

NWX-US DEPT OF COMMERCE

**Moderator: Gregory Pewett
May 13, 2019
12:00 pm CT**

Coordinator: Welcome and thank you for standing by.

Participants are in a listen-only mode until the question-and-answer session of today's conference. At that time you may press star, then 1 on your touchtone phone to ask a question.

Today's conference is being recorded.

Dave Waddington: Okay. Go ahead. Everybody, good afternoon. We're going to go ahead and get started. Thank you.

So if you didn't hear Trudi a little bit earlier, we do have some water and cookies for those in the room over there in the corner, so help yourself throughout the afternoon to those.

Again, good afternoon. I'm Dave Waddington. I'm the chief of the Social, Economic and Housing Statistics Division. I'm happy that you all were able to make it today, for those in the room as well as those on the phone. I appreciate you calling in to participate in this meeting.

Today, again, we're here to talk about the expert meeting for (unintelligible) changes in the current population survey, the Annual Social and Economic Supplement. Before we get started, I wanted to do a quick introduction as well of Tori Velkoff here to my left. She is our associate director for demographic programs, previously division chief as well for Seaside. So want to give her a chance to say hello quickly before she has to run off, unfortunately. Yeah.

Tori Velkoff: Thank you Dave. Yes. Welcome. I'm glad everyone is here. Unfortunately I have to go downtown to a meeting at 1:15, so I'm going to miss this meeting. But I look forward to hearing about the results. And again, I appreciate everyone coming out.

Dave Waddington: (Unintelligible). All right. So as you all know, the Current Population Survey's Annual Social and Economic Supplement is one of the key sources for many indicators and statistics. It's the longest running demographic survey that we have here at the Census Bureau for over 50 years providing key statistics.

As you can see on the screen here, the Income and Poverty Report, Health Insurance and Supplemental Poverty are some of the key ones that we put out each September. But there's also many other statistics and indicators that are put there such as the Educational Attainment tables and where we track movers and other things in times of migration flows as well.

So today we're going to talk about our new processing system that we're rolling out. This has been a two-stage process where, back in 2014 and '15 we've made some changes and updates to the questions in the CPS ASEC. In particular with respect to income, health insurance, and then in '15 some

questions on relationship. And starting that first year that we implemented those questions and so we started using that information in our old processing system, we sort of rearranged things to process and then used the information, and they were part of our releases all along, since we've been collecting the data. But now we've completed work on a new processing system, and that actually fully utilizes all the information we're collecting in terms of imputation and calculating health insurance coverage and such, which is what we're going to be talk about here today.

So it's - again, it's sort of a two-stage process. The questions were made. We've been using the data, and now we're implementing and rolling out the processing system that'll be a part of our release this September.

So this slide shows our agenda for today. We have our - of course our welcome and introductions, and then we'll have - the first presentation will be (Ben) over here talking about household and family relationships. We'll move on around 1:30 to income and poverty. And then 2:45 to health insurance. And there'll be an opportunity at the end for questions and wrap up at 3:30.

So that will be our day today. I - again, I appreciate you all coming out and participating. If you have questions, I think you'll be able to ask them as you go. We do ask you, if you are going to ask a question, to turn your mike on. A little red light will come on, so you'll - the folks on the phone can hear as well.

Is there a way for the folks on the phone to let us know if they have questions or we'll have a time at the end for that?

Man: Yes. So whenever you all would like to take questions (unintelligible).

Dave Waddington: Okay.

Man: And then that will be (unintelligible).

Dave Waddington: Okay. So for the folks on the phone, probably what we'll do is we'll have a little question time at the end of each presentation where you'll be able to chime in with questions for that. So all right, with that, I'm going to turn it over to (Ben).

Tori Velkoff: In the interest of full disclosure, we are taping this webinar today so that we want to be able to make it available on the Internet for people who weren't able to come or to call in. So as long as your mike is on, you could be taped. So just know that.

Ben Gurrentz: Thank you Ashley. Oh, there we go. All right. My name is Ben Gurrentz. Good afternoon. Today I'll be talking about the upcoming changes, the household data and the measurement of same-sex couples in the Current Population Survey's Annual Social and Economic Supplement or CPS ASEC.

In order to meet the Census Bureau's mission to provide quality measures about the nation's people and households, we've been working over the last decade or so on the improvement of families, specifically same-sex families. We have undertaken extensive research to develop and implement changes to the relationship question and methods for processing these data in Census Bureau surveys as well as the Decennial Census.

Today I'll be talking about the changes we've made in the household relationship data in the CPS over the last decade or so, including the expansion of husband or wife in unmarried partner categories and to opposite sex and same-sex categories, changes in our editing procedures for same-sex

couples and changes to the parent identification questions. First I'll describe what has changed, and then I'll show you the estimates and the research file, where those changes have been implemented.

All right. So a little background information before we delve into the nitty gritty. In 2010 we revised our editing procedures for those reports as same-sex spouses. In the past we changed the sex of one spouse and showing them as opposite sex married couples. Beginning in 2010 we instead changed the relationship from spouse to unmarried partner, retaining the information on sex.

So when we're showing the estimates in the research file and comparing it to the production file, we're going to be using an extract file where that file reports the original responses of spouses and their same-sex reports. So when we're talking about those comparisons, we're going to be using that extract file.

All right. So as most of you know, we've been working to improve the measurement of same-sex couples by revising the relationship to householder questions specifically. And this has been a focus so we could address some important data quality challenges. And it's a statistical one. If you have a low rate of random measurement error in a large group -- like in this case opposite sex-married couples -- it can create problems in the estimates of smaller groups like same-sex married couples and inflate those numbers.

So in 2017 there was a reported 57 million opposite sex married couples. But if you just take a small, very small portion of those cases where they mismark the reports it can just inflate the number of same-sex spouses. So that's something that we want to avoid doing.

So in order to address this, we have revised the relationship question by expanding the husband or wife in unmarried partner categories into same-sex and opposite sex categories. The revised relationship question has been phased into the CPS starting with the incoming rotation group in May of 2015 so that by the time of 2017 ASEC all cases received the revised question.

And in the beginning of the interview, we ask how each household members related to the householder, also called the reference person. This is the person who owns the home or whose name is on the lease. Here you see a comparison of the relationship categories, and what's on the production file, which is just spouse and unmarried partner, and then on the research file you see it expanded into opposite sex and same-sex categories up top.

In this slide, we've added the percent distribution of all people across the relationship to householder categories, for everyone except for the householder, that is. But we'll talk about the top categories later on in the presentation. But for the bottom categories, you see that most of the categories actually didn't show a significant difference, with the one exception being the housemate/roommate category, which you see third from the bottom.

You see a significant decrease, and a part - potential reason for that is that household members that report being a roommate or a housemate get changed to an unmarried partner if they identify as an unmarried partner in the direct cohabitation question. So someone felt, this person's my roommate, housemate, and then later through a direct cohabitation question asked, "Do you have a boyfriend, girlfriend or partner in the household," they said, "Oh, yeah, that person is my boyfriend or girlfriend." Like, oh, okay, so it's an unmarried partner relationship. So now that's included and that explains that difference there.

All right. In terms of editing changes, by adding the response categories to distinguish opposite sex and same-sex couples, it's now possible to see reports where the relationship category is inconsistent with sex reports. So same-sex couple, but then you see opposite sex, you see different sex reports there and vice versa. That creates a challenge in our edit.

So for example of what we're calling mismatched households where relationship and sex of both are inconsistent. And the couple on the left you see a female householder and then a household member that's reported as male. But the relationship to householder category is same-sex spouse. On the right you see two male household members, but the person on the right is indicated as a opposite sex spouse. So this becomes an issue, right? Is this a same-sex couple? Is this a opposite sex couple? We don't really know because there's a inconsistency there.

Now often these types of mismarks are inadvertent mistakes. Sometimes they might also reflect cases where transgender people tried to do the best with the limited sex categories that were offered on the questionnaire. But in any case, we don't output the data this way. We try to make them consistent.

So our rationale for resolving these inconsistencies in the CPS is based on data from the largest test conducted in Decennial program. The 2015 National Content Test would reassign sex for the mismatched couples in those test data based on the First Names Index. The mismatched couples are opposite sex about 70% of the time. So based on that rationale, we randomly assigned this mismatched cases to be opposite sex couples about 70% of the time in CPS. We don't have a names index in CPS that we were not able to use that to inform our decision. We're informing our decision based on the results from the 2015 National Content Test.

All right. Let's move on to the parent identification questions. Previously these questions identified a mother and a father, but now they're gender neutral asking whether a parent is present and whether a second parent is present. And this allows us to accurately reflect same-sex as well as opposite sex parents.

All right. Here you see the production questions is, "So-and-so's mother a member of the household, is this person's father a member of this household?" But this is an issue because let's say there's someone that has two mothers or two fathers. There would - someone might report a mother but then the other mother would not be accurately reflected in the data, and we might put them into a one-parent household or something along that lines.

So we wanted to be able to get a better idea of same-sex couples that are parents. So now we have it - does this household member have a parent in this household? Do they have another - do they have another parent in the household? So this allows for two mothers or two fathers.

So in addition to easily reporting same-sex parents, we also adjusted the way we edit parent identification questions in two ways. So now we're able to identify cohabiting parents in the edit who are not measured in the direct cohabitation question. So there are specific cases where someone might not get a direct cohabitation question, where they ask if a member of the household has a boyfriend, girlfriend or partner. It has to be an adult who lives in the household and who is not related to them.

So you can think of a case where you might have a householder, a child of the householder, and that child has a partner, and that partner gets labeled an other relative of the household. That person would not actually get the direct

cohabitation question. But if those two people have a child in the householder and we assume the child according to both of them, we can identify, "Oh, wait. These are two unmarried partners living together. They share a child in the household." So it's a way to kind of get around those types of challenges where someone may not get that direct cohabitation question, but using the child who identifies the parents, we can say, "Okay, these are unmarried parents living together."

And then another difference is how we've edited the parent identification questions. In order to keep it parallel to what we had before with the mother and father identification questions, Parent 1 is always the mother if the mother is present. And Parent 2 is always the father if the father is present, the one exception being if you have two fathers or two mothers you're going to have a father that's in the mother category and vice versa. But in order to get a better idea when you're looking through the data, most of the times you're going to see Parent 1 is the mother and Parent 2 is the father.

All right. So let's look at some results in the 2017 research file. So as I mentioned before, in order to get the same-sex married estimates, you have to use that extract file. All right. Now that we have that extract file and the production file so we compare it to the revised categories of the research file. We see overall the number of coupled households is higher in the research file, an estimated 68.9 million compared to 68.1 million in the production file. And on this slide we've actually indicated significant differences using the arrow, the green arrow significant increase, a red arrow down a significant decrease.

So you'll notice that in terms of the percent distribution of coupled households in the research file unmarried couples have a larger share, and we believe potentially one of the reasons for that larger share overall might be because of

the editing where we're able to identify unmarried parents that may have not gotten the cohabitation question and we make them cohabiting. So potentially it could be that is why we're seeing more unmarried couples in this file.

And then we also see showed same-sex married couples decrease. And again, this is actually one of the things that we - one of the reasons why we revised the relationship categories, because of those inadvertent mistakes that were inflating the same-sex married categories from the opposite sex married categories. So again, these changes make sense with what we would expect.

We also looked at characteristics of coupled households by type and we examined whether both partners had a bachelor's degree, both partners were employed and whether partners were an interracial couple. We found no statistical differences at the 90% confidence level. And you can see. I have some supplemental slides if you actually want to see those estimates, but it's basically a lot of null results which are not really fun to look at.

But for the sake of keeping things concise, we don't include that here. But we didn't find any significant differences between the research file and the production file when you look at characteristics of coupled households. And now keep in mind that in order to have a significant difference, it has to be a large difference for the same sex categories because they're fairly small. So that might explain why we get null results there.

All right. So let's look at adults who are parents and how they differ between the research and production file. So here when you're looking at just the estimates or the totals of ages - people that identify who are parents 15 and over who have a co-resident child, there's not a statistical difference between the production or research file, either in terms of just all for resident parents or parents that are specifically living with co-resident children under 18. We

don't see a significant difference between the production and research file estimates.

However, we're specifically interested in improving the measurement of same-sex parents, and that's why we implemented those gender-neutral parent identification questions. So when we break it down by couple type, we actually see that we have an increase in same-sex married parents and same-sex unmarried parents. So again, this is something we would expect giving our gender-neutral parent identification questions that we've implemented.

