Slide 1 – Intro

- Thank you and good afternoon, everyone. Thank you so much for joining us to hear about the updates to our methodology for the Vintage 2021 estimates series. My name is Christine Hartley, and I am the Assistant Division Chief for Estimates and Projections in the Population Division at the U.S. Census Bureau.

- Today, I’ll begin by sharing some background on the Population Estimates program and briefly reviewing our processes and methods at a high level. Then I’ll describe how the challenges that we encountered for developing our Vintage 2021 estimates inspired us to manage those challenges in special ways, including the development of a blended population estimates base. And finally, I’ll share our plans for future research and vintages.

- We hope that this information will be interesting to you, as well as helpful for planning for your own research which may rely on our estimates.

Slide 2 – The Population Estimates Program

- The Population Estimates Program is responsible for developing annual estimates products to satisfy a mandate to Congress as per Title 13, Section 181. These are the official estimates of population and housing units for the Census Bureau and they are used to distribute federal funds, develop weights for demographic surveys, facilitate planning at the state and local levels, and to generally inform the public about changes to the nation’s population. Over the past couple years, for example, the estimates were used to distribute over 200 billion dollars to states and cities for the 2020 CARES Act and the 2021 American Rescue Plan.

- The latest estimates that we’ve published are for July 1, 2020, although we have estimates on our website dating back to 1900. And of course, presently, we are preparing for the release of our new July 1, 2021 estimates.

Slide 3 – PEP Estimates Geographies

- Every year, we release estimates for over 80,000 areas in the United States and Puerto Rico. You can see here that our geographic detail covers the nation, regions and divisions, states and Puerto Rico, metropolitan and micropolitan statistical areas, counties and municipios, and finally incorporated places and minor civil divisions, also referred to as cities and towns.

Slide 4 – Annual Estimates Detail

- The detail that we publish varies by level of geography, which is summarized here in this table: we release estimates down to the county level by demographic characteristics—age, sex, race, and Hispanic origin, and estimates of total population down to cities and towns. We also publish estimates of the group quarters population by age, sex, race, and Hispanic origin for the nation,
total group quarters population for states and counties, and housing unit estimates for the nation, states, and counties.

Slide 5 – Release Schedule

- The estimates are published on a rolling basis beginning in December of each year with the total population, components of change, and voting-age population for the nation, states, and Puerto Rico.
- In March, we release county, metro and micro, and municipio total population, and components of change for counties and Metropolitan and Micropolitan Statistical Areas.
- National population by single year of age and sex comes out in April.
- Total population for cities and towns, as well as housing units for the nation, states, and counties, are published in May.
- And finally, our estimates by demographic characteristics for the nation, state, counties, Puerto Rico Commonwealth, and the municipios are published in June.

Slide 6 – Methods

- For population estimates at the county level and above, the Pop Estimates Program utilizes the cohort-component method to measure population changes since the last census through the use of a variety of administrative records on births, deaths, and migration. Sources include the National Center for Health Statistics, Internal Revenue Service, Social Security Administration, Defense Manpower Data Center, and survey data from the American Community Survey and Puerto Rico Community Survey. We also work closely with state demographers through our partnership group, the Federal-State Cooperative for Population Estimates, or FSCPE.
- At the subcounty geography level, the population estimates for cities and towns are created by distributing the county population estimates to each place within the county, based on the place’s ratio of household pop to the total housing units. We add that to an estimate of the population in group quarters to get the total resident population.
- And finally, our estimates of housing units are also produced with a component-based method that starts with the most recent decennial census and uses a variety of survey and administrative data to estimate change in the housing stock, specifically in terms of building permits, estimates of non-permitted construction, mobile home shipments, and estimates of housing. Data sources include the Building Permit Survey, Survey of Construction, Manufactured Homes Survey, American Community Survey, and the Federal Emergency Management Agency individual records on requests for disaster assistance.
Slide 7 – Annual Updates to the Estimates

- Our general methodological approach each year is to revise our entire time series of estimates beginning at the date of the last census and up to the vintage year, which represents the last year of estimates available. So for example, our upcoming Vintage 2021 estimates series contains estimates for April 1, 2020 through July 1, 2021.

