THE BROOKINGS INSTITUTION

FALK AUDITORIUM

MODERNIZING THE CURRENT POPULATION SURVEY
CO-HOSTED BY THE AMERICAN ENTERPRISE INSTITUTE

Washington, D.C.

Thursday, October 29, 2026

This is an automated transcript that has been minimally reviewed. Please check against the recording for accuracy. If you find any significant errors of substance, please let us know at events@brookings.edu

WELCOMING AND INTRODUCTION:

JOHN SABELHAUS
Senior Fellow, Economic Studies, Urban-Brookings Tax Policy Center, Retirement Security Project, The Brookings Institution

PRESENTATIONS AND Q&A: OVERVIEW OF CPS MODERNIZATION:

MODERATOR: JOHN SABELHAUS
Senior Fellow, Economic Studies, Urban-Brookings Tax Policy Center, Retirement Security Project, The Brookings Institution

KYRA LINSE
Survey Director, Current Population Survey and American Time Use Survey Team, U.S. Census Bureau

NICHOLAS JOHNSON
Chief of the Division of Labor Force Statistics, U.S. Bureau of Labor Statistics

PANEL: EXPERIENCES WITH CONVERSION TO INTERNET SELF-RESPONSE (ISR):

MODERATOR: DAVID S. JOHNSON
Visiting Fellow, Economic Studies, The Brookings Institution

JOELLE ABRAMOWITZ
Associate Research Scientist, Survey Research Center, University of Michigan, Co-Director, Michigan Federal Statistical Research Data Center

TOM EVANS
Assistant Deputy Director, Labour Market Division, Surveys and Economic Statistics, Office for National Statistics (UK)

WILBERT VAN DER KLAAUW
Economic Research Advisor, Federal Reserve Bank of New York

PANEL: WHAT ELSE DO WE NEED FROM A REDESIGNED CPS?:

MODERATOR: MICHAEL R. STRAIN
Paul F. Oreffice Chair in Political Economy and Director of Economic Policy Studies, American Enterprise Institute

KATHARINE ABRAHAM
Distinguished University Professor of Economics, University of Maryland

DIANE WHITMORE SCHANZENBACH
Nonresident Senior Fellow, Economic Studies, The Hamilton Project, The Brookings Institution, Professor and McCourt Chair, McCourt School of Public Policy at Georgetown University

ALEX MARTIN RICHMOND
Head of Labor and Workforce Economics, Economic Research, OpenAI

* * * * *

SABELHAUS: So welcome everyone. My name is John Sabelhaus. I'm a senior fellow here at the Brookings Institution. And along with my co-organizers, David Johnson and Michael Strain, I'm really happy to welcome you to this event on CPS modernization. So in addition to thanking all the participants who are- will be joining us today, I, I wanna especially thank the Alfred P.

Sloan Foundation for their generous support of this event and all of the under activities under our Economic Indicators Initiative over the past year. So it's really fabulous to see such a great turnout for an event like this. Modernizing the CPS is clearly a priority issue for the US statistical agencies and for all of us who rely on the data for our research and policy analysis.

So many, or maybe most of us in this room, have relied on the CPS for some aspect of our work, everything from monthly labor force statistics to building long time series on income distribution or health insurance coverage. And while the agencies are embracing new ways to measure economic activity using big data, administrative data sources, the CPS is a good example of why we also need to talk to respondents.

There are a lot of things that you can only get from a survey. We, we can measure what people are doing, but we often need to know why they're doing it, and that means we need to keep the surveys like the CPS up and running. So as we'll hear, the CPS is facing new sorts of challenges, but as we'll also hear, the agencies are working together and developing new strategies to deal with those and and prepare the survey for the next generation of economic statistics.

So we have three great sessions lined up. Our first session I'll be moderating. We have Nick Johnson from BLS and Kyra Linse from Census to, to give you an overview of what the, what they've been doing in terms of developing ISR. I'm gonna moderate, and I am going to be strongly encouraging... I have a few questions, but I really wanna encourage you to be thinking about your questions as they're giving their presentations, and we can call on you right away.

So the last point of business we'll then have two more sessions. David Johnson will lead a second session on experiences of other organizations with with ISR in their surveys. And Mike Strain will lead a third session on what else might we want from a, a redesigned CPS. After the three sessions, we will have a reception, and we hope you can join us, so please stick around.

So with that, I welcome Kyra to the stage to provide a Census perspective on CPS modernization.

LINSE: Good afternoon, everyone. I'm really glad to be here to discuss all the hard work s- the staff at both BLS and Census are doing to modernize the CPS. My name is Kyra Linse. I am the survey director for CPS at the Census Bureau, and I'll be starting the discussion with reviewing the steps we've taken up to now in our effort to modernize.

The current population survey, as I think everybody in this room probably knows, is the national-- is the nation's primary monthly source of labor force statistics jointly sponsored between the Census Bureau and the Bureau of Labor Statistics. It provides the key measures we rely on, like unemployment rate, labor force participation rate, and the unemployment population ratio.

Each month, about sixty thousand eligible households are interviewed. The first interview is usually conducted in person with follow-up interviews typically done over the phone. Depending on the household, the survey takes about ten to thirty minutes to complete. Data collection for the CPS occurs every t-- in a-- over a ten-day window every month.

The households are interviewed for four months, rested for eight months, and then return for another four months. This design allows us to reduce variance in both over-the-month and over-the-year changes while minimizing the respondent burden. So we're facing three major pressures. First, response rates are declining steadily.

The chart on this slide shows that the trend is clearly falling from the low ninety percent range in 2010 to the upper sixty percent range today. This decline mirrors many of the other federal statistics surveys as well. Second, as the response rate falls, the cost of collecting high quality data continues to rise.

More effort is required to reach the households and complete interviews, which increases the overall costs of our field collection. And finally, these factors together put pressure on the reliability of the estimates. Lower response rates and higher level of non-response can affect data quality and can make it more challenging to produce those timely, accurate labor force indicators that policymakers and researchers like you count on.

So

CPS modernization is a high priority as we work to counter those declining response rates and those raising data collection costs. Our initial focus has been to develop an Internet self-response option, or ISR. The goal here is to add flexibility for the respondents by supplementing rather than replacing our existing in-person and telephone interviewing.

We plan to continue prioritizing in-person interviewing in the month one, since the first contact is important to legitimizing the survey and to build a rapport with respondents. After that, web-eligible households will be offered the opportunity to respond by ISR-ISR, giving them a convenient alternative to help reduce follow-up burden.

Integrating ISR into the survey at, that operates on a strict ten-day collection window does have extreme operational challenges. We need to be carefully manage the flow of cases between the response modes, but overall, introducing ISR is an important step toward a modern, resilient ARC CPS, one that expands on respondent choice and helps maintain the data, data quality in a challenging environment.

In addition to CPS Basic, we have the Annual Social and Economic Supplement, also known as ASEC. It is a key part of the CPS program. Completed every year in February, March, and April, it is the main source on census reporting of poverty, income, and health insurance coverage estimates. As CPS is the base of this survey, it has the same issues as the main CPS survey is facing with lower response rates and decline reli-reliability and granularity of the estimates being reported.

If CPS is to have a web mode, then ASEC must have one as well. As part of modernization, Census has been reviewing and developing and testing the ASEC questionnaire to using the same methods as CPS. Due to the length of the survey and the questions that are more complex in reporting, such as income and the amounts of money paid towards items such as healthcare, their testing needs to consider ways to reduce burden in their design to combat break-off

This slide lays out the overall timeline for the CPS modernization effort, beginning with the work that started in 2022. From 2022 to 2024, the focus was on the core development and testing. That included Census building a new enterprise data collection system, converting the CPS instrument to that platform, and conducting extensive content review through expert reviews, crowdsourced testing, and cognitive testing.

This phase was foundational, ensuring that both the underlying systems and the survey content was ready for more advanced testing in the public. In 2025, we ran the first limited field tests, and this allowed us to test the data collection approach and produce the first full sets of outputs from the new system and use those results to refine both the instrument and the operational procedures.

We'll discuss the field testing more in d- detail in a couple minutes. This year, we evaluated what we learned from the 2025 field test and launched a second field test. The goal of the 2026 field test was to fine-tune the ISR contact strategies and logistics of integrating web responses into CPS operation.

Beginning in 2027, we plan to introduce the next field test, focusing on the ASEC supplement and beginning on an experimental parallel sample, and Nick will talk about that more in his next presentation. Overall, this timeline reflects a very careful, phased-in, slow

approach, moving from development to major testing to validation, so the modernizing of CPS protects the data quality.

So the '25 field test had one main goal

is it was for a proof of concept. Can we collect the data by self-response on web quickly and of quality while supporting the field operation? In our research and review of the other surveys, there weren't surveys that interviewed household as many times as CPS or had to get responses in such a short interview period.

So there were many unknowns on whether this would work for CPS. We focused on how to get the respondents to sign on to a website and answer the survey, and for them to do it quickly and consistently every month. Within the goal of assuming we could, we would be successful, we also ran some experiments on types of contexts, incentives, questionnaire warning, et cetera, to see what works best in the contents of a, context of ISR.

So we ran the test from May to September of 2025. We used a sample of eight thousand cases that were interviewed. A household could be interviewed up to four times, just like production. In addition, we chose about eight thousand cases from the current production to invite to a web response for a ninth interview.

This was not to test response rates, but just an attempt to collect more data from those who have already likely to respond and see what they would do by web. Based on the initial development testing, we decided the best procedures to keep overall response rate high and to legitimize the survey would be to reach out by our CAPI personal interview again in the first month.

In the following months, the cases in the ISR would be available for the first three or four days, and then they would be moved to a non-response follow-up by CAPI for the last seven days. No, not all cases were eligible for ISR. We need to have enough information for the respondent to answer correctly for dependent interviewing, a key to CPS.

We also needed to be able to protect their data from anyone who might want to sign on that wasn't them. So to be eligible for ISR, they had to have completed the previous year interview and provided good names as well as a cell number and/or an email to verify when they sign in

So

So what were the results? In the end, about 70% of cases were actually eligible for ISR. Of those, about 15% of cases answered in the first three days before we had an interview start, interviewer start contacting the household. By the end of the data collection window, the web response averaged about a 21% response rate.

We found of those that did respond, they were choosing web more often than CAPI, about 57% to 42%. But all together, in the end, about 14% of the entire sample answered by web each month. While we didn't have many other surveys to compare to, this exceeded our expectations and was enough of success to deem the web a viable mode of collection to move forward if looking at the efficiency of data collection alone.

It also showed promise that there were many other options that we could do in, to improve these response rates in the future.

While we were concerned about response rates, we also needed to consider representativeness. One of the hopes was that we would be able to reach households that are typical non-responders or people who are high effort to reach in a CAPI environment. So we reviewed the respondents to web to see what type of households or respondents tended to answer by web more.

We found that large households, those likely-- more likely to be married, and households with higher incomes, as well as respondents that were more educated, answered online at a higher rate than CAPI. These are not people who are typical non-responders, so more research will need to be done to see what we could to encourage a broader range of households to respond by web.

But no matter what, the good news is these respondents will now no longer need a CAPI interview, and hopefully that allows them to give our interviewers more time to answer the harder-to-get respondents and do more interviews with them.

As I said earlier, we have some secondary evaluations that are linked to encouraging response and specifically a web response. For the first time, CPS test- tested offering incentives to encourage the sponsor to answer by web within the first days of data collection to avoid the need for CAPI follow-up.

This was the form of a digital gift card emailed to them immediately after completing the survey. In the first month, a respondent was eligible by web, we saw 3% more answered by web with an incentive than those who didn't. Of note though, it didn't improve web responses as they did more interviews, showing that interview fatigue is still an issue even with an incentive.

This is promising enough that more research is being done. We also tested sending an email to invite the respondent to web in connection with their letters, and while it did push respondents to be more likely to answer by web, it did not affect the overall response rate. Much of the data is still being reviewed from the 2025 test, and results have not been yet approved for public dissemination.

So far, we have found that there are no significant issues as to quality as of today, and we're focusing more on the data evaluation in the 2026 test

Now let's talk about the 2026 test that we literally just completed. It was sampled out of 10,000 cases. We followed the same formula as the last test, where we contacted them the first month in person and followed up in the next months with web and CAPI. Some cases of the, from the first field test were interviewed again to replicate the yearly interview cycle of CPS.

For those returning cases, if the case did not complete by web in the past, they were invent-invited to answer by web, but also contacted by CAPI on the very first day. We did not give them a few days to answer by ISR first like other months. This month is too important to establish legitimacy and rapport and a and extensive quality is needed to continue.

