

2010 Census Coverage Measurement Pre-Release Webinar

**May 17, 2012
11 a.m. ET**

Coordinator: Welcome and thank you for standing by. At this time all participants are in listen only mode until the question and answer portion of the conference. If you would like to ask a question at that time please press star then 1 on your touchtone phone.

Today's conference is being recorded. If you have any objections you may disconnect.

Now I'd like to turn the call over to your host Ms. Stacy Vidal. Thank you, ma'am, you may begin.

Stacy Vidal: Thank you, good morning. I am Stacy Vidal of the Census Bureau's Public Information Office. I'd like to welcome everyone participating in today's Webinar about the Census Coverage Measurement Survey.

I'm joined today by Pat Cantwell from the Census Bureau's Statistical Studies Division. Pat will provide an overview of what the Census Coverage Measurement Program is what it measures, and what will be released next week.

This Webinar is designed to help you prepare for next week's release. On Tuesday, May 22 we will release the results of CCM which provide estimates of the coverage of the 2010 census. Essentially the coverage measurement results provide one way of analyzing the quality of the 2010 census count.

On Tuesday we'll host a news conference releasing the results and then we'll dive into a deeper discussion about the results, their implications for planning

for the next census among other topics in the afternoon with a panel presentation and some discussions.

I'll provide more logistical details at the conclusion of today's Webinar. In the meantime I'll turn it over to Pat for his presentation. After he concludes his presentation we'll open up the Webinar for questions and answers.

We'll first take questions from members of the media and then we'll open up for questions from others who may be on the line.

Pat?

Pat Cantwell: Thank you, Stacy.

Today I'll cover three topics. First I'll define what is coverage of the census? Then I'll look into net coverage error and components of census coverage error, what these things are and how we describe them. Next we'll look at results that we'll be releasing this coming Tuesday. And finally I'll say a few words about the approach we take in order to derive estimates and also the design of the Census Coverage Measurement Survey.

There are several ways to measure the quality of a census. We've talked about several of these before, for instance census process indicators. Examples of these are response rates to the mail out questionnaires that we provided two years ago, also rates of proxy response in the census.

The second way is through demographic analysis. The Census Bureau through our population division provides estimates of the population to demographic analysis. In this they use vital records such as birth records, deaths, estimates of immigration/emigration to estimate the population. One key feature about

demographic analysis, it produced results only at the national level and by several demographic characteristics.

A third way of measuring coverage is through a false numeration survey. And our house survey in 2010 was called the Census Coverage Measurement Survey. An advantage of a coverage survey is that we can produce estimates at the national level but also break it down by various groups whether by demographic characteristics, by geography, or other things.

And first I want to define net census coverage. We define it as the estimate of the population from the survey minus the census count. And if this number is positive we refer to it as a net undercount. If it's negative we refer to it as a net overcount.

Now because many of the numbers that we'll be dealing with are quite large we often talk about percent net coverage. For example, the census totals are up in the range of 300 million. For percent net undercount we take the net undercount estimate, estimate from the survey minus the census count and divide it by the estimate and multiply by 100.

So once again if this percent is positive we refer to it as a percent net undercount. And if it's negative, a percent net overcount.

This gives you a picture in this chart. For example in 1990, the Census Bureau measured a net undercount of 1.61%. So you see that number above the line of zero. In 2000 we measured a net overcount of 0.49%.

You'll also notice in this chart there are confidence intervals around the estimate itself. So in 1990 you see that estimate 1.61% and those little, little bars on each side of it. This is an estimate of the confidence interval. It shows you that there's error due to sampling and other features in our estimate.

Let me say a few words about our coverage in the 2000 census. After the original survey we conducted extensive analyses, other research, evaluations, in fact we did a nationwide computer matching. Based upon these we estimated a small net overcount that you saw in the prior page of about half a percent overcount.

We also estimated a large number of omissions and erroneous enumerations even though the net undercount itself is very small.

Well what lessons did we learn from 2000? First we realized we had to improve better determination of census day residents. And for this we've changed the questionnaire for our 2010 census coverage survey, also some of the procedures.

The second lesson we had to improve our techniques to identify and resolve duplicate enumerations. A duplicate enumeration is a case where we have one person on two or more census records.

I'll go into the purpose of the 2010 Census Coverage Measurement. First we want to provide estimates of net coverage. This is what we've done the past several decades. But what's new for the 2010 coverage program is we're also going to take this net coverage and break it down into components of census coverage. Now in a few slides I'll tell you exactly what I mean by these components.

