attention. I hope this was useful. And now I'd like to pass this over to Matthew Spence, who is a Senior Adviser in our Population Division. Matthew.

Matthew Spence: Hello everyone. Thank you for joining us today. Today I'd like to talk about the 2020 Census's Disclosure Avoidance System.

The Census Bureau has a deep commitment to protecting the privacy of our respondents and the confidentiality of their data. It's not only a legal obligation under Title 13 and other laws but it's also a core part of our institutional culture.

But protecting privacy doesn't just mean keeping individual responses confidential. In fact, every time we publish tabulations or statistics we leak a little bit of private information. In fact, if we publish enough data tables we risk allowing the identification of individuals. For example, if we release detailed data for small areas we could potentially identify individuals albeit unknowingly.

So Census has continually improved the methods of privacy protection over time. In 1930 Census stopped publishing small area data. In 1970, we pioneered the use of whole table suppression. In 1990, suppression was phased out in favor of data swapping where individuals or whole households would be swapped across geographic boundaries. And then in 2020 we're introducing - we're proud to be the first decennial census to use differential privacy.

So why differential privacy? Really, there are two reasons. The first is that there has been a big proliferation of other databases with individual level information that could be used in conjunction with our census publication to identify individuals. And although, you know, we can't control that, we need to be extra careful.

The second reason is that faster computers make reidentification attacks much more feasible. These attacks are basically a series of calculations, so faster computers with more memory make those attacks easier and more feasible.

All traditional methods of disclosure avoidance like suppression, swapping or rounding try to make these attacks more difficult by reducing precision, removing vulnerable records and adding noise or uncertainty. Indeed, all statistical disclosure avoidance methods involve a trade-off between privacy and accuracy. And all of them have their own way of quantifying that trade-off.

The nice thing about differential privacy is that it's a framework for defining and quantifying privacy protection. Differential privacy allows precise control over that balance between privacy and accuracy while preserving mathematically provable future proof privacy protection.

And speaking of tuning over the past two years, the Census Bureau has released five sets of demonstration data products that run 2010 Census data through the software for the Disclosure Avoidance System. And we published that data to show, you know, the fitness for use of this data.

We took feedback from external stakeholders and data users. And we made significant improvements to increasing the accuracy, reducing bias and reducing the number of outliers. This graphic on your screen for example shows how many incorporated places in the country had a larger than five percent change in the total population due to the application of disclosure avoidance.

As you can see in October 2019 our very first demonstration data product nearly 8,000 of the 19,000 incorporated places had a larger than five percent

error in the total population. But now all the way on the right side of the screen the production settings run it's about 500. So that's a significant improvement in reducing those outliers.

We've also heard from stakeholders about, you know, things that were biased that we've addressed by tuning our queries, by tuning the parameters. And because of that there's been a substantial reduction in bias and a substantial increase in accuracy. And all of that is thanks to feedback from our data users and our stakeholders.

On June 8, the Census Bureau's Data Stewardship Executive Policy Committee met to set the disclosure avoidance parameters for the 2020 redistricting data. They set a higher privacy loss budget. Now the privacy loss budget is a single number that sort of represents to a great extent the amount of that privacy versus accuracy trade-offs.

And what they did was they set a higher privacy loss budget to increase accuracy for the total population and queries for race by ethnicity. They wanted to improve it for the block group level and above. Now currently online we've got newsletters, blog posts and a detailed summary metric spreadsheet that allows data users to assess the accuracy, bias and the number of outliers for those production settings. Again, this is using 2010 data but it's the exact same software that was used for the 2020 production. And in fact, a new demonstration microdata file will be released along with the 2020 redistricting data.

So we have a couple notes of caution related to the Disclosure Avoidance System. The first is that small areas like census blocks may have noisy, fuzzy, inconsistent or improbable results. We've got blog posts and newsletters

explaining this. But just as an example there may be blocks where there are zero occupied housing units but there are people in households.

This is a result of how the Disclosure Avoidance System is applied separately to unit files and person files. But really what you need to know is that as you aggregate blocks together, as you group them together into larger geographies such as block groups, tracts or counties those go away. Those inconsistencies are substantially attenuated and do go away. You also get increased accuracy.

So that's our recommendation. We encourage data users to as much as possible group blocks together to create those sorts of aggregate statistics. And you can see the director's blog for more information on these results.

And lastly, I would like you - I'd like to encourage you to visit our Web site, search for disclosure avoidance in the top search bar and you'll come upon our 2020 Data Products Disclosure Avoidance Modernization page. Including at the very top right of the page a new video that explains in very clear language how this Disclosure Avoidance System works.

And with that I'd like to conclude and introduce you to Rachel Marks, who is the Chief of Racial Statistics Branch in the Population Division.

Rachel Marks: Thank you Matt. And hello everyone. Today I'll walk through the improvements to the 2020 Census race and Hispanic origin question design, data processing and coding procedures. These changes provide important context as we prepare to release the 2020 Census redistricting data.

The U.S. Census Bureau has collected data on race since the first census in 1790 and on Hispanic or Latino origin since the 1970 Census. How these topics are measured and statistics on them are collected and coded has

changed nearly every decade throughout the history of the census reflecting social, political and economic factors.

This slide shows a screenshot of our updated data visualization titled, US Decennial Census Measurement of Race and Ethnicity Across the Decades 1790 to 2020. This data allows users to see how the race and ethnicity categories on the census have changed throughout the decade, including in the 2020 Census. You can explore the data viz on [census.gov](https://www.census.gov).

Within the context of the United States decennial census, the questions and concepts of race and ethnicity follow the standards set forth by the U.S. Office of Management and Budget or OMB. The OMB minimum categories for data on race and ethnicity for federal statistics, program administrative reporting and civil rights compliance reporting are defined as follows.

The OMB requires two minimum categories for data on ethnicity Hispanic or Latino and not Hispanic or Latino. And the OMB requires five minimum categories for race American Indian or Alaska Native, Asian, Black or African-American, Native Hawaiian or Other Pacific Islander and White and people may report multiple races. The Census Bureau is also required by Congress to use the category Some Other Race.

Displayed on this slide are the OMB definitions for the minimum race and ethnicity categories. I present these definitions to underscore the point that the Census Bureau follows the OMB definitions to classify and tabulate data on race and ethnicity.

The race and ethnicity question format designed for the 2020 Census adhere to the 1997 OMB race and ethnicity standards. You can see the 2020 Census questions on this slide. The Census Bureau did not use the combined question

format for collecting race and ethnicity as the OMB standards require the use of two separate questions on ethnicity and on race for self response.

While a separate Middle Eastern or North African or MENA response category was not included detailed MENA responses were elicited, collected and coded. Per the 1997 OMB standard, MENA responses are classified as part of the White racial category.