- From vintage to vintage, there are three types of updates that we may incorporate into the time series:
  - First, the base population is updated each year to reflect legal boundary changes or other geographic updates. It may also be updated due to Count Question Resolution changes.
  - Then, as more recent or complete data become available—such as births or deaths, or the components we use for our housing unit estimates—we use those to update the estimates of the components of change. We may also revise the methods of how we estimate the various components. And what you’ll see is that data updates are more likely to affect the estimates for the most recent years of the time series, while method revisions generally affect the time series cumulatively from the census date forward.
  - And finally, we may improve our method of tying all the components together and generating the estimates. This type of change also affects the series from the census date forward.

Slide 8 – Challenges for Developing the 2020 Estimates Base

- A key detail regarding our methodology is that, as you’ve already heard me say a couple times, traditionally, we use the results of the latest decennial census to serve as the base for our estimates, and then annually we measure change since the last Census to update our entire time series. However, early on as we began to witness changes to the 2020 Census operational schedule resulting from the COVID-19 pandemic, it became apparent that downstream delays in the availability of the census data were likely. For months, it wasn’t clear what data would be available and when—impacting our ability to plan for the Vintage 2021 estimates base, as well as other major products such as our 2010-2020 Intercensal Estimates which rely on the Census data to serve as the end point of that time series.

- And because of how the pandemic impacted Census field operations, there were also many questions about quality—so if we had the 2020 Census data available to us, how much time would we need to evaluate the results and determine if they were suitable for our specific use? And given the Census’s Bureau’s new disclosure avoidance strategy, would it be possible to undertake an evaluation of this magnitude and to finalize a method to make the Vintage 2021 estimates differentially private?

- All we knew for sure was that we needed some kind of solution which did not fully rely on 2020 Census data.

Slide 9 – Blending the 2020 Population Estimates Base (County+)
• And what we came up with is to create what we’re calling a “blended base.” This method begins with the April 1, 2020 value from our already-published Vintage 2020 estimates series, which is based on the 2010 Census. Then, to create a plausible base, it controls this April 1, 2020 value to our national 2020 Demographic Analysis estimates distribution by single year of age and sex, and county total population from the 2020 Census PL 94-171 redistricting data. With the time we had available to us, this represented the most detail that we could confidently incorporate into the estimates base. For those of you who may not be familiar with the Demographic Analysis estimates, or DA as we call them, these are national-level estimates for April 1, 2020 produced using current and historical vital records, data on international migration, and Medicare records. So the primary difference between the vintage estimates and the DA estimates is that we don’t use Census data as a base for DA—the DA estimates are entirely independent of the Census, and as such, they’re used as one of the official measures of coverage for the Census. We have a very high level of confidence in these estimates, which is why they were deemed a reliable data source to pull into the blended base.

• Back to the base: the outcome is that for Vintage 2021, the April 1, 2020 estimates of total population for counties and higher levels of geography will match the published results of the 2020 Census. Demographic detail in the estimates will come from a combination of the Vintage 2020 estimates series and the 2020 Demographic Analysis estimates.

• So a possible limitation of this approach is that where we see gaps in the detail that we’re getting from the controls, like for race or Hispanic origin, we’re filling them in with the detail from our Vintage 2020 estimates—although this isn’t necessarily a negative thing. But it does mean that any potential issues from Vintage 2020 get carried into Vintage 2021 unless we have a separate data source serving as the control.

• However, this approach has numerous benefits as well: particularly, that whereas it doesn’t rely on the 2020 Census, it provides some level of consistency with it. Then, it is an adaptive framework, meaning that we can continue to use this approach as long as necessary, but we can also improve upon it for future series of estimates by pulling in additional data from the 2020 Census or other sources to serve as controls. This is contingent upon the results of research over the coming year, which I’ll be getting into in just a bit.