We want to provide these estimates for persons in housing units and also for housing units themselves. Now for the persons this means we're not including, we're not measuring the coverage of people in group quarters. Group quarters are structures such as college dormitories and living in other cases.

A third purpose of the 2010 Census Coverage Measurement Program is to use the results to improve the 2020 census. And I'll discuss that a little in the future as we look at some of the results we'll provide for operations.

Finally all throughout the decade there's been no intent to adjust the census counts.

Okay let's talk about the components of census coverage. First we'll look at the census count itself. And this is the actual number of records in the census either as a whole or for sub-domains.

First we have the correct enumerations. Those enumerations that we determined through our survey were correct. Second erroneous enumerations, these are cases that erroneous due to several reasons. One important reason is due to duplication as I mentioned a few minutes ago we have one person but there are two or more records.

Other cases for erroneous enumeration other examples, people who might have died before census day or born after census day. Also somebody who's just visiting temporarily from another country but is put down on the census form. There are several other possibilities.

The third part of the census count are what we call census imputations. These are records in the census, actual records for which we have no name and all the characteristics are imputed or statistically inserted.

Now what's important about these is that for more than 80% of these cases we actually have a count during the census. For example we might have gone to a housing unit and not been able to reach anybody. A neighbor tells us there are

three people living there but we have no further characteristics. So we have a count of three but no name and all the characteristics had to be imputed.

This gives you a picture of what it might look like. And this is only an example so I've intentionally exaggerated the proportions here. We have at the top of this you'll see the census imputations. Sometimes they call them whole person census imputations. Then you have the erroneous numerations and the correct enumerations.

Now typically the correct enumerations will be most of this bar so I've just made it so you can see it a little better.

Now we also have an estimate of the population from the survey. We break this estimate into two parts. First we have the correct enumerations. And this is the same thing that you saw just two slides ago, the remainder of the omissions, what we call the omissions. This is made up really of two groups, people who were completely missed in the census. But there are also people for whom we can't verify that they were actually enumerated correctly.

For instance I mentioned those census imputations. Because we don't have a name for those census records or other characteristics, they've just been imputed statistically. We can't really verify that they were correctly enumerated. However in most cases we believe they were.

And this gives you the picture of the omissions and the correct enumerations that make up the estimate from the survey. And again I've exaggerated the dimensions and proportions here.

So together you can see the census count on one side, the estimate from the survey on the other side, and the picture could look like this. Now again this is

just an example, they aren't real numbers. And you can see the census count and the estimate from the survey. They're made up first or primarily of correct enumerations. And that's the blue that you see at the bottom of each bar.

Then you have a census count in the left hand side, the survey in the right hand side, I've given you an example where there might be a small undercount, so at the census count is slightly less than the estimate from the survey.

And what you can see is that you have the omissions in the estimate that we derived from the survey but this can be offset primarily, in the most part by these imputations and erroneous enumerations. And of course the pictures would be similar but a little different if you had an overcount in the census the left hand bar would be a little larger.

Now for housing units we have just a little different situation. You can see the components in the left hand and the right hand side census count, and an estimate of housing units from the survey. But on the census count you'll notice that there are only two components here. There's no third component and this is because for every housing unit we try to identify whether it's correct or erroneous through our sample in the survey.

I'm going to look now into some of the tabulations, some of the results that we're going to be providing next week. And there are three major groups. First we have groups that are defined by demographic or housing unit characteristics, those that are defined by geographic areas, and finally by census operations.

So the first of these three coverage estimates by demographic characteristics it will be generally at the U.S. level. And some of the characteristics that we use

are race, Hispanic origin, that's whether a person is Hispanic origin or Hispanic or Latino or not, owner or renter what we sometimes call tenure in the census, and also by age, sex groups.

We have several groups for children by different age categories. And for adults we break them up into the age categories and by male or female.

Coverage estimates by geography, we're going to release estimates for the U.S. as a whole and later we'll release estimate for Puerto Rico. This won't come this coming Tuesday for Puerto Rico.

For regions of the U.S., the major regions we'll have estimates for the East, the Midwest, the South, and the West, also for the 50 states and the District of Columbia. We'll also provide estimates for large counties in large places. A place is a census term for city or town.

So what do we mean by large? We have thresholds of the counties or places to define what is large. For estimates of net error, and this is what we talked about as net undercount, net overcount, or percentages. One hundred thousand or more in census population, so any county or city or town that has at least 100,000 in their census population, we'll provide an estimate of net error.