So note we are not proposing a change to the Census Bureau's definition of family, which is a group of two persons or more residing together related by birth, marriage or adoption. However, the changes to the relationship categories and gender-neutral parent identifiers will result in changes for who is included. So married couple families now include same-sex married couples. You also notice changes in our labels and our tables from female householder no husband present, female householder no spouse present. So you'll see those changes throughout our products as well.

In conclusion, the changes we've made to the relationship to householder and parent identification questions just allow for a better measurement of the specific type of family and living arrangement, especially for same-sex couples and their children, right? We're trying to get a better idea of family diversity, of groups that tend to be harder to get an accurate statistical report of. So here we're trying to get more accurate estimates. But at the same time we don't see why differences in the characteristics of couple households in the research file compared to the production file.

All right. So this is my contact information. I appreciate speaking today, and if you have any questions, please feel free to contact me or visit the CPS Web site for more details about the survey.

Tori Velkoff: I think we have a few moments for questions if anybody has some.
(Unintelligible) here. Oops.

Man: Thanks for your talk. There used to be a step in the waiting where there was like a spousal equalization step. Is that - does this affect that at all?

Ben Gurrentz: Yes. It does.

Man: Okay.

Ben Gurrentz: Yes.

Man: Can you speak about how?

Ben Gurrentz: Can I? Ashley might be able to answer this one.

Ashley Edwards: Well, effectively the idea is that if you're - there is like a spousal equalization. So if you select one spouse, you know, it should be representative of that household. So I believe that's functionally what's happening. I know I can't speak to the magnitude of that impact, but...

Lisa Cheok: So before we used to equalize...

Tori Velkoff: We should have people introduce themselves.

Lisa Cheok: Sorry. I'm Lisa Cheok. I work in the -- that's a long name -- Associate Directorate of Demographic Program Survey Office on the CPS team, and I work with all of these people by hand in a lot of different areas.

We used to assign the female's weight of a couple, of an opposite sex couple, to both the female and the male, and then any leftover weights were distributed to single males other than that. So now with same-sex married couples, if we have two females we just put them together and average their weight and assign it to both of them. And we do the same for males. So we did have to change it, but we didn't change the way that it works for the people who were already going through that stuff.

Tori Velkoff: Operator, are there any questions on the telephone?

Coordinator: Presently no, but at any time if you have a question over the phone, please press star, then 1. Presently I have no questions over the phone lines.

Tori Velkoff: Thank you. It might be helpful if we went around the room and introduced ourselves. Yes?

Man: (Unintelligible).

Tori Velkoff: Yes. Do you want to start, Dave? And then we'll go this way.

Dave Waddington: After I changed seats, I thought about that. And I thought, "I don't know who's on the phone." We don't know who's on the phone. So it'd be a good idea to do that. So we'll go around the room here, and then on the phone we'll do a quick rollcall on the phone as well. So I'm Dave Waddington, mentioned earlier, the chief of Seaside. I want to go to John.

John Czajka: John from Mathematica Policy Research.

Amy Steinweg: Amy Steinweg, Health and Disability Statistics Branch.

Ashley Edwards: Ashley Edwards, Chief of Poverty Statistics Branch.

Sharon Stern: Sharon Stern, Assistant Division Chief for Employment Characteristics.

Yerís Mayol-García: Yerís Mayol-García, Fertility and Family Statistics Branch.

Heide Jackson: Heide Jackson, Health and Disabilities Statistics Branch.

(Laryssa Mykyta): (Laryssa Mykyta), Chief, Health and Disabilities Statistics Branch.

Edward Berchick: Edward Berchick, Health and Disabilities Statistics Branch.

John Creamer: John Creamer, Poverty Statistics Branch.

(Jon Rothbaum): (Jon Rothbaum), chief of the Income Statistics Branch.

Trudi Renwick: Trudi Renwick. I'm assistant division chief for Economic Characteristics.

Liana Fox: Liana Fox, Supplemental Poverty Measure area.

Jessica Semega: Jessica Semega in the Income Branch.

Lisa Cheok: Lisa Cheok, CPS team.

Lisa Clement: Lisa Clement, Survey Director for CPS.

Dorinda Allard: Dori Allard, Bureau of Labor Statistics.

Michael Karpman: Michael Karpman, Urban Institute.

Melissa Kollar: Melissa Kollar, Income Statistics Branch.

Maggie Woodward: Maggie Woodward, the Energy Information Administration.

Arloc Sherman: Arloc Sherman, Center on Budget and Policy Priorities.

Eamon Molloy: Eamon Molloy, Congressional Budget Office.

Michel Boudreaux: Hi. I'm at the University of Maryland, College Park.

Lynn Blewett: Lynn Blewett, University of Minnesota State Health Access Data Assistance Center.

Brett Fried: Brett Fried also at SHADAC.

Jennifer Madans: Jennifer Madans, National Center for Health Statistics.

(Sage Arcu): Hi Jessica. This (Sage Arcu).

Robin Cohen: Robin Cohen, National Center for Health Statistics.

Dean Resnick: Dean Resnick, NORC.

Tokunbo Oluwole: Toks Oluwole, Social Security Administration.

Dave Waddington: Okay. So we've gone around the room. So can we go on the phone? Yes, the operator, I guess you'll need to unmute the phone and we'll see how many folks are on there, try and speak at once.

Coordinator: All confirmed on the phone all lines are open at this time.

James Ziliak: James Ziliak, University of Kentucky.

Joyce Morton: Joyce Morton, the Urban Institute.

Paul Jacobs: Paul Jacobs, the Agency for Healthcare Research and Quality.

Dave Waddington: Are there any others on the phone.

((Crosstalk))

Martha Heberlein: Medicaid and CHIP Payment and Access Commission

Man: (Unintelligible) (CLS).

Man: Hi. Can you hear me?

Denilo Trisi: Denilo Trisi, Center on Budget.

Robert Hest: Robert Hest, State Health Access Data Assistance Center.

Jasmine Tucker: (Jasmine Tucker) National Women's Law Center.

Margaret Weant: Margaret Weant, Urban Institute.

Joyce Morton: Joyce Morton, Urban Institute.

Dave Waddington: All right. Any others on the phone?

Hector Rodriguez: Hector Rodriguez, Bureau of Labor and Statistics.

Dave Waddington: Is that it for the phone? All right. Thank you, everybody on the phone and around the room. I guess operator you can put them back on mute, and we'll move to our next presentation.

Jon Rothbaum: Yes. So I will be presenting on the changes to the processing for the income statistics. I may move over there so I don't have to keep looking back.

So I just want to do a little figure to make it a little bit easier to understand this two stage implementation of the redesign, for income and poverty particularly. So in 2014, in Income Year 2013 we redesigned the questionnaire and we had a split panel sample where a portion of the sample was given the new question, the portion was given in the old. And there we have a bridge where we have two estimates for the different questions at the same point in time. And you can see that as the break in series on the figure.

And we're going to be releasing from Income Year 2018 forward in September estimates with the new processing system. But we've also released two years of research files or one year we call the research and one year we call the bridge file. But two years of files where the data has been edited with a new processing system to go with the releases we already had under the old processing system so that there's a two-year overlap now where you can calculate the same statistics under both processing systems.

And you can't really see on the figure because it just covers the old - the dark red, the blue covers of the dark red, but we have two years now to bridge this change in the processing system as well.

So I'm going to go through a little bit of background for the income side of what we - why we redesigned the survey and talk briefly about the survey instrument changes that we made. We had another expert group meeting in 2015 to cover that, but I'm happy to answer questions about it. I just don't want to spend a ton of time here on it. And then talk about what we did with the processing system and why and then go through some results, what did it do? How did it affect our estimates of income statistics?

So why did we do the redesign? Probably this is the case for all redesigns, but we're trying to improve the data quality, particularly for us about misreporting, item nonresponse and errors resulting from respondent fatigue. And then there were some areas where we were particularly targeting like underreporting of means tested benefits and means tested programs and underreporting of income amounts. More generally just to take better advantage of an automated instrument.

Another big focus of the redesign was the nature of the retirement has changed a lot since the CPSA 6 started. We've moved much more toward the world of defined contribution benefit or defined contribution pension plans, so things like 401Ks from defined benefit pensions, traditional pension. And we're also hoping to improve the reporting of asset income, interest in dividends in particular.

So what do we actually do? One change we made was a dual pass approach. So on the first pass we identify all income sources, and the second pass we identify amounts. And here we're trying to avoid the possibility that a

respondent could learn that if you answer yes you get a bunch of follow up questions, but if you answer no you don't so that you could avoid follow up questions by saying no.

We also removed the family income screener. So if you said up front that your family had income in a high enough bin we wouldn't give you all the questions, and there was evidence that people were being screened from getting questions to which they've had benefits or income. We changed the skip patterns, so that means the order of the questions, the income questions we ask so that the more likely relevant ones will get asked first. A big change we made is we added income range follow ups so if you were unable or unwilling to tell us the amount that a person had in income we would try to get a range from you so we have some information about your income, a lot more detail on the retirement interest questions and some changes to how we handle Social Security and SSI questions for disability.

So those were the - those were sort of the main areas of the changes, and we saw big increases, for example, in retirement income. And when you look at the two files in the bridge year, you see changes in various income types like interest and dividends and retirement income that were targeted by other redesigns.

In terms of what are we doing now, I'll go through them really quickly and then talk about them in more detail. We're using these ranges in the imputations so when someone gave us a - you know, they didn't give us an exact income amount, but they gave us a range, that's information we now are able to use in the imputation. And I should say, actually, when we had the new questions before, we were sort of pushing everything into the old processing system. So some of the new information we couldn't use because

it wasn't being used in the old processing we're using now. So ranges are one of those.

We completely overall how we handled other incomes, so anything but earnings, how we impute it. And we also use ranges, and then there's some smaller changes we made like how we impute mortgage, presence of a mortgage, how we handle mean tested benefits and the caps that were on there in the old processing system. We've tried to improve our income allocation flag, update top codes and various small fixes.

So I'm going to go through these in more detail. The earnings ranges, if you ask - well, you know, if we were asking, "What did you earn in your primary job," if you were unable or unwilling to respond, we would then say, "Could you tell me if you earned one of these three categories, 45,000 and under, 45 to 60, 60 or above?" If you answered in that first bin, 45 and under, we would then expand it and say, "Well, was it 15 and under, 15 to 30 or 30 and above?" So for 73% of people who didn't answer the earnings question, they gave us a range. So this is a lot of information we're getting from people about their earnings that we didn't before and we weren't using in the old processing system.

We've actually - let me - we've actually also done some work - (Adam B's) done some work building on some work that was done by Mathematica on look at the - let's look at the W2s for those people. And you actually see that if you just look at the ranges versus the bins we would have put someone in if we had used our (Gusting) hot-deck method, there's essentially more information just in these ranges than in all the demographic and such economic characteristics we're including in the hot deck. You're more likely to be in the bin - your range is more likely to be in the bin that your W2 was actually in than we would have put you with all the other information we're

using in our hot deck. So that's evidence that the ranges are providing us a lot of information that we weren't getting otherwise for most people who don't answer that question.

For other income, we - with the old processing system, we were handling their redesign questions by running them through the old processing system, but we've actually sort of expanded for certain income types the number of questions we would ask about things. So if you had multiple questions on interest, if you said, "I don't know how much I got from - I got \$12 in interest from my savings account. I don't know how much I got for my checking account," we didn't have a way of imputing them separately, so we would essentially be treating that second nonresponse item as a zero. That's something we don't do anymore.

Another big change is we're using the ranges that were provided for that as well. And just to give you a sense, there the probability that somebody would give us a range if they didn't give us an amount is a little lower, so it's - you know, it's 41% for unemployment insurance, 40% - 39% for Social Security, et cetera. If you just go across all of the items, all of the possible income questions for which someone didn't respond, in about half of the cases when they didn't respond to the question, they gave us a range. So again, this is a lot of information that we weren't using - able to use under the old processing system that we're now using.

Another thing we wanted to do was update the hot-deck model. There's been critique of the hot-deck approach used under the sort of language of the term match bias and the economic literature or noncongeniality in the statistics literature. But the idea basically is if you don't have something in your imputation model, it's not going to be - some relationship in your imputation model, it won't be there in the data that you impute. And so to the extent that

we could within the current hot-deck framework we were trying to increase the number of variables in the model so that the data would reflect more relationships and have multiple levels of matches so that if - so to give you an example, for earnings we have 16 variables in our first match level.