Slide 10 – Privacy-Protected 2020 Population Estimates Base (Subcounty)

• So what I described on the previous slide neatly summarizes our approach for our county-level estimates and higher levels of geography. But if you recall from earlier in the presentation, our subcounty estimates are handled with a separate methodology, and subsequently, a different approach was necessary for our subcounty population estimates base.

• The subcounty methodology relies on a place’s ratio of household pop to the total housing units. As it is, the 2020 Census housing unit counts are invariant, which means they don’t get infused with differentially private noise. So to retain our current methodology, we only needed a solution for the household and group quarters populations. What we determined is that the most straightforward and effective way to produce these subcounty estimates is to use the
results of the 2020 Census which are available to us on the internal, unprotected Census Edited File or CEF, and then infuse noise into the household and group quarters data, so if we do this in the estimates base, the rest of our methodology can remain unchanged, and then the subcounty estimates get controlled to the county level blended base, and by starting from the CEF instead of the public census data, we maintain the ability to apply annual geographic updates to the data, something that is critical for producing accurate estimates of these low levels of geography, and which is not possible with the information available on the public census data files, such as the PL 94-171 redistricting data.

- To that end, we collaborated with the bureau’s Disclosure Avoidance staff in the Center for Enterprise Dissemination to develop an appropriate differential privacy mechanism, and our proposal received approval from the Data Stewardship Executive Policy Committee earlier this month.

**Slide 11 – The Impact of COVID-19**

- For the Population Estimates Program, the impact of the COVID-19 pandemic was not limited to our April 1, 2020 estimates base. After all, our Vintage 2021 estimates series extends from April 1, 2020 through July 1, 2021—a period that falls entirely within the pandemic. As such, there were also effects on the components of change that we use to update the base population, so each had to be evaluated in order to determine whether adjustments would be needed to more accurately capture the impact of the pandemic.

- Ultimately, for births and deaths, we were able to pull in the most current provisional data made available by the National Center for Health Statistics. These data account for the increased mortality and changes to natality that took place over the course of the pandemic.

- For international migration, we typically rely on the American Community Survey to serve as an input, but of course the ACS had its own suite of issues and we could not use the 2020 file for this purpose. So instead, we used current data from the Department of Justice, Department of Homeland Security, State Department, and the Institute of International Education to calculate an adjustment we could apply to the 2019 ACS estimates.

- And finally, for domestic migration, historical trends in the IRS tax return data that we use in our method were examined to determine if the extensions to filing deadlines affected data quality in any significant way; and we also examined benchmark data sources, such as the US Postal Service’s National Change of Address File. And we were able to determine that no adjustment to our source data or method was needed.

- All of the changes that we applied will be described in more detail in the Vintage 2021 method statement and release notes that will be published in December with our first data release of the vintage.
Slide 12 – Next Steps for Vintage 2021

• Regarding next steps, we are in the process of producing and reviewing the estimates for counties and higher levels of geography, followed closely by production of our subcounty population and housing unit estimates.

• And in the meantime, we’ll be continuing our stakeholder outreach on the finalized methodology, such as doing webinar like this one, and also some other engagements.

Slide 13 – Beyond Vintage 2021

• And then finally, once we have everything settled for Vintage 2021, we shift gears and begin planning for Vintage 2022. At this time, there are still many unknowns for the Pop Estimates Program, particularly regarding the timeline for receiving 2020 Census data—and of course, this influences our options for the Vintage 2022 population base, as well as our ability to produce the 2010 to 2020 Intercensal Estimates, which is generally a popular product.

• We also have questions regarding whether the 2020 Census data will be sufficient for our use in the Vintage 2022 estimates base. Our intention is to use coverage measures such as Demographic Analysis and the results from our robust internal Estimates Evaluation project (or E2) to make these determinations. E2 makes a series of comparisons between our Vintage 2020 estimates and the results of the 2020 Census and then evaluates those differences. Typically, these differences are regarded as error in the estimates and are used to inform research and method improvements over the coming decade. But this time around, given all the complexities of the 2020 Census, the evaluation of differences is expected to be more nuanced. And will help us to determine if there are any specific geographies or characteristics that could introduce challenges if used for the population estimates base.