So we didn't want to limit interviewing time. We are interested if the interviewer had a positive effect to push the web as well. For other months, we attempted targeting ISR cases more for an easier to manage workload on the CAPI side. The result from this test will help us better model what kind of cases need to be sent to ISR and what kind of cases should go right to CAPI for the most efficient use of our 10-day window.

In addition to testing these longitudinal cases for the first time, we'll be testing some updated questions on earnings, relationship, as well as texting to invite respondents to the web, and we will also be continuing our research on incentives. The goal of this test is to increase the web response, review the data quality more closely, test the possible question changes, and hopefully revi- refine to final procedures.

As mentioned, the ASEC is also in development and very much following the same path as CPS Basic. They've used many of the tools available to test the questions and design of the survey in the web mode for usability, clarity, and accuracy. They are on the same development schedule as shown earlier, just one year behind, and they are wrapping up development and initial testing stage

For 2027, AZEC is ready to begin its first field test. This will be the first opportunity for a test in a more realistic CPS environment. And while the previous test was looking to test pushing to web, AZEC is more focused on making web collection the most efficient and accurate in terms of AZEC needs.

Knowing they are a longer survey, they are attempting to streamline the questions for easier and faster responses. Many of the concepts can be difficult for the respondent to report, like income sources and the amount paid for certain items. AZEC is testing how to make it simpler to report with more specific questions, and as always, find ways to see if there's new or more specific questions improve that reporting and the quality

So the A-- the test will consist of two sets of cases. The first set will be internet only, and the goal of this is not to test the web uptake, but more on the changes to the questions for self-reporting and see what the results are. These will include cases that have been previously respond, have, that have preasively res- excuse me, previously responded and are retired test or production cases.

The second set of q-cases will get ASEC in the second month to replicate how ASEC will be experienced for most CPS respondents. They'll get new cases interviewed with just basic CPS the first month, and then at the second month, they'll receive the ASEC content, and then ha- receive follow-up by CAPI to look at the overall response in this new mode.

All cases, regardless of mode eligibility, will be randomly assigned a control versus test version of the questionnaire. The test is currently scheduled for early twenty seventy-seven to best replicate when they, a respondent would normally report on ASEC

This is the webpage that is about to be updated and includes all of the items I covered here as well as a lot of the research. It's being updated on November 10th along with a webinar that Census and BLS is hosting. Please feel free to reach out to our contact email as well for any thoughts or questions.

And now I will turn it over to Nick to go over BLS's efforts and future planning. Thank you.

N. JOHNSON: Thank you, Kyra. My name is Nick Johnson. I am her counterpart at the BLS. I oversee-- I'm the program manager for the CPS. So Kyra laid out what we've done so far and, and what's in the near future. I'm going to try to walk through everything until we get to the finish line for getting the ISR into official production.

So to do this, I'm gonna walk through the limitations of the previous field tests, the ones that Kyra just discussed. We accomplished a lot with those field tests, but there are limitations in, in answering all of the questions. I'm gonna go over what we know and what we don't know about CPS ISR so far, talk about our plan to evaluate before transition, and then although the majority of what we're discussing is the ISR, at the very end, I, I will talk about our other plans for modernization beyond ISR.

So first, the limitations. So again, we learned a lot from the, the, the first field test and are in the progress-- process of learning a lot from the second field test, but there are limitations you need to know. So the first is that the resources per case were very limited, far below what we would use in production.

So we're seeing lower response rate. There was a lot less follow-up. We really, our goal was to get as many people in the ISR as possible to, to meet all the objectives that Kyra discussed. That, that was the priority. So lower resources, we had to use a different reference week, a shortened collection period in order to fit the collection into the CPS collection schedule.

So these are all sacrifices we had to make. And relatively modest sample sizes, at least by CPS standards, especially if you account for the, the lower response and the, the fairly short running time on both of these. So what have we learned? So as I said, we, we've gotten a lot out of this very encouraging.

The instrument performs well, okay? That was, as Kyra said, that was kind of our first focus. The second is respondents showed a willingness to report online. If you take at the-- take a look at the chart up here this is from the second field test, so the one we just completed. If you looked at the first field test, it's essentially the same.

This is the response of respondents who reported in the first month, who were then web eligible for months two and three, and how they reported. And as you can see, forty-two percent came in by ISR, thirty-one percent by CAPI, and the remaining twenty-seven percent were non-respondents. So we're, we're definitely seeing that willingness to report, so that's a great sign.

We made progress in collection strategies and field logistics. There were a tremendous number of experiments that were embedded in the first field test and, and continued through into the second field test. This is one of those things that'll appear both on the what we learned and what we need to learn more about slide because we will continue to look at this really even after production.

This is the kind of thing you, you-- we'll continue to focus on. And then finally, we identified some questionnaire changes that were needed and, and ruled out other changes. Very important, none of these field tests that have been conducted so far are mode effect tests. We, we can't pull that out of them.

Because of the sacrifices and the limitations I talked about in the previous slide, that, that-- we don't have that right now. But what we were able to do is for the ISR questionnaire, we were able to essentially A/B test this with two alternatives of the questionnaire, one that we call core and one that we call alternative.

The core is very close to our existing questionnaire. It was the minimum amount of changes we could possibly make. And then the alternative was, at least by CPS standards, a little bit bolder take on some of the, the questions as if it was a little bit closer to something that was web native. In the end, we did not see much of a difference at all between those two questionnaires.

So that doesn't speak to the mode effects of what the survey will look like with an ISR necessarily, but it did give us a little bit of confidence that the, the difference between the questions, there's not so much sensitivity. But there's a lot of detail in there to dig into. What we still need to know.

So the overall impact to response, w-we can't tell you what the response rate will be going forward. The hope, as Kyra said, is to s- make, make sure we can stop that decline, and if, if we're lucky and, and we do this the right way, maybe we can even start to move response the other direction. But we don't know that from the field test because they didn't have the same resources, a shortened collection period.

That's something we just don't know. How to optimize collection logistics. Again, this is the one I told you you'd see on both, both slides, that that's something that we continue to work on. Because it was a shortened collection period, we were using existing sta- Census Bureau was using existing staff.

Not all of the logistics got, got a full run, but we did improve the things that we saw from the first field test. And then this is kind of the big one. You know, this is the one that could be highlighted or in bold or something, is the impact of mode effects. As I said, the field tests won't tell us that.

So that's, that's really what we need to figure out. Before we put this into production and we're, we're calculating the unemployment rate, I think everyone in this room wants to have some idea of, of, you know, what impact that will have, and, and we're, we're part of that group too, of course. So what is our path to deploying the ISR?

Well, we wanna do it safely. Obviously, that- that's, that's at the top of our list. And so our goal is to set up a parallel survey that al- allows us and you as data users to see what, what the estimates would look like with the ISR incorporated. And that reduced risk really on, in two different dimensions.

The, the first one is that, you know, if, if we're as realistic as possible, so we try to take as many of those limitations from the field test away, we'll have the opportunity to test everything from start to finish all the way up to estimates to, to know that the process works. So that's the first risk we're looking to mitigate.

And then the, the second one is the mode effect risk, right? We don't wanna put out estimates when we can't tell you how they changed because of introducing a new mode. And this, the idea of parallel testing is not unprecedented. That's exactly what we did when

we did the 1994 redesign. The design of that parallel m-maybe is a little bit different than what we're planning now, but it's the s- the same basic idea, and it helps to give data users some confidence in what they're seeing when it comes, comes out.

And then our recent internat- you know, our conversations with our, our colleagues internationally, I know we're gonna hear from ONS a little bit later today that's certainly part of, part of where our discussions have been, has shown that as they move to, to ISR collection tools, they saw mode effects initially and were able to start working through those with questionnaire changes.

'Cause we need the questionnaire to, to explain to the respondent as much as the interviewer does now Okay. There's a lot of, a lot on this slide. Bear with me as I, I walk through it a little bit. But, but this is our, our phase-in of, of web collection. So if we look at this, this red box first on the left side, this is what we consider the evaluation phase.

So that top blue banner, important part here is the official CPS continues going on just the way that it is now. No change there. All of the official estimates, the employment situation is, is unchanged. While we're doing that, that's when we're going to conduct parallel testing. So in that evaluation phase box at, at the bottom, if you look at, at number one, a limited parallel test, this would be phase one.

And this is something we're already working on and planning to kick off next summer, and it would be in six to eight states, and we would conduct a high-quality field test where we knock out a lot of those limitations that I talked about before and, and we'll walk in-- walk through some details in a minute, where we really have an opportunity to produce estimates and start to see mode effects.

This will be challenging at this kind of limited six to eight state level, but we do think that we maybe will be able to see some of the larger mode effects, and that'll give us an opportunity to work through changes to the questionnaire before we go forward to the, the next step. And this is about a nine-month process.

So then the, the second step, which I have in this, the, the blue box there, the number two, that has dashed lines around it because right now this is just a proposal. So our proposal is to expand that limited test nationwide, and that would give us an opportunity over an eighteen to twenty-four-month period to not only produce estimates t- but to make estimates available to the public so you can see the difference between our official estimates and those that come from a, kind of a new CPS that has the ISR.

During this time, there will still be changes that we imagine we'll be making to mode effects, and so there will be changes over time as we move towards that production date.

So once this gets the, the, the go ahead, we kind of move here to the box number three, which is when we incorporate the ISR into production. So the, the official CPS, the, the sample will continue and we will phase in ISR over time So what are the goals of that nationwide parallel test? Well, a-as I've said a few times, it, it's really largely around mode effects.

As a, a larger sample nationwide that allows us a high quality test that allows us to produce those estimates will help us and you identify mode effects. Now, the key is what we mitigate.

A-and this is going to be the real challenge. As we see mode effects, we will have to determine which ones are ones that we think need to be mitigated and work through the questionnaire to do that.

And then there will be other mode effects that we identify that, that we decide we don't want to, that we may be getting something more out of the ISR that we're not currently getting. So that's an oversimplification of how difficult that process will be. That's why that parallel testing phase is, you know, a nine-month limited followed by eighteen to twenty-four you know, f-full test.

These estimates and this ex-- these experimental estimates will also serve to provide a bridge between the existing survey and the future for those mode effects that we either decide not to mitigate or are unable to mitigate. This te-- nationwide test also allows us to further refine the collection logistics ahead of production as we get as realistic as possible.

Okay, so I started off this presentation talking about all the limits of field testing. So I'm gonna try to end it towards the other direction here. So the, the goal of the parallel testing is to be as realistic as possible. So we're using... The goal would be to use identical sampling procedures, comparable amount of resource collect- resources per case which is really the key one.

We will see what this looks like in real life. And Kyra mentioned something earlier about how, so far in ISR, we're seeing respondents that are, are, are already our good respondents are the one that are, ones that are showing up on ISR. That's not necessarily a bad thing. Every, every case that reports on ISR allows us to have resources to then repurpose for those tougher to reach cases.

And, and I can tell you right now, it, it is a challenge. I mean, the reason ISR is our priority is in, in order to the, you know, the CP-- collection of CPS is very expensive. We have a large sample. It's getting more and more difficult to knock on someone's door or call them and have them answer and respond.

So having more resources to get to those tough cases is really a key part of this. So we really win either way. If we see ISR cases are ones that we don't normally get to, fantastic. If we see cases that we normally get to, great, we can then repurpose for other cases. We'll be able to use the same reference week and collection period because there will be dedicated staff available for the parallel test.

That's a very key one because if we're gonna produce estimates, we wanna make sure we're measuring the same time period. And then the last one is, is our biggest challenge. So as Kyra said before, sample rotation in CPS is household is in for four months, they're out for eight months, and then they come back in for four months.

On our timeline, that's, that's gonna be a difficult thing to replicate, so we're going to have to have a different rotation schedule. So that's something we will need to adjust for.

So now that we get to my final slide, this, this is the first one beyond ISR. So a-as I've said, the, the focus has been on ISR 'cause that's really critical to the sustainability of the survey, that we get to a successful ISR as soon as possible. But we have a lot of other goals with modernization. It's, it's not simply that.

So most of these other projects are things that we have, while we've maybe taken a look at them, we have not given them the full effort that they will eventually get. The first one is to research the use of alternative data. Important to know, of course, we already take the, the CPS data and extend it for local area unemployment statistics using alternative data, b-both administrative and other survey data.

So that's not new to us. We do, however, wanna take an additional look at it for the CPS overall. But importantly, we see that mostly, right now at least, as something that would replace or help out non-response weighting adjustments, things along those lines, as opposed to, like, a direct variable replacement.