In our hot-deck model there's 600 billion cells that somebody could be in, and I'll go through an example. But not everybody is going to find a match. But to give you a sense of how hot-deck works - (unintelligible). We're trying to match non-respondents to similar respondents along a set of characteristics. So if you thought the only thing that mattered for earnings was your race and your gender, just as a sort of toy example, you could define your hot-deck model.

So there's two race categories and two gender categories, and you have two variable - you have two - yes, so we have - everybody is either a white male, white female, nonwhite male or nonwhite female in this sort of toy example. And you say, "If you're a white female non-respondent, we're going to randomly pick a white female who did respond and give you her earnings response." And so we're drawing from the distribution of observed responses for people like you under the terms of the model.

For earnings we actually have 16 variables in the model. There's 620 billion cells or 660 - 600 billion cells, basically. There are not 600 billion respondents to the CPSA 6 so there are people who are in a cell for which there is nobody - there's a nonrespondent cell for which there is no respondent, and so you have to then decide what will we take out of our model to find a match for someone. That's the challenge of a hot-deck and we're trying to sort of work within those constraints to make our imputations better.

How do we choose what to put in our model? The - we did this recently when this was done in the past there wasn't a lot of the work that's been done on machine learning that's available to us now. So we decided to try to take some of that work and apply it to this by using a random forest that's sort of a supervised machine learning technique, not to get over-buzzwordy. But the basic idea is you can sort of pick a set of candidate variables that you might think would get into your - that would be a good predictor for, say, interesting (comps) and then you let this random forest choose which one - tell you which ones best predict interest income.

So if you have - if you randomly include or don't include certain variables, how much of the variation in interest income are you predicting, and then you let it sort of order them from most predictive to least. And so that's how we set up our models.

And I just want to give you some examples - an example for rental income of what this does. So the old system and the new, you can see there are more variables in the new imputation system. So we're now getting better matches because we're conditioning on more things. We're also including things by changing how we sequenced it that couldn't have been in the model before about - basically the way it was done before it was done in one whole block where all the other income items were imputed sort of at once. Now we do it sequentially so that once you impute interest income it can be a candidate variable for rental income or public assistance income because (unintelligible) are correlated. And so those are sort of the two big changes that we made.

Now I'll get to some that are sort of down - yes, some other changes we made. Another one was mortgage imputations. So we actually used to link the CPSAs second statistical match to the American Housing Survey. The HS is connected every two years. There's a lag between when they conduct the

survey and when the data's available for us for processing. And so you can imagine if we're three years or more years behind in the HS where using that can be a problem if the housing market is moving up and down. We're going to be imputing housing information from the HS maybe from 2007 and 2010 and the housing markets have changed - would have changed a lot between those two periods.

And we were doing it to add characteristics from the HS that were very frequent - infrequently used in the AC - in the CPS ASEC. And so we decided we would - we have these questions on mortgage in the CPS ASEC now that we should just impute directly from the CPS ASEC and get better information for the questions we have. So that's one change.

Another thing we did was we changed how we cap - I mean the ability for households to get means tested benefits. So in the old system there was actually a fixed nominal income cap above which you could not be a recipient of energy assistance even if you told us that you had received it and so we took that out. And for other means tested benefits, generally the rule is if your household income was above a threshold you couldn't be imputed those benefits.

And from our research on means tested benefits, that actually doesn't really match what we see in the administrative records when you link them at the individual and household level or responses that we got on surveys like the American Community Survey. So now we allow for those values to be imputed because you do see households with, you know, \$75,000 of income who are in the administrative records for having received, say, SNAP benefits in the prior year. And we aren't then subject to this fixed nominal cap which is sort of declining in real value over time.

And you should see -- and this is what we do see -- that, you know, high income households don't receive these benefits at high rates, but that will be reflected in our responses and in our imputation.

Another change we made that hopefully is of use to data users is how we flag allocated income. In the old system you actually need to check two variables for any income item, and the first is to sort of flag for that item. So I underscore earned value for earnings. And then you also need to check (FL655), which is our supplement nonresponse flag, which means somebody didn't give income information for the whole - they basically didn't give us enough useful income information overall so we imputed all of their income information. This anecdotally, I have found it comprehensive people either weren't aware of or at some point in their research they realized that they needed to be looking at the second flag. It wasn't, I think, as well documented as it could have been.

So now we've changed the flag so that every imputation flag has a code which tells you was this imputed because of an item nonresponse or because of supplement nonresponse? If you just check the imputation flag is greater than zero your code should still work, but you now only have to check one thing if you're interested in a imputation for a single item.

We also are trying to provide more information so I've told you that I think ranges are very useful. They give us much better information. I have - we're also trying to communicate to data users when the ranges were actually used in the imputation. So if you're trying to adjust for nonresponse by potentially you have information about who did actually give us ranges in (unintelligible).

Also the lower the value of the imputation code the more detailed the hot-deck model we used, again, to try to give you information about the quality of the

imputes for any individual person or household for data users to make better use of the data. We also have more composite variables now so we've tried provide codes that say interest now is a composite of multiple questions. Were all of those questions responded to or not? Or if it's a value sort of relatively how much of the dollars that you see in interest were imputed versus reported?

Another thing we did, which you'll see in the results, is we increased top codes for some income types. So for a long time our top codes had been fixed at particular nominal values which had been decreasing in real terms over the years. So a small but increasing share of responses were moving above these top codes. So we increase them for particular items like interest dividends and rental income, for retirement income, et cetera. And we also changed how we handled the top codes so that we revisit them every year so that over time we won't go as long between top code changes.

We made some other minor fixes, some really simple ones like we added age to variables that were household variables that didn't have them so it's easier just to look at the variable name and understand what it is. There were some very, very minor errors on how we allocate market value of SNAP benefits across the subfamilies in a household or how income losses in a primary family affects subfamily income. It could affect some subfamilies by dollar.

So what I do want to spend a little bit of time on is we fixed an error where the redesign instrument was causing interviewer respondent confusion about multiple pension sources. So there was a - essentially what we were seeing was a big increase in the number of people who reported two pensions with exactly the same value, and we sort of tracked it back by looking at the trace files, the interview files to how you - how a interviewer would report no on other income items for sources was different for pensions. And so you would

see what looked like accidental reporting of secondary pensions, and then when the - it was hard to go back and take them out and so a respondent might say, "Well, I had \$10,000 in my first pension." And they said, "What'd you have in this other pension?" And they'd say, "I had \$10,000 in my pension." And so you see this - it increased by about 50 times the number of pension - people who'd you see with two pensions that had exactly the same amount.

It didn't have a huge impact on - surprisingly, actually, on sort of mean income of household 65 and over, but it was something we also wanted to fix. And you're going to see it when I show you income statistics by income type. You see declines in the amount of pension - the number of people receiving these secondary pensions.

Well, I'm just going to go through - they're odd numbers. I'll focus on particular things. This is person income, so it's total income, for example, at the 25th percentile at the person level or earnings. And I'm focusing here. So for earnings the only thing that would have affected this is the only change we really made for how we edit earnings is using ranges in our imputation model.

So you see earnings which is about 80% of income, of all income on the CPS ASEC is up by about 1% to 2%. So for each of these I've got the two files. It's - the 2017 research compared to the 2017 production for income year 2016, the 2018 bridge compared to 2018 production for income in 2017. So income is up about - earnings are up about 1% to 2% at the sort of left of the distribution because of this change in how we impute primarily. That's not true for all income items. So for Social Security, SSI and public assistance, again, the primary changes we made to these were how we - our imputation models changed and we use ranges. So you see less income at the bottom of the distribution in some of these categories.

Overall we see more income by about 1-1/2%, and you see a lot more in certain categories like interest and dividends. This is primarily a top code thing. So if you increase the top codes, the average is going to go up and the total of total dollars are going to go up. And interest and dividends are - they're heavily skewed. There's a lot of people with low values and then sort of skewed distributions.

And then you see this for retirement income. There are more people with retirement income. Again, the changes we're looking at people ask the same questions, only process differently. There's more people with retirement income and there are more dollars but there is a lot less retirement income - there are lot fewer people with certain kinds of retirement income. It's specially secondary sources of secondary by they would be potentially the second source reported like federal government, military or state and local government. These are the - this is the place where you see this sort of interviewer-respondent confusion error coming in. There are more people who had this reported, and when we corrected for it you see a decline in that - in those reports.

But you see a big increase. So the other thing we totally overhauled how we handled defined contribution, how we edit and impute these - the old processing system didn't have any - this wasn't separate from the retirement values and often would be left sort of - if you have multiple sources of retirement income. If you had three sources, we don't actually do an imputation for that third retirement source under the old processing systems, so these were sort of sometimes - so you see the amount - number of people with these values, these sources and the amounts are up.

In terms of what does it do the income distribution I'm going to show you a lot of versions of the same table which is, again, comparing 2017 research to

2017 production, 2018 bridge to 2018 production and then different points in the income distribution of household income for the same subgroups. But the point generally - for the - for most of the distribution it looks like this, which is there is no statistically significant differences in either year for most of the subgroups or overall. That's true with the median. It's true at the 10th percentile although you do see particular subgroups with differences at the 10% level. It's true at the 25th percentile for most groups.

As you start moving up, especially at the 95th percentile, here the top code change starts to matter a lot and you see more income at the top of the distribution of household income in the new files versus the production, the previously released production files.

With that, I will leave you with my contact information and - oh, should we do questions on this now or should we do...

Woman: Yes.

Man: Help me out. Can you hear - mike?

Susanne McCartney: Susanne McCartney. I'm with HHS. I just wanted to ask about your Slide 54 where you looked at SSI income. And could you just say a little bit about - I know there's a couple of categories, I guess public assistance and SSI, it really changes a lot of the lower end of the income distribution. Is that what this is showing?

Jon Rothbaum: Yes. So - and the changes we made that would have particularly affected the distribution of income in these categories is primarily in how we do our imputations. So how we match. We use more information to impute values for nonrespondents and we use the ranges that people provided us. So if

somebody didn't respond to that question but they gave us a range -- and the ranges do vary by income type. So I showed you earnings before we use ranges that are lower in value for income types like SSI where 60 and above might be a - too high of a - which isn't too high a value with the range. But those are the things that would have caused SSI at the bottom of the distribution to be different one file on the other are all - almost all in the allocation how we handle nonresponse.

Arloc Sherman: Arloc Sherman, Center on Budget and Policy. Was I seeing in the last two slides low declines systematically for the younger households?

Jon Rothbaum: So you do see it. So for 25 to 30 household's headed by a 25 to 34-year old. At the median you see about a 1-1/2 point decline in their income, or lower income. I guess it's not declined because it's the same data for the same year. And yes, 3% for one year at the 25th. It's not consistently (unintelligible) interview look across some. But you do see differences for that group.

Arloc Sherman: And thoughts about what the chief mechanism is?

Jon Rothbaum: I mean again, it's the primary things. I mean for one - I mean the primary things would have been, I think the allocation system, how we handle imputation for that group earnings there's even more disproportionately the largest share of their income. So that would be my sense. We're also - when we change the imputation system we're essentially also randomizing again (unintelligible) files that may be just (unintelligible). But my guess for this would be that it's about - primarily about using earnings ranges.

Man: You might have said this, but are the characteristics of the people that provided the ranges similar to the people that didn't?

Man: I don't know how much we've looked into sort of prediction. I don't know if we've run models on sort of what explains your probability of having provided a range (unintelligible). What we focused on was looking at the ranges of how they matched to the W2s and seeing (unintelligible).

Yes. So we've always - for decades, I mean if somebody's not responded, we have some method for trying to match a nonrespondent to a distribution of categories.

Woman: (Unintelligible) more information.

Man: Yes. So essentially, yes. And that was when we looked at the W2s. I don't think I have it in these slides or even (unintelligible). We sort of looked at the - if you're in one of the five bins in the W2s where do we put your earnings - where did you report them in the survey either with ranges or in respondents. And so respondents are usually not with 100% probability I think it's like somewhere like 70% probability in the same bin (unintelligible) survey and in the W2s. And for people who provided a range, it's a lower but then for the people who didn't give a response it's even lower when we impute your value are you in the same bin as the W2 that we see for you? And so the range is...

Woman: (Unintelligible).

Man: Yes.

Woman: (Unintelligible).

Man: No. And we haven't looked. We actually have administrative information on SSI payments. That's something we do want to study but I don't think it's

something we've looked at yet because we have information on interest dividends, SSI and Social Security (unintelligible).

Woman: (Unintelligible).