For components of census coverage we have a different threshold, it's 500,000. This is much larger because we use different estimation techniques for providing net error estimates for components of census coverage.

Now there are many operations which we'll break down and then we're into components. I'm just going to show you a few of these.

And the first example here, we'll provide components by type of enumeration area. For example the first one being mail out/mail back, that's the predominant type of enumeration area in the country.

What this means is that about 90% of the country we mail out a questionnaire and we hope that respondents will send back the questionnaire completed. And for those that don't we have a non-response follow up operation where we go out and knock on doors.

There are two other major types of enumeration area, a second where we leave the form and we hope the people will mail it back and a third which is very small part of the country, where we enumerate the people in person.

Another type of operation is what we call bi-lingual mailing areas. For instance some areas in the census received a bi-lingual mailing where they have the questionnaire in English and in Spanish. We'll provide estimates and break down the components for those in these bi-lingual mailing areas and those that were not.

Here are some other examples and they have a little difference in these examples. I'll mention the first one, we'll look at mail return by weeks of return. For those people who sent back a valid return we can look at the various weeks and look at their components.

For people who didn't send back a mail return they were put in the non-response follow up operation. And for those who are hit in the non-response field operation we'll provide estimates by month of interview.

A third case, a third operation is a coverage follow up. This was an operation in the census that we used to try to provide better coverage, sometime

recontacting people who had sent in a form. We'll break down results by reason that they were put in the coverage follow up operations.

Now for some operations such as the ones listed on this page we cannot estimate the net error or the omissions. But we can provide proponents of census coverage. So we'll break down the census count into the three parts you saw before.

Now we'll also provide estimates for housing units. For the housing units we'll have groupings such as some of the ones I've shown the prior several pages. On the other hand we'll also provide results by some characteristics that are much more important for housing units for example by whether the housing unit is occupied or vacant. And for the occupied housing units we break that into whether the housing unit is owned or rented.

For housing units type of structure is also very important. For example is the structure a single unit or is it a small multi-unit that is two to nine units, or a large multi-unit, then or more units. Many apartment buildings or condos would fall in the categories of small units or large multi-units. A trailer is another category. We'll break down results by type of structure.

Next I'll provide a little information on the statistical approach that we use to estimate the size of the population. And from this we can derive an estimate of the coverage of the census.

This is an approach that's been used in other population studies for many, many decades. So this is nothing that's new even though we've derived new techniques to apply it. Dual system estimation is what's sometimes used to label this type of approach and estimation.

For some of you some of these slides may be a little more technical, more technical than you like to see. For others this might just be basic statistical material so I want to give you a little overview of what we do here in order to apply this approach.

First we take the census. Then we take a post-enumeration survey, a survey that's conducted after the census. And this survey's conducted completely independent of the census. Then we interview the people in this post-enumeration survey.

We try to match the people from the survey to the census records. This helps us try to determine whether or not they were captured in the census. Then we follow up as necessary, following up cases that might still be unresolved after our matching or might need further review. Finally we try to estimate how many census records are correct and how many the people in the survey were enumerated correctly in the census.

In order to do this we take two samples. First in this post-enumeration survey we take a sample of block clusters. Now these samples, within these block clusters we list, we actually go into the block clusters and list the housing units independently of the census. We call this first sample the P sample.

From the census we take a sample of records or enumerations in these same block clusters. We call this the E sample, E for enumeration. Now what you can see here is a picture trying to show you what it might look like. Each person can fall into one of these four main categories you see in the shaded cells.

Cell person, was the person enumerated in the census, yes or no. In the same way was that same person enumerated in the post-enumeration survey which

we call the P sample, yes or no, while there are four possibilities and you can see these in the middle of the table.

Now I've given you the same table but I've shaded a little differently. This time I've shaded it so that what you can see are those cells for which we have an estimate from the survey and from the census.

So you can see down the first column of symbols or numbers which is shaded because we can estimate the population that's in the P sample of the post-enumeration sample. Across the first row again you can see it's shaded. Those people we can estimate from the sample of records from the census.

If you look at that original two by two table in the middle you can see one cell that's not shaded. We have N22 in there. This represents people who did not fall, were not captured in the census also didn't fall in the, weren't captured in the post-enumeration survey, they were missed in the survey.

Well because they were missed in both we can't estimate them directly and therefore we can't estimate that bottom right cell which is capital N, the total, into population. And that's what we want to try to get to.

