

DARYL FOX: Good afternoon, everyone, and welcome to today's webinar, Update on the National Crime Victimization Survey Subnational Program: First Release of Direct Subnational Estimates hosted by the Bureau of Justice Statistics. At this time, it's my pleasure to introduce Grace Kena, Statistician with the Bureau of Justice Statistics for welcoming remarks and to begin the presentation. Grace?

GRACE KENA: Thank you, Daryl, and welcome, everyone. Thank you for taking the time to join us. I'll be moderating today's session. We're honored to be joined by the BJS Director, Dr. Alex Piquero, who will give opening remarks. Dr. Piquero is a nationally and internationally recognized criminologist with wide-ranging research experience. Alex's wealth of knowledge from his academic career and passion for using data to improve the field of criminal justice has been instrumental in his leadership of the agency. Join me in welcoming Alex.

DR. ALEX PIQUERO: Thank you, Grace. Thanks, everybody, who's in the world virtually with us on this Monday afternoon. And kudos for those you all who are on spring break and tuning in on your beach or wherever you happen to be at because this is very, very cool stuff. You know, for my entire career, I've always admired not just BJS but BJS's one of its principal data collections is the NCVS which, of course, is celebrating its 50th anniversary this year. And it's one of the most well-respected, looked to, regarded, cited, used, downloaded datasets that exist for criminologists, and sociologists, and researchers around the country with respect to the range of questions that it has and its breadth of coverage. But what's really cool is the work that BJS has done with RTI to develop the first ever release of subnational estimates that recently came out. You know, the victimization team at BJS is great. You know, we have a lot of fun in the hallways. But they are really great statisticians who are always thinking about what can be done with the NCVS data better. And then, of course, when we go forward with our products and our data releases, we want to make sure that the world uses them to the best of their ability. So, I can talk on for a while, but I'll stop. I'll turn it back over to Grace and her team to go through the BJS report on the subnational estimates. Grace.

GRACE KENA: Thanks so much, Alex. Before we begin, in addition to the support that BJS leadership--the BJS Victimization Statistics Unit and Rachel, my co-author on the inaugural report have shown for this work, I also want to take a moment to recognize a few former BJS people who made key contributions. Doctors Jim Lynch, Lynn Langton, Mike Planty, Allen Beck, and Barbara Oudekerk. Turning to the presentations for the day, here's our agenda. We will first have just a brief overview of the subnational estimation program. With the NCVS, Rachel will go over the--or Andrew rather will go over the validation of state-level estimates. Rachel will cover the data from the first release. And then Andrew will take it back over to cover the user's guide material and Rachel will share with us how to access the restricted-use data. So, we have a great panel lined up for you today. Before I introduce everyone, if all of our guests could just answer one question for me in the chat, what is the most important thing that you want to get out of today's webinar? I really appreciate that.

So, Dr. Rachel Morgan is a Statistician at the Bureau of Justice Statistics whose research, interest, and work focus on criminal victimization, stalking, financial fraud, and subnational estimates of crime using data from the NCVS. Rachel is also involved in BJS's Victim Services Statistical Research Program, which includes the first ever national data collections on victim service provider provision in the United States. Andrew Moore is a Research Statistician at RTI International with 14 years of survey research experience. His areas of interest include imputation, weighting, data analysis, and SAS programming. For the past 10 years, he has served as a Statistical Task Leader on the National Victimization Statistical Support Program, or NVSSP, where he has contributed to numerous substantive and methodological tasks aimed at enhancing the NCVS. Dr. Marcus Berzofsky is a Senior Research Statistician at RTI with over 20 years of experience designing, implementing, and analyzing complex survey data. For the past 10, he has served as the co-principal investigator on the NVSSP, which has helped BJS to review and improve the methodology used for the NCVS. He has authored several federal and peer review publications on the NCVS methodology, including reports for producing subnational estimates via direct estimation, small area estimation, and generic areas.

Like Rachel, I'm a Statistician at BJS for the past six years, I was a Project Manager for the NVSSP while I worked on multiple components of subnational estimation using the NCVS. So, as Alex mentioned, we think, of course, humbly that the NCVS is a very important survey. It's one of the nation's two key sources of violent and property crime data collected from people aged 12 or older as part of a household sample. Survey interviewers collect a great deal of information about victims and victimization incidents, including whether crimes are reported to police. The data collected are then weighted to produce victimization estimates that represent the nation, other areas of the country, or parts of the population. The NCVS has been in the field since 1973 and it's celebrating its 50th year. So, while the NCVS has been a very valuable resource over the years with interest in being able to produce reliable estimates for areas below the subnational level, such as for states, has been there for a long time.

Over the past 10 years or so, BJS has worked with its partners, including the U.S. Census Bureau, RTI, and Westat to develop ways to produce these data considering both the additional cost required to generate more granular data for areas and the need to make sure that these estimates are statistically valid and reliable. After years of preparation and testing, BJS planned to produce estimates directly through the survey for the 22 largest states, using at least three years of data. As mentioned, BJS worked through several different strategies to produce subnational data. Information about the entire research program is available at our website. Today though, we're focused on the direct estimates produced by boosting and reallocating the NCVS sample to generate estimates for the 22 largest states. This work has allowed us to provide reliable estimates for these states just as we have traditionally done for the nation as a whole. This timeline shows major milestones in producing direct estimates using NCVS data. And I've talked through some of these already.