Man: Although because our earnings is 80% of income, even small differences in what happens to earnings can have a big impact on - as big of an impact as a large difference in another income type (unintelligible). But that's - it is a good point. I take your point - it is - it's something we're interested in looking at, having (unintelligible).

Dean Resnick: I'm not so sure if you just answered this.

Dave Waddington: Please introduce yourself please.

Dean Resnick: Dean Resnick, NORC.

Dave Waddington: Forgot to do that for me.

Dean Resnick: Thank you. I'm just - is it - is the decrease have to do with the fact that when you were doing a hot-deck imputation before you were drawing from the whole range and now you're drawing from just the low income because they - the people bend themselves in the low ranges?

Dave Waddington: So yes, if you gave us a range and you said it was between 0 and \$15,000 per earnings, we're only going to draw a response rate. That's a match variable in the hot-deck. So we're only going to draw value from the people who are like you on the other characteristics who also reported between 0 and \$15,000 in earnings. And before we would have only matched on the people like you in the other characteristics and some of them might have also been

between 0 and \$15,000 but some of them would have been in the other four ranges.

So we're constraining the distribution of potential earnings values or SSI values to the range that you told us you were in.

Dave Waddington: And again on the phone, as the operator said earlier, if you want to have a question at any point in time, go ahead and I think it was star 1?

Coordinator: Yes. It is star 1, and I have no questions at this time.

Ashley Edwards: I'm Ashley Edwards. I'm the chief of the Poverty Statistics Branch here at the Census Bureau. I'm going to be presenting some work that myself and John Creamer and Liana Fox have done looking at how the official and the supplemental poverty rates are impacted by these changes in the CPS.

So again this is sort of all background aware that this is the expert meeting. But just want to, again, make it clear that one of the reasons the Census Bureau has been so thorough in documenting these changes, both the questionnaire and the processing system, is that CPSA seg does serve as the official measure of US poverty going back to 1959. So certainly while changes become necessary over time, it is important that we're making clear why these changes are being made and how you might expect them to impact the measurement of poverty.

Sort of summarize the previous two presentations, the two changes that we'll mostly be talking about are related to the efforts to better measure things ex-couples as well as income and program participation. And here you sort of get that laid out. There's data collection changes. There's data processing changes. And so if you see the table above, we expanded the relationship

categories to include the explicit same and opposite sex relationships as well as collect income data using new techniques designed to both reduce nonresponse bias and to make it easier for respondents to report income. So those are changes done on - sorry, the data collection side.

And then by collecting the data this way, the Census Bureau then needs to process it differently, and so that's what we've been discussing, taking advantage of this new content through things like assigning families using the more inclusive definition and including new income content to improve our imputation methods. And so this is just sort of an alternate way to lay this out. You know, how were these income and demographic changes implemented over time, and note that these changes are a little different than the process for the health insurance changes which will be discussed later today. But those changes don't affect official poverty. So here we're really focusing on the changes to the income and the demographics.

And so first you see the questionnaire design changes with the income changes implemented in the 2014 CPSA sect split panel design. So a portion of that sample received the old questionnaire, a portion received the new questionnaire and that allowed us to measure sort of forward and backwards across that split panel. And then changes to the reporting of the relationship status actually started in the basic monthly CPS and then by the 2017 CPS ASEC all respondents were getting that expanded relationship question.

And so for changes in both the collection of income as well as relationship status, those changes were implemented in two steps like Dave had mentioned. So first implementing the changes in the questionnaire both in the basic and in the CPS ASEC, but effectively converting the responses back into the old categories and not taking advantage of this new content because we're still running that data through the legacy processing system.

So here's what you're seeing now is, you know, this sort of second and last step, where we're updating the processing system to take advantage of that new content. So most of the data I'll be presenting today and what you've seen previously is from the 2017 CPS ASEC referring to calendar year 2016. And so we'll be comparing data for 2016 across the legacy and the updated processing system. And again, there's two files. There's the research file and the bridge file. Both of those are reflecting the updated processing system.

And then in September -- this is sort of what this is all leading up to -- we'll be releasing one file, the 2019 CPS ASEC and it will only be processed under the updated processing system. So for the year to year changes that we'll be presenting in September 2019 we're going to be comparing to the 2018 CPS bridge file by using new processing system to new processing system.

So briefly just to summarize why we'd expect estimates to change in terms of poverty based on these updates, changes to the demographic content are allowing respondents to report as same-sex married couples and the new editing procedure -- so this is new -- the new editing procedure keeps those respondents for those individuals remain in a married family relationship as opposed to the process beginning in 2010 where they were treated as unmarried partners. So again, getting at what (Ben) mentioned. The family definition is not changing. So we're keeping that as first marriage or adoption. But these families are now married and so they are treated as a single family unit.

And then similarly the income changes are a little more straightforward to understand. They're designed to better collect personal income, both improving how we're asking the questions as well as how we're allocating income to nonrespondents.

And so this is sort of the official poverty measure. Again, it's based on pretax, cash income and that birth, marriage, adoption, family definition. I'll also be presenting some data from the supplemental poverty measure. And so how is this impact sort of different based on the measure that you're using? So in the (SPM) on the threshold side it's unchanged. The (SPM) has already been treating unmarried partners as a single family unit, and so they've historically been treated as members of the same family.

But there are some additional impacts on the resource measure in addition to the changes to cash income. Changes in health insurance coverage -- which will be discussed after the break -- have led to changes in the reporting of medical out of pocket expenses. And the removal of maximal income thresholds that John mentioned for reporting as well as allocation a noncash assistance will allow more people to receive those benefits and that will only be captured in the (SPM) as the value of SNAP and housing assistance isn't reflected in the OPM.

Further, although the (SPM) treated unmarried partners as a single family, they were still run through tax simulations as separate filing units. So that's something to keep in mind. Treating these couples as a married family will also lead to changes in their tax units and liabilities. So just to give you a sense of how we may be expecting those changes to manifest in the different measures.

So to measure the impact on poverty we're first going to look at through this total impact comparing the official poverty estimates from the 2017 CPS ASEC across the processing systems. We'll get into a little bit more detail here. We wanted to just look at the impact of the demographic edit. So allowing same-sex married couples to be treated as a single family. We apply

these new family assignments but we keep personal income unchanged from the legacy processing system. And then we simply resum personal income based on these new family groupings, and that gives us some sense of the isolated impact of the new family definition.

I'm going to do the same thing for income, holding family assignments constant but we're measuring poverty based on personal income as calculated in the updated processing system. And then we're similarly resuming to the family level but the family is constant. So that sort of allows us to break out sort of these dual impacts that we're thinking about here.

We'll also present some estimates for 2018 ASEC. There we're just going to be looking at the overall total impact. And then I'll also look at the supplemental poverty measure as well.

So this is really the big takeaway for the official poverty measure. So looking at all of the changes made in the updating processing system, both demographic and income, we see that top line overall poverty rates in 2016 are not statistically different across the editing procedures. We do see some significant changes in poverty by demographic groups. Under the new processing system we're seeing higher poverty rates for Hispanics, the elderly, those living in the South, they're outside of MSAs, and for people aged 25 and older without a high school diploma.

We do see that poverty decreases under the new processing system for individuals with at least a bachelor's degree, and this was the group that was already among the lowest poverty rates of the demographic groups shown. So now once we've seen this total impact I'll go into sort of the more detailed evaluation of - starting with the demographic changes.

So to start in 2016 there were 937,000 spouses who reported themselves in a same-sex marriage. Those individuals were more likely to be female, more likely to have a (unintelligible) education. And as (Ben) mentioned, he's got some supplemental statistics on their characteristics.

And so the legacy demographic processing system did not take into account the questionnaire changes that included the explicit opposite and same-sex couples. And we were editing those individuals to be unmarried partners. So because the official poverty measure treats unmarried partners as two separate families, you can see here poverty rates in 2016 for this group were high compared to spouses who had reported being in an opposite sex marriage.

But now in the updated demographic processing system we're treating these spouses as a single family unit, and you see their poverty rates fall in line with those of opposite sex married couples, not statistically different under the new processing system.

So here we're looking at the isolated impact of just that demographic processing system change. We find that poverty declines 0.05 percentage points. And will statistically significant, when you round the poverty rate to the 10th digit it's unchanged. The impact of the updated family edit is largely as you'd expect given the characteristics of the same-sex married population. We find that poverty's declining for females, individuals aged 18 to 64, those with a bachelor's degree and those who live in principle cities, all of which are more prevalent among the same-sex married population.

Oh, okay. So the impact of the demographic edit by family type is a little bit more nuanced. Poverty declines for individuals in primary families, but we do see increases in poverty for unrelates sub families and unrelated individuals. And so here I'm not showing same-sex couples as an individual comparison

group because that group didn't exist in the production processing system. So people who are moving into same-sex families under the new processing system would have been in any of these family types previously.

And here you get a really good sort of illustration of the complication movement of people into and out of those family groupings, unrelated individuals who are getting assigned into primary families under the new editing system, as you can see primarily same-sex married families have lower poverty rates than unrelated individuals who remained classified as unrelated individuals. You see similar differences in people who are moving in and out of unrelated subfamilies, but it's really not driven by people moving predominantly into same-sex married couple families.

So here we're getting at the income edits. So here we're looking at holding demographic and family units constant but looking at new personal income. And so when we hold family assignments constant we see no change in overall poverty rates but we do see increases in poverty for a number of groups. This includes Hispanics, males, people aged 65 and older, those living in the South or outside of MSAs and people with lower levels of educational attainment. One interesting thing to note is that people in the south were the only group to have significant and conflicting changes in poverty across the demographic edit and the income edit.

And so counterintuitively, even though we're seeing some increases in poverty, average household income, as John discussed, does increase under the new processing system, although as John also discussed that impact varies by income sources and by your place on the distribution. As we saw on the last slide, it's pretty consistent with what he'd previously presented. Only seniors age 65 and older saw significant increases in poverty under the new income edits. And that's supported here by the decreases that we're seeing in

average Social Security, although retirement income outside of Social Security was not statistically different.

And this is where it gets - becomes valuable to look at the distribution. So John mentioned in terms of how these changes may be pushing people into poverty, we see that most of these declines in income among the retirement sources are happening at the low end of the distribution among people who would be most vulnerable to falling into poverty. And one thing I just want to sort of reiterate is, you know, although the goal of the redesign was to better capture income data, and so while you might think at first glance that this doesn't look encouraging, this is only the impact of the new processing system.

So as John mentioned, new allocation procedures, new imputation, editing. When we looked at the impact of the questionnaire change itself across the 2014 split panel, we found that there were increases for both Social Security and retirement in terms of both reciprocity rates as well as aggregate income. This is just the impact of the new editing procedures.

So far we've just been talking about the official poverty measure in the 2017 ASEC. This is just upgrading that to the 2018 CPS ASEC. Again, the impact overall both income and demographic edits. And as you can see again in Calendar Year 2017 we're similarly seeing no significant change in overall poverty rates across the legacy and updated processing systems. Like in the 2017 ASEC we continue to see increases in poverty among those age 65 and older and declines in poverty for those with an advanced education.

In the 2018 CPS ASEC we also see declines in poverty for white non-Hispanics and those living outside principle cities. And we also see fewer

instances were poverty increases under the new editing procedure, limited here to blacks and individuals age 65 and older.

And so now we're going to sort of switchgears again back to the 2017 CPS ASEC but now with the supplemental poverty measure. So unlike the unofficial poverty measure the SPM does show a significant change in poverty rates across processing systems and poverty rates are lower under the new processing system for the overall population as well as for major age groups.

And so to look at this a little bit further, here's the impact of various resource additions and subtractions that push people either into or out of SPM poverty based on age group. We find that there's no significant difference in the impact of refundable tax credits or SNAP benefits in moving people out of poverty but there are significant difference when you subtract medical out of pocket expenses from income. And so we see fewer people who are moved into poverty based on medical out of pocket expenses in the updated processing system. And so certainly you can imagine that would be more impactful among the elderly.

So here we're just evaluating the relationship between the official poverty measure and supplemental poverty measure. In 2016 under the legacy processing system -- so the historic relationship -- the SPM was 1.3 percentage points higher than the OPM and this gap narrows to 0.7 percentage points when using the updated processing system. Again, because we saw no significant change in the overall OPM rate while the SPM poverty rate declines.

This overall convergence is really driven by changes in the 65 year and older age group. We saw OPM poverty, again, increase for this population under the new processing system while the SPM poverty decreased.