But there's plenty of research to do that's ahead of us. We wanna consider additional improvements to collection. We're not stopping at ISR. We want to get the process as efficient as possible to collect the most cases of high-quality data. That will continue. And then down the road, preparation for a larger redesign to modernize the questionnaire.

You know, the, the survey hasn't been redesigned since 1994. There have been only a couple very small changes to the basic questionnaire. So that's something on our radar to finish up with. So that's all I have. Thank you very much.

AUDIENCE QUESTION: Kari is going to join us a- also?

SABELHAUS: Great So as promised, we have lots of time to to chat with Kyra and Nick. I will seed with a few questions, but how are we doing at raising hands to get, to get a mic? Okay. So I'll, I'll keep my eye out after I ask you a couple basics. So I sort of wanna go back to where you started with a dec- a chart of declining response rates-

that we've all seen many, many times. And overall it seems the impression I'm getting is that ISR is part of the toolkit for- Mm ... for combating that. And I just, I wanna try and summarize what I think I learned, which is that 57% of people, y- y- you get to the door, you knock on the door 57% of the people who say they're gonna participate chose ISR and 42% chose CAPI.

Right. Did you find out anything or did the interviewers ... Is there any metadata that might tell us something about why people ... I mean, you gave us an indicator that the, the households are larger, married, higher income, educated. I would sum that up saying they're busy. ... And I, my reluctance to answer some surveys, never government surveys-

But would be based on being busy. I- is, is that the, the sense that you got from the interviewers, that they, they just wanted to be able to do it on their own time, or is there something else more subtle going on?

LINSE: We did do a focus debriefing with our interviewers after the first test and even as part of development during the, the before to see where our interviewers struggle in the, in the interview.

We've been asked by interviewers for years really even long before we thought about developing a t- of adding an ISR mode. I... the respondents are just flat out telling them, "Can't I just do this online?"

SABELHAUS: Right.

LINSE: So- Which

SABELHAUS: is probably what I

LINSE: would say. So- Right ... so I think there's a real interest of, as you said, all qualitatively, but you know, interested in doing it when they have the time, not talking to somebody, right, in person or at their house, you know, privacy, that kind of thing.

SABELHAUS: Right. And y- it wasn't on your list, but was there an age component to the ISR versus CAPI choice?

LINSE: I would have to re-look, but yes, we did measure age. I don't remember anything being that significant. Okay.

SABELHAUS: That'd be an interesting one to know also. I would, I would imagine, for example it, it would fit with your, your profile, but parents of young children I think would have especially tough times, but, but may be happy to go online after the kids are asleep.

N. JOHNSON: I, I, I will say for the age thing something we did in the field test is our researchers added some, some questions at the end of the survey to ask about burden, and the, the one group where we s- we saw a little bit more interest in from people who completed web, which means they would've also talked to an interviewer the first month, the older age group was the only area where we saw some interest in some preference for an interviewer, while the other age groups were kind of strongly in web was great, you know.

Right. Re- really worked. So-

SABELHAUS: So, so- ...

N. JOHNSON: somewhere, so.

SABELHAUS: And, and that again, consistent with my priors- Yeah ... and makes a lot of sense. And again, I, I'm I really think ISR is the future and this is the results we're seeing here are very, very promising. But I guess I'd, I'd love to flip it around and say why might some people be reluctant to agree to ISR?

So for example, if I have to give you my email address, is that a limiting factor? And, and I heard the word text in your, in your word. Mm-hmm. So is, are there some people who are... they're, they'd be happy to give you a phone number or maybe not happy- ... but they'd be willing to give you a phone number but not- Yeah

An email address?

LINSE: Yeah. I think actually we see both. You know, it varies depending on the type of person. But that is one of the key factors in why people aren't eligible for ISR is they don't want to provide a ph- a cell phone number or a email 'cause they have privacy concerns.

SABELHAUS: And I guess as, as you were speaking, I was trying to think of some way around that, but I guess I, I can't.

So- No.

LINSE: We have very strong privacy laws- Right ... at the Census Bureau we're very proud of, and so that is- Right ... one of the ways we verify that we're talking to the right person and that nobody else can get to their data.

SABELHAUS: Right. So, makes a lot of sense. The the other aspect of the sort of overall survey response is the incentive question.

And so, is there anything Nick, if you wanna take the first stab at this one, anything about... So when I think about incentives and incentive experiments, I guess there's different ways to do it. Mm-hmm. You can have a small you know, token appreciation. People can be entered into a lottery for a bigger prize.

Is there do you have sorta any insights that you learned about what might be the appropriate both level and design for incentives? I'll hand it to Kara.

LINSE: It's all right. So the Census Bureau has done quite a bit of research across different surveys, and we have discovered that there's different motivators depending on the population you're trying to reach.

We did... I didn't mention it here while we did provide a post we also did attempt a pre in the first field test, and that was not a driver in any way and not successful. So that was interesting. But I think it varies greatly, and a lot of people are playing around with things and items and that kind of things.

But those can get as costly as money very quickly with mailing and all of those other logistics that the... of handling those items. And so it seems that a lot of cash or a cash equivalent is still king.

SABELHAUS: Okay. So, so... That's good. And, and full disclosure, I worked for many years on the Survey of Consumer Finances and cash was king then- -as well.

So, I have one other sort of set of questions. In your presentation today, you talked a little bit about... or I guess Kara mentioned it, but it also came up with you. Some types of questions are harder to answer when we think about we think about particular ASEC type questions. And I think one of you mentioned dollars are harder to collect.

I would think that in ISR, that would be something if people are provided with ranges that it would be pretty straightforward to do. I, I was a little bit surprised when I heard that.

LINSE: I think it's more about the a- ... So for ASEC, it's very important to try and get an exact number for a lot of the things that we're, we report on especially when you're measuring income and, and government subsidy incomes and those kind of things.

And people will not necessarily have the records or know those things off the top of their heads. And so it's it's difficult to encourage them to go get their records if there's no interviewer there at the house to say, "Can you take a moment and go get your records?" Right. Right. And so you wanna try to avoid ranges.

But yes, so we can... You know, those are the things that can be backups.

SABELHAUS: Right. Well, anyway- We have a... Well, go

N. JOHNSON: ahead. Oh, yeah, I was going to say, before field testing, we, we did a lot of series of cognitive tests and focus groups with the interviewers. A-and particularly the latter in focus groups with the interviewers, we focused on what questions are difficult for you to ask as they're currently in the script, which ones require follow-ups.

And so, a lot of the, the things that we have tested in the field test are, are driven by these are questions that they thought were, were difficult to answer. And so something in the, the second field test which we didn't fully explore in the first field test are earnings And, and I think that that'll be very interesting to see if on the internet where you do have a little bit of time that you could maybe pull up a pay stub or, or ask somebody as opposed to something just over the phone- Yeah

where you're, you're pressured for time. So I, I think importantly with mode effects is that we, we can see them go both ways. You know, mode effects aren't necessarily a bad thing. We, we may see some good come out of this. So it- Right ... it's something to continue to explore.

SABELHAUS: So I did not hold up my end of the bargain.

We're down to about two minutes for questions, but please let me, Go ahead, Andrew. And please introduce yourself.

AUDIENCE QUESTION: A- Andrew Rimmer, George Washington University. Can you give us some context about the kinds of folks who are increasingly refusing to answer the CPS? You had... You started off with the falling response rate, but demogra- demographic characteristics, who's less likely or more likely to

respond?

LINSE: Your typically more transient type o- of people, younger minorities, et cetera we have a little bit of a higher non-response rate at. We also have an attrition issue, obviously, the 484. We-- I think we're a little different from some other surveys in that we probably have a broader range of the typical non-responder just because attrition is, is a lar- becoming a larger problem than it used to be.

So, and that's much more of a broad population.

SABELHAUS: Michael Strain.

AUDIENCE QUESTION: Thank you. I had a question about why with ISR you can't keep the four eight four structure?

N. JOHNSON: Sure. Let me clarify. That's, that's not ... Once we go to ISR, we will continue four eight four. So the, the structure and, of the survey is not changing. Simply during the testing phase, the four eight four for the first panel takes us 16 months to get all the way through.

So when we're doing the parallel test, we just don't have enough time to get enough panels in and get to the sample we need. So that, that is only during testing.

SABELHAUS: We have time for one quick one. Catherine

Can we turn her mic on? It's

UNKNOWN: on.

SABELHAUS: Here you go

AUDIENCE QUESTION: One thing I was... Almost every month on the CPS has a supplement. It's not just the ASEC.

LINSE: Correct.

AUDIENCE QUESTION: And I have been wondering what the plan is for those other supplements.

LINSE: Yeah. So we've let ourse- our sponsors, for those other, for those of you in the room a lot of other federal agencies use CPS as a way to do a larger sample size and get their information that they need.

And so we do supplements almost every single month. And we've been working with the sponsors, letting them know that this is coming and working out what kind of testing that they would like to do as part of this, this process.

SABELHAUS: Great.

So join me in thanking the panel. And we'll welcome David Johnson to moderate the second group

D. JOHNSON: Okay, thank you. Thank you all for coming. So we've heard a lot about what BLS and Census are doing. This panel, we're gonna have three experts talk about their experience with three other surveys that use ISR. In fact, some have converted from CADI to ISR. So we'll first hear from Joelle Abramowitz.

She's from the University of Michigan and the Institute for Social Research, and she's an investigator on the Panel Study of Income Dynamics, a survey that's near and dear to my

heart. And so she'll talk about their experience with ISR and the PSID. Then we'll hear from Tom Evans, who's been thirteen years at the Office of National-- the Office for National Statistics in the UK, so he came a long way to be with us.

And they're working on a transformed labor force survey because many of you have heard that they were in the news a lot in terms of their labor force survey. And then finally, we'll hear from Wilbert Van Der Klaauw, who's an economic research advisor and the director of the Center for Microeconomic Data at the Federal Reserve Bank of New York.

And he's been working with the Survey of Consumer Expectations actually since its inception, so he'll talk about their experience with ISR. So I'll let... So we'll, we'll go through each, have about eight-minute presentation, then we'll have, again, I'll have some discussion questions, and hopefully have time for some questions from you.

So think of your questions. Joelle.

ABRAMOWITZ: Thanks, David, and thanks for including me on the panel today. So as David mentioned, I'll be talking about the PSID's transition to web that took place in the twenty twenty-one survey wave of the PSID. It is a biannual longitudinal survey. It includes over nine thousand families, nearly twenty-five thousand individuals as of twenty twenty-one.

Prior to twenty twenty-one, the primary mode of data collection was computer-assisted telephone interviewing, CATI. In twenty twenty-one, the PSID introduced the self-administered web mode, and about forty percent of interviews were completed in the web mode. The survey randomly assigned families to either a web first or CADI first protocol, 40% and 60%, in order to be able to assess these mode effects.

And I should mention, though, that that web first sample excluded recontact, split-offs, institutionalized or proxy interviews. So these are the, the continuing core respondents. In the next wave in 2023, a larger share of the sample was assigned to web first resulting in about 73% completing via web So here I show some response rates by mode.

And what we can see, this is in the randomized sample. When we look at the share completing the interview, we can see that it is high and similar across both modes. We're seeing ninety-two percent for the CADI first protocol versus ninety-one percent for the web first protocol. When we look at the completion mode, right, to see how many individuals switched from their initial assigned mode to the alternative mode, we can see that eighty-six percent of those that were assigned to CADI completed in CADI, and seventy-two percent that were assigned to web completed in web.

So we do see some switching in both directions but the majority completed in the mode that they were assigned to

So we've done some analysis to look at the effect on close-ended responses, and overall, we see only small effects by mode. So exploiting this random assignment we can look to see effects on item non-response and the question responses. We see that the effects are concentrated in certain topic areas for example, expenditures and philanthropy.

We did identify modest social desirability bias in the CADI mode for certain topics on embarrassing or socially undesirable behavior. So in those cases, we're seeing more response on the web mode. Focusing on the employment section, we see that mean item non-response rates were a little bit higher for web, four point nine percent versus four point three percent for CADI.

Of the one hundred and eighty-six employment variables, eighteen showed statistically significantly different non-response by mode. So that's about nine point seven percent of the employment variables. I've done some work looking at the effects on open-ended questions in the PSID, focusing on those for industry and occupation.

So I'm again exploiting the random assignment to the web mode in twenty twenty-one, and I look at the effect on the non-response rate, the response length in words, and trying to understand how this affects the quality of the assigned industry and occupation codes that we're collecting these open-ended responses for.