While the Census Bureau tested an alternative question designed in 2015, we must ultimately follow the 1997 OMB standards and use two separate questions to collect data on race and ethnicity. Our testing however did show that we could make improvements to the 2020 Census race and ethnicity questions within the OMB guideline.

Overall, several significant changes were made for the 2020 Census questions on race and ethnicity and I'll go through this in the next few slides. The 2020 Census Hispanic origin question included the same three detailed checkboxes that were included in the 2010 Census of Mexican, Mexican Am., and Chicano, Puerto Rican, Cuban along with yet another Hispanic, Latino or Spanish origin checkbox.

There were two changes to the 2020 Census Hispanic origin question. The instruction to print origin for example was revised to print for example. In our research we found that this revised instruction allowed respondents to understand what the question was asking them to report and it did not limit their write in response by confusing the instructions with terms like origin that mean different things to different people.

We also updated the example or groups. They were revised from Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard and so on to

Salvadoran, Dominican, Colombian, Guatemalan, Spaniard, Ecuadoran, et cetera, in order to represent the largest Hispanic origin population groups and the geographic diversity of the Hispanic or Latino category as defined by the OMB 1997 standards.

We also made several design improvements to the race question for the 2020 Census based on our research over the past decade. In response to community feedback over the past decade, we added dedicated write-in response areas and examples for the White and the Black or African-American response category.

We provided six example groups for each of the White, Black or African-American and American Indian or Alaska Native racial categories. These examples represent the largest population groups within each of the geographically diverse population groups for each category, as defined by the 1997 OMB standard.

Based on successful previous testing, the term Negro was removed from the 2010 Census by updating the category black, African and/or Negro to black or African Am on paper questionnaires and black or African-American on electronic instruments.

We reordered detailed Asian and Native Hawaiian and Other Pacific Islander checkboxes by population size. And we changed the checkbox category Guamanian or Chamorro to Chamorro based on research and positive stakeholder feedback. We also updated the write-in instructions for the Some Other Race category to better solicit detailed reporting.

The 2010 Census form included instruction to print race. When we updated the 2020 Census instruction to read print race or origin to correspond with the overall question instruction to mark one or more boxes and print origin.

This slide provides a summary of processing and coding improvements made to the 2020 Census race and Hispanic origin question. Coding is a process we use to assign numeric codes to write in responses in our questions and we use the numeric codes when we process and tabulate data. In the 2010 Census we only captured the first thirty characters of written responses to the race and ethnicity questions and coded up to two write in responses in each write in line.

Our research over the past decade found that people were reporting longer and more detailed responses to the question. For the 2020 Census we wanted to reflect more fully and accurately the complex details of how people identified their race and ethnicity.

Based on further research, testing and outreach throughout the decade we changed how we captured and coded responses to the 2020 Census race and Hispanic origin question. We increased the number of characters captured from 30 to 200, which allowed us to capture and fully recognized longer write in responses instead of prioritizing multiple responses into only two codes. We coded up to six detailed codes for each write in area.

This increase effectively gave all the responses an equal opportunity to be coded into one of the six major race categories. It also simplified our coding work to remove duplicate or repetitive terms to reduce response to two codes. We fully tested these coding design and question changes in the 2015 National Content Test and finalized them in the 2018 Census Test.

This example illustrates how a response was coded in 2010 versus 2020 based on the differences I just described. The first example shows that in the 2010 Census the response of Mexican American Indian, Portuguese and (sic: African) American (sic: ~~Indian~~) was not fully coded because it was longer than 30 characters.

Only the text outlined by the red box was captured and coded in the 2010 Census. The additional text was not captured therefore the responses of Portuguese and African-American were not able to be coded.

The second example shows that the same write in response was provided in 2020, all of the text outlined by the red box was captured. This enables all three terms of Mexican American Indian, Portuguese and African-American to be recognized and coded.

In another improvement for 2020, we used a single code list for coding data from the Hispanic origin and race questions. In previous censuses we used two separate codes one for Hispanic origin and one for race. Previously the code list focused on providing codes for either detailed Hispanic origin groups or detailed race group.

By combining these code lists we expanded the number of detailed groups that could be coded in each question. We also expanded our code list to include additional detailed White and Black or African-American groups as the race question elicited the collection of detailed White and Black or African-American responses through dedicated write in lines for the first time.

During the 2010 Census, if someone provided more than two write in responses in the Hispanic origin question write in area, we prioritized coding Hispanic groups over race groups. In the 2020 Census, subject matter experts

coded what they saw, coding up to six responses from left to right regardless of the Hispanic origin or race group. This enabled all responses to be treated equally.

The following examples illustrate this coding change. A write in response of Black Peruvians as coded as Black and Peruvian in both 2010 and 2020. In this example although there's no Hispanic response to the write in because there are only two responses both the Hispanic and the race response were coded in 2010 and 2020.

Similarly in the next example a response of White Puerto Rican was coded as White and Puerto Rican in both 2010 and 2020 because there were only two responses. And the third example of Spaniard white and Honduran the response of White was not coded in 2010 because of the prioritization of Hispanic origin responses over race responses when there were more than two responses in a single write in line. In 2020 all three of these responses were coded.

In the last example of Mexican, Black and Colombian the response of Black was not coded in 2010 because of the prioritization of Hispanic origin responses over race responses when there are more than two responses in a write in line. In 2020 all three of these responses were coded.

After coding up to six responses on the Hispanic origin question write in line like in the 2010 Census only one response is permitted to be tabulated for Hispanic origin in accordance with the 1997 OMB standard. This coding change does not impact Hispanic origin data for the 2020 Census redistricting file.

As we just saw in 2010 this example write in response of Black, Colombian and Peruvian was coded as Colombian and Peruvian. And the response of Black was not coded because it's a race response. For the redistricting data this response was tabulated as part of the Hispanic or Latino category.

In 2020 we code all three responses. And all of these codes are retained internally for research purposes. For the official redistricting file tabulation the response is tabulated as Hispanic or Latino.

Following the 1997 OMB standard respondents can only be Hispanic or not Hispanic. As long as the respondent provides at least one Hispanic origin response they are tabulated as Hispanic.

The 2010 Census used a complex series of coding rules to determine how to prioritize and define up to two codes for each unique text string. In 2010 if more than two groups were part of a write in text string on the same line in the race question we prioritized coding race groups over Hispanic origin groups because they were limited to only coding two responses.

In the 2020 Census our subject matter experts coded what they saw coding up to six responses from left-to-right regardless of race group or Hispanic origin enabling all responses to be treated equally. So we have some examples here on this slide – a write-in response of Hispanic and White was coded as Hispanic and White in both 2010 and 2020.