So just to sort of tie this all together, a few takeaways, when updating the CPS processing system to take advantage of this new content, we're seeing no significant change in the percent of people in OPM poverty in both 2016 and 2017 although both years did show significant increases in the official poverty measure for people age 65 and older. The impact on the supplemental poverty measure varies. SPM rates declined overall under the new processing system as well as for each major age group and the subtraction of medical out of pocket expenses is moving fewer people into SPM poverty.

Just to kind of revisit our primary goals, sort of the beginning of this redesign, we're identifying 1.2 million individuals who are living in same-sex married couple families, and we see large corrections in their poverty rates. We also see increases in official poverty across a number of demographic groups based on these income processing changes and while average household income is up, we see declines in Social Security and retirement income at the 10th percentile. Although again this is only based on the changes to the editing system, so changes to imputation, allocation. It's independent of the earlier changes based on data collection.

Just to sort of give you an update of what's coming next, in September 2019 we'll release one file, the 2019 CPS ASEC and this will reflect all questionnaire and processing changes that we've discussed here. And year to year changes will be evaluated based on comparing 2018 bridge file to the 2019 CPS ASEC. And certainly we're encouraging data users to prepare for understanding those changes. We have the public use 2017 research file, public use 2018 bridge file, currently on our web site. You can get a sense of

variable name changes, categorical changes and just sort of get familiar with what you'll see in the 2019 CPS ASEC.

Great. And look forward to any questions.

(Brett Fried): Did you look at this by state at all? This is (Brett Fried) at SHADAC.

Ashley Edwards: So that's a good question. So typically for the CPS we would do state estimates by pooling multiple years of data. Right now we have two years with the '17 research and '18 bridge. I think the expectations is that for the release in 20 - in September of this year that the SPM is planning to release state-level estimates using those three years of the pooled redesign data. But it's important, I think, to make clear that when pooling multiple years of data, the guidance is to use one processing system. So, you know, only use pooled legacy data for the years that are relevant to that and only used pooled updated processing system. Don't sort of like combine different processing systems just because you want a three year state average.

(Lynn Blewett): (Lynn Blewett), University of Minnesota. Could you describe again what you did with the out of pocket, medical out of pocket for both the OPM and the SPM?

Ashley Edwards: Yes. So I probably wouldn't be the best person to speak on that. I know there's going to be a presentation this afternoon...

(Lynn Blewett): (Unintelligible)...

Ashley Edwards: ...on the health insurance changes. Does anybody want to talk about that now? (Unintelligible) to be honest, certainly, would be the expert on the SPM.

Woman: I can just say we just - so we don't - the medical expenses aren't part of the OPM at all. It's just part of the SPM and we just subtract the value that people report in terms of the out of pocket expenses for premiums, copays, other necessary expenses, over the counter drugs.

(Lynn Blewett): And prior you didn't subtract them?

Woman: No, we always did. The only change is in the way that we do the imputations primarily of the medical expenses. So the way that we subtract it has not changed at all. It's just the way that we have imputed the values. So we have lower values of medical expenses now.

Heide Jackson: I'm going off again. My name's Heide Jackson. I'm with the Health and Disabilities Statistics Branch. So actually one of the things that we've been doing is actually improving our imputation of medical out of pocket expenditures, and we've been doing that in part by if people don't report any health insurance information we jointly impute their health insurance and medical out of pocket information. And we've also been adding additional detail to our hot-decks to hopefully improve our estimates of medical out of pocket expenditures.

That evaluation is ongoing. But that's something that we definitely have been working on and is new to the new processing system and will be covered later today, but I can definitely reach out to you with more information.

Woman: Maybe this was covered earlier, but with the change in the same-sex definition of family -- I realize you're not changing the definition of family -- but is it actually changing the number of unrelated individuals that you're counting?

Ashley Edwards: Certainly, you know, at least the changes there. Most of the unrelated individuals who were unrelated individuals and no longer are moved into same-sex married couple families. I'd have to look at the exact numbers but certainly we did take a look at, you know, why is poverty - why are poverty rates changing based on who's moving into or out of certain family types. But certainly the demographic edit doesn't just move people into same-sex families. It's that there's lots of different kind of changes that are happening there.

And I guess we'll move it to John for like sort of a formal discussion at this point.

Dave Waddington: Before that, just one more check on the phone to see if there's any questions on the phone.

Coordinator: We do have a question. He's been in queue for a little while. (Jim Ziliak) of Kentucky. Sir your line is open now.

Dave Waddington: Thanks (Jim).

(James Ziliak): Sorry. I pressed the button just as you were transitioning to Ashley, so this is really for John, but - so if it's too late we can postpone. If I can go back I'll do so.

Dave Waddington: No bring it on.

(James Ziliak): Okay. So first of all, you know, as one of those people nagging about the former hot-deck procedure I think these are really good changes. One thing that - so last week at the (unintelligible) meeting (Catherine Abraham) presented her joint work with (unintelligible) where they're linking the ASEC

to the (DUR) and they show really high discordance between self-employment between the ASEC and the (DUR). And so that was using data prior to the redesign. So it'd be really, I think, useful to see whether or not your new processing - but your - both the redesign as well as the hot-deck, you know, once you link up, you know, post redesign (DUR) data to see whether or not we're, you know, doing better on those scores.

The second thing is, you know, John, I'm just a little, you know, concerned about, you know, kind of the decline and income that we're seeing in TANF and SSI and the new approach, you know, how this squares away with (Bruce Meyer) and his related, you know, series of papers on underreporting at transfers.

John Czajka: So if I know the work you're talking about with (Abraham) and all, the discordance is on whether somebody reports self-employment income in the survey and self-employment income to the IRS?

(James Ziliak): Right.

John Czajka: I don't think that the - just the sort of conceptual differences in how people view the income when they report it on a survey versus how they report to the IRS, I don't think our edits are going to take change that discordance and it's something we've seen too that if - you see a lot of people who report their forms they file to the IRS which would indicate self-employment but then they say they have wage and salary earnings and it seems like that income is being reported in one place or the other. They report it as self-employment but as wage and salary to the IRS.

In terms of underreporting of program benefits, there's - we have a lot of active research in how we would use the administrative data to handle the fact

- this is a well-known fact that people who have in the administrative data are recipients of program benefits don't necessarily report them as well as we would like on the survey. And so we're working on how to handle that also in terms of we don't have data from all 50 states. We presented it PAA on that.

This - are imputation improvements would sort of take the survey responses we have and better match nonrespondents to respondents. That doesn't necessarily address the mire criticism that people...

(James Ziliak): Right.

John Czajka: ...(unintelligible) reporting those benefits, so in theory we'd like them to have increase in benefits, but if we're changing our imputation system, but we also want the right types of households to be getting benefits, households that are likely recipients rather than - even if it doesn't necessarily move us towards the aggregates that we're also trying to match because we can always match the IRS. We could just give people money in the survey, but that doesn't necessarily make our data better, so that's...

(James Ziliak): Right.

John Czajka: ...something we're definitely working on, though. It's an area we're very interested in.

(James Ziliak): Good. Thank you.

Coordinator: Thank you I have no further questions in queue.

Woman: You have time for one more quick one? As I understand it you'll be - this fall showing kind of the results of this in a bridge table. Down the line, is there

any capacity for some - probably can't do it for all the questions changes but to go back and show the effect of at least some of the survey question changes and add a procedure changes together, is there any way to reconstruct some part of it because it seems like you raised the income. It was a little hard to tell what was happening some of the time in 2014. Any way of seeing the combined effect of any of those changes?

John Czajka: So you're saying like to put together the two stages of the implementation of the redesignment. So the only way to do that would be to reprocess the 2014 data where we have the old questions and the new question, and I think right now it's something we would love to do. It's just something that if (unintelligible) the function of the resources that we have available. I don't know that we have plans right now to come back.

Woman: And the only thing I'll add to that is...

Man: (Unintelligible).

Woman :part of the process of system change was for the demographic changes and those that that data wouldn't be available in 2014, but...

Woman: I just wanted to add for all the users out there that if you use the STM variables there used to be a separate research file and now they're going to be on the after text file for the person level CPS ASEC. It's also going to have a tax unit identifier for the people who want to try to recreate that. So that's there.

Woman: Hi. I just wanted to ask Ashley, you did mention with the process changes that poverty increased among those 65 and older and among Hispanics. Can you, you know - is - can you say any more about that, about whether that's

related to the program participation or just about the way folks reported income?

Ashley Edwards: I can speak about that a little bit. So for Hispanics that increase was not consistent in both years. But for 65 and older we saw that in both the 2017 research file and the 2018 bridge file. And so when we looked at changes in their income, you know, what it came down to for the 65 and older population was the changes in the retirement income. But the - you know, going into more detail by demographic group is certainly something that we could look into.

Woman: Plus changes in how you're doing the out of pocket, the medical out of pocket too.

(Crosstalk)

Woman: So that would be an FPL. So yeah, in the FPL they - sort of both impacts. The change to the cash income and retirement and then the change and sort of the non- cash deduction. So, but again there you see poverty declined. So it's sort of this offsetting change.

Dave Waddington: All right. So let's move on to (John) here and give you an opportunity to discuss it and then we'll take our break after that.

(John Czajka): Thank you, Dave. Several years ago, shortly after the bureau completed the split panel test of the new income question, I had an opportunity with census bureau funding to look at the split panel. And one of the things we learned right away was that - oh, well we haven't changed the processing system so all the changes we made to the questions aren't necessarily in the data.

And, this is sort of a familiar note I think from the Census Bureau. (Unintelligible) spent decades trying to move to their new processing system. They've finally done that. So it's good to see you finally have gone the rest of the way and incorporated all the changes. It does mean that it's difficult to separate out exactly what changes had what impact.

For example, with respect to income there were some new questions that clearly weren't incorporated. There was a question about capital gains and we're not going to presumably use that as part of the poverty measurement. But since there was no variable that can be squeezed into, that didn't show up until the new processing system.

Some of the other changes were less clear. The new questionnaire separated questions about interest and dividends from non-retirement accounts and retirement accounts, and I think that was squeezed into the interest and dividends variables back then. But it meant you couldn't separate them. Now you've got separate variables to measure those two.

I'm not sure if the poverty measure is including interest reported in retirement accounts. I hope not, because that's not really something people put in their pockets and can use, but that's a question to be, to be resolved.

I don't have much to say about the family relationship changes except it's great that the CPF caught up to society and now we'll be able to track these changes over time, which should be pretty important. There's one thing that wasn't clear to me. A number of years ago, there used to be only one parent pointer in the CPS and if you had an unmarried couple, and they had a child, one of the parents would not be identified as the parent of that child.

And it also meant that if, when you're constructing poverty units, you'd have one parent and the child in one unit and this other unidentified parent was being a totally separate unit. It's not clear when you say that you haven't changed the family definition if you fix that problem or not. So you can get back to me with an answer when I'm done.

The changes to the income questions, especially with regard to retirement, were something that people have been pushing for, for a while. And this partly reflected, as (John) mentioned, a change in our retirement system that we've moved from a lot of, you know, entirely defined benefit pension plans to defined contributions where the way that people realize the income that goes into the 401(k) and IRAs is to make withdrawals from their savings.

And that was never considered a part of the income concept that the CPS is based on. And the fact that the bureau was willing to make changes is a significant change in the income concept that hasn't really gotten much attention. A wording change that was pretty important and something that a few of us were advocating for a number of years ago was to go from asking people about regular payments from 401(k)s and IRAs, and who knows what that means.

I mean, what you get from these accounts or withdrawals and distributions. But the term regular payments is not something that people really thought of. And so the sense was that these are really being estimated. With the changes to the questionnaire, people are being asked for distributions, everything. And so hopefully we're doing much better. I don't recall seeing big changes back when we looked at this, but it may take a while for that to work its way through the system.

The processing system changes that allow improvements in imputation and the use of the brackets - the brackets are something else that we were disappointed about back then. We made these changes and then we had to wait four or five years to actually see them. Are important because the fraction of the total income that has to be imputed in our government income surveys has been rising at a really scary rate.

Maybe it's leveled off recently, but the last time I looked at numbers something like 30% of CPS income was imputed. And it means it really matters how good you do those imputations. And the hot tech was always a real limit to this because, you know, as (John) says it's based on a big table but what actually gets done for particular income sources is a small number of variables.

You may potentially have a large table. It gets collapsed. You have no way of knowing actually how any particular person's income was imputed from that combination of variables because you don't know what variables ended up in that collapsed matrix.

The improvements look good and they have had some anomalous effects, like the increase in poverty among the elderly, which I think is due to just better imputation. If you think about - if you didn't have a great way of imputing income then you're kind of distributing the total population of donors income across these cases that are getting imputed. And the better that you can control on where people fall into this distribution, the better you're going to get.