So just a little bit of background on these open-ended questions. Respondents are asked to describe all of the work for money they and their spouse have done since January first of the prior wave year, including current main jobs, which I'll call CMJs, and other jobs, non-CMJs. And so they're asked what kind of business or its industry is that job in, and your work for your employer, what is your occupation, what sort of work do you do, what are your most important activities or duties?

And the reason for collecting these open-ended responses is to assign census occupation and industry codes. The CADI interviewers are directed to probe for clear, complete answers and specifics to be able to distinguish among occupations. So converting this to the web version, each of these three occupation questions appeared individually and on separate screens.

So we find increases in item non-response rates and decreases in response length associated with the med- with the web mode. So focusing on item non-response rates, when we look at the current main job occupation, we see one point four percent for CADI, so really very small, and we see a substantial increase for web of four point three percent.

We see similar rates for non-current main job industry and occupation, slightly smaller for current main job industry. When we look at response length conditional on responding at all we do see a, a substantial drop. So for current main job occupation, for CADI, we were seeing almost twenty-four words, and we see a drop to fourteen words.

Now, I should say my understanding is that for many federal statistical surveys, the response length is much shorter to begin with. So, so we are seeing a large drop, but it's still somewhat of a, of a longer response. Similar for industry, those tended to be shorter even under CADI, but we see a drop again from almost seventeen words to about six words via the web.

And for the non-current main job responses, they tend to be much shorter. We're not really seeing much difference there. So the main question that we would ask for these shorter response lengths is, is that affecting the quality of code assignment? And so we look at three different dimensions to try to assess this.

First, we look to see, are more general codes being assigned? And the answer is no, we do not see significant differences in the more general codes. We do look to see how many narratives ultimately remain uncoded. That includes the non-response as well as sh-shorter responses that might not be codable.

And similar to the non-response rates we're seeing large gaps there. And we, then we looked to see in two different ways, effects on the variation in code assignment. So we looked to see for codes that were coded by multiple coders did we see more disagreement among the web responses or the CADI responses?

And we do not see significant differences there. We also used a machine learning model to code industry and occupation, and then we compared the match rates for web and CADI, and we looked at the probability scores of the the top codes. And again, we're not seeing significant differences there either.

So taking this together, the findings suggest that the web mode produces fewer codable narratives, but comparable mode assignment conditional on being codable. So I will stop there.

D. JOHNSON: Thank you.

Thank you.

EVANS: Great. Thank you very much, and good afternoon, everyone, and thank you very much for the invitation to come and speak to you all. So yeah, I'm Tom, and I'm gonna talk to you about the UK experience of transforming our labor force survey. Sorry to hear that some of the news of the trials and tribulations of the LFS have made, made their way across the Atlantic, but hopefully some useful reflections of everyone in the room.

I'm gonna kind of skip straight to the transformation and not, not labor too much on the background of the LFS. I'm happy to touch on that if it's helpful later on, but we'd be here for more than eight minutes if I go that far back. So the intention for the TLFS is to replace our traditional labor force survey with what we refer to as a mixed mode, but online first TLFS.

And I think that's perhaps a, a crucial difference to, to the situation you guys have here with the CPS. So this is perhaps a, perhaps like a more fundamental redesign. We've been dual running the LFS and TLFS since early twenty twenty-two, so it's been quite a journey with quite a lot of development to the TLFS since, since early twenty twenty-two.

And, and through that time as you would imagine, we've encountered a few challenges some things to to talk to you about today which include but do are not limited to how we're addressing things like the survey burden, where again we have perhaps a slightly different starting position, which I'll touch on but the impact that that has on data completeness.

Talk a bit about complex variables, which is very much linked to what Joel has just spoken to you all about. A bit about understanding differences, which is obviously a, a kind of key challenge in replacing one survey or adding a mode is, is how do we explain to our users, policymakers what, what the differences that we are, are seeing really kind of mean, where are they coming from.

And if I have time, I'll also just touch a little bit on how, kind of how we set ourselves up at the ONS to kind of work through this challenge and, and a little about how that has evolved over time as well. So if I start off with the sort of survey burden point. So I don't think anyone in the room will really need telling this, but unsurprisingly, kind of moving to a self-completion online first mode means that we do suffer from higher levels of item missingness in, in, in our data.

So we see that respondents are more likely to potentially skip questions 'cause it's a bit easier if there is not somebody sat in your in your house with you or potentially to drop out the survey entirely if they kind of lose patience and just think, "Oh, no, this is, this isn't for me." So we've taken a lot of action to try and kind of improve this.

Now, just for context this is where I said we're probably starting from slightly different positions. The current Labor Force Survey, so this is the survey that's been, been running for decades in the UK takes roughly f- on average around forty minutes for its first wave. So it covers an awful lot of topics beyond just the labor market content and has some probably somewhat sort of older sequences of questions that I think you, if you were designing a survey now as we are, you, you wouldn't start off that way.

As the most recent kind of published response rate is around thirty-three point six percent. So again, operating in slightly different territory. But I will say that has improved quite significantly since around two, three years ago when you were perhaps most likely to have seen some of our challenges in the news on the LFS.

So moving into the TLFS world prior to kind of July last year we were around kind of thirty minutes at wave one. So this is off the back of sort of a decent amount of survey development since twenty twenty two, but perhaps not too much development that was really targeted at shortening the survey or reducing burden.

The development to that point was more about kind of clarifying questions and things like that. Since then, we've kind of made two perhaps significant design changes alongside a, a whole a whole host of other design changes around the edges. So in July 2025, we actually adopted a two-survey model on the TLFS.

So rather than having one survey that goes to the entire sample, which is kind of consistent across the board, we now have a two-survey model, which we refer to as a core survey and a plus survey. The core survey is effectively a lot shorter. So we have tried to strip it back to the kind of core labor market questions that address policymakers' needs, whilst the plus survey has remained somewhat longer and captures some of perhaps those broader social topics.

Then in April this year, so we implemented well, one, one kind of development and then made one change. So the, the change is fairly straightforward to say that we just increased the incentive for the survey, so that went from ten pounds to twenty pounds. And then the development was to implement data rotation.

So this is about rotating forward certain elements of the survey between the collection waves, so that if we're returning to the same occupants in the same household they do not have to completely fill out the household grid again, which particularly for larger households can take a certain amount of time.

We also rotate forward certain sort of demographic variables, but then continue to re-ask the important things, the time-varying things. So where's this ended up? So again as, as of around April this year so we're now down to about seventeen minutes at wave one and somewhere between nine and twelve minutes for the subsequent waves on the TLFS.

The subsequent waves improved much more because of that data rotation point. And, and over the piece, we're seeing somewhat improved response rates with the current LFS. But I should say that this isn't this hasn't been a silver bullet. We do s- still have kind of missing this and, and that thirty-nine point nine percent that I quote on the slide will be inclusive of some cases that perhaps haven't made it all the way through the survey.

So, we are exploring imputation strategies to help try and address missing this where it is a more more sort of clear issue and likewise weighting strategies as well. Complex variables. So this is where we have some challenges with classifying respondents who are in employment to the correct industries and occupations, exactly the, the problem that Joelle was describing.

So I won't redescribe it just to s- but just to say some of the solutions that we are exploring, so some in the collection space and some in the processing space. In the collection space we are-- have just concluded a f- sort of a fairly large field test of a search as you type function. So rather than having those open text box questions having something that a respondent, as we all do for various forms that we fill out online you start typing your occupation or where you work, and it starts bringing you up suggestions.

So we're looking to implement something like that. There's also been some exploration, early exploration, I should say, of potential AI solutions. Those range from enhancing a search as you type function just to kind of help it produce more more likely matches to what the respondent is intending, through to potential use of kind of LLMs where they're sort of almost sort of substituting for an interviewer and asking probing follow-up questions.

But those are, I think it's fair to say, further off in the future for us. They are contingent on a whole range of different sort of testing and, and thinking about the risks of, of including those things in the survey as well. In processing space, we are making iterative improvements to the, like the matching algorithms that we use to code the free text responses that we currently get, and we expect they will continue to be a fe- a feature.

Even if we implement some of these other solutions, there will probably always be some cases who end up with a free text box, and so we'll, we'll be making sure that those improvements kind of go all the way through. And we've just started looking also at admin data linkage. So some colleagues at the ONS recently published some research on admin data linkage in the, in the weighting kind of space.

But with that linkage in place, we're also just looking to see if if we're able to make use of that in terms of coding industry, because once we get through to tax records, that becomes easier to link to to business business records which contain industry as well. I would touch briefly on our approach to then quality assurance.

So this is this is perhaps where I spend most of my time in the, in my, in my day job. So three broad components of things that we have focused on in, in the UK. So the first of these

is, is, is assuring the questionnaire. So we've had perhaps a bit more redesign of the questionnaire, not wholesale redesign, but more redesign of the questionnaire.

So it's really understanding about where the questionnaire has changed and making sure that that actually has im-been implemented correctly through the collection, but also the processing of the questionnaire. Assuring the estimates themselves. So we have, have ongoing production of the estimates within the ONS.

We do not publish them at the moment, but a lot of thinking about when is the right time to do that. But internally, we compare obviously between the two surveys that we're running, but also third sources where we have them to help us understand where, where we think the kind of strengths and limitations of the data are.

And all with a view to working towards that kind of crucial explanation that we know we will need with our stakeholders when we move towards a transition into live production. And having touched on stakeholders a lot of engagement with stakeholders. So we run a kind of working level technical group with sort of four or five of the key kind of government departments in the UK who have an interest in this data.

We run a stakeholder advisory panel, which includes more academics. And then my teams run deep dives with kind of working level analysts across those two audiences as well. And all of this feeds into our kind of measures of success, which are our kind of self-stated aims that we need to meet in order to transition to the TLFS on the point that has been raised particularly by Nick on the kind of understanding changes particularly on the questionnaire and in a sense mode changes we did have a spell where we were kind of fairly repeatedly saying if only we could get the same respondents to complete both surveys, wouldn't that be very useful?

And eventually in twenty twenty-five, we did actually manage to conduct what we refer to as a crossover survey. So this is where the final wave five LFS respondents, and we run five quarterly waves on the LFS were then invited to complete a one-off TLFS. And this, this has been useful.

Again, we've not yet published much in the way of findings from this but it has internally given us s- more insights than we would have had certainly on how the mode and the questionnaire can influence responses. Things as simple as reordering questionnaire responses can have quite profound effects.

And this has really improved the evidence base that we have to explain ourselves. And we'll be looking to make sure we incorporate this a lot when we get to that kind of public facing part of, of our program of towards transition. I would say as a point, something we didn't do but if this is something that you were considering doing, if, if you are able to consider a qualitative follow-up to this, then I would certainly recommend that because we still do have a few why questions.

So this tells us perhaps where the differences are coming from, but not always why. And then finally just a bit on organizational setup, which I appreciate may not be of interest to absolutely everybody. But, but I think is important. And I say that coming from a place where we're doing this across just the one organization let alone two.

So I would explain briefly that in this world in the ONS, I sit sort of in where these labor market teams are towards the right-hand side. So in the outputs and the analysis space is, is where I focus most of my time and my, my teams focus their efforts are really on understanding the data.

And if I just take that central labor market analysis and evaluation box I can easily think of all of the different things that we do on a day-to-day basis, linking up with every part of the end-to-end of the transformation. So requesting changes to the survey over time speaking to collection colleagues, processing colleagues.

I won't go through every single arrow, but you get the point that, that this has definitely been an important part of making, making that work. I think that was all of my reflections, so happy to leave it there. Thanks, thanks for your attention, and we'll move on.

D. JOHNSON: Thank you.

VAN DER KLAAUW: Well, thanks very much for for having me here and giving me the opportunity to talk a little bit about the Survey of Consumer Expectations. So this is an internet survey from the beginning. It's been in the field since two thousand thirteen. And so this is more about some reflections about actually running a survey for a long time on the internet.

And first to, you know, start off with the usual disclaimer, these are my views, not necessarily those of the Federal Reserve or New York Fed. And also, we're still in FOMC blackout, so I cannot talk about monetary policy. Not that I was going to. So, so there's been this huge rise of internet survey panels over the last two decades.

It provides this additional timeliness and flexibility. Some of the biggest panels that have been running are for, at USC, the Understanding America Survey, Rand has the AOP, and then Pew, NORC, Ipsos, SSRS. There're now, you know, growing number of internet panels out there where, you know, you know, the, the, a lot of value coming out of those surveys and a lot of experiences.

And of course, during the pandemic also, the Census Bureau introduced the Household Pulse Survey. And other annual conferences that are really are to me, like, fantastic because people share their experiences, talk about all the recruitment challenges that are faced, representativeness, biases ways to, you know, reduce retention.

