

DARYL FOX: Good afternoon, everyone, and welcome to today's webinar, "Analyzing Data from the National Crime Victimization Survey: A Workshop for Data Users," hosted by the Bureau of Justice Statistics. At this time, it's my pleasure to introduce Dr. Rachel Morgan, Victimization Statistics Unit Chief within BJS, for welcome remarks and to begin the presentation. Rachel?

DR. RACHEL MORGAN: Thank you, Daryl, and thank you all for joining today. We're very excited to have this NCVS webinar, and hopefully teach you how to use the NCVS, and so you can go out and answer your own research questions, and publish research on this topic.

So, we are a small but mighty team at BJS in the Victimization Statistics Unit. We are Emilie Coen, Erika Harrell, Susannah Tapp, Lexy Thompson, and Jenna Truman. Next slide.

And so our agenda today, the first 30 or so minutes, we're going to talk about background of the NCVS, the methodology, how to access our data and statistics, data file structure, weighting, and unit of analysis. Then, we're going to go over a quick demonstration using our N-DASH, or our data dashboard, followed by a data analysis demonstration so you can follow along with us and learn how to replicate estimates from our Criminal Victimization bulletin. And then finally, we will have closing remarks and Q&A. And as Daryl said, please feel free to put Q&A questions along in the Q&A box anytime, and we will be monitoring those. So without further ado, I'm going to turn it over to my colleague Susannah Tapp to kick us off.

SUSANNAH TAPP: Thank you, Rachel. I'm just going to talk a little bit about the National Crime Victimization Survey. So what is the National Crime Victimization Survey? The National Crime Victimization Survey, or the NCVS, is the nation's primary source of information on criminal victimization. The NCVS is currently sponsored and directed by the Bureau of Justice Statistics, which is us. It started in 1972 as the National Crime Survey and was redesigned and renamed the National Crime Victimization Survey in 1992. The NCVS is one of two national measures of crime and was developed to complement and serve as an independent calibration of the other of the nation's measures of crime, the FBI's Uniform Crime Reporting Program, or UCR program. The NCVS measures crimes that are both reported and not reported to police, which is what makes it unique, and it serves as an index of change in reporting to police. The NCVS has been collecting data since 1973, so it recently celebrated its 50-year anniversary.

There's been many developments over time, including the redesign in 1992. From 1993 until now, the content of the NCVS has remained relatively unchanged, but that is about to change. The sample is redesigned every 10 years to align with the decennial census, and there have been other changes over time, including the addition of supplemental surveys on specific topics.

So, looking at how the data are collected, the NCVS is administered by the U.S. Census Bureau. They send interviewers to a nationally represented sample of households, and all persons age 12 or older living in selected households are surveyed. Interviewees are surveyed a total of seven times over a period of 3.5 years, or about every six months. Whenever possible, interviews take place in person, and subsequent interviews can be conducted generally over the phone, but in person, if the person desires. The NCVS is a self-report survey, and the persons in the survey are asked about criminal victimization experienced during the prior six months. The NCVS is also incident-based, and information is collected about every crime victimization incident.

So, looking at what data are collected. The NCVS collects information on violent crime, including rape and sexual assault, robbery, aggravated assault, and simple assault; property crime, including burglary and trespassing, motor vehicle theft, and other household theft; and larceny, which includes pocket picking and purse snatching. No information is collected on homicide because the NCVS is based on interviews with victims. And if you'd like to know more about crime classification in the NCVS, there is additional information on the supplemental slides at the back of the presentation that you received.

The NCVS also collects additional information, such as victim demographics and information about the specific crime incident, such as whether a firearm was used. The NCVS is the only source of national data on several topics, most notably crimes that are not reported to police.

And so this webinar will be a brief summary of the NCVS, but if you're interested in learning more about specific topics, you might be interested in our Knowledge Corner video series. The videos are about five to eight minutes long and are informational videos that's focused in on specific topic, for example, weighting the data. So now I will turn it over to Lexy to talk about these survey instruments.

ALEXANDRA THOMPSON: Thanks, Susannah. When we talk about the broader NCVS and the instrument, it's actually made up of a couple different instruments, which I'll go over in this section.

On this slide is a flowchart that details the different steps of a typical NCVS interview. And we start with the NCVS 500, or Control Card. And there's more information on all of these different steps at the back of this slide deck if you'd like to learn more.

The main purpose of the Control Card is to get a roster of who lives in each household. And this is only filled out by the household respondent. I could also maybe refer to them as the head of household. And so they detail everyone who lives there, both over and under 12, so that we know who to administer the NCVS to. Then we go to the NCVS-1, which is the basic screen questionnaire. This is filled out by both the household respondent and everyone else in the household. And the purpose of this questionnaire is to see whether there was a victimization in the last six months, or if there was no victimization. The household respondent answers whether there was a crime committed against themselves, like a personal crime, like a robbery or an assault. But the household respondent also answers if there was a property crime that occurred, which we record at the household level. And then every subsequent person in the household is asked questions about whether there was a crime against their person, like a robbery or assault. And then whether there was no victimization reported, or whether there was a victimization reported, changes what the next step is, but assuming there was a victimization reported, the next is the NCVS-2, or Crime Incident Report. We'll sometimes refer to this as the CIR. And the CIR is to gather more information about the incident being reported. So this includes our questions about whether the crime was reported to police, whether the victim accessed victim services, whether information about the offender—other information about the incident, like if there was an injury or other subsequent questions about why they didn't report to the police, et cetera.