The ranges have to be a huge help and it does show from (Adam Bee)'s work that the ranges that people report, while they're not as good as the detailed income that people give in terms of matching up with the IRS data, they are

better than what was being imputed previously. And so that is a real improvement.

I'm glad to see that this flag for the people who were total person imputes in the CPS ASEC is finally incorporated into these regular allocation flags. This is a group as (John) says was often a mystery to CPS users. It wasn't helped that this flag was put in a variable called FL_665, which – and nothing in the description is that variable explained what this really was. You had to have inside information to figure out how to take this really, you know, it's become quite big now.

Is it something like 15% of the CPS.

Man: The 20.

(John): The 20, yeah. You know, and years ago it was a few percent. And so that's important that users know what's really going on there. The top coding changes, I'm not entirely clear about those implications. I know many years ago for earnings, the bureau switched to not just giving a flat value, you know, the cutoff point for the top codes, but using the average of people and something like eight or 15 groups. And then the bureau switched to swapping, which works out better in terms of matching distributions over time.

But I don't know how far below the earned income sources this goes. I think that there are sources other than earnings where you impute mean values, but - and if that's the case than changing a top coat isn't going to...

Man: So the top coating is actually on the internal file and on the external there is a hard cap. So even with the swapping, you would still have for certain items, like \$100,000 would be the maximum that we would have on the file

internally or externally. And then you could be swapping at that value or under, we changed the top coats of earnings, it's like 1.1 million is the top code for interest and dividends it was \$100,000.

And so then if you reported \$500,000 of dividend income, we would cap it at 100. And then the swapping would happen. So this was – and that value had been fixed for decades. So it was essentially declining in real terms what the maximum amount you could have at any point on the file.

(John): Okay. Thanks. And then the last area, which I think will get addressed before, but I noticed the inconsistency between poverty going up for the elderly with the standard measure and going down with the CSPM. But I gather you'll get to that when you talk about the medical expended, out of pocket expenditures. If you think that's really where the difference is, and that's presumably better imputation. But that's something to be determined.

With regard to the income sources, where I'm not clear. You added a question about rollovers because if you just ask people about how much they withdrew from an account, they may have rolled over that into another type of retirement account. You know, it's pretty common. People move money out of 401(k)s into IRAs or others such moves. And you certainly wouldn't want to count that as income.

I didn't see a variable on the research file for the amount of the rollover and it's not clear if you subtract the amount of the rollover from what you report.

Man: We do subtract it. I think it's – so the net is what's on the public use file. We have internally as a gross distribution and rollovers as well though. And both are imputed separately too. One was responded to and the other wasn't.

- (John): Okay. That'll be something important to document to users as well as the question of whether the interest that comes from retirement accounts is being counted as income that people received in measuring poverty. And I think that's all I had.
- Man: So for that one, under the old and new processing system, they were both getting sent - they're both getting put into (unintelligible) variable. Turns on mic. But we do now have a variable which is non-retirement interest on the file. So you have the overall interest and then the non-retirement on the public use or internal only.
- Woman: I think we only have the retirement interest, which you can subtract out of the total interest. So.
- (John): I did have one other question. Dividends were something that you asking with interest and you were using the source of the interest or dividends to decide if something is interest or if it's dividends. And you have a variable on the file for dividends. It's not clear if that variable is coming from both retirement and non-retirement sources. Or if you even attempt to pull out dividends from retirement sources because it's not as simple as for the non-retirement sources.
- Man: We only do that for interest. I don't think we have a retirement account dividend. We don't have a retirement account dividend value. And there are separate questions for your interest and dividends.
- (John): But not for retirement. The question is interest or dividends.
- Man: I know that we - we don't have a separate variable for it. We don't have a separate value for retirement account dividends. Retirement interest, or value for both.

Dave Waddington: All right. Thank you. Is there - let me just do a quick check on the phone.
Is there any questions on the phone?

Woman: Thank you. I have none on the phone lines.

Dave Waddington: All right. That's good. That'll be perfect. So we're a little behind our planned break. So let's plan to come back at, let's say, what is it, 2:54. We'll still get you 15 minutes. So 2:54, for the folks on the phone, please hold on. We'll be back in 15 minutes. There are - for the folks in the room, there are bathrooms around the corner, out to the right. Or if you go out toward the cafeteria and turn left and go down the hall, there's restrooms down there and there's a little shop down there if you want to get coffee or something else. We do have

****BREAK****

Dave Waddington: ...54, we're back in the room and we're going to go ahead and get started, hopefully on the phone that you all can still hear us. We're going to start the section now here on health insurance and Edward Berchick and (Marisa Makita) are going to speak, and I'll turn it over to them. And then we'll still have about 20 minutes or so at the end for wrap up and questions as well. And of course the discussion too.

Edward Berchick: Hello, I'm Edward Berchick. I'm going to talk about some of the background from the improvements to the CPS ASEC, just to give a high level overview of some estimates using the traditional processing system – the legacy processing system and the updated processing system. And the discussion is based on joint analyses with (Heidi Jackson), who's to my right. And then after that I'll turn it over to Marisa, who'll talk about some of our exciting new

measures and some key takeaways for data users. Oatmeal raisin cookies are not the best call.

So to look at some background, as you heard throughout the day or throughout the afternoon, that does improve and is part of a two stage process. In the first stage of the process, we introduced a new questionnaire. So unlike income, for those who don't know the health insurance questions, we introduced the entire - the full sample of who received the new question in 2014 to establish a strong baseline for 2013 calendar year estimates forward.

And we also had to ensure a timely release, so missing and incomplete data were handled through the traditional legacy processing system rather than through the updated processing system. But the bulk of what I'll be talking about today is the redesign of the processing system. As you heard before the break, it debuted earlier this year. And right now we have the 2017 research file, which was released in January, the 2018 bridge file, which was released in April. And then in September we'll have the 2019 CPS ASAC and subsequent ASAC releases.

So the next two slides, I'm going to focus on the right column, which is features of the updated processing system. And the text on the left is just to give some sort of comparison to what the features that the legacy system were. So one of the updates we did is when we extract - is how we extracted information from the instrument.

So one of the motivations behind the questionnaire was over a decade of research had shown that this - done by a lot of people sitting directly across from me - had shown that the CPS ASEC had captured less coverage compared with other federal surveys. And top of that there is some concern about their - reducing respondent burden. So, as a result, the redesigned

questionnaire gave over 600 unique paths to which a person could report coverage, which is great for those two aims.

But in terms of turning it into microdata that we can use for analysis, that required extracting and reformatting those data. So we improved the way we did that. We used some annual information, coverage collected on a monthly level to improve edits. We used, and I'll talk about this more in a few slides, a lot - some households - no household member reports any of their coverage. And the way we handled that imputation for these whole unit imputes, was also substantially revised and substantially improved.

So we also changed our universe. So previously we, if you were not alive in the previous calendar year you could still be insured or uninsured. So now we fixed and you have to be alive to have coverage or not have coverage. And then another change - the last change that we're really excited about is the new questionnaire allowed us to distinguish between types of military coverage.

So before we could say you had military coverage or didn't, but now we can say if you had Tricare, if you had VA or had Champ VA. And what that's allowed us to do is to put Tricare in the category of - rather than treat it with other types of military coverage, put into the bucket of private coverage for our recode. So the other two types of military coverage are public coverage.

So as a result of these changes, as you can see, they're pretty fundamental from the ground up. So this is a version of what (Ashley) showed you a little bit ago, but for health insurance. As you can see - do I want to try it with the laser pointer? So you can see, here's where we introduced the full sample to the new questionnaire. And, I'm going to today talk about some comparisons for 2016, comparing production file and with the research file.

And as you can see here, I have an X instead of an arrow. And that's because we view this as a start of a new time series. And we do not - we urge caution in making comparisons across. This is one time series and this is another and then it's not showing up as long as I had hoped. but if you look really closely, you can see that green arrow there is a slightly different shade of green. It has stripes rather than a solid fill. It looks a lot different on my screen.

And that's to say there are a few differences between the research file and bridge file and we have a very handy, very detailed variable specific set of user notes on our website. So if you're going to compare between the research file and the bridge file, we urge you to look at that to make sure that you're making a true apples to apples comparison.

So with that in mind, I just want to compare some key estimates for 2016 and again, we have the 2017 ASEC, which uses the legacy processing system and the 2017 research file, which uses the updated processing system. And once again, don't make these estimates, or these comparisons. I'm doing this today to show you some features in the new processing system. But if you're trying to make substantive claims about health insurance coverage, don't make that comparison.

So here's the bottom line. So you can see one of the reasons I've been emphasizing this is we have a nine-tenths of a percentage point change in the overall coverage rate. As I said before, the CPS ASEC tended to have lower coverage compared with other surveys. Here you can see that coverage increased, and mechanically the uninsured rate also decreased by nine tenths of a percentage point.

You also see changes to all types of coverage. You see employer coverage, direct purchase coverage in particular I'm going to point out. There is a 4.4

percentage point decrease for that, and the other changes that you see on the screen.

So now I just gave you a previous slide that said overall coverages up and now I'm giving you a slide that says for a lot of types of coverage, the coverage estimates there are lower. So what's going on? And so I think this is helpful to look at reports of multiple coverage types. So if on the top line you can see any combination of coverage was 19.6, so this is people who had more than one type of any - more than one cover type. It decreased 6.1 percentage points to 13.5.

If you look at Medicaid overall, we saw a five tenths of a percentage point decrease. But if you look at the percentage of people who have Medicaid alone, that's actually increased by 2.2 percentage points. (Unintelligible) for direct purchase - sort of similar but opposite, which is, direct purchase coverage decreased by 4.4 percentage points. But the percentage of people with direct purchase alone changed a relatively modest three-tenths of a percentage point. And this is people were getting assigned more than one type of coverage.

So what explains these differences? So the one I want to focus on in particular right now is this change to our imputation. So previously for people for who no member of the household responded with the coverage information, imputation was done individually for each person without respect to other people in the household and without respect to other types of coverage.

But now, we group people into something called health insurance units. This group of people towards my left have developed that and worked with the Census Bureau in the early 2000s to define these. And it's - the idea is

groupings of people who are eligible to share coverage with one another. And so we impute for the entire HIU at once.

So, we're looking at people who are likely sharing coverage and thinking about them together with respect to not just one coverage at a time but together.

So this is sort of a high level view. If you want more details about what I talked about, there's this working paper that's available on the census website that goes in more detail. If that's not enough, we have I think three more working papers in the website where we go into even greater gaps about, you know, if you restrict the universes the same, how did the comparisons change? You look at which subgroups have larger changes, smaller changes. That's all there. What, what are the nuts and bolts of imputation against paper that goes into that? So I urge you if you're interested to also read, those working papers.

(Laryssa Mykyta): Okay, now that (Edward) has kind of talked about the effects of the processing changes on improving our estimates of health insurance coverage to address some of the previous critiques of the CPS measures, I'm going to discuss some of the new measures that are going to be available as a result of the coupling of the 2014 instrument redesign with our updated processing system.

So our updated processing system is going to give data users the ability to analyze health coverage, including marketplace coverage, different forms of military coverage, different types of military coverages as (Edward) mentioned, as well as kind of sub-annual insurance coverage and providing also - and current coverage. So we'll have (Richard) detail across the whole healthcare, health coverage landscape.

So in terms of market coverage, for example data users will be able to distinguish between different types of direct purchase coverage, whether the coverage was purchased on the health care marketplace or not, whether there's a premium that was subsidized or not. So that's a new level of detail that we didn't have or couldn't take advantage of under the legacy processing system.

In addition, the updated processing system will enable us to examine sub-annual coverage, but specifically whether individuals had any coverage, had private coverage or public coverage for any part of the year - for the whole year, for part of the year or not at all. And we're still evaluating the measures of sub-annual coverage to make sure that they meet our disclosure standards. As (Edward) mentioned, we also the new processing system lets us identify different - distinguish different types of military covering, separating Tricare from CHAMPVA and VA Care.

This gives users kind of more flexibility, this is consistent with other Census Bureau survey. But it also gives users flexibility to analyze these variables separately or to construct categories of coverage that fit their analysis.

And then finally, in addition, we also have type of coverage at the time of interview. We used to ask - we asked about current coverage before, but now we also can identify the type of coverage that someone has at the time of the interview - at that point in time.

