

2010 CENSUS BUREAU PRESS BRIEFING
SEPTEMBER 22, 2010

STAN ROLARK: Good morning. I'm Stan Rolark. I'm Chief of the Census Bureau's Public Information Office. I would just like to offer a warm welcome to everyone who is attending this press conference today. We're very happy to have you. Today Census Bureau Director, Bob Groves, will give an update on census operations. This is the next in our continuing series of briefings on our census operations. Today, he'll talk about the processing of data. And then, he'll discuss release of 2010 census data, which is coming later in the year.

Before I begin, let me just go over a few things and let you know of a few resources that we have. Once the presentation concludes, we'll have a question and answer session. We'll ask the members of the media in the audience for questions. And then, we'll go to a phone line. Then we'll go back and forth. We'll limit it to one question per individual. Of course, if you have an additional question, you can dial back in and get back in line to ask additional questions.

Second, let me point you to some resources that we have. If you go to www.census.gov, you can get a copy of our electronic press kit. And certainly, if you have any questions afterwards, you can call the Public Information Office: 301-763-3030, and we'll be happy to get an answer for you.

So, with that, let me bring up Dr. Robert Groves. Dr. Groves.

ROBERT GROVES: Thank you, Stan. Good morning. Great to see you. Today, as Stan noted, I'll update you on where we are. And, in a nutshell, I can say that we have completed the phase that relies on the American public to produce a successful census. And it is truly now in the Census Bureau's hands to follow through. The American

public, as we've said over and over, came through in an amazingly powerful way to make this a good census. And we thank them again.

We now begin a long set of checks, making sure that the aggregation of the answers that people gave us are done well, that we have data as reported, and the data makes sense. I can tell you, given our early processing results, I am very confident that we will return to the American public a good census, that we will meet our deadline of December 31st for the apportionment counts and for the state level counts.

So, let me kind of take you through. I'm going to take you through the operations that we've finished recently, give you just a little update on how we think they went. Then I'm going to say a little about the events we're going to be doing, the things we're doing inside the Census Bureau over the next few months.

Then, I want to give you a sense of the public releases of data that are going to occur this fall. You're going to see a lot of data from the Census Bureau. And we want to make sure that the American public understands what the data are and what the data are not when we release them.

So, let me go through the operational update. During our processing of the data, we discovered something. We discovered about 700,000 cases where a mail-back return actually had no data on it, or just answered-- or the form just contained a population count for the housing unit.

Then we noticed, in our nonresponse follow-up operations, that we identified some households that were judged to be occupied, but we couldn't get a population count for the household. That was about 700,000 addresses in total. We went back out. We had time. We had money. We went back out to those housing units to see if we could get better data.

We've completed that effort now. I can tell you of the roughly 700,000, about half were classified as occupied households. And we picked up population counts on the vast majority of those. This operation-- So, this operation is finished. It was over budget, as it turns out, about \$42 millions we spent on this versus the \$31 we thought was going to be spent. Basically, the productivity was lower than we thought. The mileage was higher. People had to drive more miles to get these houses classified.

The data operation is finished. Let me turn to another operation that we have completed. We mentioned this last time, called the "Vacant Delete Check" in Census Bureau jargon. This is where we doublecheck about 8.7 million households that were judged to be either vacant, or we couldn't find them when we went out on site to visit the address during the nonresponse follow-up stage.

We also added about 200,000 late additions to the address list. Some of these were new construction houses, newly constructed homes. Others were houses we just missed. That operation is finished successfully. About half of those houses were judged vacant. About 25% were judged to be occupied. And we have population counts on most of them. And 24% were deleted, as it turns out.

That operation, too, was over budget. The same diagnosis. Our interviewers had to travel more miles. There are two reasons that occurs. Either the cases were more dispersed spatially than we thought in our cross-modeling, or they had to go back to these houses more than we thought in order to get a final disposition. But that operation is now completed.