• Or, possibly, whether there is additional demographic detail that we feel we can incorporate into a blended base for Vintage 2022.

• As I mentioned previously, a major benefit of the blended base is that it’s adaptive, so we can use it as long as we need to, or even improve upon it from vintage to vintage as we explore the possibility of switching over to using 2020 Census data as the base. But the findings from E2 are really going to be the driving force for determining what direction we take for future vintages. And that research is ramping up as we speak, so we expect to have findings we can share over the course of 2022.

• And of course, our intent is to keep all of our customers and data users informed as decisions are made so that you all may plan accordingly.

Slide 14 – Q&A

• And to that end, we’re about to open up to questions, which you can enter into the Q&A tool, but if you need any more information at any time, you can reach out the Coordination, Dissemination, and Outreach Branch, which is located in the Estimates & Projections Area of the
Population Division, and the contact information for that branch is available on this slide, along with the link to the Population and Housing Unit Estimates website where you can access the schedule for our release, data releases, and also our method statement and other technical documentation.

- Thank you very much for joining us today. We are very happy to take any questions you may have, and I have a number of members of my team here to assist as needed.

Post-Presentation Q&A

- Thanks, Christine. Before we go into questions, I will say that there was a question regarding slides, so the slide deck along with the webinar recording or in transcript will be available within about two weeks, and that’ll be posted on census.gov. Our first question comes from a data user who studies African immigration, and they wanted to hear more about the international migration component in the estimates.

- Great. I'll turn that right over to Jason Schachter, who's in our Net International Migration Branch.

- Hi, everyone. Could you repeat the question again, for me, please? Just more information about how we did the adjustment?

- Absolutely, so this data user specifically they're interested in African immigration, but they just wanted to hear about that international migration component in the estimates.

- For the COVID adjustment or just in general on the estimate?

- In general.

- All right, in general, yeah, we have a -- we use the American Community Survey primarily as our major data source. We have a number of different components. One is the foreign-born immigration component. Another piece is foreign born emigration, so an estimate of those leaving the country. We also have a net native component for US-born, and we also have a Puerto Rico piece as well as a military, net military movement piece, so those are the five general pieces that go into our international migration estimate. I guess that would be my summary for that.
And there are actually lots of details on how we generate our net international migration component in the method statement that's available on our site, so that's probably the best source to go to for more details, and then if there are follow up questions, you can email the address on the slide with your specific questions.

Thanks, Jason and Christine. Our next question is asking for more detail in how it was decided that 2020 census alone cannot underpin the estimates. Specifically, some users develop cohort component projections relying on five-year age, sex, decennial enumeration.

So the primary factor that drove us to developing the blended base was the availability of the data. So you know, we have our own estimates, project -- estimates production schedule, which we follow in order to meet our legislatively mandated deadline to release estimates by the end of the year, and the input data that we would need from the census was just simply not available to us internally in time to be able to evaluate it and use it for the production of the estimate base, so there's nothing at this point that we can say about comparisons, you know, in terms of specific detail, like age detail. We didn't necessarily evaluate the data and determine that it wasn't good enough quality. It was just really that we didn't have the time to do that evaluation, so that's what we're going to be focusing on over this next year with our Estimates Evaluation project, and then that will be where we actually are looking at the detail and the differences between the estimates and the census and deciding like, if it's up to the specific standards at that fine level of detail that we need in order to use it in our estimates production.

Thanks, Christine. Next question. Are there changes in the structural outputs of ACS PUMS products that we need to reprogram around? Or do the outputs adhere to prior parameters despite methodological changes?

I am not specifically familiar with ACS PUMS products. Is anybody who is on the Q&A panel familiar with them?