So just briefly some key takeaways. With the updates to this processing system, data users can take full advantage of the richer detail that we collect in the redesigned instrument for the first time before. And this was true - we actually discussed this as well, talking about income and poverty. We really weren't taking advantage - full advantage of the instrument design until we updated the processing system.

Now we've done that so we can move forward. So changes to the processing system and this includes data extraction, which (Edward) talked about, imputation, and waiting have resulted in higher quality estimates of health insurance coverage more in line with the ACS and other Census Bureau surveys.

Second, the new health insurance coverage variables enable data users flexibility in how they want to analyze or define different types of coverage. And finally we're able to capture new measures of health insurance coverage including sub-annual estimates as well as more detail on different sorts of plans and marketplace coverage which was previously unavailable. The files produced using these data - so that's the 2017 research file, the 2018 bridge file and what will be the 2019 production file - marked the beginning of the new series, which (Edward) - the graph showed - with new measures that offer a more nuanced understanding of our current health insurance landscape.

And I just also wanted to point out that just like the SPM used to be a separate extract file, the separate extract files that we had for employment-sponsored coverage and these other types of coverage will be included in the production file. So there won't be all these separate data sets that you need to merge to do the analysis that you want to do. All that details will be on the 2019 production file. So thank you. and I'll turn it over to (Brad).

(Brad): Thanks so much. You know, I mean, thanks for inviting me to this meeting and for asking me to be a discussant here. I mean, it's super exciting and we have actually a very big change for the population in that you have that 0.9% drop in the insurance rate, so a really large change.

But I also wanted to make sure that we thanked Census for creating the research file and making it publicly available, for all the great work on comparing the research file to that production file, and for convening the meeting. And I would definitely encourage anybody that's interested in health insurance to, you know, read the papers that are out there. Certainly these are very exciting papers.

And this type of research, it falls within a long tradition of Census constantly working to improve the quality of health insurance information. It also falls within the Census tradition of partnering with researchers outside of Census. For SHADAC's part in this effort. I have to say it's exciting to see research that SHADAC has contributed to translate it into action.

And so when I was reading through those studies, there were five research studies that were referenced, by (Jackson Burchuk) here, in their papers by SHADAC authors. and these were mostly written prior to when I joined SHADAC. However, two people from a SHADAC who are authors on some of these papers are here with us today. Professor (Lynn Fluet) and then at least I thought remotely that (Kathleen Call) would be on as well.

And another contributor who's a former member of the (SHADAC) team - now a professor at the University of Maryland, (Michael Boudreaux), is also with us today. And I always think it's a good idea to keep reminding ourselves also of how important the CPS is as a source of health insurance information. So this change in the uninsurance rate, this change in other covered shots - that makes a big difference. It is routinely used to value policy options and to project the cost of proposed legislation.

It is also the only survey with publicly available state individual health information. That includes variables on health status, medical out of pocket

expenditures, dependent and policy holder information, EFI offer and eligibility and take-up rates. And now with the research file, health insurance marketplace coverage. And then as mentioned before, it's a really long health insurance series going back to 1988.

And as recently as in 2011, CPS - in the 2011 CPS haystack, the Census Bureau implemented a new imputation routine to improve the quality of health information. And according to a 2011 brief by (Dro) and (Turner) where they compared health insurance estimates in the research file and the production file, the new imputation routine resulted in a reduction in uninsurance by .5 percentage points and an increase in any private coverage by .5 percentage points.

They also, besides comparing, you know, what we're doing here is comparing the research file to the production file. They also compared it to the SHADAC enhanced CPS data file, which was a data file where they stripped out full supplemental imputations and then re-weighted it to populated totals and they compared it to estimates there too.

Although they didn't find significant differences in terms of uninsurance overall, and only small differences for any private and any public, they found large differences relative to the SHADAC-enhanced CPS when they dug deeper into some sub-populations, such as private-only coverage was higher for individuals in poverty, and any public coverage was lower for individuals in poverty. And then relative to the SHADAC-enhanced CPS, the uninsurance rate was significantly lower in two of the six states analyzed for Hispanics and significantly higher for children and for whites.

They also found that employer cover for dependence was higher in the research file. In addition the 2011 new imputation routine did not address the

issue of higher than expectance prevalence of individuals with more than one health insurance type among fully imputed cases.

So it was sort of in this sort of historical context that I was reading the papers. And one big difference here between the sort of previous research, you know, and the current we're looking at here between - the research and production file - is this whole 2014 CPS redesign. Now since this has all this additional information that they got through the redesign that they can use for cleaning and weighting and imputing. As mentioned in their papers and as (unintelligible) was sort of talking about, they have new variables related to the types of coverage they purchased and held at the time of interview and marketplace coverage. And probably more importantly, they have all the sub-annual information that they can use.

And this new information that's only available in the redesigned CPS. But, you know, the census didn't just stop there. They created – they used a new imputation routine this time using the health insurance unit. And, you know, so you have also done improving their imputation routine. And there's a great paper on there, that they - about, you know, how that changed in terms of the difference between the fully imputed and not fully imputed file. You know, where they actually, you know, after they controlled for the covariates weren't finding significant differences in health insurance estimates between the fully imputed and the not fully imputed files.

So I want to, you know, I wanted to, you know, make sure that nobody's forgetting here that we have this big change from 8.7% uninsurance in the production file and 7.9% in the research file. Any direct purchase, 16.2% of the production file and 11.8% in the research file. So on direct purchase, you know, that really, that's, you know, what seems like a good result from looking at other research, because you know there's other research that shows

that, you know, that people are - report more coverage types, you know, multiple coverage types more often than is actually true.

So, that seems like a really good result. As I said, once you, you know, take that and you look at it in terms of direct purchase only, you know, that narrows that gap and you have, I don't know, 6.5%, in the research file and whatever it was, 6.8% in the production file.

And then I wanted to go back to what you were, what these showed about Medicaid. You know that, so you found any Medicaid in the production file, 19.4% versus 18.9% of the research file. But that seems, you know, everybody here has heard of a Medicaid undercount. That's probably not a result that you wanted to see. But it is interesting that when you, you know, take that, you know, look at it in terms of Medicaid alone, that change to 15.1% and the research file in 2012 .9% in the production file.

So a lot of this reason is because you see as they had been showing that you have this any combination of coverage at 19.6% of the production file, 13.5% in the research file.

And you know, I mean a lot of this could be a result of the - using the HIU in the imputation process, but also could be the result of this additional information, particularly probably just sub-annual information that you now have available. And you know, what SHADAC is particularly interested in of course is the subgroup analysis. There was some subgroup analysis that was done in the papers by age, poverty rate levels and by expansion status, as I asked that question before.

So, you know, obviously we're very interested in state level. It does seem like you have two years of data that could be combined to do state-level estimates,

which would be super interesting. It was interesting thought that, you know, that people in poverty had higher rates of public coverage and lower rates of private coverage in the research file versus the production file. Another sort of, you know, positive result that - we've seen prior research has shown that, you know, those tended - that the CPS tended to produce estimates of lower rates of public coverage and higher rates of private coverage for people in poverty than expected.

Yeah. The, the other - the only other thing that I was going to mention here is that it would also be interesting and maybe somebody's looked at this, and this is something, you know, that we've used a lot at SHADAC is to look at like, you know, ESI coverage for dependents versus policy holders. You know, that's something that, you know, we looked at before when we did the 2011 information, that'd be great to see that in the research file versus the production file. That's all I've got.

Man: Those are great questions and there are more questions than we have time to answer. And so one thing I encourage you guys to do is download this data and answer them, save the best ones for us. That was a joke. But yeah, so yeah, I think they're fantastic questions that we have not had time to go to look at the answer. But go to our website, download the data and if you see anything that you don't quite understand or want to change, email us.

(Brad): No, I want to say that, you know, this whole meeting and everything's got at least me super excited about, you know, downloading a file and going through it.

Man: Does anybody have questions?

(Jessica): Hi Jessica (unintelligible) from HRQ. Just curious remembering all of the - the long conversations that we had about how to define marketplace coverage with our data years ago, where you ended up with your edits. Was it a complicated process or did you believe what they said, or how did you determine marketplace from the data – without going into great detail.

Man: The short answer is it's complicated. And so we had drafted at three different points in the way we edit. We look at what people say. If people report both Medicaid and marketplace coverage, we send it to a hot deck and (unintelligible) assign one or them to one type or the other type or both. Yeah, I'll have to – offline we can talk - it's a very detailed, gory answer. But we, I think we are currently evaluating it and looking at it. Yeah, I'm putting in (Amy Steinway), who spent a lot of time working on that part of the edit. And yeah, so our aggregate estimates of marketplace coverage are in line with aggregate estimates from CMS.

So, I guess I should say that too. It'll be from the aggregate level that we at least seem to have hit the right target.

(Robin Cohen): (Unintelligible). It keeps going on and off. (Robin Cohen) at CHS. How did you define an insurance unit in terms of that we've seen over time - that there's more mixtures of coverage within families and our current system kind of even encourages that. That you may have a child on CHIP, a parent on the marketplace, another parent pulling it from their employer. So how do you deal with that when you're using this health insurance unit to impute for those who haven't answered the health insurance coverage question?

Woman: Good afternoon. So everyone is looking at me, so I guess I will attempt an answer. So I think this is also a case where it's complicated. And there was some excellent work before I joined the branch, actually in collaboration with

SHADAC looking to define the health insurance unit and find likely combinations of coverage based on eligibility.

And actually (unintelligible) can actually probably speak to that better than I can. But in terms of our imputation, one of the things that we did is we constructed different health insurance units of people who are likely to share coverage. And we did that imputation simultaneously. So if we believed that someone was in that eligibility unit, we imputed all types of health insurance coverage and medical expenditure information if they were missing all health insurance information. And so we did that kind of simultaneously and matched them to health insurance units that were observed in the data.

And so in that way, we hoped to preserve the joint sort of information of people who likely shared coverage, policyholder-dependent relationships and match those across the eligibility criteria.

(Brad): Yeah, we're always looking at sort of updating the - our health insurance unit. We did it in - I think in 2016 and we're looking at it again, actually.

(Robin Cohen): That would be fantastic. And we'd love to learn more about what updates you're thinking of making, and making sure that when we are thinking about the health insurance unit about how that aligns with your current thinking. That would be great.

Woman: (Unintelligible) center on budget, I'm definitely not a health insurance expert. So I'm just curious what the type of coverage is for someone who has a premium subsidy. Are they public, private? Both?

Edward Berchick: So if they receive marketplace coverage and a subsidy. So for the purposes of our classification, we put them into the private coverage bucket, because the

insurer themselves are - it's a private insurer. But if you're unhappy with it, we can also give you the ability, the flexibility to recode it yourself. But this is in line with other surveys. So that's why we made that point at that juncture.

Woman: Have you compared those estimates with the administrative data on the marketplace enrollment with subsidy?

Edward Berchick: It's in our wishlist with future projects.

Woman: Yeah. I was just curious what kind of sub-annual variables. Are they going to be summary variables or the monthly variables.

Edward Berchick: They're going to be summary variables. So on the file there'll be five of them for no coverage, any coverage, public coverage, private coverage or Medicaid, whether you're covered, part, all or none of the year. So we're also currently evaluating the quality of the monthly data. So I presented a paper at APAM and that work's ongoing as we're sort of figuring out the way to best preserve confidentiality of respondents, because there are a lot of unique combinations of one month as well as to assess the quality of this data.

So that work is ongoing. But at this time the plan is to continue releasing this – these re-codes for those five variables.

Woman: What does the percentage that you impute insurance for?

Woman: So it differs across files. But just to give a ballpark in the research, probably leaves about 20% goes through the full health insurance imputation process. And then a smaller share will go through some item level imputation.

Woman: Thank you.

Woman: Can you just say on the part-year coverage, what the definition is.

Edward Berchick: One to 11 months of coverage. So basically one to 11.

Woman: Or 12.

Edward Berchick: Or 12, are the three categories.

Woman: Full coverage or not full. Full year coverage or not full, you don't have any indication of months.

Edward Berchick: Just one to 11 is how we - how we created that re-code.

Woman: Thank you.

Dave Waddington: So I'll remind if there's any questions on the phone, remind folks to state your name before they ask the question so you can keep up with the folks on the phones. They know who's talking as well.

(Dean Resnick): (Dean Resick), NORC. Do you provide any type of aggregate statistics on the total number of months of coverage?

Edward Berchick: We do not.

(John Czajka): (John Czajka) from Mathematica. I forget what year (Kathy Swartz) wrote her paper that was the first one to publicly report that the CPS limits of the uninsured – supposedly the uninsured through the whole year looked an awful lot like point in time estimates. Well now you've got point in time estimates,

but I didn't see any reported numbers in any of the materials. Do you happen to know what they are and do they look like say, HIS.