The remaining speakers will focus on the work to validate the state-level estimates and share findings from the inaugural statistical report using the first three years of state-level data. Looking forward among other activities, BJS plans to continue work in developing subnational estimates. As you saw on the previous slide, BJS is also in the midst of an instrument redesign, which is a key effort that is expected to add significant value to the already exciting development of producing direct victimization estimates. Using the expanded instrument in the coming years and other existing data sources now, analysts can use the NCVS data to explore research questions across states or within a particular state, keeping in mind that the data will need to be aggregated or pooled over multiple years to get enough samples to produce reliable estimates. Later in the session, Andrew will delve into things you need to consider when conducting state-level analyses. And Rachel will cover the process for requesting access to work with the data. This slide includes a few resources. Be sure to keep up with future BJS releases. Engage with BJS on social media and let Heather Brotsos, the Chief of the Victimization Statistics Unit, or me, know of any questions. Andrew will now share with us the work that he oversaw to validate the state-level estimates. Andrew?

ANDREW MOORE: Thank you, Grace. So, as Grace mentioned, we'll next talk about this validation study that was done prior to the release of estimates. So, before 2016, the primary objective of the NCVS was to provide national estimates of criminal victimization. When the goals of the survey were expanded in 2016 to enable the creation of victimization estimates within the 22 most populous states, many changes were required. These changes included a new sample design, an inclusion of new geographic areas, an increased sample size to ensure estimates could be produced with adequate precision, and changes to the weighting procedures to allow representative estimates to be produced within subnational areas as well as at the national level. Given these changes and evaluation of the effectiveness of the revised procedures and the quality of state-level estimates was undertaken before the release of official state-level statistics, the goals of this evaluation were to, one, evaluate the redesigned NCVS from a Total Survey Error perspective to identify potential sources of error that may compromise the validity of state-level estimates. And the second was to evaluate how NCVS estimates compare with and augment estimates from the FBI's Uniform Crime Reporting Program. I will mention that the purpose was not to assess specific estimates for particular states, crime types or domains, but instead, we were focused on evaluating the sample design and data processing procedures to identify overarching issues or data patterns that may undermine the validity of estimates for all or for a large number of states.

Total Survey Error refers to the accumulation of all errors that may arise in the design, collection, processing, and analysis of survey data. And these errors can lead to differences between survey responses and their true values and diminish the accuracy of conclusions based on survey data. Total Survey Error can be decomposed into Sampling Error and Non-Sampling Error. Sampling Error arises from the fact that estimates are based on only a fraction of the population of interest. And then Non-Sampling Error encompasses all other error sources, including coverage, nonresponse, measurement, and Data Processing Error. So, the first of these we'll talk about is

Coverage Error. And this occurs when certain groups are under or overrepresented on the sampling frame. And coverage area can be a source of bias. To evaluate the potential Coverage Error, we calculated coverage rates by comparing the NCVS population totals for various groups to the same population totals from the American Community Survey. A ratio of one indicates these totals are the same, while ratios above or below one indicate over or under coverage. At the person level on the right, the overall coverage ratios by sex, age, race, and marital status, are generally close to one for all states. The biggest deviations are seen for the Federal Poverty Limit ratio categories, though the NCVS and ACS measure income differently, which is likely contributing to some of this variation. At the household level on the left, the overall coverage ratios are near one for all states. The ratios by characteristics of the household tend to vary more than the ratios for person level characteristics due to differences in how person and household weights are constructed and calibrated to ACS population totals.

The next type of error we looked into was Nonresponse Error. And this quantifies the error due to a portion of the sample not participating in the survey and can be a source of bias while also negatively affecting the precision of estimates. Generally, NCVS response rates are pretty high. The household level response rate was 73% nationally during this time period. And across the 22 states, it ranged from about 57% in New Jersey to 81% in North Carolina. Person level response rates are higher given that they are conditional on household response. So, within responding households, the person level response rate ranged from about 75% in Indiana to 90% in Michigan. In addition to the overall response rates we looked at on the previous slide, we also examine person level response rates among various subgroups as seen here on the left. Response rates tended to be lower for persons under 25 years of age, that is with less than a high school education, and single persons. So, we also looked at another way to assess Nonresponse Error by looking at the percent relative bias. And the relative bias quantifies the difference between those who responded to the survey and the overall sample. And we see kind of a similar pattern where younger respondents tend to be underrepresented among respondents relative to the overall sample.