And then as Susannah mentioned, they are asked all of these questions for every incident that the person reports. So if one person in the household reports two different victimizations, then they will fill out two CIR questionnaires. So that's kind of where that possible additional step comes in. But then at this point, once they filled out all the CIRs or they reported no victimizations, we then move to some follow-up demographic questions. We do ask about sexual orientation and gender identity. We also ask about disability status of every person in a household. For the SOGI measures, that's only 16 and older. But then that includes what we would call the core NCVS interview. And then if the person is participating in one of our supplements—like the identity theft supplement or the school crime supplement, depending on, again, their demographics, because some supplements are only administered to some people—then they would complete the supplement questionnaire after completing the full core questionnaire. So if you have a project in mind or you're interested in analyzing more data using the NCVS, there are a couple different ways to access estimates and statistics.

The four main ways to access data and statistics are BJS statistical reports; the NCVS dashboard online data tool, or N-DASH; there are different types of data files that we publish, both public-use and restricted-use; and then also the NCVS API. BJS statistical reports are a good starting place to see what we've published on as an agency, and also they're available for everyone on BJS's website.

Every year we publish the *Criminal Victimization* bulletin, which you can see on the left-hand side of the slide. And this features a lot of high-level estimates each year in regards to crime and criminal victimization in the United States. But we also will publish reports on other topics, like carjacking, crimes against persons with disabilities. We published our first report with data by state, which is pictured here. And there's a couple of other topics we've covered over the years as well. And so if you're interested in something, just take a look on the BJS website and see if something's already been published in regards to that topic.

We also have the NCVS Dashboard, or N-DASH, which is great if you are looking into something, but we haven't published a report on it yet. So there are graphics available. There are tables available on the N-DASH, and you'll actually get a demo later today in this webinar.

So, I won't spend too much time on this, but it contains a subset of questions, or variables, that we ask on the NCVS. You can look at victimization by age and race, whereas maybe in our reports, they're broken out individually rather than together. We have both quick graphics, which are the same every year. And then custom graphics, which you can create on your own, and we'll see more in that demo later today. Here are some examples of different graphics you can create on the tool. There you go.

Another way to access the data are the public-use data files. We also refer to these as the PUF. These are published on the National Archive of Criminal Justice Data, or NACJD, which is housed at the Inter-university Consortium for Political and Social Research. And the PUF is a de-identified version of the final data file for the NCVS. So it actually includes three different data files: one for households, one for persons, and one for incidents. So there's a record of every victimization that we recorded and information for all those subsequent questions that we ask on the CIR. The main barrier to using the PUF is that it does require some knowledge of a type of statistical software like Stata, SPSS, R. There are different file options to download these data. And so, if you're not familiar with software already, it could be a bit difficult to analyze the data this way. But that's why we do this webinar is to help you learn more about how to analyze NCVS data. But the PUF gives you the most flexibility in terms of analyzing different

characteristics alongside other characteristics of interest. This is an example of how the data are displayed on the website.

We also have the restricted-use data files, or the RUF as we sometimes call it. And these are housed specifically in the U.S. Census Bureau's Federal Statistical Research Data Centers. And because they are restricted data, they can contain more information that could possibly identify our respondents. But there are cases where people are able to access those data, and they do go through an application process in order to access them. The main reason that people might wanna access these data are for lower levels of geography, like doing victimization estimates by state or by city, or for other smaller geographic areas. There's also some information on attempts to contact respondents and the field representatives who interview our respondents.

And here's an example of how it's displayed on ResearchDataGov. ResearchDataGov is where you would go to look for and apply, to access these different datasets.

And then we also have the NCVS API, if you've used a different type of API, it probably functions very similarly. But the API is a nice blend of the PUFs, but also we've done some of the work for you already.

So the API is structured similar to the PUF files, in that there's a record for every incident or person. But some of the commonly used variables, like race, or victim sex, or age have already been recoded for you and displayed in kind of a more easy to understand way. So sometimes if you're analyzing data by race with Hispanic origin, that information is contained in a couple different variables on the PUF data file, but on the API, it's already been recoded and combined for you. So that's one benefit of the API compared to maybe the PUF file. But I'll turn it over to Jenna to talk about structure and weighting.

JENNIFER TRUMAN: Great. Thank you, Lexy. All right. So I'm going to start talking about the data file and structure weighting. And so the NCVS data are annual data. And data users can look at individual years of data. The code books include detailed information on individual NCVS years. The data downloads can be downloaded in multiple formats depending on what program you use, so if SAS, SPSS, Stata, so on and so forth, and the files at NACJD, which is what Lexy told you just how to access those data files, includes five files. There's an address record, household, person, incident, and then the incident-level extract file, and I'll be talking a little bit more about those here in a second. So those are the annual data files. There's also concatenated NCVS data files. So those include all years of data from 1992 through the most recent

year, so the most recent year available is 2022. And these codebooks include fewer details because the majority of details are in the annual files.

So if you want all of the details, refer to the annual codebooks, and the concatenated files will have details still but fewer. Again, still able to download in multiple formats. And then those concatenated files—there are three files—so there's household, the person, and then the incident extract file.

The NCVS data are hierarchical. So we have sample addresses, as Susannah told you. So we have address local information, and then it goes to the household, so each household will be on the household files. And then within those households where samples target everyone age 12 or older, those persons will be on the person-level file. So if we sample three persons in household one, they will all be on the person file, and on household two, if there was just one person, so that'll be Person Four on that person file. And for that first household, only Person One experienced an incident, so both of those two incidents, both those incidents will be incident-level file. And then in the second household, that person also experienced an incident, so that will be on there, and you'll see the three incidents and then the four different persons in two households.

As I mentioned, these are the separate files that are available on the annual basis. There's address-level, which generally you won't need to use. The household-level, person-level, and incident-level files are what you'll be interested in using.