Well we have these two samples which allow us to try to estimate that fourth cell and therefore the capital N. We have the E sample and this is used to measure the portion of correct enumerations in the census.

So we start with each census record in the sample and we try to determine if it's correct or erroneous. Is it a duplicate? Is this the correct or incorrect part of the duplicate? Is it fictitious? Is it correct?

A second sample is the P sample. We try to use this to measure the rate of coverage or capture in the census. So for each person in this independent or P sample, we interview them in the field. We try to match them to the census afterwards to determine if the census captured them correctly or didn't capture them.

Now here's the picture again. And if you look at the first column this is the one shaded in red, you can see the first cell, N_{11} in the table, this represents the people that were enumerated in the P sample and then they were matched to a correct enumeration in the census.

Down the bottom you see N_{+1} , that's a total of people enumerated in the P sample. Well the ratio of N_{11} to N_{+1} , as you can see the arrow defines this ratio this gives us an estimate of the rate of capture in the census, rate of coverage in the census to the post-enumeration sample.

Now on the right hand side you see in light blue those people who were enumerated in the census. Now after we remove the erroneous enumerations we have an estimate of the number of people who are correctly enumerated in the census.

What we can do is take that ratio that you see in the first column that's in red and inflate the N_{+1} by the inverse of that ratio. Now under the assumption of independence which is how we conduct our operations, this inflated number will be an estimate for the capital N in the lower right hand corner.

Now I've oversimplified many of this, much of this, but this is a general idea. I'll say a few words now about the sample design of the independent or P sample. It's a sample of block clusters. That's a term that we use but a block

cluster just means one block or perhaps two or more contiguous blocks put together.

We put these together so that these block clusters average about 30 units. Actually they can vary quite widely from below 30 to above, much above 30. As we did back ten years ago with the 2000 coverage survey we exclude the group quarters in remote Alaska from the coverage.

Now we select a sample from subgroups otherwise we stratify using several characteristics. For instance we stratify by using the states and the District of Columbia. We also use other characteristics such as is this area part of an American Indian reservation? And what is the size of the block cluster, what is the tenure otherwise? Is this block cluster made up primarily of owners or renters?

Here are some numbers that describe the size of the independent or P sample. Now the first number is exact block clusters. This is for the United States not including our Puerto Rico sample. In the United States we had 6148 clusters. Now approximately we had about 170,000 housing units in the P sample and this gave us about 390,000 persons.

Finally there are many operations that are conducted during this post-enumeration survey. I'm just going to talk about a few of these here. For instance I mentioned that we took a sample of block clusters and we went out and listed all the housing units in the block clusters. This was done about a half a year before census day back in late 2009.

There are many operations that had to be done after, following this listing. But then following the census we interviewed all the people in these housing units

in the post-enumeration survey. This was done during August through October of 2010, several months after census day.

After we collect the information the person interviews we conduct person matching. And this consists of extensive automated computer matching as well as clerical matching to help resolve many of these cases. And combined with follow up this was done shortly after the person interviewing ended and all the way through April of 2011. And there were many other operations including final housing unit operations to follow that.

So we completed the field work last year and we've been processing data and preparing estimates and measures of error in the interim. So this Tuesday we look forward to their release.

Now I'll hand you back to Stacy.

Stacy Vidal: Thanks, Pat.

Before we go to take some questions in just a moment I'd like to give you a little more logistical information about Tuesday's release. As I mentioned we will host a news conference from 10:00 to 11:00 in the morning. We'll host the news conference physically at George Washington University. Certainly anyone is welcome to watch online on our UStream channel. You'll see the link there on our slide.

Following that news conference in the afternoon we'll host a technical briefing where we'll have a panel of Census Bureau experts as well as a couple of discussions for outside experts, many others in the room as well to dive a little deeper into those results.

Media are certainly welcome to attend that. Anyone as well can again watch both sessions, the news conference and the afternoon technical briefing on our UStream channel. If you do plan to attend the event either morning or afternoon at George Washington we'd ask that you RSVP to pio@census.gov.

And you'll find more information about the exact address and some dial in information if you want to participate in that news conference over the phone. And you can find that information on our Web site from the media advisory that we sent out on Monday.

And I think just also stay tuned we'll, Dr. Groves posted a blog last week that describes some of the things that Pat referenced today that might be another good reference resource for you. He will probably also be posting another blog the next day or so that will kind of describe the whole process of how we conduct this post-enumeration survey in the next day or two so stay tuned for that as well for a little additional background.