The next source of error we looked into was Measurement Error. And this occurs when the construct being assessed, such as whether respondent was a victim of a crime, is incorrectly measured during data collection. And in a panel survey, such as the NCVS, two ways in which Measurement Error can be introduced are through telescoping, where a respondent reports an event that occurred outside the six-month reference period, and respondent fatigue or panel conditioning, where a respondent chooses not to report incidents to shorten the length of the interview. Variations in the time in sample, which counts the number of times an individual is eligible for the survey, could indicate differential impacts of telescoping and fatigue on victimization estimates. But as we see in the figure on the left, between 2017 and 2019, the distribution was relatively consistent across the 22 states, and the percentage of cases in their first interview wave was consistent with levels prior to the redesign, which was about 30% in 2015. The increase in sample size and the introduction of new geographic areas as part of the 2016 redesign also resulted in an influx of new interviewers, which previous analyses

have found to be associated with increased reports of victimization. And so this could also be another source of Measurement Error. However, between 2017 and 2019, more than 80% of person-level interviews were conducted by experienced NCVS field staff in all 22 states. And this percentage exceeded 90% in all but three states. And this staffing, we made sure, was also consistent with the distribution prior to the redesign.

The next source of error we looked into was Data Processing Error, and this occurs during the data processing stage after the conclusion of data collection and can be often introduced while editing or weighting the data. Variations in the weights can impact the variance of estimates through design effects, and this can vary across states. The variability of person weights is shown here on the left. These violin plots are similar to histograms, and generally the shape of the distribution is similar across states. Though as expected, the more populous states have, on average, larger weights. We also looked at the percentage of extreme weights within each state, and this ranged from about 1% in California to 6.5% in Indiana. While this isn't necessarily a problem, it's just something analysts should be aware of, for the potential impact of extreme weights on the variance of estimates, as well as the potential for victimization estimates to be influenced by a respondent's large weights, particularly if you're looking at rare crime types or small subpopulations.

As mentioned just a minute ago, telescoping and fatigue have the potential to bias estimates produced with NCVS data. To mitigate these effects, the NCVS employs a bounding adjustment and a time in sample adjustment to adjust the victimization weight for some respondents. And the figure on the right, the violent crime rate for each state with and without these two adjustment factors, are calculated. The distance between each point and the 45-degree reference line represents the impact of these adjustments on victimization estimates. And as seen in this figure, this impact is relatively homogeneous across states. And the average decrease is approximately 15.5%. There's also an adjustment made to the victimization weight to account for series victimizations. And series victimizations are incidents that are similar in nature but occur with such frequency that the victim is unable to recall the details of each individual event. The series adjustment increases victimization rates and the impact of this adjustment is more variable across states. So, for example, in the figure on the left in Arizona and Washington, the series adjustment increased the violent victimization rate by about 40%, while in Minnesota and Virginia, the increase was less than 10%.

So the other part of total survey error or sampling error and, again, this arises when a sample rather than a census is used to estimate characteristics of a target population, because the true population parameters are unknown, it's not really possible to estimate the contribution of bias to the sampling error. However, the Relative Standard Error or RSE is one way to quantify the sampling error. The figure on the left shows the RSE of the violent victimization rate for each state, along with reference lines for the national estimate, and the goal that was based on the sample design. Although only five states achieved an RSE less than 10%, the violent victimization rate had an RSE of less than 30% for all states, and most states had an RSE of less than 15%. In the table on the right, we provide summary statistics for the RSE across all 22 states for various crime

types. And we've also indicated the number of states whose estimates were flagged as unreliable because the RSE was greater than 50% or the numerator of the rate was based on 15 or fewer sample cases. As expected, rare crime types such as rape or sexual assault and motor vehicle theft are the crime types for which estimates are most often flagged. In addition to overall crime type estimates, we also examined violent crime rate RSE among various subpopulations, similar to the pattern we just saw for rarer crime types. Estimates for rarer or harder to reach subpopulations also tend to be flagged as unreliable more often than other subpops, and this includes persons classified as non-Hispanic other, persons aged 12 to 17, and those aged 65 or older.

Most national surveys, including the NCVS, incorporate a complex sample design with oversampling and clustering to increase cost efficiency and to ensure a diverse and representative sample. However, this complex design can be less efficient from a precision standpoint due to an increase in the sampling error. The Unequal Weighting Effect or UWE can be used to quantify the sampling error introduced by a complex sample design. And the UWE measures the variability of sampling weights and is an indicator of the loss of efficiency from this complex design, with larger Unequal Weighting Effects being associated with larger variances and less precision. So these figures show the Unequal Weighting Effects of the household and person level weights by state. At the household level, the UWE for all states except one is less than or equal to 1.15. Person-level UWE tend to be slightly higher, although the values are generally less than 1.25. At both the household and person levels, the Unequal Weighting Effects for each state is less than that for the U.S. overall, as weights within states tend to be more homogeneous than weights across states. So we just covered the first part of the evaluation, which was looking at the survey from a total survey error perspective.