And so for example, the Cypher conference that's held actually, you know, a couple of blocks away from here every year is is full of people with experience, you know, showing re-results from experiments where they're trying to improve response rates and representativeness, and it's really, I can recommend people going there.

Now, a big advantage of online surveys is that you can get very quick feedback, which is very useful for measuring things that you don't typically get in a traditional survey or in administrative data. In particular, people's beliefs and perceptions, people's expectations and how they experience shocks, you know, things that they didn't expect.

And now there's this growing literature on using subjective expectations in economic research. There's a Handbook of Economic Expectations, and there are these internet panels now run by the New York Fed, but also Bank of Canada, Bundesbank, and the ECB is running this in eleven countries. So yeah, the SCE at the New York Fed, as I mentioned, it started in two thousand thirteen.

It's a monthly survey. It's actually fielded for us by Nielsen IQ, although we developed the questionnaire. That was preceded by a five-year experimental phase, you know, really a design phase where we we did cognitive interviewing. We did a lot of randomizations, including the incentives to see, you know, how people responded to incentives.

So based on that, you know, we started fielding in, in two thou-two thousand thirteen. It's nationally representative. It's a rotating p- panel. People come in and stay for twelve months, and then they rotate out. It's about thirteen hundred households every month, but it's very stable because of this rotating nature, n-nature, which is very important.

Initial recruitment is like the sur- like the CPS is doing. It starts off still from address space sampling and contacting people, in our case, by, by mail. There's a monthly core survey that ex- collects expectations at the household level, but also about the economy, about inflation, unemployment, wage growth.

There's a lot of labor market questions, but also income spending. And then a lot of the questions are probabilistic in nature. So we a- ask a- about, about people's expectations in the probabilistic way. So what's the, the percent chance that this would happen or that would happen? And I'll show you some examples.

And then we have triannual modules that rotate in and out so on credit access labor and housing markets, and spending and public policy changes. The incentives are fixed. It's about fifteen dollars per survey, and all the microdata is available online. So this, this is an example of a question we ask about earnings growth on the job.

So, you know, how d- how much do you expect your earnings to increase or, or decrease over the next twelve months? What is the percent chance that you will lose your job in the next twelve months? Then we also ask about quits and job finding expectations. And then every four months, we also ask about probabilities of, of all the transitions in the labor market also ask about job offers, retirements current labor market status, earnings, job change, job search behavior, and also ask for reservation wages and job offer realization.

So we collect a lot of data coming in at the monthly level. Now, what do we get? This is just an example. If you look at earnings growth expectations on the job, and you look over people's life cycle, you see this declining pattern, which is also there in the CPS. Now, the CPS measure is not earnings growth on the job.

It's just earnings growth. So that includes increases or in your earnings by job switches. So they're slightly different measures, but you got- you get the same patterns over the life cycle in the SCE, we also look at uncertainty pe-- that people have about their earnings growth, right? They we ask about the whole distribution, and that also is declining over the life cycle.

But also, as you see in the twenty-fifth and seventy-fifth percentile, there's a lot of heterogeneity, a lot of differences across people in these expectations. This is when you

look over time, and you see that earnings growth expectations were pretty stable up to the pandemic, and then they, they dipped.

And after that, we've seen a, a steady increase in earnings growth expectations, just as you also see that in actual realizations. Another measure we developed with the survey is a measure of la-labor market tightness. So in our measure, it's the mean expected probability of finding a job if you were to, to lose one, minus the expected probability of sorry, of finding a job minus the probability of losing a job.

And we compare that to the unemployment rate, so the red line is the invert-inverted unemployment rate, and the blue line is the Conference Board Labor Market Differential. So that's the proportion of people saying jobs are plentiful versus hard to get. And you see that you find very similar patterns in these measures.

And the SE measure, because it's forward-looking, it actually leads the other two as well a little bit. So, so this is just to, to show that, you know, the, the data you can get from the, from the on, on the internet, you know, you get monthly data that's, that, that looks very reasonable.

Now, what are some of the advantages that we've experienced and that people are telling us they, they like? Well, of course, first for us, it's cost-effective to do it to... And especially when you run a panel once people get into a panel, they tend to stay. And the incentives matter a lot there. They, they help a lot.

But people, you know... And if you change the topics, as I show-showed you, we do some of the modules are changing every month people really like that so people stay in. So a lot of the work is recruiting people to get into the panel. Once they're there, there is some attrition, but it's, it's it's not very high.

Convenience. Of course, people can fill it out whenever they want. They can do it over the phone, on their laptop or desktop. The, you know, phones are becoming more popular. And for us, of course, the data comes in directly. You can use it almost immediately. We can change the questions very quickly.

So for example, during the pandemic, you know, to really understand what's going on, you know, a lot of people are on furlough, and if you don't typically, you know, elicit that kind of information, you know, it's something we can put in very quickly. Also during the pandemic because people can access the internet wherever they are, we didn't have a big drop-off, not as big a drop-off as in other surveys where people moved.

A lot of people moved during the pandemic. And when the address changes, it's sometimes very hard to find them again. So again, that's an advantage. And then the visual tools that you can provide. So when you ask about, for example, a probability, and you show them a scale, a scale from zero to 100 and explain what it is with an example that's very helpful to people.

And then, of course, there are-- we discussed already the potential for interviewer biasing effect and that's also reduced by having, you know, a similar format for everyone. If you want to emphasize certain things in the question, you can do it with a bold font or with a color but at least it's consistent across interviewers.

Okay. And then what are some of the drawbacks? Well, of course, not everyone uses the internet, although now smartphones are, I think eight... It was ninety-one percent last year. It's now h- even higher. So smartphones are there, and then of course, yeah, people can do it on their, on their laptop or desktop as well.

Like other surveys, we have experienced drops in response rates especially among young and lower-income participants. From what we are seeing so far you know, the, it looks like they've stabilized. They, they're not falling any, any further. And the evidence that's out there shows that there's not really a strong relationship between a decline in response rate and a worsening quality.

There doesn't seem to be any implications for the quality of the data that's collected. Of course, another thing that you have to worry about on the, on the, on a web survey or on a phone survey especially that it fits on the screen. So there are these things that come in that, you know, only with testing you realize some of the questions don't map very well to a, a, a screen, so you have to make some changes.

And then you know, another risk is that if you don't have the in- the, the interviewer helping out in answering questions, there are more entry errors, more outliers. But of course, you could also then program that in, that if somebody gives a very extreme answer, you could follow that up with a, a clarification question.

And another risk we already talked a little bit about are, are AI chatbots people straight lining or people looking up answers on the internet. And then, you know, you can have a debate of that, whether that's good or, or bad. But you know, there are a lot of... there's a lot of effort in, in trying to mitigate these problems.

So you can look at how long people take in the survey and see if there's any big changes in that. You could build in attention checks. You can look at mouse movements. You can even see if people leave the page to go somewhere else. We also look for rounding and typically, you know, people round in the survey when you ask for probabilities.

And so if you start seeing very precise answer, that might be an indication that there's an AI chatbot there. But then, of course, also like the CPS, it's still very important how you recruit people. So if you recruit them by mail or in person, you know these people actually exist. So that really reduces a lot the chances that you have chatbots completing your survey.

Last one comment I have is about the risk of changing internet modes. So of course, we used internet mode from the beginning. Transitioning has a lot of risk, and there are a lot of surveys out there at Health, HRS, PSID, the Consumer Confident Survey, just to mention a few of them. They all experience breaks in series.

Now, typically, most series are very stable. You don't really see anything. But in most of these surveys, there's like one or two questions where you see a clear break. And so, another, and exam- another example is the Michigan Survey of Consumers. They really recently switched from phone to web. Suddenly the proportion of Democrats to people affiliated with the Democratic Party jumped up a lot.

And that's not a... You know, you don't see that in other, like in the Gallup or in our survey, and you don't see that anywhere else. So there's suddenly a composition change. And

another thing that you have to be very careful with is that when you start, when you switch mode, there co- could be a change in the distribution of tenure, of how long people have been in the panel, especially if you have a four eight four.

You know, you have to check on how many, you know, times people have already completed this. You start now putting in a lot of new people for whom it's the first time they respond, and a lot of questions have, you know, the panel conditioning effect. And especially in our survey, you see that when you ask people about inflation expectation.

The first time you ask about it is, and you get more extreme answers than the second time you answer it. So if your composition changes in the sample because you're switching modes, then you, you can, you know, lead to spurious changes that are not, not real. So let me stop there.

D. JOHNSON: Thanks. So I'm gonna build a couple questions off of the previous presentation. So response rates is a big deal, but what we learned here, which is more troubling, is the item non-response rate actually increases. So Joelle showed it for occupation. Now, PSID has low item non-response, and I don't know what the CPS item non-response rate, but moving up a few percentage points or that PSID is like a tripling.

If it tripled in the CPS, that's a problem. And then you showed that, you didn't really talk about it, but do you-- have you looked at how to mitigate this? Are there ways to change the survey questions, or how do we deal with this item non-response increase? Anybody?

Okay. You

EVANS: want to- Happy to make a start.

Yeah, I mean, I th-- as, as I sort of touched on on the slides, I think for us the step one in the UK context was, okay, let's-- we've gotta make this shorter. And I think particularly the concerns around drop, survey drop-offs, so people dropping out being linked, obviously linked to the number of questions that a respondent faces, and the fact that we ask a lot more questions of employed people because we want to know lots of details about their employment versus somebody who is unemployed or inactive.

So im-imputation then once we've sort of brought the survey down as much as we think we can, how do we, how do we mitigate? So im-imputation is one area where we're looking at essentially sort of versions of like nearest neighbor imputation which is a significant improvement on the LFS, where we just have roll forward, so we take the response from a previous a previous sort of wave and bring them forward, which is not always appropriate.

And then weighting. So we, we can identify certain points through the questionnaire where drop-off is perhaps more likely and say, "Okay, well, from this point we will apply a different weight to the survey. We'll recalculate the weights." Now that will present some challenges for our users that we'll need to sort of help them with when, when we go live in terms of how they kind of change their use of the data compared with the LFS.

But that, that's the sort of things that we're looking into.

ABRAMOWITZ: I would

D. JOHNSON: s- Yeah,

ABRAMOWITZ: go on. I would say, you know, especially for the industry and occupation questions, thinking about how do you translate that from an interviewer, you know, having a conversation with you about your work to what this looks like when you're completing it online and, you know, should it be multiple screens?

Should there be some sort of automated probing, whether that's AI or, or something else? And so I think when we were talking about the modernization of the CPS and, you know, testing what is an equivalent question, it might not be the same questions that an interviewer would ask.

VAN DER KLAAUW: So in our survey, n-- item non-response is very low, about, you know, one or two percent.

Partly it's because if people try to skip it, they get a, an, another reply saying, "Please," you know, "give your, give your best answer." So that, that, you know, that encouragement works. We did run an experiment on the Dutch panel, the LIS panel and there we found you know, item non-response for certain questions, like especially financial questions, were actually lower on the web compared to, like, in person.

So you ask about personal finances or you ask about even the inflation expectations, the, actually, the non-response was lower on the web. People feel more free to give an answer there than in person, either because of embarrassment or privacy issues. So I, I think it, it really depends on what the question is.

Yeah.

D. JOHNSON: Okay, great. So I have other questions, but I wanna give some time. But I want... One, one thing I want, since you have this great audience who can probably solve all the problems in the CPS or make all your wish, dreams come true, I have a two-part on opportunities and challenges. So if you were on one of the advisory committees that doesn't exist anymore, what would you tell BLS and Census is the biggest obstacle or the thing to watch for?

And then alternatively, if you could tell BLS and Census, what would be the wish of a new question you would like to add to this new ISR? So two-part. Want to go?

VAN DER KLAAUW: I c- I can start off. You answer one. The, the, the last, the last one of course, that's the easy one. Like you know, I would like a lot of other questions, but right now, of course, AI and how it's changing what people do on the job is, like, really crucial.

So to understand how it's changing the activities, you know, that people are engaged in on the, on the job, to be able to track that over time seems really important. And also, you know, non... I mean, a lot of focus is always on, on the earnings on the job, but non-wage benefits, non-pecuniary benefits also increasingly important and trying to measure the willingness to

pay of how much people are willing to give up for those bene- I think is also something that, you know, would be nice to have.

EVANS: Great. I, I think in terms of what to watch for, and I think you did allude to this, Nick, in your presentation, it's when you get to that parallel run and you decide you might wanna make a change, it's how long you'll have to wait to find out whether your change had the impact you were hoping it would.