And with that I think we'll go ahead and go to some questions and answers. If you'd like to follow up with anything after the presentation here's some contact information for you. If you're with the media please feel free to contact us at 301-763-3030. If you're not with the media please feel free to contact our customer services center at 1-800-923-8282.

And I have put the wrong press kit link here for you but we'll get that updated and we will post a press kit with some additional background information for you within the next day or so.

So we'll go ahead and take some questions. Again we'll first take questions from members of the media and we do ask that just to allow everyone to have

a chance to ask their question that you ask just one and then one follow up if needed.

Operator?

Coordinator: Thank you, Ms. Vidal. If you would like to ask a question please press star then 1 on your touchtone phone. You will be prompted to record your first and last name. You will be announced into conference at your turn.

One moment, please while parties take time to queue up.

Once again if you would like to ask a question please press star then 1 on your touchtone phone. You will be prompted to record your first and last name.

Stacy Vidal: While we wait for folks to queue I'll just let you know that today we did release population estimates for the nation, states, and counties by demographic characteristics. Specifically we included estimates by race, Hispanic origin, age, and sex. And these were the first demographic population estimates that we've released since the 2010 census.

So they'll, they examined population change for these groups nationally as well as within states and counties. So please feel free to check that out on our Web site this morning.

Coordinator: Ms. Vidal, at this time we have no one in queue. We'll just remind them that media or non-media can now queue up with star 1.

We do have some questions coming up. We will take one from (D'Vera Cohn), Pew Research Center.

Stacy Vidal: Okay thank you.

D’Vera Cohn: Good morning and thank you for doing this. I just have a logistical question. When will the actual report be posted? Will it be before the news conference, after the news conference? I missed the very beginning so you may have stated this.

Stacy Vidal: Sure no I, you're right. I didn't actually state the exact timing. We plan to release the news release which summarizes the high level findings at the national level right at, around 10:00 am once the news conference gets underway.

We do have a series of memos that we're also planning to post that day I think at least a couple of them will be posted around 10:00. The others that get more technical we may decide to hold those until the afternoon's technical briefing or we may just go ahead and post them all but we'll try to make that clear on our press kit page once that's live.

D’Vera Cohn: Thank you.

Coordinator: Chris Williamson, City of Oxnard, California, your line is open.

Chris Williamson: Thank you, good morning. I'm a planner in a city with a high excuse me, a high population that might have been undercounted, Hispanics and farm workers and so on.

Thinking back to 2000 just wondering if there's any plans to generate adjusted counts at any level. You know, maybe not block level certainly but counties or track some kind of adjustments down the road, thanks.

Pat Cantwell: Yes now I can't speak for population estimates or anything like that going between censuses but we have no plans to adjust the numbers based upon this survey.

Chris Williamson: Okay thank you.

Pat Cantwell: You're welcome.

Coordinator: Next we have Karl Gregory of Karl D. Gregory and Associates.

Karl Gregory: Yes I noted from the presentation that there was an overcount of 1.61% in 1990 and an undercount of 0.49% in 2000. Could somebody review what the highlights or the major factors that accounted for this shift in performance between 1990 and 2000? And what changes that made in approaches so that the count could be more accurate in 2010?

Pat Cantwell: Okay I will say first of all that in 1990 we had an undercount of 1.61%...

Karl Gregory: Oh I...

Pat Cantwell: No that's easily understood because when the number is positive we call that an undercount and when it's negative we call it an overcount. So if you look at the slide that's up there now you can see that yes, there was an undercount in 1990 and an overcount in 2000.

What we've done since then is we've tried to improve our estimates, our methods for estimating duplicates and other erroneous types of enumerations. The census itself has also worked on many of their procedures including I mentioned a coverage follow up operation which tries to go back to recontact

housing units if there was some questions based upon their response on the census form that could help improve the coverage from that.

Coordinator: Any further questions, Mr. Gregory?

Karl Gregory: No, no thank you.

Coordinator: Thank you. And at this time we have no further questions in the queue.

Stacy Vidal: Okay then I will just put up our contact information one more time. So if you do have any questions following the Webinar please feel free to give us a call or send us an email. Again we will be posting some additional background information on our Web site which will hopefully serve as a good resource for you.

Otherwise let us know if you plan to participate in person on Tuesday or feel free to watch us online. Thank you.

Coordinator: This does conclude today's conference call. Thank you for your participation and you may disconnect your lines at this time.

END