So, the household-level file includes information about the household and household respondents. Those variables are named at the 2000-level variables, so it'll be V2116, and so those are the variable names. The person-level file has information about, again, each person age 12 or older within those households that were sampled. Those are 3000-level variables, so that'll be the V3000. And then there's finally the incident-level file. So this information has everything about the criminal incidents that were reported by a respondent. Property crimes are reported by household respondents and personal crimes are reported by everyone 12 or older. These are 4000-level variables. And what's really important here is the extract file—which is what, I think, most people want to use—also includes the household and person variables, so data users are able to analyze victim demographics.

The NCVS public-use file structure. So again, as I mentioned, they are annual files. The collection year data from January 1 to December 31st, so the information is based on the date of the interview, not the date of the incident. Most sampled persons or households should be included twice in the person/household level file, but that's not always the case. So their first interview would be January through June, and the second

interview would be July through December, but once they hit those seven interviews, that household phases out, so there's a seventh of the sample that phases out during July through December, and then a seventh phases out, and there are also replacement households coming in. So that's why that most people will have, or should be included in the files, but that's not always the case—their house may have phased out, they may not have reported, so on and so forth, so. Just keep that in mind.

In terms of a sample weight—so NCVS data are typically weighted. Sample weights are adjusted, excuse me, are adjustments that are applied to sample data to make the data representative of a specific population of interest. For the NCVS in particular, sample weights are applied to make these data representative of the U.S. population of households or persons age 12 or older, because again, that's who we're sampling.

So sample data that are not weighted, or unweighted data, can be biased as a result of who did and did not complete the survey. The unweighted data are not representative and cannot be generalized about specific populations of interest. So BJS typically does not report out on unweighted NCVS data. We as statisticians do examine unweighted data to understand the sample characteristics and assess sample sizes to estimate reliability, and all NCVS statistical estimates in BJS reports are weighted to be representative of the population.

So the NCVS has four types of weights: household, person, incident, and victimization. And I'm going to talk in detail about them here in a second. And really knowing when to use each type of weight and what's really critical to your analysis.

So, first, the household weight, which are located on the household data files. There are two weights. So there's the V2116. This weight adjusts each six-month period to count the U.S. households. So again, because we're interviewing everyone in the household every six months, we're able to adjust the weight to account for the total U.S. population within that six-month period. The adjusted household weight or the WGTHHCY, divides the original household weight, so the V2116, by two for an annual count of U.S. households. So this is the weight that we would use to generate the annual count of U.S. households when calculating property crime rates. So if that's what you're looking for, this is the weight you want to use. And as you can see, this is an example from the 2022 household file where you can look and see that when you compare the V2116 weight to the WGTHHCY weight, they are cut in half, so you'll see that comparison here.

For the person weights, [these are] set up the same way. So there's an original person weight, so the V3080 that adjusts a six-month period to the U.S. population count of persons, again, age 12 or older. And then the WGTPERCY, or Weight Percy as we'll

sometimes say, adjusts that person weight. Again, so divides it by two, and then it will give you the annual U.S. population of persons age 12 or older. So you want to use this weight when you're generating annual counts of persons age 12 or older, when you're looking to calculate violent crime or personal larceny rates. And again, same example here, where these are a couple weights that we pulled from the data file, and you can see there's the first weight is 1243, the second—the Weight Percy rate—is 621, so it's cut in half.

And then we get into the victimization and incident weights. So both of these are on the incident data file. There's two ways to look at incidents or victimization. So an incident is the number of specific criminal acts involving one or more victims. So if every incident has one victim, the number of incidents would be the same as the number of victimization. If there's more than one victim, the incident estimate is adjusted to compensate for that possibility that that incident could have been multiple times and over-counted, whereas the victimization weight is the total number of times that people or households were victimized by a crime.

So, for violence and personal larceny—so personal crime—the number of victimizations is the number of victims of that crime. And then remember, we told you that the property crimes are at household level, so each crimes against a household is counted as having a single victim, the affected household. Only respondents who reported an incident or victimization are included in the incident data file, so there's nothing else included in that file. And then really, it's up to the data users to decide what you're interested in and whether it's analyzing incidents or victimizations, and then apply the appropriate weight from them.

So in terms of the victimization weight, again, this is in the incident data file. There are two variations. So there's the original weight that's on there, so the WGTVICCY, or Weight Vicky, if you want to go shorthand. Those provide the total count of victimization, and there's no adjustments for series crimes. So series crimes, when I say that, I mean that crimes that are similar in type to one another but occur with such frequency that a victim is unable to recall each event or describe in detail, so something like domestic violence or intimate partner violence. Someone may have had that happen multiple times and can't distinguish the event, so we just ask them about the most recent. These crimes are weighted to account for the high frequency repeat victimization or series victimization, and that is the series weight or the SERIES_WEIGHT, and that weight counts series victimizations as the actual number of victimizations up to a maximum of 10. And the reason we capped it at 10 was to avoid that extreme outlier effect, so if someone was reporting 30, we're not getting extreme outliers. So if it's not a series victimization weight, the two weights are going to be equal.

So, you'll see in this example that's highlighted here. The first line, that are non-series, so the weights are the same. The second line is a series victimization, and that person reported 12 victimizations, so it was capped at 10, so you'll see that the first victimization weight is 750 and then it's multiplied by 10 to get the series weight.

The NCVS incident weights are set up similarly. So the difference is what they're counting. So the incident weight is the V4527, and that provides the total count of incidents. Again, no adjustment for series crimes. And then the SERIES_IWEIGHT accounts for those high-frequency repeat or series incidents. And same thing, the weight counts series as the actual number of incidents up to a maximum of 10 so they'll be equal. So same thing here. You'll see that the non-series, the weights are equal, and then the series for the reported 12 incidents were capped at 10, it's multiplied by 10. And again, just remember that the difference between victimization and incidents is the number of victims involved.