The second part of the evaluation focused on external context and how the NCVS state-level estimates compare with and augment estimates from the FBI's UCR Program, and whether any of the differences can be attributed to design or other differences. This table highlights some of the key differences between the two data collections, and these primarily relate to the population of interest, the types of crimes measured by each data collection, how estimates are calculated, and whether crimes not reported to the police are included or excluded from estimates. So for this analysis, NCVS victimization estimates for crimes reported to the police from the 22 largest states were compared to estimates from the UCR Summary Reporting System. Even after trying to account for many of the differences mentioned on the previous slide, there were still key distinctions between the two collections, so complete agreement wasn't expected. So the first way we compared estimates from the NCVS and UCR were just in absolute terms. So, here we have the victimization rate, a 95% confidence interval for violent crimes excluding simple assaults and the NCVS in blue, and then the victimization from the UCR shown in red. So in 14 of the 22 states, the 95% confidence interval for the NCVS estimate did include the UCR estimate, but there wasn't really a clear pattern for whether the NCVS estimate was higher or lower than the estimate from the UCR.

The second way we compared estimates from the NCVS and UCR was in relative terms. And we did this by examining the rank order of states across the two sources.

So, in this relative-term setting, household-level crime type showed better agreement than personal-level crime types. So when grouping the states in the high, medium, or low rank category shown by the area is shaded in blue. The NCVS and UCR estimates matched in 11 to 15 states for the four household-level crime types shown on the left. And for the four person-level crime types shown on the right, the groupings matched in six to 11 states depending on the type of crime. So this table summarizes the overall results from the validation study. The primary concerns we identify were related to subpopulation estimates within states. Some subgroups have much lower response rates than others, which could affect the main estimates, resulting in lower precision and estimates that are more susceptible to being influenced by a small number of respondents. But despite these concerns, no issues were identified that would also call into question the validity of state-level victimization estimates produced from the redesigned NCVS. Although there were some initial rare irregularities in the first half of 2016, as the new sample design was rolled out, these anomalies were largely abated by 2017 and should have negligible effects on state-level estimates produced from the 2017 to the 2019 NCVS data. In terms of external context, rather than undermining state-level estimates produced with the NCVS, the lack of agreement with the UCR likely reflects the difference of strengths and limitations of these two complementary data collection methods and the significant contribution each one makes to a more complete understanding of patterns of criminal victimization. And now I will turn it over to Rachel.

DR. RACHEL MORGAN: Thanks Andrew. All right. So I am going to talk about the First Release of Statistical Estimates for the 22 Largest U.S. States. And these were included in our BJS report that was released last week. So presentation overview, I'm going to talk about statistical estimates here is the front page of the report that's on the BJS website. So these are the 22 largest states that are included in the NCVS direct subnational estimation program. These states accounted for a majority of both the U.S. person population and the violent victimizations in 2016. The states represented 79% of the U.S. person population and 75% of violent victimizations captured in the NCVS in 2016. So the rate of violent victimization, so in the figure on the right, the U.S. rate of violent victimization is designated in gray. And then the lightest-color green represents states that were not significantly different from the U.S. rate, the medium-color green significantly lower than the U.S. rate. And then the dark green were states that represented estimates that were significantly higher than the U.S. rate. And you can also see the different statistical significance levels. So, this would be easier to see in the report. It's kind of hard to see in the figure, but just wanted to let you know that's what we're looking at here. So violent victimization includes rape or sexual assault, robbery, aggravated and simple assault.

During the aggregate period of 2017 through 2019, the U.S. rate of violent victimization was 21.6 victimizations per 1,000 persons ages 12 or older. The violent victimization rate exceeded the U.S. rate in three of the 22 most populous states. So that was Colorado at a rate of 45 per 1,000, Arizona and Washington, and both were at a rate of 36.8 per 1,000 persons aged 12 or older. Seven states have lower rates than the U.S. rate and the remaining 12 states were not statistically different from the U.S. rate.

And then looking at the rate of property victimization, so property victimization includes burglary or trespassing, motor vehicle theft, and other types of household theft, during the aggregate three-year period from 2017 to 2019, the U.S. rate of property victimization was 105.9 victimizations per 1,000 households. So the NCVS measures property victimization at the household level as opposed to the person level. The property crime rates were higher than the U.S. rate in six of the 22 largest states, so that's Washington, Colorado, Arizona, California, Indiana, and Texas. And then property crime rates were lower than the U.S. in 12 states. I didn't list these out, but you can see them in the figure and in the report. And then in four states, the property crime rate was not statistically different than the U.S. rate, and these states were Maryland, Minnesota, Missouri, and Tennessee.

So next, the rate of violent victimization excluding simple assault, so this includes rape or sexual assault, robbery, and aggravated assault. During the three-year period, the rate of violent victimization excluding simple assault was 7.7 victimizations per 1,000 persons ages 12 or older. Arizona, Colorado, and Washington had rates that were higher than the U.S. rate and seven states including Florida, Georgia, Maryland, Massachusetts, New York, North Carolina, and Virginia had lower rates of violent victimization excluding simple assault compared to the U.S. rate. And then the rates were not statistically different in the 12 remaining states. For burglary victimization, so burglary is the unlawful or forcible entry or attempted entry of a place where there was a completed or attempted theft. During the three-year period, the burglary victimization rate in the U.S. was 19.7 victimizations per 1,000 households. Four states had higher rates than the U.S., Washington, Arizona, Indiana, and Tennessee. Eight states had lower rates and then the remaining 10 states were not statistically different than the U.S. rate.