In this example although the Hispanic response in the write in because there are only two responses both responses were coded in 2010 and 2020. Similarly a response of Black Latino was to Black and Latino in both 2010 and 2020 because there were only two write in responses.

In the third example a response of Hispanic, White and Chinese was coded as White and Chinese in 2010 because of the prioritization of race responses over Hispanic responses when there were more than two responses in a single write in line. In 2020 however all three responses were coded.

And then the last example a response of Spanish, Mexican, Samoan, and Chamorro was only coded as Samoan and Chamorro in 2010 and the responses of the Spanish and Mexican were not coded because of the prioritization of race responses over Hispanic responses when there were more than two responses in a single write in line. In 2020 all four of these responses were coded.

On this slide we can see how this improvement to the coding rules impacts the final data by recognizing the rich and complex detailed identities reported by respondents. Looking at the example response of Cuban, Thai and Filipino we see in 2010 even though Cuban was written first, as shown in the green text, the responses of Thai and Filipino were prioritized over Cuban because they are detailed race group. Cuban was not coded or included in race tabulations because the two race groups were prioritized over it in coding.

The redistricting data tabulates major race groups. So this response was tabulated as part of the Asian race category representing Thai and Filipino in 2010. In 2020 all three groups and the write in response were coded and the tabulation would be for Asian representing Thai and Filipino and some other race representing the Cuban response.

Improving the 2020 Census questions on Hispanic origin race along with our coding procedures enables us to have a more complete picture of the detailed identities reported by the US population in 2020. We expect that the Hispanic origin race statistics in the upcoming 2020 Census redistricting data will not

only reflect demographic changes but also improvements in how we ask the questions and caption and coded the responses.

These improvements will accurately illustrate the richness and complexity of how people identify their race and ethnicity in the 21st century. So thanks for the interest on this important topic. And now I'll turn it over to Eric Jensen, Senior Technical Expert for Demographic Analysis in our Population Division.

Eric Jensen: Thank you Rachel. Before I begin I want to note that for those who are dialed in that you can follow along our slide deck that is available in our redistricting press kit.

So good afternoon. I'm going to talk about the Census Bureau's work on measuring racial and ethnic diversity in the 2020 Census. As Rachel mentioned in her presentation, we follow standards on race and ethnicity set by the U.S. Office of Management and Budget. These standards guide how the federal government collects and present data on these topics.

The Census Bureau diversity calculations I'll present today require that we use mutually exclusive or nonoverlapping racial and ethnic categories. To do this we cross tabulated the race and Hispanic origin data so the Hispanic or Latino population of any race is it's own category.

Then each of the race alone non-Hispanic groups are their own individual categories. And finally the multiracial non-Hispanic groups as a distinct category. So we end up with eight groups, White alone, not Hispanic or Latino, Black or African-American alone, not Hispanic or Latino, American Indian and Alaska Natives alone, not Hispanic origin or Latino, Asian alone, not Hispanic or Asian or Latino, Native Hawaiian and other Pacific Islander

alone not Hispanic or Latino, Some Other Race alone not Hispanic or Latino, Two or More Races not Hispanic or Latino and finally Hispanic or Latino.

The concept of diversity refers to the representation and relative size of different racial and ethnic groups within a population. Diversity is maximized when all groups are represented in an area and have equal shares of population. We have been working on several different approaches to measuring racial and ethnic diversity. These include the Diversity Index, prevalence rankings, and the diffusion score and prevalence mapping. In my presentation I'm going to go through each of these in more detail.

The Diversity Index is a statistic that social scientists regularly use to quantify the racial and ethnic diversity of a population. People generally like this measure because it's relatively easy to interpret. The Diversity Index score shows the probability that two people chosen at random will be from different racial and ethnic groups.

The Diversity Index is bounded between zero and one with a zero value indicating that everyone in the population has the same racial and ethnic characteristics while a value close to one indicates that everyone in the population has different characteristics. We convert the probabilities to percentages to get the percent chance that two people chosen at random will be from different racial and ethnic groups. The Diversity Index can be used at different levels of geography and for the upcoming release we'll be showing the Diversity Index for the nation, states and counties.

To illustrate how the Diversity Index works we compare three hypothetical populations and their distributions. In the first example the population is made up of two large and even groups. The Diversity Index score for this population

indicates there's a 50% chance that two people are chosen at random will be from different race and ethnic groups.

Second hypothetical example shows a population with four equally sized groups. And here the Diversity Index score is 75%. The chance that two people come from different races and ethnic groups is increased even though the size of each group is smaller than in the first example.

The final example shows a hypothetical population with four equally sized groups and a Diversity Index score of 70%. Comparing the second and third examples we see that the relative size of the racial and ethnic groups affects the Diversity Index score by decreasing the probability when some groups are larger than others.

This graph shows the Diversity Index score for the United States using actual data from the 2010 Census. Here we see how the metric can vary based on the distribution in the population by race and ethnicity. In 2010 there was a 54.9% chance that two people chosen at random from the US population would be from different race or ethnic groups.

In 2010, the Diversity Index varied greatly by state which we did not show here but among all states the Diversity Index ranged from a low of 10.8% in Maine to a high of 7.5% and Hawaii. We'll be providing a Diversity Index score for states and counties in 2010 and 2020 during the upcoming release of the 2020 Census redistricting data.

Another method we'll use to measure racial and ethnic diversity is prevalence rankings. With this method we focus on the most common groups in an area. We look at patterns and percentages of the population that falls into the largest racial ethnic group, second largest group and third largest group.

The prevalence ranking approach uses tables or graphs to show the percentages of the largest groups. For 2010, looking at the 2010 prevalence rankings we find that the White alone not Hispanic population was the largest racial or ethnic group in the state at 63.7%. The Hispanic population was the second largest at 6.3%. Black or African American alone not Hispanic population was the third largest at 12.2%.

We also show the diffusion score along with the prevalence rankings. The diffusion score measures the percentage of the population that is not in the first, second or third largest racial and ethnic groups because it's combined. This calculation tells us how diverse and diffused the population is relative to the largest groups. The higher the score the less concentrated the population is in the three largest racial and ethnic groups.

For example the diffusion score for the United States was 7.7% in 2010 as 7.7 % of the population was not in one of the three largest racial or ethnic groups. When we look across the country we see a lot of variation in diffusion score by state in 2010.

The diffusion score was highest in Hawaii at 20.1% indicating that 1/5 of the population was in a racial or ethnic group that was not one of the three largest groups for the state. The next highest diffusion scores for Alaska were 15.1% followed by Oklahoma with 14.3%. The lowest diffusion score was in West Virginia at 2.1%.

The final conceptual approach that we use to illustrate the racial and ethnic diversity of the population is to map the most prevalent racial or ethnic groups for all counties in the United States. These prevalence maps show the geographic distribution of the largest or second largest racial or ethnic group.