As you'll see from my examples here, the weight can vary quite a bit. So in 2022, for example, the household weight ranged from 83 to almost 5,300. The person weight ranged from 98 to about 12,000. And the victimization weight ranged from 197 to over 47,000. So what that means is that a weight of about 47,800 victimization could be based on an unweighted sample of one or hundreds. It depends on the characteristics of the victim. So it is always important to make sure that you're checking the unweighted counts to make sure to understand what you're dealing with the sample.

And again, as BJS statisticians, we always ensure that we are publishing estimates that meet our data quality standards. So we look at the unweighted estimates as well as the coefficient of variation to estimate to decide if the estimate needs to be flagged or suppressed if it's not meeting that minimum sample size. So it's up to you as the data users to take a look at that, as well when you're doing the analysis. Okay. So I'm going to turn it over to Emilie for an N-DASH demonstration.

EMILIE COEN: Great. Thank you. So as Lexy said, the NCVS Dashboard is one of the avenues for completing analysis off of a smaller set of variables. And so in order to show that, I will be going to the website outside of PowerPoint. Okay. Great. So if you go to the NCVS homepage or even to the BJS homepage, and there's a data dropdown that can take you to the N-DASH, you will see a variety of options up at the top. So when you see the original—this is the landing page you will come to, and I can show you what you can get to. So the main things that you'll find are an overview of the tool, what we call Quick Graphics and then Custom Graphics as well.

So as I said, there is an overview that gives you an idea of the types of analyses you can do, the crime categories, the units of measure that are specifically available, as well as the types of comparison views. So whether you want to look by crime type or characteristic, and those characteristics can include incident-specific ones or demographic characteristics, as well as a lot of the display and filter settings. So do you want confidence intervals? Do you want estimates flagged? What are you looking at specifically? Do you only want to see within those who identified as females versus male to female? And so, just explaining a lot of these steps for you.

There's also a User's Guide that goes through a lot of these steps and gives you more information there. And Terms and Definitions that are very helpful when thinking about how we design the age brackets or the certain type of crime.

So as I mentioned, one of the first options are Quick Graphics. And these are representations of data that remain the same year to year, more or less, but with the addition of the most recently published year. So as you can see here, it has been updated to include the 2022 data, as *Criminal Victimization 2022* was released this fall. And you're going to see here trends over time. And so, you can look year to year, and when you hover over the year of interest, you can see the rate for all violent victimizations here, as well as your confidence intervals and standard errors. You have the option to show the table behind it, download it, and also download the image.

So they get broken down a little bit more within type of crime or within some of the victim demographics. Here, we have rates in Hispanic origin. These non-alterable graphics also include similarly within Property Crime, Reporting to Police, the Victim-Offender Relationship, Weapon Use, or Victim Service Use. So these are some of our larger and most popular variables.

Now, what is probably of interest to a lot of folks is looking at the Custom Graphics where we have a select number of variables that you can choose from in order to complete some analyses online.

And so there are three types. You can look at Multi-Year Trends. So how has this crime type or characteristic changed over time? What do those rates look like? Percentages? For those sorts of things, you can look within a single year. So a Single-Year Comparison, again, they're all by crime type or characteristic. Or you could do a Year-to-Year Comparison. And this would be looking at, for example, if we wanted to look at characteristics, say year-to-year—you can see the filters I used last time. And looking at, say, we wanted to compare 2019 to 2022 within Personal Crime, we're looking for a rate, not a count or the number. And if we wanted to then see, say, Reporting to Police

by Victim Sex for these two years, you can see that there isn't a large amount of change. And this is for all violent victimization.

So here, this circle is going to be 2019 and you can see that the rate for males who did report to the police did go up from 2019 to 2022, whereas for example, within those who identify as female, that rate actually went down a little bit. You can then get even a little bit more specific. If we wanted to look within those who report a Robbery victimization, you're going to see a pretty different picture here. And so within men, within those who are male, the reporting rate went up slightly, but within female, you can see that there are much larger changes from 2019 to 2022. And again, if you hover over them, you're able to see the percentage of those who reported, the change, those confidence intervals, as well as the standard errors.

So, there are the variety of variables within Personal Crime that you can look by as well as obviously the type of crime. If you get within Property Crime, you will see slightly different comparison characteristics. And these are collected at the household level and so being able to, again, do slightly different analyses here specifically and directly on the NCVS webpage. All right. I am going to pass it off. This is an opportunity to look at the Dropbox files and to work through an example.

ERIKA HARRELL: Thank you. Thank you, Emilie. I am going to do a demonstration. And even though your materials may reference two examples, I'm only going to be looking at Example 1. Before I start with the demo, I wanted to show you this slide, which has resources that you can use to generate estimates from the NCVS data. Most of these files should be available in the Dropbox that was mentioned. The Dropbox link should be in the chat or you can find them on the BJS website or through NACJD.

This slide contains two tables from our *Criminal Victimization* report. And the circled estimates are the estimates that I will be replicating, the number and rate of total violent crime for 2022, and the number and rate of total property victimization for 2022.

What you'll need for today's example basically are three files. The NCVS syntax example is a .sps file. Actually, you'll only need Example 1. You won't need Example 2 for any of this. I'm just using just one example. The NCVS replication Example 1, which is an Excel file, and another Excel file with the GVF parameters and rhos.

Now, to begin, I want to open up the Excel file, the NCVS workshop replication example 1.xlsx and go to the 'output' tab for the SPSS code and data tables. And I will go there.

I am on the 'output'. You should be on the 'output' tab for the Excel file. Basically, what's showing in this worksheet is what you would get if you ran that actual .sps file that I mentioned earlier. This is the output. And I'm going to go through this output and show you how we generate our estimates and actually generate standard errors as well.