So one advantage of the NCVS is that we can look at victimizations that were reported and not reported to police, so about two in five violent victimizations or 43% during this aggregate period were reported to police nationwide. And then across the 22 states, 34% to 58% of violent victimizations were reported to police. The percentage of violent victimizations reported to police was higher than the U.S. in three states. These include Massachusetts, 58% of violent victimizations were reported to police, 55% were reported in Florida, and 51% in Ohio. And the percentage was lower than the U.S. in two states, and this is Maryland and Wisconsin, and it was 34% were reported to police in both of those states. And then looking at property victimizations reported to police, about one in three property victimizations, 34%, were reported to police during the three-year aggregate period. Between 28% and 44% of property crimes were reported to police across the 22 states. The percentage was higher than the U.S. percentage in five states, including North Carolina, Michigan, New Jersey, Georgia, and Florida. And then the percentage was lower than the U.S. percentage in five states as well, California, Indiana, Washington, Maryland, and New York. And for the remaining 12 states, the percentage was not statistically different than the U.S. percentage.

So, again, all of these findings and more details are included in the criminal victimization statistical report that was just released. If you go to bjs.gov, there's a link to it on the

main page. And then also for more information on our subnational program, you can go to the BJS website as well. And then as Grace mentioned at the beginning, please sign up for JUSTSTATS because that's how you can get notifications about BJS publications and products. And then be sure to follow BJS on Twitter and Facebook. So I'm going to turn it back over to Andrew to talk about the user's guide.

ANDREW MOORE: Thank you Rachel. So now we'll talk about a user's guide that was created to assist analysts with generating state-level victimization estimates with NCVS data. There were three primary goals of the user's guide. The first was to provide guidelines for producing state-level victimization estimates and highlight some of the key differences between the public-use data files and the restricted-use data files available within a Census research data center. The second goal was to provide some recommendations for researchers, including how to identify and resolve potential issues with state-level estimates. And finally, the user's guide includes some starter code in SAS for performing standard analyses at the state-level.

This table summarizes some differences between the NCVS public-use files and the restricted-use files, with respect to the availability of subnational identifiers and using the data to produce victimization estimates. In terms of geographic identifiers, the restricted-use files contain state and substate identifiers not currently available on public-use files. However, although geographic identifiers are available for all survey years within the RDC, the sampling and weighting methodologies implemented prior to 2016 were not designed to produce representative state-level estimates. The 2016 data is also not recommended for direct state-level estimation due to the introduction of the new sample design. Therefore analysts are discouraged from using data from survey years prior to 2017 to produce direct state-level estimates with the sample boost data. For a similar reason, it's not recommended to generate estimates for states that were not part of the sample boost, as direct estimates of victimization within non-boost states would be subject to bias and low precision. Another key difference is that estimates produced within a Census research data center must be approved by the Census Disclosure Review Board or DRB prior to release.

And then the last key difference that I'll highlight is that while national-level estimates can be generated using a single year of data, state-level estimates for the 22 most populous states should be based on a minimum of three years to ensure estimates achieve adequate precision. Although the validation study discussed earlier found no overarching issues, it's not really possible to anticipate every type of analysis that may be conducted by data users. Therefore, researchers are encouraged to assess the potential impact of the different error sources for the specific state-level estimates they choose to produce. For example, coverage error, as we talked about, can be a source of bias in state-level estimates. So analysts are encouraged to examine coverage quality for the population and time periods of interest by calculating coverage ratios. Nonresponse error can also lead to bias, so researchers are encouraged to calculate response rates and percent relative bias, particularly for subdomain estimates. The unweighted sample sizes, standard errors, and relative standard errors should also be calculated. The standard reporting convention used by BJS is to flag estimates with an

RSE greater than 50% or estimates that are based on 10 or fewer sample cases to alert readers that estimates may be unreliable. And data users are also encouraged to follow the same practice, though some modifications are required due to routing rules imposed by the DRB for unweighted counts. And then the last row is--this is just another action that's recommended when producing at state-level estimates. And that's to look for outliers which are loosely defined as individuals that report a large number of victimizations, yet they also represent a large number of individuals in the target population.

Although data users are not expected to re-weight the data, users are encouraged to evaluate whether a single individual or a small number of respondents are accounting for a large proportion of an estimate. So if you do come across one of these issues, what can you do about it? So, if the recommended steps and considerations we outlined on the previous slide illicit concerns about a particular estimate or analysis, there are a few different steps that can be taken to address those concerns. The first option is to include more years of data in your analysis to increase the number of respondents used to produce estimates. This strategy can be useful when faced with small sample sizes, low precision, or estimates that are being driven by a small number of cases. The second option would be to collapse the domain categories to increase the sample size in the combined category. And then a third option is to exclude the particular state, victimization type, or subgroup from the analysis that was kind of raising concerns to begin with. Although this is an option of last resort, the redesign of the NCVS was focused on ensuring the reliability of estimates for the entire population of interest within the 22 largest states. So analysis of some subdomains and rare crime victimization types may not be feasible with current sample sizes and the number of years of data available. You'll also notice that in addition to listing the strategies, we've also proposed an order of implementation indicated by the sequence of numbers in each row. However, this order is kind of dependent on the analysis. So, some modification may be needed depending on the goals of your particular analysis.