And then, we have one other big field operation. And this was, again, a quality check. If someone picked up what we called "Be Counted Forms" from a local convenience store, or a library, or a school, and filled it out, and wrote in an address that we couldn't match to our address list, we wanted to make sure that we could find those houses. We'd both have to count people and then put them in space.

So, we sent out, for a double check, houses or addresses that were not matched to our list. We found, when we went out there, that we verified 49% of them. We added those to our group. The others were judged either to be duplicates of existing addresses, or we couldn't find them at all. And, that operation was, as it turns out, quite under budget, relative to our expectations. And that's completed.

So, at the end of all this, we have completed all the field operations that contribute to the count. We do have other operations going on. And they are operations that tell us how good that count is. One of them is called the Census Coverage Measurement Operation. That's a large sample survey of about 187,000 households. Interviewers are contacting those households as we speak.

This field activity will be completed in early October. This is a key component to judge how good this census is. We'll end up matching the addresses that we visit in the sample survey to the census base, in order to estimate matches and mismatches.

Then we have one other evaluative research project going on called the Census Quality Survey, that is asking about 286,000 households around the country to fill out a census questionnaire again, either on paper or on the Internet. This is a big test for us to see, to judge how we can use the Internet wisely in the 2020 census.

Now, there are a bunch of smaller evaluations that are going on, that may be contacting a small number of households. But basically, as I said at the beginning, the phase that relies on the American public to produce the population counts is over and over successfully, I'm happy to note.

We are beginning to close out our local census offices. We have 494 offices around the country. They are being shut down and closed up. We'll close about 128 of those by mid-October and the rest by mid-November. There may be a couple of stragglers if we have problems.

We also, as you may recall, have 12 regional census centers spread throughout the country. And those are going to continue to be open for a few months because we have a variety of personnel issues that need to be dealt with. For example, if an enumerator doesn't get a W-2 form or doesn't understand a W-2 form, they can call the local census office, and they'll be connected to the regional census center to get relief on that.

And just a note, we have permanently in place 12 regional offices. Those will continue to be there to support the variety of survey activities we do, including the American Community Survey that I'll talk about in a minute.

Okay. So, what are we going to do now? We have data in house. And we are beginning a long set of quality checks and editing and careful scrutiny of the individual returns, to make sure that our aggregates, the final population counts, are the best they can be. All of the questionnaires that we receive have to be processed through optical scanning machines in three processing centers.

We then need to make sure that for every form we have, we can place it in a geographical unit appropriately. Because that's part of our obligation. We will, in that phase, notice that some addresses have duplicate returns. We have to deal with that. And then, we have to make sure that, in inspecting each completed form, the characteristics, the answers to the questions that we pose to the American public are, indeed, filled out in ways that make sense. And then finally, we strip identifying information from every data record before we do the final aggregation for the population counts.

So, where are we in that big process? As of September 7th, we have completed all the data capture of questionnaires from our processing centers. That means that those paper forms that you mailed back or were completed by an enumerator have been scanned and turned into numeric data, almost completely. In some cases, as I said, we'll have duplicate forms when we look at an address, a housing unit. And, we will have to resolve those duplications. We've done this in every census. There are a set of algorithms that are in place to identify which form is the preferred form to keep for a particular address.

And then, as I said, we have a big set of resolutions to make sure we have geographical boundaries right, to make sure that we place each form's data in the correct census block, and that we understand how blocks aggregate up to tracts and to school districts and so on.

Now, in your press kit, you have a little example of another problem we face. And that is, on the write-in for race and ethnicity and a couple of other questions, we sometimes see text written in that our optical reading machines can't pick up. And so, we have a large set of data entry clerks in our Jeffersonville, Indiana operation, that are right now, as we speak, completing the transference of these sorts of written in answers into a coding system. This is mainly affecting the race, ethnicity questions. And they will end up assigning numeric code values to all of those other specify, as we call them, answers.