And this is similar to the prevalence ranking approach that I spoke about earlier.

This map shows the most prevalent racial or ethnic group for each county in 2010. For most counties White alone non-Hispanic was the most prevalent group, and that's the orange color. However we see some regional variation.

In the South Black or African-American alone, non-Hispanic was the largest group in some counties. In the West and Southwest, Hispanic or Latino was the largest group for a subset of counties. And then Alaska and parts of the Midwest and Southwest, American Indian, Alaska Native alone non-Hispanic was the most prevalent group.

This map shows the second most prevalent race and ethnicity group for all counties in United States. The number of racial or ethnic groups represented the map increases. For example, the Asian alone not Hispanic population and the multiracial not Hispanic population are now represented in some counties on the map as the second most prevalent group.

As with the first map we also see regional patterns in the racial and ethnic distribution of the population. In the past the Census Bureau has sometimes used the concept majority and minority for measuring diversity. This approach has several conceptual and practical challenges that limit its ability to illustrate the complex racial and ethnic diversity of the US population.

For example, how some people classified multiracial individuals such as white and Hispanic, White and Black and White and Asian as part of the majority population, others classify them as part of the minority population. The dual identities of these groups highlight the social, political and economic complexities of race and ethnicity in 21st century U.S. society.

Also the inclusion of certain groups as part of the majority or minority has become more complex and contested in recent decades especially as many people may not identify with certain population groups even if that is how they are classified and tabulated per federal standards. The majority minority approach is ambiguous and it is further complicated by complex demographic and social realities.

To overcome these limitations we chose to highlight some of these alternative racial and ethnic diversity measures to illustrate the complex racial and ethnic composition on the 2020 Census results. And we plan to explore other diversity measures as part of our future research with the 2020 Census data.

We're excited for the upcoming release of the redistricting data because this will be the first results of the 2020 Census by race and Hispanic origin and our first opportunity to share with the public measures on racial and ethnic diversity in the 2020 Census.

Yesterday we released the Random Samplings article about our plans to measure diversity in the 2020 Census products. Around the time we release the redistricting data we will also release an American Counts story on racial and ethnic diversity in the United States using data from the 2020 Census and 2010 Census.

This story highlights national level data for all 50 states and the District of Columbia. Data for states, counties and Puerto Rico will be available in interactive data visualization. We are very excited about the data, about the diversity data visualization and encourage you to explore it once it's available.

In conclusion the analysis we release this month will provide a complimentary perspective to the statistics in the 2020 Census redistricting file and will help the public to understand the racial and ethnic makeup of the US population and the diverse identities that people share. With that I'd like to turn it back over to Michael Cook, Chief of the Public Information Office.

Michael Cook: Thank you Eric. We will now open up the phone for credentialed media to ask questions. Operator, can you provide the instructions for asking a question?

Coordinator: Thank you. If you would like to ask the question please press Star 1. Your name is required to introduce your question. If you need to withdraw your question press Star 2. Once again Star 1 to ask a question.

Michael Cook: While we await questions, I'd like to highlight some of the products we will have available for the upcoming redistricting data release. We will host a news conference that will provide analysis on population change, race and ethnicity along with our findings on diversity.

The release will also include one national news release and associated products to include interactive data visualization and America Counts stories that provide analysis. Members of the media who have covered us in the past may remember that last decade we produced state by state news releases.

We will not be providing these on release date, instead there will be a series of America Counts stories featuring state level analysis that we'll provide - that will post rather after the initial release of their digital data and through September.

For this release we will have data tools that you can use such as the 2020 Census data map and data visualizations that you can use to write your stories. Later this month we're going to release four interactive data visualizations.

We will have one data visualization on race and ethnicity in the United States, then a second data visualization will focus on racial and ethnic diversity in the United States, and yet another data visualization is about the adult and under the age of 18 populations in the 2020 Census.

The final data visualization, which isn't shown on this graphic, focuses on housing and population change. The data visualizations are quite extensive. They include maps, charts, ranking tables and detailed tables. These resources let data users explore the demographic and housing characteristics found in the redistricting data.

For example the data will be presented at the national, state and county levels in such a way that you can easily move between geographies or quickly zoom in on a specific county. The data visualizations include data from the 2010 Census and the 2020 Census so you can look at trends over time.

We are very excited about these products. We think they will be available - be valuable rather resources for state officials, researchers, media and the public at large. We hope to have these products available either on, or as close as possible, to release date.

And as James mentioned earlier, in addition to publishing the redistricting data on our FTP site we will also have the statistics available through our QuickFacts application and through the demographic data map application. All of these tools will assist reporters with writing stories about demographics

and local level population changes across the country. Operator, do we have our first caller?

Coordinator: Yes. Our first caller is from Hansi Wang with NPR. Your line is open.

Hansi Wang: Thanks very much. Hansi Wang with NPR here. I have a question about the coding for the census, 2020 Census race and Hispanic origin responses. Do the 2020 Census coding rules take into account the area of the questionnaire where write in response to the race question is provided. You know, for example, if there's a write in response of Thai provided in the write in field for the white category how would that be coded hypothetically?

Michael Cook: Thanks for that question on coding Hansi. I'm going to pass that over to our Population Division SMEs to address that.

Rachel Marks: Hi, thanks for that question. So for the 2020 Census we made a lot of improvements to our processing and coding procedures. And the goal is to respect respondents full racial identities while still following the '97 OMB standards.

So if a respondent selects a checkbox and write in a response, both of those responses are coded respecting the respondents full identity. So in that particular example a respondent who would select a White checkbox and provides a write in response of Thai would be coded as both White and Asian representing the Thai response in order to respect all of the responses that they provided.

So really regardless of where response is reported we will code it still following the OMB standards. And we code any additional responses that are

provided with those write in responses including checkboxes and also other write in responses. Thanks for the question.

Hansi Wang: Thank you.

Michael Cook: Yes, thanks for that Hansi. And for our reporters who are dialed in and not on WebEx that was Rachel Marks, our pop division. And her contact information or titles in our press release and press kit if you needed to write down or to jot it down.

I wanted to also announce or to make mention that our slides are available in the press kit. And we'll also have a recording and transcript up soon. I know that it looks like we are maxed out in our WebEx application. This is very popular Webinar today. Thanks for that. Operator, do we have our next call?

Coordinator: The next caller is Mike Schneider with Associated Press. Your line is open.

Michael Cook: Hi Mike.

Mike Schneider: Hi. Thanks for having us. Two questions, the first one is quick and easy. I don't think I missed it but did you guys say an official date? Is it going to be before August 16 when the release is?