In addition to differences between the public-use and restricted-use files discussed earlier with regards to the types of estimates that can be produced, there are also some key differences in the structure and content of the restricted-use files. For example, although both file types are separated into household, person, and incident files, within the Census RDC, the annual NCVS files are further divided by typically either a quarter or half year. And so those must be aggregated prior to analysis. The files within the Census Research Data Center also use different variable names, such as those shown here for identifying unique household or person-level interviews. Another difference is that on the restricted-use files, the household and person weights will need to be modified to create the equivalent of the adjusted collection year weights, and further adjustments may also be needed depending on the type of analysis due to aggregating multiple years of data when producing state-level estimates. And finally, you should be aware that sometimes the incident file will contain unclassified crimes for some years. And those will need to be excluded from analysis when producing victimization estimates.

As we've mentioned, the variance or standard errors of state-level estimates is an important indicator when evaluating the quality. There are two general methods for calculating the variance of estimates that account for the complex sample design of the NCVS. The first is generalized variance functions or GVFs, and then the second is direct variance estimation. So with the public-use files, there are two direct variance estimation approaches that can be used, and that is Balanced Repeated Replication and Taylor Series Linearization or TSL. Within the restricted-use or on the restricted-use data files, the BRR weights are not currently included. They are actually on a separate file and they're working on adding those or making those available. But for instance, when we did this analysis, those weren't available. GVF parameters are also only available at the national level and shouldn't be used for subnational estimates.

And so, in the example we'll go through in a minute and the user's guide, we focus on Taylor Series Linearization variance estimation. One drawback of this approach though is that it does require some manipulation of the data files, and how those files need to be manipulated can vary depending on the type of estimate being produced. And then also the complex sample design information must also be specified when calculating TSL standard errors. So there are specific packages or procedures within a software that also must be used for analysis. So the user's guide includes example codes for performing many standard analyses with NCVS data and those examples are provided in SAS. This includes calculating victimization totals, rates, and proportions, along with the TSL standard errors, as well as code for performing statistical tests to compare subgroups, states, and changes in victimization over time.

So now we'll run through a quick example of how to calculate state-level victimization estimates and standard errors. This is kind of a flow chart of the process. For person-level crimes, that process is kind of shown in blue and then for household-level crimes, the process is the gray boxes. And while this might look a little complicated at first glance, for the most part, this process is the same as calculating victimization rates for subgroups, such as age, using the public-use files. So the specific example that we will look at today is calculating the violent victimization rate in California for 2017 to 2019. So the first step of the process, we bring in the concatenated incident file that includes data from 2017 to 2019 and we create an indicator for the crime type of interest, in this case violent crime. We also create an indicator for crimes that occurred outside the U.S. so we can exclude them when creating estimates. And, finally, we create a count of the number of incidents for series crimes. So if the incident isn't a series crime, then this variable we're creating, Series Weight, will just be equal to one. In the second step, we need to summarize the series adjusted number of violent incidents reported by each person within a reporting period.

Because the NCVS is a panel survey and because we're going to be aggregating multiple years of data to produce state-level estimates, most respondents will be on the files more than once. And we need to keep the incident count separate for each of those interviews because there may be a different weight associated with each interview. Since violent crime is a personal crime, we're summarizing here by year quarter, control number, and line number, which uniquely identify a person-level interview. And we are

also excluding crimes that occurred outside the U.S. on the where statement, as well as any crimes that are not classified as a violent crime. We also add the victimization weight to the ID statement so that it will be included on the summary data set for the next step. In the third step, we're merging the summary file we just created with the person-level file. Respondents who did not report any violent incidents will not be included on the summary file. So we're going to set our series adjusted count for the number of victimizations to zero. And then step three, in this case, because we're estimating a personal crime type, there's an additional step that needs to be done. We need to pull over some information that is only included on the household-level file. This includes the complex sample design information, the pseudostrata, and half sample code that we need for estimation of TSL standard errors. And we also need an indicator of the state, which is the state FIPS code in this variable called State.

You'll also note that I'm creating numeric versions of many other variables on the internal files within the RDC are character, and so I'm just converting some of them here and renaming them to a numeric variable. Unique household interviews are identified by the variables year quarter and control number. So that's what we are using to merge information from the household file onto the person file. And the person-level file only contains persons within responding households, so we're also going to drop records for nonresponding households when merging. And so that's what we're doing with the "if in 1" statement. In the fourth step, we do a few things to prepare the file for estimation. The first, I mentioned earlier, the person and household weights might need to be adjusted. So here we are dividing the person weight by two so that we have the adjusted collection year of person weight. Because we can only use one weight when calculating estimates, we also need to figure out how to incorporate the victimization weight. And so we do that by creating an adjustment factor that's equal to the ratio of the victimization weight and the adjusted collection year person weight we just created. If a respondent is not a victim of violent crime, in this example, then this adjustment factor will just equal zero.