Well, if you're talking about the ACS controls to the population estimates that we make in the Population Division, so if you're -- like to any extent that they use our estimates as weights, then yeah, this would affect their PUMS data. However, I can't think of a specific reason why this particular methodology would affect PUMS data any differently than just a regular census base would. Does that make sense?
• What we're able to provide to the ACS for the survey weights should be unchanged, even though it's developed from a different type of estimate base, but if they are making any changes to the structural output of the ACS PUMS product, that would be a better question for somebody from the ACS program. If you aren't unable to find a good email address for that, you could, again email pop.cdob@census.gov and we could provide to you an email address to reach out to the ACS program.

• Thanks, Ben. Thanks, Christine. Next question is if the methods we spoke about are going to be included through 2030.

• That is actually the big question, so at this point, what we know is that we're going to be using it for this Vintage 2021 production. It seems extremely likely that we'll be doing some kind of version of a blended base for Vintage 2022, but we are going to be relying on our research, the Estimates Evaluation research that we do over the coming year to decide whether or not there's going to be a point at which we can, you know, use the 2020 Census data as the base or if it's always going to be, you know, up until 2030 some kind of blend but at this time we don't have that information but we are actively investigating.

• I would add to that that we also think, given what we know so far, that including the Demographic Analysis estimates to adjust age and sex is an improvement to the quality of the data, so at least I, personally, I'm hoping we're able to maintain that, but it's really going to depend, like Christine said, on what the Estimates Evaluation project shows about Decennial.

• Next question asks --

• Can you place the link to the methodology for additional information?

• Thanks, Amel. The next question asks if we're able to share the blended base data, and when will the July 2021 data be released?

• So right now our release will not be before December 21st. So that'll be the point in time where the July 1st, 2021 time series, as well as the updated methodology statement and release notes which cover the development of the blended base.
Great. Next question asks at some point in the future, will you compare data from the blended base method to what you would have gotten if you had used the 2020 Census data as the base for different demographic groups like minorities or children?

That's a good question. I mean, it kind of assumes that we'll have a method to create the base from the 2020 Census data to compare to. If we did go down that path, you know, we were evaluating the census data. You know, it met all of our needs, and we decided to use it for the base. I definitely expect that we would be making comparisons between that and the latest iteration of the blended base, probably as part of our review process, you know, to examine the differences by, you know, the different demographic characteristics, but whether we ever end up with that 2020 Census base, you know, that will definitely be contingent on the Estimates Evaluation research. But yeah, if that happens, we'll be looking at those comparisons.

Next question is about DA or Demographic Analysis. How do the limitations of the Hispanic data in the DA given that Hispanic data are available for only the population under 30 affect the quality of the Hispanic data and of the blended base?

So actually, at this time, we're not using the Hispanic data from Demographic Analysis. We're only using the data by age and sex, and the reason we made the decision was because of the age data. The same can actually be said for the race data. We only had the black, non-black race groups, and so we didn't have enough time to investigate how we could account for those limitations for the race and Hispanic origin data, so that's why we incorporated this vintage, but that's something that we'll be looking into.

Actually, to that end, I'll add that so the Hispanic origin data that you see in the blended base are coming from the Vintage 2020 estimates.

Next question asks about the new data sources. Are those new data sources more or less likely to pick up people in households who are often left off the Decennial Census forms such as doubled up households?

I'm not so sure.

Are you talking about specifically for the blended base? Oh, I forgot to introduce myself last time. My name is Ben Bolender. I'm a Senior Technical Advisor for Population Division and co-lead of this team for the last few years. It shouldn't necessarily -- like there shouldn't necessarily
be a direct interaction between picking up combined households and that kind of thing, because we're not using, at least for the county level, we're not using housing unit data in order to make the blended base, so to the extent that the Demographic Analysis control adjusts age and sex, or that the population estimates for the administrative records was picking up on a different number of people in a housing unit, or in housing units are in an area they were never before, it's possible, but again, it's not -- because this wasn't a direct survey method with Demographic Analysis and census totals and the Vintage 2020 Population Estimates, it's really impossible for me to say anything about that level of detail.

- Thanks, Ben. Next question is asking if the new methodological updates, will they impact the undercount of black and Hispanic populations or other minority groups?