And then my second question is for Mr. Whitehorne. I wanted to ask you about item non-response, you know, the range of questions that ranged in the 2020 Census from 10% to 20%, you know, across states and across questions and, you know, that's typically less than 5%.

I wanted to ask you what are your theories for the high item non-response?
And then also does it raise questions about the quality of the data especially
when it will be used for forming districts to comply with Voting Rights Act?

Michael Cook: Thanks for that line of questioning, Mike. Before I turn it over to James to address your questions about item non-response I would just want to remind media members who subscribe to our email releases and information that immediately following this Webinar I ask you to check your inboxes because we plan on making two very important announcement which will give you some insight into when that next statement will be made.

The question you had on item non response - can I just confirm I have the first part. James might not be the one who needs to address that. Can you repeat by chance, Mike Schneider, your last part of your question?

Mike Schneider: Yes, I mean I was just wondering first of all with the high item non-response what are the theories on why it was so high? And then what are - what if your concerns about the quality of the data because it was so high?

Michael Cook: Well as you know -- and I'll jump in right here instead of kicking this over to James -- item non-response is part of an ongoing litigation that we have. And it's something that is actually being litigated now. (Please note that this statement is incorrect and was clarified by Michael Cook further down in this document on page 38.)

So if you could Mike reach out to pio@census.gov, the Public Information Office pio@census.gov and we'll address those lines of questions to the best of our ability given the current situation when it comes to trial litigation. Hope that helps. Operator, do we have a next caller?

Coordinator: The next caller is Ashley Tarver, Labor Market Information. Your line is open.

Ashley Tarver: Hi. I am not the media. I will - is that okay if I still ask my question?

Michael Cook: Well actually we are holding this Webinar to give media the opportunity to ask their questions. What I'm going to ask you to do if you could please call 303-763-3030, 303-763-3030 and we'll make sure and get your questions answered. Thanks for calling and thanks for your interest in this topic.
Operator, can we have our next caller?

Coordinator: Next caller is from Anthony DeBarros The Wall Street Journal. Your line is open.

Michael Cook: Hi Anthony.

Anthony DeBarros: Hi. Good afternoon. Thanks for having this call, really appreciate it. Two questions, I want to confirm, I believe I saw it in the early slide, but just to be sure that the data release files are going to follow exactly the format from the prototype data release. That's my first question.

And question two is, are you releasing or have you released files that would allow us to compare 2020 tracted geography data to 2010 tracted geography data?

Michael Cook: Thanks for those line of questions on the prototype files. And I'm going to - actually this time I will turn it over to the James Whitehorne who will addressed those questions. Thanks James.

James Whitehorne: Yes. So we have produced a prototype data set and it's in the exact same format as what you can expect for the 2020 Census data. So if you're working to develop a system that can ingest that data and working to develop queries to be able to extract from that data they will work with both the prototype data and they'll work with the 2020 Census data. The formatting is the same simple text files with pipe delimitations for each field that's within those files.

Michael Cook: Thanks for the questions. Do you have any follow-ups?

James Whitehorne: The second part of question I don't know for a fact that we've created tract to tract relationship files that would be something we'd have to check with our geography division. They have a centralized location where they post geographic information like that, relationship files.

We have a block to block relationship file that is produced to support the redistricting efforts. And so there are similar products like that but I can't say specifically whether or not that product has been developed and published.

Michael Cook: Right.

Anthony DeBarros: Thank you.

Michael Cook: Thank you, James. Operator, do we have our next caller?

Coordinator: The next caller is Geoff Hing with the Arizona Republic. Your line is open.

Geoff Hing: Hi. Thank you so much for holding this Webinar. My question is that in 2020 there are a number of factors whether it's natural disasters or the COVID-19 pandemic and just different attitudes about participation in different communities. How should we think about the way in which different

communities were counted and their participation and how that affects the accuracy of numbers?

And what measures of that accuracy and quality are available now and which ones will be available in the future? And finally, what sort of recourse do (unintelligible) have if they believe that the - enumeration in 2020 did not accurately count their community.

Michael Cook: Thanks for that question about data quality. I know that before I allow some of our estimators to chew on this, I'll mention and Geoff, I'll encourage you to reach out to PIO if you can't find these breadcrumbs that I'm about to give you.

We have released a couple of data quality measures. We have some blogs that are out that are available on our random sampling site. And we have others that will be forthcoming in the coming weeks that will take another bite, if you will, and give some insight. And we will explain how we feel and what our position is on data quality when it comes to the 2020 Census.

But I'll turn it over to the SMEs to see if they have anything else they'd like to add. But I encourage you Jeff, to reach out to PIO if you can't find what you're looking for in those random sampling blogs. And we'll definitely make sure that you're plugged in and subscribing to all of our notifications so that when we do release our impending next round of data quality assessment measures, you'll be able to get that.

And I'm looking now at our our SMEs. I don't think I have anybody adding to that. No, I think we're good. Thanks for that Geoff. Operator, do we have our next caller?

Coordinator: The next question. I'm sorry, the next caller is Alison Dirr from Milwaukee Journal Sentinel. Your line is open.

Alison Dirr: Hi, thanks for taking my call. I guess I'm wondering, I cover city and local or city and county government. And so I'm wondering regarding the Disclosure Avoidance System, are we going to be able to see demographic changes down to the neighborhood level in a city of about 600,000, or, you know, I guess, would we risk inaccuracies that at that level? Thank you.

Michael Cook: Thanks for that question about disclosure avoidance. And I'm going to turn that over. I'm looking now to see if I can get some insight from Matt on that. Matt?

Matthew Spence: Hi. Thanks for the question. Yes, there - we've done a lot to tune additional accuracy for block groups and above for total population and race by ethnicity. In this case it's, you know, Hispanic origin by race.

So we are confident that there is a great deal of accuracy for example, tract level data. We have published statistics that show things like the mean absolute error which is essentially a measure of like, you know, how off are you on average?

So, you know, for example if it's a tract level, you know, data for, say, the white alone population say a mean absolute error of five, that means, you know, on average that tract is about plus or minus five of the true value. And so, you know, that's an example of course.

And so, you know, we've got these metrics available on our Web site and we've got lots of supporting documentation describing the improvements that

we've made to the accuracy at these lower levels of geography and for those queries in particular total population and race and race by Hispanic origin.

And so, you know, and then we will be releasing the demonstration data so that people can create their own tabulations to really, you know, zoom in on a particular area to see well just how accurate, you know, are these data. So that is that is, you know, forthcoming along with the redistricting data.

But yes, we in general, you know, feel that at - looking at tract level and even block group level should be, you know, overall very accurate. Individual blocks, though, may still be fuzzy. And that's just that's just our word of caution.

Michael Cook: Thanks for that Matt. And before - and it looks like I need to toss it over to James for a little bit more..