These data, the race, ethnicity data, as you know, are used by a variety of federal programs to monitor compliance and antidiscrimination provisions of the Voting Rights Act and Civil Rights Act. And then, state governments also use these data for redistricting purposes.

After we get the data in numeric form, we then sometimes note that there are missing or inconsistent responses on one form. For example, if people leave off their age or sex, or they mark a five year old as a parent-- For example, imagine that you have a two-person household, a mother and a small child.

The mother, on the questionnaire, fills out the person one column for the form. And the person two column is filled out for the child. But, in the relationship for the person two, the relationship to person one, the mother erroneously writes in "Parent." That is, it appears that this five year old child is the parent of person one.

In those kinds of cases, we examine and edit those records to resolve that kind of inconsistency. So, that would be changed to a child designation for the person to reflecting the five year old.

And then, the final step of the data examination is to strip out all the personally identifiable information. That allows us, then, to create census data products in aggregate form that will be used both for the release before December 31st for reapportionment, and all of the other releases-- for example, the redistricting work.

I want to now turn to something else. This is a problem that we are posing. And we need your help on this. We're going to release a lot of population data over the fall. This is the first census where we've had the American community survey going on at the same time as the census. And there is a potential confusion that we will create in the American public's mind, about whether these are Census 2010 data we're releasing or some other data.

So again, in your press packet, you have a little map, a time map, that looks like this. Take that out. And let me help you through our request to you, to make sure that we describe this right. And in your stories, you understand the differences.

Next week, we're going to release estimates based on the American Community Survey for 2009, for larger areas, areas with 65,000 or more population. These will describe America as it was in 2009. These are not 2010 census data. These are very valuable data, as those of you who've watched the American Community Survey evolves. It gives us socio-economic data and housing characteristics of households. And these data are used by a lot of state and federal programs to administer funding decisions. But these are 2009 sample survey data. So, that's next week.

Then, we're going to go quiet for two months. And then, December, we're going to hit you-- boom, boom, boom-- with data. The first sound of estimates we're going to give

you will be for 2010. But it won't be 2010 census data. These are demographic analysis data.

What is demographic analysis? This is a way of estimating national population totals based on birth records, death records, immigration and emigration estimates. Every country has such a system for national population estimates. And we do too. In the past, we've used demographic analysis to evaluate the census. As we've grown more and more insightful about demographic analysis, as our demographers and other demographers around the country have inspected demographic analysis, we realize it has various weaknesses in it.

For example, there are no direct estimates for emigration and immigration that are judged by, in a consensus way, to be accurate. There are multiple estimates of immigration into the country, the number of immigrants the country, that have logical bases to them, but differ from one another. They're based on different assumptions.

Secondly, the race classification on death certificates and birth certificates, this is a different classification scheme than we use in the census. So, for those two reasons, using demographic analysis on the national level, to answer the question, "How good is the census?" is becoming less and less appropriate.

We had a conference, a great conference in January, 2010, where we brought together the best minds and demography of the country. There was a consensus there that what the Census Bureau should do honestly is to produce multiple estimates of the population, based on demographic analysis. We're going to do that. We'll give you multiple estimates, based on reasonable but different assumptions, about components of immigration and the race, ethnicity classifications.

So, we will use less than in prior decades, demographic analysis to measure undercounts in the census. But we'll be fully honest and transparent in this to tell you what the best

minds in demography say, given this way of estimating the population, about the range of likely estimates.

So, that's going to happen in December. So, what is it again? They are 2010 estimates. They're not census estimates. They're estimates based on wider registration systems. And they will only be at the national level. Then, if you look at your chart, the next thing that's going to happen are again, an American Community Survey release. These will be descriptive of the 2005 through 2009 period.

Now, how do you understand this? Remember, the American Community Survey measures the kind of variables we used to measure in the long form. In the long form of the census, we gave the country once every ten years, estimates at very low level. The American Community Survey is better than that. We're going to give people at all levels, or units at all levels, data about themselves every five years. And this is our first chance to do this.