And that, that maybe speaks to the fact that we've been parallel running for quite some time now. I- in terms of questions, I mean, I guess different contexts and I probably shouldn't be telling anyone to add any questions to make the survey longer. But for us, I guess a key, a key area that we've sort of-- we're hoping some of the redesign questions will help with and has obviously risen up at the sort of levels of interest i- is around kind of precarious forms of employment and more marginal attachment to the labor market.

So any questions that help us understand that more, I think are, are likely to be kind of, useful in a, in an ongoing way.

ABRAMOWITZ: Yeah, I would echo that, thinking about more non-traditional work arrangements and better capturing those. And in terms of advice, I think one takeaway from the PSID experience is that having this randomization was really critical because, you know, now we can look and see really narrow in on the effects of mode.

A benefit of PSID being a panel is that we can also look for the same person if they changed, if they were assigned differently to mode from twenty nineteen to twenty twenty-one and look across that assignment how that reporting changed for that person. So, utilizing the panel structure of the data could be really valuable for that.

D. JOHNSON: Great. So I'm gonna answer my own question. So the big challenge when the PSID converted was the non-response follow-up. We thought we'd save lots of money on doing internet, and it took following up with getting people to complete it on the internet s- ate up a lot of the cost savings. Mm. So you see people switch.

So if you're gonna look at something that you might not have looked at, I would look at the non-response follow-up. Okay, questions. Wait. Up here in the front.

AUDIENCE QUESTION: Thank you. Louise Sheiner, Brookings. So you just brought up AI and the possibility of using AI, and this whole thing was conceived, the CPS modernization, before we realized how powerful AI was. And I just wonder if like kind of going along and doing what you plan to do instead of shifting gears and saying, "We should figure out how to use AI to do this better."

And you can think of a million things as you were talking about the occupation, like making sure you get it right by doing the right follow-up questions. And like there's just a million ways. And so I just wonder, it's really almost a question for the prior panel, but like sh- isn't this something where AI can make the thing much better, much cheaper?

D. JOHNSON: Anybody want to answer?

VAN DER KLAAUW: I mean- We

D. JOHNSON: use AI, right?

VAN DER KLAAUW: We use AI for the open-ended questions to analyze. So AI is really good in finding category, you know, categories, and then you have, like, another thousand independent agents rating those categories or how close your answer is to each of these categories.

So it's very useful for analyzing open text responses. The other use of AI that I think is, is, has a future is chatbots. The chatbots who are actually serve as interviewer. And so online when you're answering and you actually see a person asking the questions and be able to respond and, and you know.

Two years ago I saw an example of this, and it was already very impressive how it was able to answer questions. And now they've become very lifelike, and I think in the future, you know, people seem to like that interacting more with a person and still have the benefit of doing it whenever you want, right, online.

So I, I personally see that as something, you know, to try out in, in the future.

D. JOHNSON: Anybody else?

EVANS: I, I'd say briefly, I mean, I think one thing that gets in the way of us progressing the research further is, is quite a boring practical point, and it's about the infrastructure that the survey can sit on and how compatible that is with bringing AI in now.

So I think there is often some precursors before we can really kind of dig into it. Colleagues who are working on this more kind of day-to-day will also talk about the risks around we will need to inform respondents that this-- there is likely to be a AI in the survey and what impacts that might have, particularly if there is any shift in public perceptions for good or for worse.

So I think that's often at the back of the mind, but a, a bit of the risk aversion there maybe sometimes.

D. JOHNSON: So there was one back, yeah, that one first please, and then this

AUDIENCE QUESTION: and then- Yeah. Thank you all for presenting. My question was specifically for Tom about the crossover surveys. What were the specific differences you found in responses between the LFS and the TFLS?

Like, did you find that there was a greater willingness to share more detailed information through the TFLS?

EVANS: So, what I can say is fairly limited 'cause we haven't published a lot of it yet, but I will speak to what I am able to speak to. So one area where we've said a little bit in, in, in the public domain is around reasons for economic inactivity.

So this is telling us whether somebody is retired, a student, things like that. And so we've-- it is a quest- the question which determines this has been redesigned between the two surveys perhaps more than other areas of the questionnaire. And we've certainly seen that have a bit of an impact.

So we see, I can't really speak to them, but we do see differences in the estimates there. And then the crossover surveys helped give us a bit more of an evidence base as to, okay, is that because we've reordered the response options. So whereas I think retired on the LFS was previously perhaps the ninth option, it's now the first option because it is the most common response given.

But then you find people are getting to that first, perhaps potentially getting anchored on a certain response, not making it to another one, and they're onto the next question before you know it. So yeah, that's, that's one area where we have, where I've seen a bit of difference that we can, we are, I, I am able to speak to.

D. JOHNSON: Thank you. Why don't we do one more? It was right here. This woman here. Right there. Do you mind? Right. You are mine. Yeah.

AUDIENCE QUESTION: Hi there. I had a quick question about the usage of AI on respondents. Are you worried if respondents are using an AI filling out the surveys, that they might change their answers versus if a in-person survey or if a person surveyed them?

Are you worried that there could be a little bit of a distinct difference or some changes in how they would want to respond to that question?

VAN DER KLAAUW: I, I see that as a risk. Now if it puts them in a setting... So for example, you want to know their views about let's say mortgage rates or something like that.

If you want them in a setting where they're really deeping, you know, deeply thinking about that, and they're actually making a decision about maybe refinancing, you know, then I think it's actually okay if they go and search things up because that's what they would do if they actually make decisions. So, and it still reflects, you know, their view for whatever they see.

So it, it's not always bad, but I would say in general, yes, you don't want people to go look up. And many surveys now are, are kind of asking kindly you know, "Don't look up unless it's your own information, your own records." If you ask for your, you know, try to report their wealth assets and, and debts and so on, then it's good to ask them to look up.

But in other cases, yeah, you don't really want them to go... And of course, you can track that. Again, if you can look at mouse movements and people leave the page, if they take longer to answer questions. In our survey, we track these things. We have not seen anything happening on that front. But you know, I, I think it is, you know, it, it...

But again, it depends what you're after. You know, if you... Yeah.

D. JOHNSON: Okay. We're at the time of our session. Hopefully, if you have, still have questions, maybe at the end of the next session you can get them in. But thank you very much for the panel.

And then Michael and Dan, Alex, and Katherine can come up

UNKNOWN: Give you guys some

room.

That would be more efficient.

All right, cool.

STRAIN: Okay. Well, this has been terrific so far. This is the panel that stands in between everybody in this room and the drinks. So we have to we have to, to, to earn our, earn our keep here. I'd, I'd just like to begin by offering a word of thanks and praise to the statistical agencies- Yeah

to the representatives of the agencies who are here, to the representatives of the agencies who are not here. During my years in Washington, the job of the agencies has really only gotten harder, and they are doing such heroic work. It's often unsung. You know, I don't think enough people recognize that when they take out their iPhone and check the weather, that they're using an app that relies on government data.

I don't think enough people realize that when they go to a retail store, that the shelves are stocked based on data from official statistics. I don't think enough people realize that a small error in the consumer price index can cost billions of dollars in Social Security benefits, can lead the Fed to make a bad decision et cetera, et cetera, et cetera.

So I think the policy community as a whole would be better served by championing the work done by these agencies more often. And I think John and David are, are, are playing a b- a big role in, in, in, in pushing that in the right direction. So, thank, thank, thank you both and, and, and, and thanks to the agencies.

I think the title of this panel is, you know, what else should we, should we want from the redesign, and so, and so that could be an opportunity to have some fun. But first let me, let me introduce the panelists, who of course need no introduction. Katherine Abraham is Distinguished University Professor of Economics at the University of Maryland.

She was Commissioner of the Bureau of Legal Statistics from 1993 until 2001. She is currently the Chair of the Committee on National Statistics of the National Academies of Sciences, Engineering, and Medicine. Diane Schanzenbach is a senior fellow here at Brookings. She's a professor of public policy at Georgetown, where she holds the McCourt Chair.

She was formerly the director of the Hamilton Project at Brookings, and the director of the Institute for Policy Research at Northwestern. And Alex Richmond is the head of Labor and Workforce Economics at OpenAI. She was previously an economist at Aon, at the Burning Glass Institute, and at the Federal Reserve Board.

And so everybody, you know, direct your AI questions to Alex. We're gonna begin with some, with some opening statements and we're just gonna, I think, go in order. And so Katherine?

Great. Thank you. And let me just echo your comments about the incredibly valuable work that the people at the statistical agencies do and that, that I agree is, is under-recognized.

I want to pick up on some comments that Nick made about things that the BLS is thinking about for down the road in terms of revamping the CPS. And I'd, I'd like to mention just three things that I think it would be well worth thinking about. The first thing is y- rethinking the questions that we ask on the CPS about employment.

Those questions were designed, you know, more than thirty years ago at this point. I think they're really well designed to capture people who have traditional kinds of jobs, people who have a, an employer, one employer that they work for where they're paid a wage and salary, people who are self-employed and have a business.

But then there's all of these people who are kind of in between, people who are working as independent contractors. It's, it's just not clear how well they fit into the way the questions are, are asked. And so I... And I think it matters because, you know, from a legal point of view, if you're self-employed rather than being an employee, there's a lot of protections that you don't get.

You're not gonna be getting employer-provided benefits. And so understanding that is important, I think, for policymakers. Just many of you probably know these questions. I just want to read the question that identifies the class of worker in the CPS. Were you employed by government, by a private company, a nonprofit organization, or self-employed or working in the family business?

Well, if you're an independent contractor, and you g-get up and go to work at some company, you might well say, "Oh, I'm employed by that company." And I, I think there, there's some evidence from work that I've done that if you probe to follow up on when people have been asked that question and say, "You know, what best describes your work?

Were you an employee? Were you an independent contractor or freelancer, or were you an informal worker?" But there's a non-trivial share of people who say, "Oh, oh, I was an independent contractor." And I think that that would be worth knowing. I guess I also worry that the survey doesn't do a great job of picking up supplemental employment activities that people engage in just to boost their household income.

The multiple job holding rate in the CPS is pretty low. It's like five point five, five point six percent. There's a lot of other surveys that probe for, you know, informal work activity, and they end up with much higher numbers of people doing outside things to supplement their household incomes. And many of them say that that income is really important to them for making ends meet.

So I think it's worth knowing about. It's, it's a little bit apples and oranges. Like, you can't compare the CPS numbers to those other surveys, the numbers from the other surveys. But that's something that if there's a redesign of the questionnaire, I hope people will really look

at. Second thing, really quickly, there are some things that I think it's pretty hard to measure in the CPS, and I'm thinking mostly of the CPS ASEC as opposed to the monthly CPS.

Bruce Meyer and a series of co-authors have done a whole lot of work that convincingly establishes, for me at least, that, that folks just don't do a very good job of reporting their participation in various federal programs. So I think that's an area where pushing on linking to administrative data could be really valuable.

And then just a third thing real quick We have two surveys in the field, you know, on an ongoing basis that are collecting information about the labor force. We have the CPS, which is, from my point of view, the gold standard. We also have, you know, b- partly because of the, the participation of interviewers and, and so on.

I-- the, you know, the ISR, I, I'm confident the BLS and Census will figure out how to ask those questions well. And then there's the ACS, which is much bigger. It, it just would be-- I don't have a proposal. I don't want to pretend it's easy. But boy, it would be great if the data from those two surveys could be integrated, maybe in some model kind of way, to produce more, you know, better disaggregated estimates for some of the things in the CPS.

You know, disaggregated by geography, disaggregated to smaller occupational groups. I could go on and on, but I won't. Thank you.

I'm tempted to make a joke about how AI can just do all that, but

we won't. Yeah.

Diane.

WHITMORE SCHANZENBACH: Great. So I'm going to come as to this as a CPS user from the social policy side, where I study poverty and SNAP and food insecurity and the safety net.

So the CPS, while it's, you know, a lot of us think about it as the unemployment rate, it's so much more than that. And so of course, I want to mostly talk about the ASEC, which is where the official poverty rate, the median household income, and the uninsured rate comes from. And for almost thirty years, the December supplement gave us the nation's food insecurity numbers until they were recently discontinued.

I have four hopes for a redesigned CPS. First is to treat the ASEC and the supplements as real first-class citizens in this modernization not something that gets brought along later. I appreciated a lot of this, the conversation, and I really want to echo the praise for our great statistical agencies.

We are so grateful. Th- You've already admitted that the ASEC's the hard case. It's long, it's detailed, it's easy to break off. The federal register notice for the survey says that every single ASEC question will change. Great. Including some of the redesigned underlying con- constructs. You know, that may or may not be the right call.