James Whitehorne: Okay.

Michael Cook: ..expand on that as well.

James Whitehorne: Yes, I just wanted to point out also that we have done some accuracy studies based on the redistricting use case that involved looking at meaning specific accuracy targets that were predefined and used to help tune the system. And in one of those we've redone that paper using the current production level settings so that we could publish those accuracy. And that paper was published today. So I'm sure that PIO can get that link out to folks through the press kit or elsewhere.

But the targets they're looking at is block groups of at least 450 to 499 people and townships and places of 200 to 249 are all meeting that accuracy target.

So it's really that low population levels that we're getting to very accurate data once you assemble those blocks.

Michael Cook: Thanks for that James. And operator as our next callers are making themselves ready by dialing Star 1, I want to just step back momentarily before we go to our next caller. Looking back at my notes I believe Mike Schneider had asked a question about item non-response. And I may have misspoke and mentioned that item non-response is current topic being litigated. That's not true. Is not necessarily being litigated but it is a topic that - of much discussion in social media and the like. (This is the statement that clarifies Michael Cook's comment about the subject of item non-response litigation.)

And if you could, Schneider take a look, and if you take a look at Ron's blog, our acting director, Dr. Ron Jarmin, he has a blog where he made it clear that we're seeing higher rates, but that the rates that are being discussed publicly are - this far are not necessarily correct. And they're far - they are from long suspended stages of processing, suppressed, superseded stages, rather and that we'll see the real rate soon after the release, that the impending release that we have coming out for the redistricting data that we're preparing you guys for today on this call.

Okay. Operator do we have our next caller?

Coordinator: The next caller is Michael Macagnone with CQ Roll Call. Your line is open.

Michael Cook: Hi Michael.

Michael Macagnone: Hello. Hi Michael. How are you?

Michael Cook: (Unintelligible).

Michael Macagnone: Thank you for holding this. My question is about the race and ethnicity coding changes. I was wondering if the bureau has done any research or has a sense of the, you know, the quantifiable effect of those coding changes? So for - when we're looking at the P.L. 94 data any changes in diversity we see would be attributable to the coding change versus a change of the underlying population.

Michael Cook: Thanks for that. I want to talk that over, another coding question to our POP subject matter experts to shed some light in - or on or say what they can say about that - those line the questions.

Rachel Marks: Yes thanks for that question. So, you know, of course, we can't really share any results from the 2020 Census just yet. But, you know, what we can say is that we do expect that some of these statistics in the upcoming release will not only reflect demographic changes, but also to some of these changes that we've been talking about today due to some of the improvements to the questions on Hispanic origin, race, the way that we've process the data and also the way that we've coded the data. So we look forward to talking more about that once the data release later this month.

Michael Cook: Thanks for that Rachel. Operator do we have our next caller?

Coordinator: The next caller is Colleen O-Dea with New Jersey Spotlight News. Your line is open.

Michael Cook: Hi Colleen.

Colleen O-Dea: Hi. Thanks for taking my call. I'm going to have another question, two, one on the coding again. The simpler question I think, is just on comparing the data we're getting in redistricting with ACS data for instance the 2019 ACS, you know, race and ethnicity numbers.

But my question on the coding, on some of the slides, it seemed unclear to me, you had for instance, someone who had Hispanic plus white plus Chinese. And I don't think they - you didn't show that they were coded as Asian. Similarly, there was a Cuban, Thai, Filipino, and that was another race, but not Hispanic. I'm just - can you clarify again that coding for me? Thank you.

Michael Cook: Thanks for that. I'll have our SMEs talk to you a little bit about our OMB standards of classification.

Rachel Marks: Yes so the first, you know, to talk about the ACS is, you know, we're still evaluating that 2020 ACS. We did make these changes to 2020 ACS and Decennial at the same time so we're still evaluating that.

And to your second question about the coding changes, you know, we are - we made the same changes across the board for both Hispanic origin and race for 2020. But when we look back to 2010, what we did is, you know, we prioritized Hispanic responses in the Hispanic origin question and raised responses in the race questions because we could only code two responses.

So when we get the tabulation -- and this applies to both 2010 and 2020, we in Hispanic origin following the OMB standards, you can only be Hispanic or Latino or not Hispanic or Latino. So any response that's not Hispanic, such as an Asian response or a Native Hawaiian or Pacific Islander response, those are classified as not Hispanic or Latino in the Hispanic origin question.

When we get to the race question, you know, we classify all of the race responses within those OMB categories that are provided in one of the earlier slides I showed. Hispanic responses though, you know, the OMB standards do not state, you know, which race group Hispanics can be part of. They can be part of any race. So we classify them as part of the Some Other Race group. So I think that's what you may have been seeing there. Thank you.

Colleen O-Dea: Thank you.

Michael Cook: Thanks for that Rachel. Operator, do we have our next caller?

Coordinator: That is from Sharon Jayson, Freelance. Your line is open.

Sharon Jayson: Hi there. Thanks again for this Webinar. I'm interested in finding out whether in this housing information that'll be released if the comparisons with 2010 and 2020 will include moving and who may have moved from where or if that data comes later in a different release?

Michael Cook: Thanks for that question. I'll turn that over to our POP SMEs. And I may be able to address it if we - if I'm looking to see who wants to grab that one.

James Whitehorne: But I can say something. The housing data that's included in the redistricting data file is specifically just count of the housing units that were in the 2020 Census. And it's got them classified as whether they were occupied or occupied or vacant. So it's really that snapshot of the April 1 2020 Reference Day Collection for the 2020 Census.

Sharon Jayson: So they often Census does have information about, you know, moving. Is that going to be an analysis at a later date or are you not doing that anymore? Or is

there a way that there would be to compare where people might have moved into certain census blocks or tracts or whatever?

Michael Cook: One would ask caller Sharon if you could reach out to PIO - pio@census.gov and we will give you some insights into not only to make sure that you can ascertain what's coming out in this release, but what's coming out in forthcoming releases.

I know that over time, if you look historically at the data products that are decennial based, there are a number of releases or data products that address that. We are - it is a different product that you're referencing. And I believe that it's from one of our other program areas. But we can unpack that with you when you give us a call at [pio@census.gov](tel:3017633030) or email at pio@census.gov. And again that number if you want to call is 301-763-3030.

Sharon Jayson: Terrific. Thank you very much.

Michael Cook: You're welcome. Operator do we have our next caller?

Coordinator: That is from Chad Day with The Wall Street Journal. Your line is open.

Michael Cook: Hi Chad.

Chad Day: Hi - or hey, thanks for doing this. I just have two questions. One is just to confirm that in the summary file or in the redistricting data that the geoheader information for the congressional districts will be for the 116th Congress and not for the 117th which would also make it where we probably want to be cautious with using the new North Carolina numbers there.