- So the new methodology does impact the distribution of the population, but right now, it's primarily by the age and sex detail that we're using from Demographic Analysis. Ben or Luke, do you want to elaborate on that at all?

- Yeah, so and Luke can follow up if he has more. The tests that we did with using black and non-black from Demographic Analysis did show a bit of difference between what we would have gotten from Decennial using PL data, but there's a few questions about that. Number one, and I think Christine addressed this, the Decennial Census, the PL data does not have modified race, so some other race is an option, so it's really hard to make a direct comparison until there's a method created to convert that data into the modified race categories, and the other reason is it did seem to adjust the black-alone, not black-alone group, a sizable chunk, and we were concerned we didn't have the time to do the research about what that would do to different areas at the subnational level. So, at the national level, it made a reasonably -- like a noticeable difference, but we didn't have any way to specifically target how that adjustment was applied down through all of the lower level of geography and other characteristics, so just because the national level went up or down didn't mean that, you know, a county that had a particularly high under or overcount would have been corrected, if that makes sense.

- Thanks, Ben. We have a question asking about if the five-year estimates would still be released in December. I have a feeling that's regarding the American Community Survey.

- It sounds like it and I don't recall what the plan is for the five-year estimates. Again, is anybody on the panel more in tune with the release schedule for ACS?

- I believe it's postponed until March next year.
• Okay. Thanks, Jason.

• Thanks, Jason. Regarding the doubled-up households, there was some clarification that was sent in. They're wondering if the new data sources, because they are not household based, might pick up people that don't show up in household-based methods, such as teen mothers and young babies.

• Because the DA data is for the resident population, it is possible. It doesn't differentiate between household population and group quarters population, so how the process works is we actually create resident population, at least for this year. We create a blended base for the resident population all the way down to the full levels of detail that we need, and then we proportioned it out to households and group quarters based on the proportion that was in the Vintage 2020 estimates. So next year, one of the things that we're looking into discussing is whether or not we should create each of the universe estimates separately, so should we create the household and group quarters population separately and then add them to get resident? Or should we continue on the way the way that we're doing now? So yes, to the extent that a resident population measure might pick up on a different amount of household or group quarters population, it would be really hard to give a definitive answer one way or the other if it's actually doing that. I hope that makes sense and clarifies.

• Thanks, Ben. I'm not seeing any further questions at this time. I'll give folks a few more minutes if they're in the middle of typing. Again, you can submit your question through the Q&A textbox if you have a question. We have a question regarding differential privacy that just got submitted. How does the differential privacy instituted on the data that the pop estimates will use differ from the process used on the census data that are released to the public?

• So, at this time, we are only implementing differential privacy on our subcounty population estimates, and I think that the biggest difference between the process used for the census data and ours is that ours is for a much more limited set of data, so it was kind of tailored to just subcounty household and subcounty group quarters population, so we, you know, applied for a specific privacy loss budget or epsilon level to apply to the subcounty population estimates. Whereas for the data that were released in the redistricting data file, they had to kind of fine tune that system for a much larger amount of data. So ultimately, we’re using, you know, a similar system, because we collaborated with the same, are producing differentially private estimates for the census, but it's just on a smaller scale, and I don't know if Ben or Luke want to elaborate on that at all.
• I can add something to that. So for --

• Go for it, Amel.

• -- subcounty, we're using the same algorithm, the top-down algorithm, to introduce noise to the subcounty population. The only difference is the number of cells that are added to. So, the more -- the larger the number of cells, the more noise is introduced, and having a specific algorithm for the subcounty, it limit the number -- it limits the noise that is introduced to that.

• Yup.

• So, if we're adding noise to 100 cells, it's not the same thing if we're adding noise to 40,000 cells, so the more cells we have, the more noise is introduced, and we're trying to minimize that noise. That's why we requested a specific algorithm tailored for the subcounty population.