So, we will release, for all communities in the country, data that describe the 2005 through 2009 period based on the American Community Survey. These are not 2010 census data. Then, and only then, are we ready or will we release the 2010 data before the December 31st date. These will be state-level population counts.

I want to stop there. Your chart goes on to describe what happens in January. But I think the greatest source of confusion where we need your help to write intelligent stories about how these are different than the 2010 census, is in the fall. Because we're going to hit people with a lot of different data that describe the population.

So, why do we do all three of these methods? Why do we do the census? Why do we do the American Community Survey? Why do we do demographic analysis? These are different ways of measuring us. They're ways that have good bases. They serve different purposes.

The census is the full count of the country, the official population counts used for redistricting, reapportionment, redistricting. It will remain that. We can't use any other method to do that. The ACS is a continuous sample survey. It provides rich data on socioeconomic characteristics of the population. We need that to administer various programs that are related to that.

Demographic analysis provides national estimates. It's our source of ongoing population estimates throughout the decade. And it relies on vital statistics, vital registration systems. Together, all three of these things come together to help us understand who we are.

So, that's the basics of my remarks. We will have another press conference in October. And there, we're really going to shift attention to planning for the 2020 census. And we can also release our final mail participation rates, to give you kind of an overview, a retrospective of what that part of the census looks like. So, I'm happy to stop and happy to entertain questions.

STAN ROLARK: Before we go to questions, let me just remind everyone that we'll have one question per reporter. We'll allow you a follow-up, depending on how many questions we have. Before you ask your question, if you can give your name and your media affiliation, appreciate that. We also have folks on the phone line. So, we want to make sure that we go back and forth from inside. And then, we'll go back to the phone lines.

So, we'll start off the first question with a reporter in the room. Carol.

CAROL MORELLO: Carol Morello from the *Washington Post*. Now that field operations are over, could you give us your judgment on what impact the antigovernment sentiment had in this census? And, for example, those 700,000 cases you mentioned at the beginning, where people either gave no response or just the population count? Was that antigovernment sentiment or what?

ROBERT GROVES: Well, the concern about antigovernment sentiment is both related to the mail return rate-- that's what you focus on-- and then, I think the treatment of our enumerators when they went out. But, on these 700,000 cases that you mentioned, only a portion of those were cases where there was a form return just with the number of people in the household. That number was less than we thought it would be, frankly. And others were returns where these were second homes of people, where they returned the form either with a zero in the population count, which is exactly what we asked them to do. Or, they returned it blank.

We don't know, obviously, from our data, why people did anything. We just know what they did. Those numbers suggest to me, however, that the fears of boycotts of the census never really materialized, behaviorally. And so, that's consistent with the information we were looking at during the mail return rates. We were looking at this daily. And that didn't materialize, in ways that I can perceive, anyway.

STAN ROLARK: Okay, thanks for that question. I don't believe we have any questions on the phone. So we'll take another question in the room. And let me just remind you, too, we have a mic. So, if you could wait for the mic before you ask your question. So, do we have any questions in the room? Carol.

CAROL MORELLO: So, what are the final numbers on the number of census workers who were injured or, you know, harmed in any way?

ROBERT GROVES: Yeah, we had roughly 700 or so incidents. The modal category, from a statistical point of view, the most frequent occurrence involves some sort of assault-like event. Sometimes weapons were involved in that. There are fewer where the weapon was actually discharged. And so, if you look at these incidents, each of them is a horrible event.

If you look at it in terms of the risk of this, this is really quite a rare thing. So, I figure that our enumerators probably made 100 million visits to homes, okay, 100 million. And then, of those 100 million visits, we have 700 incidents of one sort or another. And, if you think of it that way, this is a fairly rare event. Each event is undesirable. I wish we didn't have them. But they happened.

STAN ROLARK: Okay, thanks for that question. And I just received a signal that we do have a question on the phone.

OPERATOR: The first question will come from Haya El Nasser with *USA Today*.