And my second question is, does the bureau have any documentation on how to make sure in the block crosswalk files that we're getting the correct portion assigned to the correct portion when we're dealing with mini to mini to relationships, parts of a 2010 block that go into a part of a 2020 block?
Thanks.

Michael Cook: Thanks for those questions Chad. I'm going to toss it first to James who might be able to take a - two bites - take both of those questions on the headers for the Congress as well as the block crosswalk. James?

James Whitehorne: Yes, so the geographic header file, will actually have the codes for the 113th Congress and the 116th Congress. In the Census Bureau cyclical collection of congressional and state legislative district boundaries we do not collect the one that is aligned directly with the Decennial Census. So that would have been the 117th.

So you are correct that any state that had a change to their congressional districts like North Carolina after the 116th Congress you should be careful and probably look for a different source and perhaps to rebuild that from the block level data.

I'm sorry, the second part of the question?

Chad Day: Yes on the second part of the question was about the crosswalks, the relationship files for block to block, In the mini to mini situations where we have parts of 2010 that are then now connected to parts of 2020, is there any Census Bureau documentation out there that tells us kind of what - how much of that previous block should go into how much of that other block? I understand how to marry them based on the crosswalks. I'm just trying to get the correct percentages and proportions to get - to match up.

James Whitehorne: Yes, we don't provide any specific guidance on that simply because people have different uses for that - doing that. You know, sometimes people will do just sort of an area calculation breakdown and then reassign the population based on those area calculations.

Other times they may use a weighted system where they also look at whether something has land or water components and skew towards the land. They may use other data sources that they have as far as like housing units that they have in their own data sets to help do that weighting. So because there's so many possible different ways to do that and because there are so many different uses for trying to do that, we don't really provide any specific guidance on that.

Michael Cook: Thanks for that James. Operator do we have our next caller?

Coordinator: That is from Andy Beveridge with New York Times. Your line is open.

Michael Cook: Hi Andy. Andy?

Andy Beveridge: Can you hear me now? Sorry.

Michael Cook: I can, I can, loud and clear.

Andy Beveridge: Yes, it went to mute by accident. Anyway, this is a very good Webinar. Unfortunately I was only able to see part of it slides, because you really did sell out and I got dropped by accident.

My question really has to do with an issue that's already come up which is, is there any possible way, for example, say you have changed between two tracts. We're looking at tracts at The New York Times.

If - how can we parse the proportion of difference between ten and 20 with respect to the noise versus the - with the noise versus a actual change or the reported change? You know, with the ACS or something like that, you get an air band or you get a confidence interval. Has there been any thought given to having those sorts of things actually directly with the data because as you know, every ACS statistic or estimate has an estimate of how how precise it is but unfortunately it seems with this we do not. So that's my question.

Michael Cook: Thanks for that Andy. I'm going to turn that over to Matt to address. Matt?

Matthew Spence: Hi. Thanks for your question. We have published for all of our demonstration data products over the last two years detailed summary metrics which gives data users a glimpse into the accuracy measures of bias and also counts of outliers for various geographies and the various characteristics. So if you're looking at for example total population, you can go to, you know, the most recent detailed summary metrics which is the same labeled as production settings. These are the same.

This is the same software with the same parameters the same settings that were used for the actual 2020 production run. But these were used on 2010 data so that we could publish these detailed summary metrics.

And you can go to, you know, for example, the total population at the tract level table and you can see what the mean absolute error is. And you can know that the average tract gained or lost -- I don't have the number off the top of my head but say gained or lost two people, you know, on average or for

the, you know, tract level, white alone population. They gained or lost five people for example.

And you can kind of use that information to sort of help you understand whether you know, how noisy the data are relative to the published 2010 data as a result of the application of disclosure avoidance. So these, you know, these published...

Andy Beveridge: No, I understand that.

Matthew Spence: Oh.

Andy Beveridge: My question was really like, if you're looking - if you're going to have a box or something describing the noise piece of the data - I mean with survey methods like ACS, et cetera, there's very well developed standards on how to report that kind of information. I've looked on all of your - I mean I've actually had experience with every single one of your demo products. You know, I wrote my comments -- the whole thing. So I understand what you've done and what you've produced so far.

My concern is if you're in a tract that say is 200 people, you know, what percent of that, you know, like say the tract is 200 people and they have 250 people or let's say 2000 people and they have 2500 people in 2020. The issue is see the thing about an actual error, once you have a subtraction you have a bigger problem with this. And that's effectively what you're going to be doing with comparing 2010 to 2020.

And so I was wondering if you at least would have some sort of - even if you had a - just a general paragraph on the guidance of the of the noise that was infused I think it'd be very helpful because, you know, we can't put every

single table and particularly like in a map, you know, there are 84,000 tracts. You can have 84,000 statements.

So I was wondering if there's any - been any thought given to having something or sort of a global relatively simple description of this so people, you know, like the press, what they could, you know, what they would be able to say and what kind of confidence you'd give to the readers that what you're showing is change versus noise?

Matthew Spence: Yes that's a - that is a great - a great question and a great, you know, comment. We actually, you know, I remember reading your feedback so we really appreciate your feedback and other data users' feedback, you know, as we developed a Disclosure Avoidance System over the last several years.

We, you know, will be working as we are continuing to publish additional data later in the future that are not just redistricting data. We're working on some sort of metric that will sort of give users a sense of the variability of results.

But, you know, users today or I'm sorry, not, today. When the demonstration data products go out users could download the demonstration data, you know, what's been called the PPMs as you know. Users can download that data, tabulate tract level information so something like the total population or otherwise, compare it to the actual published 2010 numbers and see what's the - what's the tract with the biggest miss, you know, what's the tract with the, you know, what's, you know, what are the 90% ranges say if you're thinking of like a margin of error? That is something that could be calculated from the demonstration data.

That said, we decided not to, you know, sort of publish anything quite like that for the redistricting data. But our guidance is what was on the director's blog which is that, you know, we have confidence in the accuracy of data for block level and above in particular for tract level and above. You know, we're very confident that the data are accurate at the tract level.

Andy Beveridge: Well thank you.

Man: Michael you're muted. I'm sorry about that.

Matthew Spence: As long as I've got a second, I just will remind folks that, you know, there's a paper that came out today that looked at very small areas that James Whitehorse mentioned earlier that looked at very small areas, aggregation of blocks, both sort of in traditional census geographies sort of like block groups and then also sort of townships and places that may be somewhat sort of off of the traditional geographic spine.

In looking at those accuracy for very small areas that are much smaller than the typical tract and we find very good accuracy. So I'd encourage people to go take a look at that paper that's just been published as well.