• Yeah, we're using the Python code that was provided to us by the Research and Methodology Division, so we're using mechanically the same thing that they're using for decennial to infuse noise into the data. Like Amel said, we just have a lot fewer cells, so it's a lot easier to optimize for it. We're also not as constrained with the subcounty estimates about all of the aggregations having to add up like, because we're only doing total household and GQ population, they only have to add up to total household and GQ population. They don't have to deal with age, sex, race, Hispanic origin, and, you know, the thousands of categories that that brings.

• The more characteristics ---.

• So it mechanicalizes the same Python code.

• Yeah, so the more characteristics you add, the more noise you're adding to your estimates or your counts, and by for subcounty, we don't have any characteristics, so just population total for cities and towns, and that helps minimize the noise introduced and getting more accurate estimates to our local governments.

• To continue with differential privacy, there's a question about if the DP methodology privacy loss budget, will that be shared in a similar manner as the redistricting product?
I don't know what manner the redistricting privacy loss budget was shared in but our privacy loss budget for the subcounty estimates was 3.0. I don't think that that's sensitive information, so.

It's 3.0 per year and the current plan is to be able to rerun that on the base every year for the decade, so a grand total privacy loss budget for the subcounty population estimates base of 30 using three per year, if everything continues to go according to plan.

Thanks, Ben. Next question is asking about when the next round of intercensal estimates will be released?

So, the release schedule for the intercensal estimates is contingent upon when internally, we have access to the Demographic and Housing Characteristics File, which will be the next 2020 Census data file. I think that the timeline for that is still kind of being worked out, but our best guess is that we will see that file sometime late 2022, so if you take production into account, I'm going to estimate that possibly early 2023 is when the intercensal will be completed and released.

Thanks, Christine.

Sure.

Next question, will total population in the July 1, 2021 estimates be consistent with the 2020 Census results for all geographies?

So definitely for counties and higher levels of geography, we're using that control from the PL 94 redistricting data file, so they'll be consistent for April 1st, 2020. In the subcounty population estimates, because we are using the methodology that infuses noise, they will be slightly different for the estimates base.

And the estimates geography is different than what decennial uses, so they would be different anyway.
• Right.

• The main difference is that census geography uses Census Designated Places, and I can't remember what the name of the other one is right now, but the census equivalent of minor civil divisions and place parts and the estimates program doesn't use that. We only, with a couple of exceptions, we only create estimates for places that have a functioning local government unit, and then the rest usually gets put into balance of county, so it wouldn't have matched exactly anyway, because the geography is different, but Christine's right. With the perturbations that we're adding with the differential privacy mechanism, it is going to be a little bit different even for places that have the exact same geography.

• It's not one to one equivalent, so it won't match anyway.

• Thank you. Next question, if a blended base with differential privacy noise is used through 2030, could a subcounty place receive positive noise in one year and negative noise in the following year?

• Yes, because the way that we add the noise, and they're technically independent from each other, so we would run this algorithm on the base every year, it's fairly likely that it would go back and forth. So, like this year, maybe it'll be, you know, plus five, and next year, it'll be a minus two or whatever. By not adding differentially private noise to the estimate itself, you don't end up with compounding error. You end up -- because we're only changing the base, it would be -- that would equate to like a plus two or a minus five or whatever for the entire time series as opposed to having a situation where you add five in the first year of the time series, and then you're adding another 10, so now you're 15 higher, and then you're adding another five, so now you're 20 higher. Because of the cross-sectional thing, just at the base and we're not adding noise all the way throughout the process, it will go -- each individual area will go up and down slightly. At least that's what we expect, but it shouldn't compound and get out of control.

• And to add to that --

• I think it's an important point to make -- oh, go ahead, Amel.

• To add to what Ben mentioned, so at the subcounty level, we're starting with the census counts and applying the noise to it, which will essentially will be monitored every run, so with every run
you're adding a random noise, which will not -- which have limited bounds to it, so it will not exceed those bounds, which we are in control with those bounds. However, with every run, it will change.