And then the last variable we create is used to calculate the numerator of the victimization rate, and it's equal to the number of violent incidents reported, and remember this is summarized by person and reporting period, multiplied by the adjustment factor or the ratio of the victimization weight and the person weight. And then we also multiply it by 1,000 so that our rates will be expressed per 1,000 persons. And then in the final step, we calculate the victimization rate and standard error. And, here, we're using the survey means procedure in SAS. On the first line, we specified the variance method, which is Taylor, since we want TSL standard errors, and then we're also specifying the statistics being requested in this case, the mean and the standard error. We next specify the complex sample design information on the strata and cluster statements. On the domain statement, we specify that we want estimates for California, so the state FIPS code for California is specified there. The adjusted collection year person weight is then specified on the weight statement. And, finally, the analysis variable created in the previous step is specified on the var statement. And this will produce the violent victimization rate in California from 2017 to 2019, as well as the standard error of the victimization rate. And now I will pass it back to Rachel.

DR. RACHEL MORGAN: Thank you, Andrew. All right. We're going to close out with how to access NCVS restricted-use data. So as of December 2022, there's a new application process for applying for NCVS restricted-use data files, or we call them RUF. So the PUFs and the RUFs. So prior to December 2022, researchers interested in applying for access had to submit an application through the Census Bureau's Federal Statistical Research Data Centers or FSRDCs. However, in the Foundations for Evidence-Based Policymaking Act of 2018, federal agencies were required to develop and implement a standardized process for access to all federal restricted-use data files. So, as I said, this new process launched in December. This effort is called ResearchDataGov. And here is a screenshot of what the website looks like. And it's a web portal for discovering data and then requesting access to restricted microdata from federal statistical agencies, and then also for a couple other federal units as well. You can see a list of all the agencies below. So the web link is researchdatagov.org. BJS will accept applications for restricted data to support projects with a demonstrated statistical or research purpose, including for evidence building.

So, to date, NCVS restricted files included for approval and access are the core NCVS from 2005 to '15. We're working on adding the 2016 and later years now that we've released this report. The Identity Theft Supplement to the NCVS, the Police Public Contact Survey, the School Crime Supplement, the Supplemental Fraud Survey, and then the Supplemental Victimization Survey on stalking. So all of these are NCVS supplemental surveys. They are also available on ResearchDataGov. So to complete an application, applicants must provide information, including but not limited to, researcher information, so this is contact information and institutional affiliation; project information, including research questions, project abstracts, and data files requested; and then dissemination information, including project publications, products, and requested output. And there is a user guide available on RDG that has more details about what's required. So if you're interested, I would suggest reviewing that. And then depending on the data you're requesting from a specific agency, additional data security requirements may be required for applicants after you've been approved. So BJS does have a few different data security requirements as well.

And then once an application is submitted, BJS has 12 weeks to review an application and issue a determination. If you've applied for NCVS restricted-use data in the past, you know that it could take a number of months to get access. It's still the case here. So just know that before going into this application process. But I must mention that there are no changes to public-use files or BJS data products as a result of this new standardized process for accessing restricted data. So all BJS public-use data files are still available from the National Archive of Criminal Justice Data through ICPSR. Interested researchers are strongly encouraged to look at these public-use data file codebooks and determine if their research questions can be answered using these files instead of restricted-use files, because you can immediately download them and analyze the data. BJS data tools, data tables, and statistical reports are still available from the BJS website. So the only thing that has changed is applying for access and gaining access to the restricted-use data.

Please reach out to BJS if you are interested in using restricted NCVS data. If you reviewed the PUF codebook and still think your research question needs the restricted-use file, please reach out to us to get feedback before you go down the road of submitting an application and waiting 12 weeks for someone to review. So we suggest emailing AskBJS@usdoj.gov with SAP. This is the name of the effort. It's the standard application process. In the subject line and then include a brief abstract about your research project. You will be connected with a statistician based on your area of interest. And if you're also interested, let me just say, not just in NCVS restricted data but other types of BJS restricted data, you are still welcome to reach out to AskBJS with a synopsis of your abstract and they'll connect you to someone that works on the other BJS restricted data as well. More information is available on the BJS website. And that is it. Here is all of our contact information. And I will turn it back over to Grace to start moderating Q&A.

GRACE KENA: Thank you, Rachel, and thank you, everyone. So thanks to our audience for all of the great questions. I will kind of group them as many have common themes. The first one that I'll speak to is just the FSRDC and the cost. That is definitely a consideration that's out of our hands, unfortunately. But I would say to consider partnering. This is a little bit of a specific question and we can certainly talk more offline with you. But partnering with other researchers--there are many veterans of NCVS in the audience today. Thanks to you all for joining us. And so I'm sure many of them would be happy to speak further and give any insights or guidance around that. Other questions relate to--hopefully, the user's guide overview that Andrew gave was helpful in how to use the data, but one thing that we also will provide some resources on are just general webinars that the victimization team has done on using the NCVS at the national level that are going to be very applicable and helpful for drilling down into this state-level work, so we'll put those into the chat as well. I will turn over to Marcus, who is also a resident expert on NIBRS, about why these estimates differ from the UCR and NIBRS estimates. And, Andrew, please also feel free to add anything to that.