To calculate the victimization estimates, you'll need to start with the incident file. That's from the dataset that you downloaded from NACJD. It has an 0005 in the name. And also, we have to apply recodes too it in order to recode the variables into usable categories. This code that I'm showing now recodes the 4529 into NEWOFF and also into NEWCRIME, which is basically our big type of crime variables that we usually use. Variable labels, value labels, this is basically what we do for most of our reports. And also, we have to exclude victimizations that occur outside of the U.S.

Now, NCVS respondents can report victimizations that occur outside of the U.S. if they meet our criteria. We do collect them, but since we generate estimates for crime that occurred in the U.S., we have to exclude anything that occurred outside of the U.S. So we do ask about the location of any incidents and we collect that information here and we exclude V4022 not equal 1. That 1 value means that the incident occurred outside of the U.S. And this is a frequency that was run just to show that that category 'outside of the U.S.' was removed. And now, in order to actually get those estimates, we needed to get weighted estimates of types of crime in order to get the number of violent and property crimes from 2022. We have to weight by series weight.

Jenna mentioned the weights that we have to use in order to get weighted estimates that weights up to the population. The report shows the crime estimates rounded to the nearest 10, but however, this will show the exact number taken from the dataset. There is violence and property crime, and this is just for 2022. The data that we referenced, that we pointed you to on NACJD is just for 2022. And also, as mentioned before, we have to flag our estimates if they do not meet the criteria of being based on a certain number of sample cases. So, we have to look at our unweighted data as well. So in order to do that, we take the weight off and we run that same frequency of the same NEWCRIME variable to get the types of crime. And here, we don't have a problem. Usually, we have to flag if an estimate is based on 10 or fewer sample cases. Here, we don't have that problem, but sometimes, if we break things down, certain types of crimes—such as rape, sexual assault—we may run into small cell sizes and we may have to flag some estimates.

Now, to calculate the rates, you will have to have populations. For the violent victimization rates, you need the person level populations. And we have to turn to a different file for that. The 0003 file has the person level population. We also have to use

WGTPERCY, the person weight, in order to get that weighted number of persons, age 12 or over for the U.S. population. And in order to do that, there's several variables that can be used. But for this example, you can use YEARQ. Now, YEARQ is a variable that has the year and the quarter that the NCVS interview was taken for each respondent. There should be four quarters. So if it's just one year of data, they should just be four categories. And this bottom number is what we'll use in order to generate the rate. That's the total for the year, the total number of persons weighted for the year.

Similarly with property crime, you'll need to get the 2022 household population. And that is a completely different file, the 0002 file, that is the property—I'm sorry—it's a household population file. And that's what we use to generate rates of property victimization. We have to use the appropriate weight for the household population file. And also, again, we use YEARQ. YEARQ is on every single—it's on most NCVS files. So that's something that we can use to get the number of persons weighted for each year. And actually, for this data set, we can use YEARQ. But this is not the number of persons. This is the number of households that were entered—this is number of households for 2022 and this will be used in calculating our rate of property victimization.

Now, in order to actually calculate that, we need to do some linking. And in order to do so, first, I'm going to start with the numbers—the number of violent crimes—basically just taking it from our output. And basically, this is what we do at BJS. We run some SPSS output and we start linking. And I will link the unweighted numbers as well. They won't be used in the calculations of the rates, but it's good to have them just so we know how big of a sample size our estimates are based on and the household population.

Now, to generate the actual rate per 1,000 is very simple. All we have to do is divide the number by the appropriate population. And since we report rates per 1,000 persons, we have to multiply this fraction by a thousand. And that is the rate, 23.5. For property crime, same thing, we take the number, divide it by the number of households, multiply it by a thousand.

Now, another thing that we do that we have to do for all of our estimates is to generate standard errors. It's a little tricky because of the structure of our sample. So the Census Bureau has helped us out a little bit. They generate what we call GVF parameters, which are used in our significance test. I'm showing you a significance test that we use. This test is only used for counts. It's not used for percentages. It's not used for rates. There is a reason for that. The reason is that the formula is used to calculate the standard errors for count. It's different from how standard errors for rates and

percentages are calculated. That calculation comes from the Census Bureau and more information can be found in our technical documentation about it.

Now, A, B, and C, these are our generalized variance function parameters. We call them GVF, we get them every year, and these are solely for the core NCVS data. They are not for any of our supplements for a variety of reasons. Basically, they're calculated differently. We get these every year from the Census Bureau, but this is for just the core only. If you are generating estimates based on supplement data, do not use these GVFs.

Now, what I'm going to do is I'm going to fill out for violent crime our GVF. For overall violent crime, we use actually this first row of GVF person. Any type of overall violent crime or overall personal larceny, we have to use this first row GVF. For property crime, any overall estimates for property crime, overall burglary, overall trespassing, overall motor vehicle theft for the entire—and what I mean by overall, I mean for the entire sample. We have to use this third row. The overall property crime estimates GVF.

Now, these domain estimates that if you are looking at something by race, by age, by type of crime, reporting to police, that's where you will use these person crime domain estimates. But since we were looking at it for the entire sample, then we'll just be using these overall estimates.

And now, we will simply link the number of violent crimes and also the number of property crimes. And over here in column L would be the standard errors that would occur. If you use the correct GVF, correct estimate, you will get the standard errors. And also these will be the same standard errors that you would get if you were generating this from the N-DASH or using data from the API. You will get these same estimates. And actually, this is how we get our estimates that we put into our BJS reports, by using these significance tests, GVF, calculating, linking all of that.

Now I'm going to work with the rates for violent and property victimization. Here we only use the B and the C parameters. The formulas are different, so different parameters are needed.

Now, here we have what's called a base. And for our rates, the base would be that population that was used, that denominator in that fraction, which would be the population that would be used.