Edward Berchick: Yes, they do. They look like HIS and they look like ACS. We had them in the working paper. I don't remember what they are offhand. They are about three or four-tenths of a percentage point higher than the calendar-year coverage. But yeah, so I don't remember offhand.

Woman: Since 2014 I think we've annually put out estimate jointly between HIS, National Health Interview Survey and CPS. And I guess that's going to end now, but we've been doing that for the past couple of years. And it's posted on the National Center for Health Statistics website.

(Michael Carbon): (Michael Carbon), Urban Institute. I had a question about - we said earlier that fewer people were removed into supplemental poverty based on out of pocket medical expenses. Is that because you're picking up more coverage in the new processing system and you're jointly imputing health insurance coverage and out of pocket spending?

(Heidi Jackson): Hi, (Heidi Jackson) for those on the phone. So our first step of our evaluation has really been on the health insurance because health insurance status is such an important predictor of medical expenditures. Our next step is really going to be looking at actually that very question, and to see how much this is really driven by the joint imputation versus other changes in the processing system as well as demographics. So that research is ongoing and definitely stay tuned.

(Lynn Blewett): (Lynn Blewett), University of Minnesota. Has Your HIU work included the demographic information that we've heard today about same-sex households?

(Heidi Jackson): Yes. yeah, it has. The demographics used to construct the health insurance units are all based on the updated demographic.

(Lynn Blewett): Okay.

Man: There's something in the survey or the new variables about some of these employment offers status in the prior year. It's all based around the day of the survey, right? Okay.

Edward Berchick: The answer to the question was yes. Any questions from anybody on the phone.

Coordinator: At this time no, I have no questions.

(John): All right. Is there any other discussion items or questions that people would like to bring up at this point? We are at our last 30 minutes of time, so we have a little bit more time if there's other questions related.

Man: Yeah, I have a more general question about - I have a more general question about imputation in the CPS, and that's, is there any thoughts about multiple imputation and if not releasing multiple imputation, perhaps having a report that would describe what added variants we would get if we did do that.

(John): I have some research on the income side on that. This is (John) (unintelligible), Census Bureau, where we look at - sort of actually related to some work that (Jim Ziliak) did. Look at using administrative records and income and multiple imputation. It's definitely in the research space right now. I think it's something like 20% of variance, on average like 20% variance increases when you sort of put all the things together.

But there's a lot of sources of uncertainty when we're doing model - regression model-based multiple imputation with - accounting for model-based uncertainty also. But that's sort of where the research is really on the income side. I don't know for health insurance.

Woman: That would definitely be on my research wish list as well. But right now there are no plans with multiple imputation.

(Brett Fried): This is (Brett Fried) with SHADAC. Can you just give us some hints for the differences between the bridge file and the 2017 file.

(Brad): Let's say you're thinking of that 2018 bridge file versus - so some of the - in terms of estimates or in terms of why we're urging some caution.

(Brett Fried): Maybe why you're urging some caution.

(Brad): So as I mentioned earlier, we changed the universe such that infants who are not born at the time of interview cannot receive coverage. And we discovered while evaluating the research file that some infants were inadvertently still left as getting coverage. So that's one element.

Some of the variables also changed in their interpretation. So the calendar year recode variable was additionally - for no coverage was conditional and having no coverage in December were you uninsured for none, part or all of the year. And so we moved that to be in line with the other, for the annual recodes.

Children, there is a chance for children with IHS coverage. I don't remember what that change was offhand. Household coverage for coverage with infants. Inadvertently the - if you had an infant born during the calendar year who did

not have coverage for being out of universe, your household could not be fully insured because your infant was not insured.

So we corrected that with the 2018 bridge file. It's those types of small changes that for many large aggregate analysis you should be fine. But if you want to start looking at children or IHS coverage, that's where that - those differences are really going to come to bear.

Woman: A follow up for that. Did you - for your point in time estimate, did you include the instance in that and just not in the full year?

(Brad): Correct. So if you were alive at the time of survey by everybody here, you were including in that measure.

Woman: On the reputation for unmarried partners, are they in the same health insurance unit or separate?

Woman: I would have to look at our editing and get back to you on that. Offhand, I think that they're separate but I'd have to verify. I will definitely look at that and get back to you.

(Brett Fried): This is (Brett Fried) at SHADAC again. Did you leave on the point in time? Are you going to do the whole coverage distribution or the just the insured on insured.

(Brad): So for current coverage point in time? So it's, you have all types including subsidized marketplace. So anything we provide in terms of coverage for the calendar year, we will also provide for coverage at the time of interview. And so it a slightly different point in time than the NHIS, because it's February, March or beginning of April weighted to March. Whereas NHIS is the first

quarter of the year and the ACS is current coverage at the time of your interview across the entire year. So it's (unintelligible).

(Trudi Renwick): I'd like to put out a question to the group, and this is (Trudi Renwick) from the Census Bureau. So in September when we throw up these numbers and our estimates for 2017 have changed from what we released last September because of the new processing system, what suggestions do you have for how we - what else we can do to prepare people for this so that it's not a shock. I can't remember the date, somebody help me, September 12 I think. Yeah. So it's not a shock on September 12. And what ideas you have for how we can manage this in September.

Woman: So we deal every year with the census, which we really appreciate, a web conference with the states who use this data all the time. And so that, you know, as much as that - and we've done it for many years. So that's very helpful to have people who can ask and there are people in the weeds. It's not always just policy. It's people who work with their own survey data and this survey data and they've explained over years how their estimate differs from the Census Bureau's estimate and all that.

Well, why is it increasing or decreasing? And then we'll get the reporter calls. So it's much, you know, talking points kind of information both at the high level and the detail level would be really helpful at, you know, for distribution and for us when we would get calls for - I remember one time looking at (John) (unintelligible) might remember when Tommy Thompson claimed credit for a decrease in the uninsured when they first added that Quebec, you know, confirmation question on the health insurance coverage.

Like did you indeed have, do you indeed have health insurance? And it dropped the uninsured rate by a significant amount and he claimed credit for,

you know, the policy. So you, but you need to, you need to talk to those people too. But yeah.

Woman: (Unintelligible), Center on Budget. What to expect from the data, you know, are they embargoed or not embargoed based on the bridge funnel, might be – just a short piece on what the previous year's data show for reporters.

(Suzanne McCartney): This is (Susanne McCartney) at HHS. (Ashley), am I right in thinking that, you know, poverty will go up or down based on the 2019 production file and the 2018 production file regardless of what we saw with the bridge file. Right? Well poverty is going to go up or down or stay the same and officially we're only going to compare, you know, what comes out in the fall with what came out last fall. No, not true.

Ashley Edwards: So we'll be looking at the bridge file. So the, you know, using the new processing system and payer to the 2019 file.

(Suzanne McCartney): So you're going to do that statistical comparison?

Ashley Edwards: Yes. Our estimates - that's what (unintelligible) is saying. So the estimates for 2017 will be different because these are going to be estimates that use the new processing system. So yeah, those time trends will be consistently based on the updated processing system. We have done research that looks at comparing time trends for the research file and the bridge file compared to the 2017 and 2018 production. And while we see like - so for overall there is no difference, you know, from 2017 to 2018. Both the legacy and the updated processing system showed a decline in overall poverty.

So, you know, feel pretty confident in that, you know, the story that we're getting across those years, difference in different, based on the files that we're

using were not statistically different across demographic groups. So, you know, I think in terms of the - certainly we need to make that clear to people. But in terms of interpretation, you know, don't expect that to be a huge part of the narrative.

Woman: Thank you. We'll try to keep that straight and make it clear at HHS. What to take credit for and what not to.

Ashley Edwards: We will be releasing a full set of detailed tables for 2017 with the bridge file. Those detailed tables will come out before September. Soon, maybe in a month. I don't want to promise a specific date but fairly - in the near future we'll be putting out for health and income, I'm sorry for poverty and income, a full new set of tables. We'll keep the old tables up as well, but there'll be just as in 2014 income year 2013, 2014 we had two sets of tables. We'll have two sets of tables for income year 2017.

Woman: Any written products to go with them.

Ashley Edwards: So we've got what we have on the website now and we've got those comparison tables on the website now. So you can look at those on the website.

(John): In the paper that - there's a paper that covered the - or the processing changes both for income and poverty. We'll talk about, like the income paper that's going to go up. We'll have an addition about the year to year changes in the production versus the new, the bridge and research file.

Woman: You're talking about the report.

(John): No, so the - to talk about the - like file about the edit, the processing changes. So for income we're going to have that, in the paper which will sort of cover it. It won't cover it in, you know, the level of detail or detailed table covered ...

Woman: We have another conference session, that's JFM in end of July in Denver. Kind of doing the same papers, but those may be - will hopefully be the final versions of those papers, that will have both 2017, 2018 and time trends.

(Brett Fried): Hi, this is (Brett Fried) from SHADAC. We do a brief where we just - we compare all the different estimates from, you know, health insurance estimates from all the different surveys, both trend and then comparing single years. You know, I would pitch for us to do like maybe the series with both estimates in there than for maybe 2017 and 2018 or something like that.

Dave Waddington: Waiting for more lights to come on. Is there anybody on the phone that has anything they'd like to contribute here as we start to wrap up?

Coordinator: We do have one request. So (Martha Haeberlin), ma'am, your line is now open.

(Martha Heberlein): Hi, thank you. This is (Martha) from Mac Pac and I just want to echo what (Lynn) was saying. I think what would be helpful for our – like, as the end user, personally, would be some sort of summation document on why you guys made these process changes. I think, you know, it makes sense on a very analytical and methodological focus, but something that's quick and easier for people to understand I think would be really helpful to sort of know that, you know, these were made for a reason.

You know, we have these survey changes and then we have these process changes and sort of this is why we did it. And more, you know, bullet, short format would be very helpful.

Woman: I'll just say it thank you for that. Historically we've done some sort of pre-release blog post to try to prime people for what's going to be released. And I think a lot of the summary information and detail about the research we've done would certainly be something that we want to try to get to a more popular audience. So that is something that we're, you know, hoping to do.

Woman: And I can echo that as well for health insurance - we're planning to have a kind of easily digestible blog post.

Dave Waddington: And I'll just toss out one more thing in health insurance related. As I think probably most of you are aware or should be aware by now is that the ACS data release this year will be delayed as part of the government shut down. The release of most of the ACS data will be, I think it's a two week delay. We will still have some health insurance data with the release, but it won't be the same level of health insurance.

So for the folks that are looking for health insurance data, all of the ACS data that we would normally put out in that report on our release day with the CPS ASEC won't be there. We're going to have state-level indicators. And then we'll have the national trend, and about two weeks later when the ACS comes out, we'll have further information at the lower levels for health insurance.

Woman: I think the release day for the ACS is September 26 and health insurance is also planning to put out a brief, you know, that includes a lot of the same information that would have been in the CPS report, using the ACS.

Dave Waddington: Hopefully it's been clear today. We're very excited about this release that's coming out. We're very excited about this. I think what a pretty big, a very big accomplishment in getting this new processing system finished, completed. A lot of work went into that and it's been - since we've come up it's been four or five years.

Maybe out of order in thanking, but I want to recognize the work of the survey team, (Lisa) and her team and (Lisa) next to her over there, the folks in our demographic systems division who worked very, very diligently with us in Seaside on updating the edits in the programs and not just the people here on these programs but anybody who touches CPS in any way, in Seaside put a hand in these specifications and has been reviewing this stuff for the past four and a half years and it's been a really a long ongoing process and they've done a wonderful job in getting us to where we are here.

We're excited about September. Special thanks again for our presenters here today. You guys did a great job and (Jon) and (Bret) too for coming. I appreciate your comments here today too. It was really good to have that extra input and thanks for everybody on the phone and everybody here in the room, especially for coming out too. And listening in on the phone. We're glad to have you here. Again, this - we'll take this recording and put it up on the website at some point in the near future.

So if you have other folks who weren't able to come, or if you want to hear what you said or if we need to listen back to (Brett) and (John) to see what your comments were, I'm sure we'll do that too. So thank you very much for coming out today and I - there's a few more donuts and cookies over there, some water. Help yourself to that. And if you have any other questions for individuals, we'll be here for a little bit longer. We're happy to answer other questions. So thank you very much.

Coordinator: As the conference has concluded, you may please go ahead and disconnect again. Thank you very much for your participation. Have a great day.

END