HAYA EL NASSER: Yes, Dr. Groves. I just wanted to clarify the two operations that you mentioned, that had come in over budget. If you could just go over that again, I'm not really clear. You mentioned the one operation, the cases of the 700,000 forms that did not have a date on it, or missing information or whatever, or came from occupied units but had no data on the form. And then, the vacant delete check, which meant double checking about 8.7 million households.

And you said that both of those turned out to come in over budget, by how much? What are we talking about here?

ROBERT GROVES: Okay. I can give you that. So, I'll do it in the same order you mentioned it. First is this revisiting of about 700,000 households. We thought that that would cost about \$31 million dollars. It cost \$42 million. The second one was the vacant delete check, where we went out to 8.7 million. We thought that was going to cost \$244 million. It cost \$282 million. We are analyzing, right now, as I speak, diagnosing those kinds of results.

This is unusual if you go back over all of our operations. These are the only two out of, say, 13 or so that were over budget. They have the markings of the same sort of problem. And that is, it looks like we underestimated the amount of visiting people required to

make contact with the households. And this is, in retrospect, understandable. These are some of the toughest cases in the country, in the toughest neighborhoods, where we went back in complicated situations to make sure we understood what was going on. But we did underestimate those costs.

STAN ROLARK: Okay, thank you. Questions in the room? Ed.

ED O'KEEFE: Yeah, Ed O'Keefe of the *Washington Post*. You asked us to sort of proactively differentiate in the coming months. To what extent might you guys end up doing that? And, should we anticipate maybe something from you guys once the big numbers come out there, the cennial(?) ones, with the potential that those same antigovernment folks, once again, raise questions and concerns about the validity of the numbers?

ROBERT GROVES: Let's see. I'm not sure I understand the second question, but let me do the first one. We're trying to get this message out to compare these multiple estimates, all of which describe the American public, in every various way we can. So, we're reaching out to our partners, to tell them sort of a heads-up, "Here's what's going to happen in September. And here's how the things are different." In every press conference and releases, when we give individual estimates, we'll say what they are and what they are not, to help on that. And then, we're trying to take every forum that we can find, to say this.

We anticipate, despite this, there'll be problems. These are unprecedented challenges we have. In a way, it's great that we're giving back to the country all of this measurement about itself. But this is the first time we've gone through this. It may be better than we think. But we're worried that there are sources of confusion. So, we want to work on that. I'm not sure I understood your second question.

ED O'KEEFE: Just, as you so effectively set up a campaign and were ready to answer the concerns when the forms went out, should we anticipate you'll be ready to do much the same once the data comes out?

ROBERT GROVES: I hope so. That's our job. And you can help us if we're not clear on that kind of communication. We need your help.

STAN ROLARK: Yeah, can we have the next question on the phone, please. Any further questions in the room?

___: [off microphone] -- would you consider revising your embargo policy so we have a day or two to look at them and analyze them?

ROBERT GROVES: I hear the question about the embargo. And I guess I can, indeed, say absolutely yes, we will consider that. And we are. I mean, this is an active thing that we are talking about.

STAN ROLARK: We do have a question on the phone. Operator, could we have the question please.

OPERATOR: Yes, Betty Ortega with Mexicano.

BETTY ORTEGA: Hi. I just wanted to see if you can clarify a little bit about the problems that you run into when it comes to race and ethnicity. Because you give some examples, but I don't think it came out very clear why. I mean, the text can be read.

ROBERT GROVES: Sure, I'd be happy to. So, as you may recall, and the press kit helps you on this, if you judged that you were a member of a racial or ethnic group that didn't have a tick box and a label, we allowed you to write in. We asked you to write in the words that you use to describe yourself.

And, as you might imagine, some people stayed within the boxes and wrote very carefully. And we have machines that can read that instantly. They really work cool. Other times, people had a very long or many words to describe themselves. And they had to write outside the boxes. The machines couldn't read that. And, whenever that happens, we kind of kick that form out, and one of our data clerks looks at it and then types in the response. And then, we code things.