For a percentage, it would just be whatever, like if you were getting a percent of robberies reported to police, that total number of robberies that you would generate that

percentage from would be the base. But for here it's just population, a persons, and households. So I am going to link the base and the rates. And again, the formulas are already implemented into the worksheet and the standard errors are being generated. Same standard errors that you would need, that you would see on the N-DASH.

Now, one thing I forgot to mention would be up at the top. If you want to compare rates and percentages, you will need to change this number from one for rates and two for percentages. And I also did not mention rhos. They're not needed here, but rhos are basically correlations. When someone responds over a period of time, over six or seven interviews, we assume that their answers are correlated. So since this also gives us correlations based on two-year and three-year sets of data, so that we can use if we were comparing estimates that were a year apart or two years apart, we would put a Y here and actually put in the rhos here, and they would be used in the calculations. But since we're not doing that, since both of these come from the same year, no rho is needed. And the same goes for this significant test here. You would actually put in a 'yes' here if rhos were needed and put the rho here.

And this is basically it. This is what we do. This is basically what we do for our reports and answering requests. And everything has to be documented, which is why we have all of that linking involved. And as I stated before, what you would get from the significance test, this is just screenshots from the N-DASH and the Excel file. The standard errors would be the same for the rates and for the numbers. And it actually shows that if you do these examples correctly, the standard errors and the numbers, they will be the same from the N-DASH to the significance test. And now, I will turn it over to Rachel.

RACHEL MORGAN: Thank you, Erika. I know folks were asking if this is going to be recorded because we did move through it pretty quickly. And, yes, it will be recorded and on the BJS website.

So, we really just want to be able to give you step-by-step instruction so that you can go back and run through it again and have all the documents to replicate these estimates. As Erika mentioned, there is also another replication example in the materials that you downloaded. So we typically do this webinar as an in-person workshop at the American Society of Criminology meeting. And then we have four hours to do it, but we were not going to do that to you guys today. And we go through two examples. So there is another example. All of the materials are in the Dropbox folder and then there are step-by-step instructions in the PowerPoint that you can use. Alright. So we'll work through some questions here. And real quick, I will end with just saying you probably found out about this webinar through JUSTSTATS, but to make sure that you stay up-to-date with

BJS data releases, statistical report releases, and updates to our data tools through JUSTSTATS, just follow BJS on X and Facebook.

And then we're going to give you our email addresses and you can reach out to us if you have any questions. But you can also reach out to AskBJS@usdoj.gov with questions and they will field those questions to the statistician subject matter experts. So you can always reach out to us if you have NCVS questions, but if you may have other questions related to other BJS work, feel free to reach out to AskBJS.

And where am I? And I'll just show real quick here our email addresses. The PowerPoint is in the Dropbox folder as a PDF. So you have all the links that Tammy put in the chat and all the step-by-step instructions. And then I'll just note that we did put a lot of reference slides here that we do present on usually at ASC just for your information if you're interested in more information on NCVS, more specific details. All right.

So, we do have a few questions in the Q&A that I will start putting on here. Yes, if the team wants to come off back onto the video so that we can field these questions, that would be great. And then feel free to keep putting them in the Q&A and we will respond, because we have plenty of time.

So, first question is for Jenna. For NCVS data. "Is the date for each offense based on the day that the person was surveyed or the estimated day month that the crime happened? I'd be interested in comparing NCVS numbers to UCR numbers to get an idea of reporting rates, but would need to compare based on estimated date that the victimization occurred. Is this information accessible in the PUF?"

JENNIFER TRUMAN: Yeah. So the date—it's based on the date that they were interviewed as we talked about in the presentation. And then we ask about the last six months, so you could potentially estimate from there, but it's based on the date of the interview. So that's going to be a little bit different when you're comparing the two data sources.

RACHEL MORGAN: "Is there a reason why it's so expensive to apply and access the restricted data files? I would love to do this, but it is cost prohibitive." Yes, we know that and we are sorry for that, but when you are applying and getting access, you're getting access to the federal research or FSRDC, Federal Statistical Research Data Centers and their staff, and so we have to pay them. So that is part of the reason that it costs a lot. We encourage folks to also think about pairing up with a university or an institution and maybe a professor working at that institution and maybe they sometimes have

collaborations or Memorandums of Understanding with the FSRDC and the institutions that may help as well. So that's an idea.

Lexy. "How often are hate crime statistics reported? Are those stats a combination of the NCVS and the UCR?"

ALEXANDRA THOMPSON: So, hate crime statistics in the—well, hate crimes through the NCVS can be reported any year or any time that we interview a respondent. The last official BJS report that we published was I believe in 2022 featuring data through 2019. But it does not include any UCR data. The way that hate crime and the NCVS overall works is that there are questions in the instruments specifically asking about whether the victim thought it was a hate crime or not. If the victim reports that they suspect it was a hate crime because the offender used hate-related words or if the victim reported to the police and the police confirmed it was a hate crime, or—there's a third one. There's three criteria that by our definition we determine whether it was a hate crime through the NCVS, but the report is called *Hate Crime Victimization 2005-2019*. If you would like to read it on the BJS website and that details more about how we define hate crime in the NCVS.

RACHEL MORGAN: Thank you. Erika, can these analyses be done online or in Excel, or both?

ERIKA HARRELL: It depends on which analyses you are referring to. What I showed can be done on Excel, which is what I just showed. You can just generate them through our N-DASH tool or you can use the data from the API actually to generate those estimates. But in terms of getting the standard errors, you can do that in Excel if you feel like programming a very, very long equation into Excel. But yeah, they can be generated, you can use Excel. You can be online using N-DASH, API data. Yes.

RACHEL MORGAN: And how about for the weighting? I'm not sure I've ever done weighting in Excel. We usually do it in our statistical software program, so I'm not sure that you could do something like that in Excel, but maybe you can. Does anyone know?