I- I trust y'all, but you need-- you just made a compelling case that unemployment rate needs a long parallel survey to bridge the series, and that you describe the BLS parallel test as covering that core CPS content. My hope is that the poverty rate income, health insurance can get the same treatment.

A real bridge published so that the twenty twenty-nine change in poverty can be decomposed into how much is due to changes in concepts versus how much is in, really in economic conditions. I also want to offer a caution about streamlining to focus on data directly affecting key estimates, which is part of the conversation.

Many of these items that aren't-- don't really roll up to move the headline numbers are really important for researchers. So detailed program participation, non-cash benefits, who in a household gets what. Those don't move the headline numbers, but they're really important. I- I think of some examples from we recently published a a annals of the American...

Just, just the, the, in, whatever, it's the whole line, non-long name is In Memory of Becky Blank. Right, right. And I remember that you and Hank Farber had a really important piece on health insurance bridging to unemployment. Mm. I have a nice paper with Lisa Barrow in that where we try to understand the evolving relationship between earnings and public benefits, all of these using the ASEC.

We learn a lot of important things from that. Okay. So second, and I'll be more quick going forward, be clear-eyed about who answers online. The field tests are, you know, are really encouraging, super excited about that but we clearly saw that these were selected. They're higher income, more educated, more likely to be married.

That's great to the extent that you can promise that you're gonna d- use that freed up money to chase down some of the people who are less likely to respond, like low-income families, renters, et cetera. My hope's that, you know, we can, you know, use those savings to improve the overall overall quality of the survey.

I also want to make sure that you at least ask you to report response rates and mode effects by income and program participation, not just the average. An effect that washes out on average could look real different in the population that I study compared to other populations. The third, just briefly I wonder how possible it might be to build administrative data into production, not just on the research side.

I say that with great trepidation, knowing the number of lawyers that would have to be involved in that. I literally recently had to take a tuberculosis test to use restricted data in our own building in yeah, so. Full employment for lawyers, I suppose. And then finally w- you know, just encourage us to protect CPS as a platform.

One of the great strengths of it is that you can ex- attach a supplement to this large, high-quality monthly s-sample. Last year, USDA ended the food security supplement that had given us thirty years of consistent data that provided meaningful in-information on economic conditions faced by low-income families.

Whatever you might think about the merits or demerits of that decision, and I can imagine that you know what I think about that- it shows that the supplement infrastructure is fragile.

And so a modern web-enabled CPS could make it cheaper to field these short, well-validated modules. And I hope that the design really keeps that flexibility.

Last point, of course, none of this is free. N-not long ago, you all were warning that you were gonna have to cut five thousand samp- households from the sample. For sure, modernization's the right move. It's gotta come with other funding or we'll trade one pr- one problem for another

Thanks.

Oh, quick, don't forget that the comment period on the 2027 field test is open until November 16th.

And I hope that other users of this data will join me in sending in public comments.

STRAIN: I think we should tell people that you don't have tuberculosis.

WHITMORE SCHANZENBACH: I don't, yes. Our whole secure data's enclave is tuberculosis-free.

STRAIN: Good. Good, good. Yes. Yes. Just in case you were worried.

WHITMORE SCHANZENBACH: Yes.

AUDIENCE QUESTION: Alex?

MARTIN RICHMOND: here and how vital they are for understanding what could be a transformative economic moment over the next several years. And the sort of points I have today, I'm planning to focus on mostly how the CPS and other surveys affect our understanding of what might be if AI's effects on the labor market.

I had less plan to say about AI as an elicitation tool, but I actually think they're linked and very important. And if we wanna talk about that after in questions, I'm more than happy to do that. Broadly speaking, when we're trying to understand and think about how AI might affect the labor market, we start to think about what AI can do and what people do, and how those may or may not overlap and differ.

Obviously, that in and of itself is an incredibly complex question that takes a lot of care. But in thinking about how that affects the US population, for example, the only real thing we know about what most people that our survey respondents do all day is their occupation. And so occupations are an incredibly, incredibly important part of being able to link who is doing what jobs and who is sitting where in the labor market and what AI is doing.

And again, it's not as simple as, you know, comparing those things one to one, but it's a start. And there are a lot of places where the current occupation taxonomy used in the CPS and in lots of other places makes that even more difficult. You know, we can't separately identify in the CPS, for example, data scientists and product managers.

Product managers we can't find just about anywhere, and even in O*NET, we're supposed to assign them to marketing manager, which is quite difficult. So there are cases where we have and we commonly use in the labor market these different occupational titles that don't have good analogs in surveys.

And if I were really making my wish list I think there are cases where perhaps, as Catherine suggested, coupled with the ACS, we could even try to get more granular. And I know that all of you... This is an incredibly expensive and time-consuming ask, but I think it's worth it. There's so much rich empirical variation in how AI's capabilities are evolving.

For example there have been pretty distinct paths for the evolution of AI's capabilities for front-end and back-end software development. And lots of people-- Some-sometimes people apply for roles that encompass those sort of two separate spheres of software development. But lots of times, people tend to be on one or another career track.

They're well-delineated. This kind of language is present in job advertisements. And if I were really making a wish list, capturing that kind of detail at the occupation level would really enrich our ability to understand how these more nuanced changes in AI capabilities are translating into changes in employment or wages in a timely way.

This, of course, is brought even more to bear even using the existing occupation taxonomy with the declining response rates and the sample size of the CPS, such that even in places Where we have relatively good occupation codes. Like I, I think, and I think software developers is a fair one that we want to use to understand what's happening in the labor market, and the software develop- tel- developers are a relatively large group.

I think there are two and a half million or so in the US. The sample size is such that it's still quite difficult to do without a lot of measurement error, track month-to-month or quarter-to-quarter changes in their employment and wages, especially if you might be concerned or want to keep a close eye on software developers who are early in their career, for example.

And so investments in the sample size and the response rate directly translate into our ability to better monitor these groups of people, which I think is valuable. And just a, a last point there. I think in the, with the lack of that granularly, granularity currently available, a lot of occupation aggregation mechanisms, I'll call them, have sprung up.

I think some people, maybe if you're reading this research, you've heard of AI exposure measures by occupation, and they're trying to get at something valuable. Some members of some people who were on my team pioneered one of the first AI exposure by occupation measures. I think that's valuable work.

But right now, when almost anyone's trying to analyze the effects of AI on the labor market, they apply an AI exposure measure to the CPS because they need a grouping mechanism by which they can accumulate enough statistical power to try to look at the effects on the labor market. There are some notable exceptions.

There are some Fed economists, Paul Soto and Leland Crane, for example, that just tried to aggregate all the occupations in the CPS that code for work, for example, and look at their labor income trajectories. But by and large, people are developing research alternatives, you

know, sort of in lack of being able to look at some of these occupations directly, which has real implications for how we understand how AI is changing the labor market.

Again, I'm, I'm really, really thankful to all the work people in statistical agencies do, and we have a lot of powerful data, but making further investments here would go a long way into sort of increasing our ability to see what's going on here. Thanks.

STRAIN: Excellent. So, you know, I think, I think we, we want a redesigned CPS to be accurate. I think we learned from the previous panel that when redesigns happen, answers change. Yeah ... the willingness to respond to certain types of questions changes and and other kind of systematic changes between the way we're doing it now and, and, and internet-based happens.

And then, you know, Katherine, just to build on your observation that you know, there are some, there are some questions about the accuracy of, of, of the surveys with respect to program participation- Mm-hmm ... with respect to some, some other, some other, some other factors that, that go beyond kind of classical measurement error.

So thinking about those three, those three observations, you know, what are, what are the implications of that for the way that the survey is redesigned and for, and for whatever ends up coming out on, on the other end? How do we, how do we know that, that the redesigned survey is more accurate?

How concerned are you about the accuracy? How concerned are you about mitigating existing bias? How concerned are you about introducing new, new types of bias or new magnitudes of bias you know, with respect to, to either the 30,000 foot level or to your kind of specific areas of research?

ABRAHAM: Maybe I c- maybe answer a s- slightly different question than you asked.

I, I learned that in-

STRAIN: That's how you know you're a pro.

ABRAHAM: My, my media, my media training at the office. Yeah. No, I... But I, I, I do wanna just make a-

STRAIN: I'm gonna be like Wilbur's bot and just ask the same question over and over again.

ABRAHAM: I, I, I do wanna make a comment about the, the, the rollout process and, and, you know, how you know what you've got at, at the other end.

I mean, partly it's about accuracy, but it's also about comparability, and I think we really care a lot about that. So for a lot of people who are using CPS data, they're, they're wanting to look at what's happened over long periods of time. So it's... You know, accuracy really matters, but it's also really important to understand what's happening with comparability.

I was very glad to see that there's a proposal for a parallel survey. But I would caution that it's not necessarily the case that differences that you see between the ISR parallel survey and the official data are all due to mode effects. I'm thinking back to, you know, 32 years ago now when the last CPS redesign was done, and, you know, what we figured out...

I mean, thanks to work by Anne Polivka and Steve Miller, who I think are, are here. You know, it, it, it turn, it turns out that the differences that we were seeing before the new CPS was introduced between the parallel survey and the official numbers were partly due to the mode, but they were also partly due to just the fact that it's a parallel survey.

It differs in subtle ways from the official survey. So I guess this is a really minor recommendation, I suppose, but it would be great if a parallel survey could continue after the new mode was introduced so that y- so that you could flip the people in the parallel survey to the old way of doing it and try to get a better handle on what the effects were.

I don't know. Y- you're, you're asking about accuracy. How do you... I mean, some of these things are pretty subjective. I, you know, I think on program participation and stuff, we... Program participation, we do have administrative data, and you could look at that. I... The other things, it's harder to assess accuracy.

We're, we're relying on the CPS

WHITMORE SCHANZENBACH: Something that I think gets lost occasionally with the administrative data is it's not that the administrative data are perfect either. Yeah, they're- And, and so, I guess we have to keep that in mind.

ABRAHAM: I do, I do think that for program participation, it's quite... That they are gonna be better than the answers that people are giving you.

WHITMORE SCHANZENBACH: Better, for sure.

AUDIENCE QUESTION: Yeah.

WHITMORE SCHANZENBACH: It's the right direction. But y- I was just using SNAP administrative data and saw there, there are huge revisions that went back to the beginning of last year.

ABRAHAM: Yeah. I shouldn't talk about program participation data- ... like I'm more knowledgeable than I am. Yeah.

WHITMORE SCHANZENBACH: No, but, but, you know, the point, the point is that there's just no perfect data.

Yeah. And, you know, so we're just, you know, trying to fit together the puzzle pieces to try to get to measure complex, important topics.

MARTIN RICHMOND: And of course, unlike many other developed countries, we have no occupations in administrative data, in at least national- Well, there's that. Yeah.

ABRAHAM: I mean, that could change.

There is, there is a big push to add occupational information to UI wage records and to start to standardize what those look like, and that would be enormously valuable, just enormously valuable.

MARTIN RICHMOND: Yeah, completely agree.

STRAIN: So Diane, in your introductory comments, you mentioned the importance of budget constraints- Oh

which was, which was just music to my ears. That's how they get me. So I'm gonna ask a, a, a, a, a, a, maybe an unfair question. And so you guys can, can, can pull Katherine's trick and not answer it if you don't want to. But in the redesign, what, what should go? I mean, Katherine, you indicated that you would like to see a, kind of a complete overhaul of the employment questions.

Do you mean that, that a big share of the existing questions should go as a, as a, as a cost saving measure? Diane, Alex, anything you would, you would, you would jettison?

WHITMORE SCHANZENBACH: Boy, that's a, that's a, a good question. And I think we were imagining bending the cost curve with you know, asking it y- online. But I'd love to pitch that back to Katherine about that.

ABRAHAM: So I, I, I mean, I... So I'll again not answer the question. But I, I, I do, I, there, there were... At the end of the last session, David sort of asked, "Well, what would you like to add to the, the CPS?" And people had various great things that it would be good to add. I do think it's really important to keep the monthly CPS focused on the core things that it's designed to produce on a month-to-month basis.

So I, I would be very cautious about adding- Yeah ... adding things in. I don't think there's a lot of fluff there now.

What about the AZEC? Is there AZEC fluff?

That's for Diane. Yeah.

WHITMORE SCHANZENBACH: No, I, I mean, I'm, I'm, I'm thinking because we've got such a complicated safety net, you have to ask about a lot of different things, and you don't have to ask everybody about all of the different topics, and I think you are already doing a good job of, you know, the skip patterns and whatnot.