Michael Cook: Thanks for that Matt. Operator do we have our next caller?

Coordinator: Yes. That is from Suzanne Gamboa with NBC News. Your line is open.

Suzanne Gamboa: Hi. I'm sorry, I think you answered this question already, but I had an audio problem, so I just want to clarify again.

The upcoming releases you're referring to it or the most immediate one, can you just clarify exactly how detailed the race information will be? Like Is it

just Hispanic, non-Hispanic? How far into the race categories you go? And then what - in the following release, what will come after that? Thank you.

Michael Cook: Thanks for that. I'm going to turn that over to Rachel Marks and see if we can answer some of those questions for you. Thanks Suzanne.

Rachel Marks: Thanks for your question. So in the redistricting data file, we have data for the Hispanic and not Hispanic population and we also have data for our major race categories along with the additional multiracial categories totaling 63 race categories. So the PL really focuses on those major OMB groups. And later on we're still working on developing our project plans for additional race and ethnicity data to be released from the Decennial Census.

Michael Cook: Thanks for that. Operator, do we have our next caller?

Coordinator: That is from Jennifer Agiesta with CNN. Your line is open.

Jennifer Agiesta: Thank you. I'm wondering if the Diversity Index scores that you talked about will be available for counties and states for 2010 or if those scores are only going to be made available for 2020 data?

Michael Cook: Thanks for that question. I'm sorry, I'm going to have to ask you to repeat the last part of it please? You went out on my end.

Jennifer Agiesta: Sure. I'm wondering if the Diversity Index scores will be available for the 2010 data for states and for counties or if you're only making those available based on the 2020 numbers?

Michael Cook: Thanks for that. I'll turn that over to Eric.

Eric Jensen: Yes the Diversity Index was going to be - we're going to be releasing that through our data visualization for 2010 and 2020 for the nation, states and counties. So it will be available for both decades.

Michael Cook: Thanks for that Eric. Operator, do we have our next caller?

Coordinator: That is from Alexander Ebert with Bloomberg News. Your line is open.

Michael Cook: Hi Alexander.

Alexander Ebert: Hey good afternoon. Thank you so much for holding this. It's incredibly helpful. I'm interested in how this more colorful and helpful data around race and ethnicity might prove possible challenges for people looking to draw specific Voting Rights Act block groups and create majority, minority or impactful minority voting map.

How will this more colorful data that's being produced, this more full enriched data impact to that ability? And are we going to see issues based on the conversations you've had with attorneys in this space?

Michael Cook: Let me turn it over to our SMEs...

Alexander Ebert: Sure.

Michael Cook: ...to chew on. But I know that anything, just so you're aware, anything that's dealing with litigation we're not going to touch it or address it on this call. But if you want to ask an official question from the media about ongoing litigation we - I please encourage you to reach out to PIO at census.gov.

Alexander Ebert: I would.

James Whitehorne: And this is James. And speaking in generalities around the use of this data for compliance with the Voting Rights Act, the work that we've done and the accuracy targets that were chosen was to ensure that the numbers that are produced from the Decennial Census, the 2020 Decennial Census are of a high enough level of accuracy when they're aggregated into those districts that if you're making decisions about a district and you've designed a district that is meant to remedy a situation that's been found under Section 2 of the Voting Rights Act or some other situation that you can have confidence that those data are accurate.

So we don't expect this data to affect the ability of the public, the government, special interest to use this data any differently than if they had used it in the - from the 2010 census. It should - showed he accuracy should be there that you need to be able to do that work.

Alexander Ebert: Yes and sorry, if I could just clarify. This isn't involving litigation with the bureau. This is in general. So one of the examples was an individual that identified as Asian or as Chinese, as white and as Hispanic. That's wonderful. You know, the diversity in the country is amazing.

But let's say you're California and you're looking to draw districts that empowers Asian and Latino voters and you have got someone who identifies also as white. You know, how does that potentially impact the drawing of districts? You have a more rich and full documentation of how people are identifying. It's not as cut and dry as it might have been in years past.

James Whitehorne: Well I know that before I toss this over to Rachel to see if she has anything to say, when you look at how we're measuring race and ethnicity, that necessarily hasn't changed. It's the way that we're looking at diversity. So

let me see, I'm looking to Rachel to see if there's anything else that she wants to add or Eric.

Rachel Marks: Yes, the only thing I'll add is that, you know, we are just collecting more data than we have before on race and ethnicity, you know, by changing our coding rules, by changing our questionnaire, improving them and, you know, trying to collect more accurate data than we have in the past. But I think that's all I'll say about that.

Michael Cook: Yes. And then lastly for this reporter, before we go to our last call just to remind you that if you didn't get the answers that you want, give us a call at pio@census.gov or we can walk you through and talk a little bit more about how we present the data to the public to use. And then the data takes that information and makes informed decisions and does things like drawing district boundaries and the like.

Operator, can we have our last call?

Coordinator: That is from (Renee Riggon) with CNN. Your line is open.

Michael Cook: Hi (Renee).

(Renee Riggon): Hi. Thank you for taking my call. My question is about the Diversity Index, the prevalence rankings and diffusion scores. So now that the coding system for race and ethnicity has been refined, will the 2020 data be directly comparable to 2020 data or to 2010 data, or is it advised that you don't compare one to the other because the underlying parameters have changed?

Michael Cook: Thanks for that question. I'm going to toss that first to Eric for response.

Eric Jensen: Yes so we have been a little cautious about comparing 2010 with 2020 because of the improvements that they've made. However with the diversity measures that I talked to today, we do a cross tabulation between Hispanic origin and race. And when we do that, it really improves that comparability between decades.

And so we feel like with how we are doing that cross tabulation, how we're comparing, we are picking up a lot of national demographic change in addition to the coding improvements of course, but more so than if you just look at race and ethnicity separately.

Michael Cook: And also before we close last caller, (Renee), wanted to let you know that when we do release our information, a big part of the analyses that I mentioned earlier off at the top will be part of those America Counts stories where we'll be able to shed some light and have a greater discussion on that topic. So thanks for that question and thanks for your call and your interest, continued interest.

(Renee Riggon): Great, thank you.

Michael Cook: So in closing again, I'd like to remind media members who subscribe to our email releases and information to check your in boxes after this Webinar for two important announcements from us. And if you've been paying close attention, you'll actually know, you know what I'm talking about.

You may contact the Public Information Office at pio@census.gov. The numbers are there on your screen. And I'd like to thank you for joining us today on behalf of today's speakers and everyone at the United States Census Bureau who worked on providing this Webinar today. I'm Michael Cook and

this concludes today's informational Webinar in advance of the 2020 Census redistricting data release. Thanks everybody.

Coordinator: Thank you for participating in today's call and you may disconnect at this time.

END