JENNIFER TRUMAN: I think you can use...

ALEXANDRA THOMPSON: I think if you...

JENNIFER TRUMAN: ...the API data, but it's going to be limited. That's sort of what the API is helpful for, but the API is limited in terms of what's available in it. And Lexy, I think you were going to add something else potentially.

ALEXANDRA THOMPSON: I think kind of what you said is that I think for maybe high-level numbers, it could be doable by summing up the weights for different categories for overall estimates. But if you want to get more complicated, I would not recommend using Excel.

RACHEL MORGAN: Yeah, I agree. Especially given how big these data files are as well, it might present some challenges. Jenna, over the past few years, have you noticed any changes in response rates and/or a change in demographics of those who choose to respond to the survey?

JENNIFER TRUMAN: Yes. Similar to all household surveys that are—we've noticed the decline on the response rates, and that's consistent with really all household surveys that are either conducted over the phone or in-person, even online surveys. And we don't do that right now. But definitely in-person and interviewer-administered surveys have experienced declines on response and we've seen the same. To address that, we do non-response bias analysis each year and take a look at that to make sure that the data are still representative. So it is something we analyze each year and take a look at, but certainly something that we have seen over time. In terms of demographics, that's part of that non-response bias analysis that we do. I would say it's been similar over the years, but certainly a different demographic may respond. But certainly a decline overall is what we've been seeing.

RACHEL MORGAN: Thank you. Emilie, do you want to show how to do a demo for a single year for a specific incident rape or sexual assault in the N-DASH real quick? Oh, I have the ball...

EMILIE COEN: Sure.

RACHEL MORGAN: ...so do I need to pass it to her, to pass it to her?

EMILIE COEN: I might be able to just share my screen. Let's find out. Of course, it might crash. Well, Daryl, do you know? Because it's saying that it's going to stop. Rachel. Oh, perfect. Okay. Thank you.

DARYL FOX: Yeah, I just passed it to you. You should be fine.

EMILIE COEN: Yup, that's good now. So if we're looking at a single year for a single type of crime, really, what you are going to get are going to be rates. So one of the things I believe that would work for you would be in the quick graphics. One of the things we do specify are breaking it out by the types of crime. And so you could hover over 2022 and it's going to give you the rate, the confidence interval, the standard error there. Another option that you would have would be to go into the custom graphics and go by crime type and essentially only pick the type of crime you are interested in and the year. And this way you could toggle between the count, the number, or the rate, but, once again, you can see that the rate is 1.9 per thousand persons 12 or older. So those would be the two ways that I would say you could get the rate, or if we wanted to look at the number of victimizations and these would be weighted. We're looking at a little over 500,000 as well as the confidence interval and the standard errors. I hope that that is what you were looking for.

RACHEL MORGAN: Thank you. "Is there a way to calculate state level rates per year?" So the state level data are behind the firewall at Census. So you would have to apply through ResearchDataGov like we talked about to get access to the restricted files. We have published one statistical report to date on that earlier this year, criminal victimization in the 22 largest states. And we are working on a variety of other subnational efforts and ways to release data because we know folks are interested in subnational estimates of crime. There is a section within the NCVS program page on the BJS website that focuses on our subnational estimation program. So I would encourage you to look at that section. Chris Campbell, thank you for your question.

Information about our response rates is, and also the GVF that Erika was using to calculate the standard errors, are in our code books and users' guides that are available for download through NACJD when you download an annual level of data file. So, say, you download the NCVS study from NACJD, there'll be a variety of documentation, download all of that documentation, and then within that is a source and accuracy statement from the U.S. Census Bureau. And that goes into details about all the different kinds of response rates that are calculated. We calculate response rates at the household level, at the person level, and then overall response rates, so they're included in there. And then the GVF for the last three years of NCVS data are also included in those code books. And more recently this year they're called users' guides, but in the past they've been called code books, so all of that information is included. If you can't find something, please feel free to reach out to us, but it is all publicly available.

EMILIE COEN: Hi, Rachel, I believe...

RACHEL MORGAN: Yeah.

EMILIE COEN: ...that says reporting rates, so it might be reporting to police if we just need some clarification.

RACHEL MORGAN: Yeah. Report rates to police, yes, crime reporting rates. Okay. I'm sorry about that. Crime reporting rates would be how we report them or we calculate them. It would be the number of victimizations that were reported to police divided by the number of persons in the population. We're looking at reporting rates by demographic characteristics. The denominator would be the population of that demographic group. Okay.

"How do we cite BJS when we do our analysis from the dashboard and produce tables based on our own chosen variables?" I believe there should be some citations in the dashboard, is that correct? Someone wants to answer that? Lexy, you look like you're about to.

ALEXANDRA THOMPSON: I was looking up to confirm it, but there in the users' guide on the tool, there's a section at the bottom of how to cite data from the tool so you can use that format for citation.

RACHEL MORGAN: Perfect. Thank you. "For anyone, so open to the group, this is a general question, but how did you learn all of these tools and strategies related to working with these data?" Heather is currently a student pursuing her master's in CJ and is very interested in crime data analysis but has had limited experience with SPSS and she's not entirely sure on how to learn more about analysis. Any advice? Does anyone want to say anything?

JENNIFER TRUMAN: I would just say it sounds like you're in the right place since you're pursuing your master's in criminal justice. I don't know where you're at and the resources that you have there, but I would suggest to maybe seek out a professor who maybe uses—it doesn't have to be NCVS. But certainly any kind of data. I know that from my personal experience, I actually used one of the supplements in my research early on. So I had some knowledge coming in, but like of the core NCVS data, I didn't really have any full on statistical analysis experience until I came into BJS and kind of dug into the data here. So I would suggest to potentially finding—that's kind of how I did. I worked with professors who use other data and, well, it's not the same. You have to learn the kind of nuances of the data. That's one way to kind of start. And, potentially, again, I don't know the resources at the university that you're at, but look, it doesn't have to be SPSS, if others know data or there are others on our team. SPSS is what I

learned on, but others on our team learned different resources, so I would just suggest that. And again, this is part of why we do this, right? Is to help people who want to use the data to start at a high level and kind of learn how to use these data.