But gosh, you know, I'm trying to think about what do I think we could get rid of, and all I can k- keep coming back to is, no, no, we need more, more, more, Yeah ... yeah, which is incompatible with the budget constraint.

STRAIN: Alex?

MARTIN RICHMOND: Lots of good ideas of things to add. Yeah. Very, very few about things to cut.

STRAIN: Is it's like asking, is...

Asking an economist what information to cut is like asking a fish to live without water, I suppose.

MARTIN RICHMOND: Yeah.

STRAIN: So, you know, I wonder, Katherine, maybe just to build on your obs... And Diane, your observations about, about administrative data. I mean, is there a role for a greater use of administrative data in the redesign?

Is there a way that, that, that a finalized redesigned CPS could make better, better use in, you know, in real time or in something close to real time of administrative records in order to, in order to improve its accuracy and, and reduce, reduce bias?

WHITMORE SCHANZENBACH: I think it's technically possible. Yeah. Do I think that there's any prayer of that g- getting through the lawyers?

STRAIN: Do you think it's advisable?

WHITMORE SCHANZENBACH: Oh, gosh. Probably not just for political reasons, but I think it would speed the collection of the data. It would improve the accuracy of the data. And what do you think?

ABRAHAM: Well, I, I think, I think it would be a good direction to push in. I had, I had high hopes when the Foundations for Evidence-Based Policymaking Act was passed eight years ago now.

W- that, you know, and with, with the, the provision that, that I thought was just so important that barring any legal prohibition, that data held by the federal government should be available for evidence-building purposes, which would include statistical production. We don't even have the rule yet that says how that's supposed to work.

And even once there's a rule, there... I mean, as you said, there's gonna be all kinds of things to work out with the lawyers. But I, I think ultimately we have to push in that direction. Yeah. I mean, I, I think the e- I think the evidence is so clear that program participation reporting is just unacceptably poor.

WHITMORE SCHANZENBACH: Yes.

ABRAHAM: And I, I don't see any solution to that other than bringing in administrative data. I don't... You know, that's not gonna be feasible with the monthly CPS, but for something like the ASEC You know, if we could improve the processing of the data and improve the, the data ingest procedures, I, I think that would be feasible.

STRAIN: Alex, in every, every room I'm in, every, every meeting I'm in, people find a way to work AI into, into the, I don't know, into the conversation. I don't know if you've noticed that yourself. Yeah. If you, if you were listening to some of the audience questions in the previous two panels, there was a mix of skepticism and, and optimism about, about, about how AI can affect the quality of official statistics, how it might you know, scare off respondents if the agencies start to use it how it, how it might attract respondents if the agencies start to use it.

I wonder if you could just kind of react, you know, broadly to all that. How do you, how do you see advances in AI technology interacting with a redesigned CPS?

MARTIN RICHMOND: Yes. Ha-happy to chat about this. I think it's, it's really important. A couple of things immediately came to mind. I think one is, when I think about ISR as a survey method, I do worry, and I'm already thinking about the fact that models are getting much better at using computers.

I mean, Codex computer use, like ChatGPT work computer use, is improving quite rapidly and can directly interact with a website, even one that wasn't designed for AI participation at all. That could have some potential benefits to survey accuracy in, in some ways. I mean, ChatGPT knows a lot about you.

In a lot of cases, people like this feature, and they like using AI to fill out forms. And AI might, AI might, like someone's ChatGPT might, in a, like, privacy-protected, secure-to-their-account way, know about their income and their finances and could answer more quickly and accurately about that. That would require both a lot of trust in our product and a lot of trust in the survey elicitation tool.

But anecdotally, it seems like people are using computer use to answer these kinds of questions and already relying on the knowledge that AI tools have of, of themselves to do it. They're starting to try to, you know, use AI to make purchases, for example, this way through computer use. And so people are becoming more comfortable with systems that can enable AI to, to communicate this kind of accurate information on their behalf.

And I think there are potential positive implications for survey accuracy and response, but again, that requires developing a lot of trust and a lot of systems along the way, and and probably an army of attorneys. I also do-- We've sort of looked a bit at this internally, and there's a lot of really interesting work from other frontier labs on this as well and using AI as an interviewer.

I think there definitely seems to be some willingness from people to respond to survey questions that way. One, I think encouraging advancement here is the advance of-- the advances in sort of voice processing and voice response. And AI can now respond to you in a fairly natural sounding way, and certainly can really accurately process your voice responses.

And when I think about like in the previous panel, some of the discussions about like occupation non-response in particular, maybe a path forward for eliciting more accurate occupation information in the future is to just ask people to talk about what they do and try to, you know, hone in on their occupation from that.

There's some really neat work that just became public, I think yesterday from Stanford, a project called TaskMet, using these kinds of assisted interviews to generate better like occupation task descriptions. They've demoed six occupations to start. I think that's really promising. So I, I think that holds promise.

I think it likely, like we were talking about, induces and reduces kinds of selection in ways we don't understand yet. And so having that linked, you know, experiments linked to administrative information or other characteristics about those folks and understanding who you additionally capture when you introduce those kinds of survey methods and who you don't, I think will be a really important part of rolling out those kinds of tools.

STRAIN: Okay, let's go to questions. Dr. Sabelhaus

SABELHAUS: Thanks, Maggie. So this is for Alex. And, and I was trying to pull apart what your, your opening statement, and you were talking a lot about occupation. At the end, we heard we need more observations which is you know, that pushes against the budget constraint. But, but as you were building up to that, it also sounded like there are things that we want to know about occupations that we don't know.

And so just now you're talking about TaskNet and some of the specific things. Is, is there sort of a, a, a direct implication for what should be done in the redesign for how to change the questions? Like, is there something beyond occupation that we could add right away? Or is it, is it gonna be something like a narrative that you described that the AI would then have to deconstruct?

And I also wonder, and this is sort of based on, on your own work experience, is there something about, for example, the way that Indeed and other you know, recruiters work, how they think about what it means to fill a job, that that could teach us something about how to do this and maybe also teach us that this is a process that's gonna keep evolving.

It's something which has certainly not reached some new static point.

MARTIN RICHMOND: Okay, this is one of my favorite topics to talk about, so I'll try to keep it brief. I spent a, a good amount of time at the Burning Glass Institute looking at job ads data, for example. And one thing that, you know, we do know is that you can classify occupations pretty well from job title and industry.

So one thing you could think about doing in a short form way would be to openly elicit job title. You already ask about industry. You could pair those and see how that matches with other occupation questions. Yeah, certainly I think the, the really valuable thing is developing as targeted of a sense as possible about what people actually do, especially in these occupations where, that tend to rely on knowledge w- or involve knowledge work and rely on computers where we think AI might be affecting them.

And those are also, unfortunately, kind of the places where the current occupation taxonomy, and this is not only the purview of the CPS, but where the current occupation taxonomy can be the most ambiguous. I think five and a half million people in the US fall into managers all other. And that's a really difficult...

It doesn't even have any associated O*NET task information really. And so it's a really difficult but large category to try to think, okay, and then w-what are these people doing and, and how might AI affect them? Certainly even knowing more information directly, again, I would hesitate to add questions to the monthly CPS, but certainly knowing who's using AI for work and how would be really important.

I also think, you know, if we're worried about and given th- given that we are worried about potential labor market disruptions from AI, any questions also that measure economic precarity, for example, would be really valuable when linked to these characteristics. So I do think we're learning about how we can better elicit occupation.

I think, again, I think getting some longer description from folks would be great, but I think even getting just a little more information or refining some of the occupation categories could also go a long way into sort of a more valuable categorization there. Mm-hmm. Yeah. Thank you.

AUDIENCE QUESTION: This question is for Alex again, but the two of you feel free to jump in as well. You mentioned some of the characteristics and attributes that, you know, maybe fall into your wish list of things that the statistical agencies could ask about. But when I talk to economists and policymakers, they're wondering a lot about sort of putting those together, what are like the leading indicators of disruption.

So do you have sort of in your mind's eye like a dream indicator that the statistical agencies could report? And then, you know, how feasible maybe it would be.

Pipa, you do.

MARTIN RICHMOND: A dream. I, I say sometimes that my desert island metric is like U3 or U6. I mean, the CPS is designed to produce these really high-quality measures of unemployment.

It does that really well. Like, they're nailing that. We need to make sure first and foremost that they have what the, the tools they need to do to keep producing U3 and U6, and we need to keep paying attention to them like we do. There are some measures that people try to use using the CPS occupations and the rotational schedule of the CPS to look at what people in particular are entering into unemployment from particular occupations.

Certainly if, you know, I could wave a magic wand and we had the kind of occupational detail you're talking about, you could imagine tracking some number kind of like unemployment for how people were entering into unemployment from different occupations or in different stages of their careers. But in general, I think in terms of...

I mean, other leading indicators I would think about are wages, and I think people are starting to talk about how AI might be affecting wages. Also something we turn to the CPS for in a timely way. And so also being able to track wages by occupation and having the statistical precision to keep a close eye on that at the occupation level would be really valuable.

STRAIN: Questions?

You guys can fight for it.

D. JOHNSON: So I want to push more on the use of admin data in the monthly CPS. So the hard thing about the unemployment rate is to figure out who's not, not working but looking for work. But all of us employed people, you would think you could just use Social Security or some data to fill in all their responses and just focus the, the cost of collecting everybody else.

I don't know how you do that, but I just don't know if that would be an option to somehow find data out there that has the employed components and just-- and then just collect the data for the unemployed.

ABRAHAM: I'm skeptical, David.

WHITMORE SCHANZENBACH: But

you wanna test

it.

D. JOHNSON: Did...

EVANS: I, I just made a suggestion. I asked Tom if it was all right if we ask how-

STRAIN: Sure ... '

EVANS: cause they're doing it.

STRAIN: Absolutely. Really? Tom? Yeah. Tom, come up on stage.

EVANS: Well, I mean, how it works in the UK at the moment is that we are matching so individual respondents to the LFS. It's a fuzzy matching process based on the personal information that we know about them, where they live, their names, date of birth, things like that through to then their tax records.

I guess one, one kind of restriction here is that this is just for employees, so the tax records are only available on a timely enough basis. I would also say that it is just in a research phase, so this is not a live- Mm-hmm ... production thing. We've published one article on this, which is super interesting.

I'm happy to kind of share a link around to those interested. So that, that's like, that's where it's at. It's very much in the bias space then the weighting. Mm-hmm. So this isn't about replacing questionnaire responses. It's much more about, okay, what, what does this tell us? We know the total number of people on payroll from the administrative data.

How, how effective are we at the capturing them- Yeah ... within the survey, and can we adjust the weighting accordingly?

STRAIN: So I know it's just one month, but how'd it go for that month?

EVANS: So what, so what we've done so far is actually backward looking. So it was looking at how the LFS performed through the aftermath of the pandemic-

STRAIN: Mm-hmm

EVANS: um, where we knew there was obviously significant changes to how it was collected. And so it was really seeing how the bias as measured by this kind of adjusted approach had evolved through both the kind of downturn in the LFS and some of those challenges, and then the recovery from that downturn.

So when we reintroduced face-to-face collection, 'cause that was suspended, what effect did that have on, on, on the, the sort of the weighting and the bias there?

ABRAHAM: I, I should have been clearer about why I'm skeptical. It's not that I don't think in principle with a different administrative data structure that this couldn't work.

I think it could. The... But I think about our administrative data structure, you know, income tax re- forms issued once a year, lags in, you know, 50 states and the District of Columbia with unemployment insurance records and lags in getting those. I just don't see how for the monthly CPS, given our current administrative data infrastructure, how, how it would work.

Though that, that probably is something we should be thinking about trying to improve too.

STRAIN: Luis?

AUDIENCE QUESTION: comment about budget constraints

STRAIN: Please find a way to end your statement in a question mark

AUDIENCE QUESTION: Isn't it true- ...

that,

that when we think about what we spend on data per dollar- We're done ... that the value, the idea that we maybe have to cut the CPS when we rely on it to understand the economy is nuts.

And so, yes, there are budget constraints, but this is not a place we should be cutting. This is a place we need to be spending more money. So just

WHITMORE SCHANZENBACH: That is true. Yes,

AUDIENCE QUESTION: it is true. Thank you for answering my question. **WHITMORE SCHANZENBACH:** I still haven't been able to come up with a single question I'd cut. I've been thinking about it for the last 20 minutes.

STRAIN: Okay, the first person who can think of a question to cut we'll buy you a drink. This has been terrific. Thank you, all. Thanks to the panel, and thanks to the previous two panels and to all of you for being here.