- Right, I think it's important to emphasize that the amount of noise that we're infusing into the estimates is very, very minimal, so even if it does switch in the base from year to year, it shouldn't be a significant, you know, flip-flop, and as Amel said, we're going to, you know, be monitoring that from year to year.

- Next question, will there be any kind of demonstration product for related to the use of differential privacy in the subcounty and population estimates so users can assess the impact of differential privacy on the estimates?

- We don't have any specific plans to release a demonstration product. I don't know that we necessarily have the resources to support something like that, but that can be something that we take into consideration as we're kind of, you know, planning out our research for the next couple of years.

- Thanks, Christine. Next question about the release schedule, what all is being released in December?

- In December, we are putting out the total population for the nation and states and Puerto Rico, as well as the components of change, so that'll be the births, deaths, and migration, and also the 18 plus, the voting age population for the nation, states and Puerto Rico, and that'll be for April 1, 2020 through July 1, 2021, so you'll get the time series.

- Thanks, Christine.

- Sure.

- Got a user who would like a little bit more definition or discussion around what we mean by noise.
• The noise is the amount of error that we're adding to a specific cell, which is -- which means it's the lowest amount of error we're adding to the lowest level of primitive geography, which is our -- the -- it's an estimate. It's the estimates geography universe that we're dealing with at the subcounty level, or places and towns, and that amount of noise is very, very small, at that primitive level, and more trying to add less than or equal to a half a point to a specific, primitive geography, which is the lowest level, like I said, the elemental level of unit of geographic unit within the subcounty that make up a place or a town, so a collection of primitive areas will make up that specific place or town.

• I'm curious if the question is even more like baseline than that. So, what you do when you do differential privacy is you take a random sample from a distribution like a normal curve or a Laplace distribution, and then you just add that to the number that you have. So, if the population of the place was 500, you could take a random sample from, say, a Gaussian or a normal distribution, and that random sample could be two, so then the population in that place is now what, 102 or 502, or whatever it is that I said. Mathematically, it's not that hard. It's getting everything else. Like it -- all you're doing is taking a sample and adding that number to it, but it gets complicated when you're trying to make everything balanced and trying to keep the overall levels of noise down, because while it's very unlikely if you have, you know, an epsilon three that you're going to add 100,000 to, you know, a cell at five, it's technically possible, it's just super unlikely.

• I think something that's important to note is that the Census Bureau adopted differential privacy as its disclosure avoidance strategy. There's a lot of information about it that's available on census.gov. Our program is expected to adhere to that because it's an enterprise strategy, so we're following along but it is, by no means, you know, our expertise in terms of relaying the details of differential privacy, so I definitely encourage you if you're not very familiar with it, or the decision-making processes that Bureau went through, to check out the information that's available on the Census Bureau's website, so at least for that.

• Thanks, Christine, Amel, and Ben. I'm not seeing any other questions at this time.

• I do know that there were some comments in the Q&A and I just wanted to assure everyone that we, you know, will be taking those into consideration and discussing them even if you didn't explicitly have a question that, you know, the information that you contributed to us and the feedback is still important. Thank you for that.

• We have a few more minutes left. Again, if you have any questions, please use the Q&A textbox to submit your question.
• I do see a couple more questions about the December release, so just to clarify, the release for December will be for April 1, 2020 through July 1, 2021. Eventually, when we do create our 2010 to 2020 intercensal estimates, you'll be able to construct a time series from 2010 all the way forward to 2021, but the new data that are coming out are specifically for April 1, 2020, which will be the blended base through July 1st, so that will all be included in the files that are available on the website in December.

• The respond to the last question, the Vintage 2021 release will have a July 1, 2021 date on it.

• And regarding what will be available in March, that will be the county level population estimates, municipios, Metropolitan and Micropolitan Statistical Areas and it will also be the components of change for the counties, Metro, micros and municipios.

• Thanks, Christine and Amel. There are no further questions at this time.

• Okay. Everyone, thanks. That concludes today's conference. Thank you for participating. You may disconnect at this time.

• Thank you, everyone.

• Thanks.

• Thank you.