That's to make sure that everybody, whether you stayed within the boxes or not, everybody's answer is captured the way they put it in. And then, we code those data. So, that's the process.

STAN ROLARK: Okay, let me remind everyone, as well, that you can get a copy of the electronic press kit by going to www.census.gov. If you have any questions beyond-- you have answers today, you can also call the Public Information Office at 301-763-3030. So next question in the room, please.

MAX CACAS: Hi Dr. Groves, Max Cacas from Federal News Radio. I want to follow up a little bit on Ed O'Keefe's question. And that is, you're a long-time statistician from the academic world. And you understand, very, very well, the challenges of trying to make a survey like this. I've heard some statisticians I've talked to say that the census is actually the ultimate statistics project because of that.

But really, people get-- Americans get thrown a lot of statistics every day. And yet, time and time again, we find that people don't really understand the statistics that they're given. And now, within the next several months, you're going to have two different surveys. And this is the challenge that you talked about. You're going to have two different surveys that measure exactly the same population but report the results in different ways.

What kinds of considerations do you have in trying to educate people about the whole field of statistics? Because the only time they ever get this is every ten years. And the only time they're ever really conscious of it is every ten years.

ROBERT GROVES: Okay, this is a rich question, so let me pick out some of the important parts here. I think it's our obligation, and I'm hoping we'll fulfill that well over the fall, to tell people what the estimates are about, what they are portraying accurately. And so here, in the fall, the big dimensions of difference across these estimates will be time. The American Community Survey will always be talking about 2009 or earlier. The census and demographic analysis is always going to be talking about 2010. So time differs.

And the other will be the level of geographical detail. Next week you're going to see data on rather large communities. That is as low as it will go. You'll have national data and then down to rather large cities. Demographic analysis is only at the national level. The five-year 2005 through 2009 American Community Surveys data are going to go down to really small levels. And then, finally, the census counts we give are only going to be state level counts, okay.

So time, geographical detail, they'll differ on that. And then finally, and this will be obvious when it is released, the American Community Survey measures a lot of different things. The census, as we've said over and over again, this little ten-question census, doesn't measure that much. The census release is only going to give population counts at a state level. No other data on December 31st. The American Community Survey data will give a lot of different characteristics of the population. And then, demographic analysis is just going to give you population counts by age, gender and black/non-black status, okay.

So, you're right. It's our obligation to say those things over and over again. I think it's important to note that although the American public is getting a lot of information about itself, it has an insatiable appetite for those data. Americans love numbers about

themselves. We're trying to fulfill that love by giving them the data they've asked for and the programs used. But it does mean, especially over the fall, that we have to be careful at how these things are different.

STAN ROLARK: Okay, thank you. Do we have any further questions in the room?

ALEX: I was just wondering if you have exact release dates available yet.

ROBERT GROVES: We do on most of these. And I don't have them in my head. We have not announced the dates, I think, for demographic analysis and for the final 2010 population counts because we wanted to make sure we hit those deadlines. But someone could give you the dates on the others.

STAN ROLARK: Yeah, Alex, if you could call, I could just give you the number again. It's 301-763-3030 or 301-763-3691. And we can follow up with you and get you that information.

ROBERT GROVES: The one-year ACS estimates are next week, on the 28th. So, that's the first thing to look for.

STAN ROLARK: Thank you. So, any questions on the phone, operator? No questions on the phone? Any further questions in the room? Okay. Well, I want to thank everybody. Dr. Groves, do you have any other?

ROBERT GROVES: No.

STAN ROLARK: Thank you very much. I'd like to thank everyone for attending today. Let me remind you, again, that you can go to our website to get a copy of the press package, www.census.gov. And just feel free to give the Public Information Office a call if you need any further information or have further questions, 301-763-3691. Thank you for attending today.

END OF CONFERENCE