EMILIE COEN: I would agree. I used NCVS data for my dissertation, but I come from the field of public health, so I didn't have any faculty that were familiar with NCVS. And so it involved with reaching out to—were there people that I cited or were citing several times, or faculty that used the data. But if you're familiar with a different software, looking at the syntax from SPSS and you can kind of see how variables are created that way. But also within the technical documentation and a couple other appendices, there are syntax examples across a couple different softwares, as well as there very well may be a repository within the school or other students, or just other folks that may have some experience using the data that could help you that way in a more kind of localized manner.

RACHEL MORGAN: Thank you both. "When you calculate reporting rates, how do you know they are accurate if police reports are based on incident date, whereas NCVS dates are based on the time of the survey and prior six months? Is there a way to adjust for that?" Does anyone want to jump in in particular?

JENNIFER TRUMAN: I mean, it's a self-report survey, so we're asking them whether or not they or someone else reported the incident to police. So we're not verifying it against police data. Again, it's a self-report survey, so we're asking them, and what the respondent is reporting is what is in the data. So I will say too in terms of, if the incident—let's say, we interviewed them today and the incident happened yesterday, potentially there's a chance they haven't reported it yet. So there are things like that that we may not actually capture in the data. So that is, certainly something to keep in mind. But again, it is what they're self-reporting, whether or not they or someone else had reported the incident to police.

ALEXANDRA THOMPSON: We do also ask about the month and the year the incident occurred, and I believe that is available in the PUF file, but I haven't looked recently in terms of whether there's a lot of missing responses or not. So maybe you check that before looking at that variable.

RACHEL MORGAN: Thank you.

EMILIE COEN: I think they also might see if you're looking at potential discrepancies reported and police came, but no report was filed or it's recorded in different ways as well. So like Jenna said, the data we have is entirely self-report.

RACHEL MORGAN: Yeah. And they're not always going to line up between FBI, UCR, NIBRS, or SRS, so keep that in mind. Typically, the data collections trend in the same manner but they're not collecting the same information. The methods are different. So realizing and noting that is helpful when interpreting the differences between the collection.

So, I would say that also BJS is a statistical agency and all of the data that we're collecting are for statistical and research purposes. So we are not a law enforcement agency. And that is not what we are focused on doing. All right. "Does BJS have guidance on how to weigh panel data for longitudinal multi-level analysis? For example, if looking at incidents nested within women, because the person weight varies by reporting period using status survey setting function for multi-level analysis is not possible." Does anyone have any guidance or suggestions for Kristen regarding this question? Looking at repeat is my guess. Repeat victimization over time.

JENNIFER TRUMAN: Yeah. I don't have an answer right at this second, but I would suggest there's a lot of longitudinal research about that there that's been done with the NCVS, so I would suggest potentially looking at some of those analyses that have been done. Most, if not all of them, have been done outside of BJS in the academic field. So you might want to suggest looking there. I don't have a direct answer because I don't typically work with Stata.

RACHEL MORGAN: Thanks, Jenna. Field supervisor [INDISTINCT]

SPEAKER 1: [INDISTINCT]

RACHEL MORGAN: ...Flint, Michigan. Thank you for joining us. That means...

SPEAKER 1: Thank you.

RACHEL MORGAN: ...Victoria works with our field representatives or our interviewers at the U.S. Census Bureau, who collect all of these data for us. So we are very appreciative of you and your work. Right now, let's see. I'm trying to think stats on Flint, they may be included—if it's near Detroit, they may be included in our MSA level file, which is a public use data file that we have available through NACJD, but we have not published on the topic. So Michigan is included in our 22 largest states report. So we can point you to that. But thank you for joining us.

"Running models unweighted and some papers out there say this is okay with the NCVS data." Yes. That is what we've heard as well. We typically aren't running regressions in different kinds of multi-level analyses. And we're weighting to the population when we're putting out reports, but yes, we have heard that other folks are doing that, researchers, and that it is okay.

EMILIE COEN: Rachel, up above the field supervisor, there's a question. "Are there data on technology fueled rape crimes, such as social media or dating sites?" So I would say within the core survey now, the stalking supplement does touch on technology stalking. It looks at both physical and technology-based measures, would be what we are able to capture within cybercrime other than identity theft.

RACHEL MORGAN: "Tribe specific crime data included?" No. We do ask if the crime occurred on an American Indian reservation or tribal lands, but tribe specific information would not be included. Last questions. No, we're not tracking tribal membership or tribal affiliation or a demographic data. We do ask about the respondent's race or ethnicity, so we are collecting if they're American Indian or Alaskan native, but that's it. No more details beyond that. Okay.

Colleagues, anything else to add, any final thoughts? No? Okay. Well, we thank you all for joining us today. You have all of the materials and if you have any questions, please feel free to reach out to us, so I'll go back—thank you, Daryl.

Go right back to that slide. Yeah, feel free to reach out to us and we will get back to you. We can only do so much at BJS analyzing NCVS data, and our goal is to get the data out to researchers and students and other folks, so that you can take it, think about research questions, and work with the data and publish on it. And really utilize the thousands of variables that are included in these data. And there's so much information we can't do it all. So we appreciate you all joining and thank you very much.

DARYL FOX: On behalf of the Bureau of Justice Statistics and our panelists, thank you for joining today's webinar. This will end today's presentation.