DR. MARCUS BERZOFSKY: NIBRS and the NCVS differ for a few reasons. First, and I know Andrew tried to account for as many of these as possible. The UCR is reported to police only and the NCVS covers both reported and nonreported. Obviously, you can subset the NCVS to reported only as well. And the second are the ages. The NCVS covers all persons age 12 and older. The UCR covers all ages. Again, you can subset that as well in the NCVS to try to make apples to apples with the UCR. And even with those two, as I guess the question asks, you will still find some differences. And the main reasons, I think, will come down to the error sources that Andrew went over in his presentation. I mean, the first and foremost being sampling error. The NCVS is a sample, whereas the UCR, to the best of its ability, is administrative data where all agencies submit reported data to the extent that they have all agencies reporting. The NCVS will have a confidence interval around it and the UCR won't. So, in many cases, once you account for that confidence interval, the estimates are more similar. That won't always be the case but in some. And in those cases where it's still not similar, then I would go to the nonsampling errors that we talked about, whether it's coverage.

Measurement is also an issue with the NCVS versus the UCR. How the different offenses are measured and defined can cause certainly some differences, as well as nonresponse in the NCVS and things like that. So, in general, most of those differences can boil down to those reasons. Unfortunately, and for every estimate every time, you can't pinpoint the exact reason. But, in general, those would be the likely culprits for why the estimates are different.

GRACE KENA: Andrew, did you have anything to add?

ANDREW MOORE: I guess I'll add, for the comparisons we did, we were using the Summary Reporting System, or Summary Reporting System estimates. We looked into using NIBRS data, but I think like 14 of the 22 states had a population coverage of less than 70% at that time when we were doing the analysis, and so we did compare it with the Summary Reporting System, which I think limited us a little bit in trying to make the estimates as similar as possible, like Marcus was saying, in terms of the age categories included and things like that.

GRACE KENA: Thank you. A question for Rachel. "Why were these 22 states picked?"

DR. RACHEL MORGAN: These 22 states accounted for the majority of the person population in 2016 and then also the majority of violent victimizations across the country in 2016.

GRACE KENA: Thank you. Please also feel free, everyone, to continue to post any additional questions into the chat. Here's a question, Andrew, for you. "What kinds of statistics, types of crime broken down by other variables, can be obtained for these 22 largest states?"

ANDREW MOORE: Sorry. Are you asking which ones are currently available within the report the BJS just released or...

GRACE KENA: I think the question is more general around what kinds of statistics can be produced, just in more detail, using the available data for the 22 largest states?

ANDREW MOORE: I think we found that some overall crime types, and for larger subpopulations, you can calculate race totals, proportions using three years of data in most of those states and get reliable estimates. I think it really depends on--and that was I think one of the points I made earlier, is that we weren't trying to assess specific estimates for a given crime type, a given state, a given subpopulation. We were more evaluating the overall methods that were used. I think it's kind of up to the data user to repeat some of those things for the specific analysis that they're interested in and see if there are any issues and whether they need to add more data or maybe collapse a category, or maybe the estimate just isn't feasible given the number of years of data we currently have available and current sample sizes.

GRACE KENA: Thank you. Marcus, anything to add? And Rachel as well. Rachel, in terms of other NCVS data that are available in specific crime types, for example, there was a question on hate crime and analyzing data on that topic for the 22 largest states. And then, Marcus, after that, any research opportunities that might exist with the NIBRS data?

DR. RACHEL MORGAN: I would think we want to do hate crime statistics at some point, but we just need to get more years of data and figure out what's feasible and how many years we need to provide reliable estimates. But I think this is our first release of this, so there's a lot we can do and a lot we want to do, so I would say stay tuned.

DR. MARCUS BERZOFSKY: So in terms, first, hate crime and then other ways that the NCVS and the UCR can be used to complement each other. Hate crime in general--and this dovetails nicely into what Andrew was talking about in his last response. It's a rare event, right? And so as Andrew was saying, with any rare event, it's going to be really challenging as sample sizes gets smaller. And that's true with the UCR too. The UCR hate crime data is also challenging to use because of reporting issues and just the rarity of those events. So I know those are even challenging within the UCR as well. So at the state level, which we're talking about here today, I think with the NCVS it's as Rachel said, I think that's--more to come on that in terms of how they pool that and make more accurate estimates at the state level. So more broadly though, in terms of how the NCVS and the UCR can be used together, I think they can actually be really complementary. So, as we said, there are distinct differences. The biggest being the NCVS being able to measure the dark figure of crime, the unreported crime. So I think, to the extent that you can take the NCVS reported data compared to the UCR and try to see if those estimates are similar for a certain crime type, you then--and let's say--just for argument's sake, say there are, that can then allow you to then sort of dovetail to the unreported and sort of approximate, in terms of these UCR counts, how those might differ. And so you can take the detail in the NCVS and apply some of the dark figure crimes to the UCR data. NIBRS, you know, obviously has a lot more detail as well, and so compared to the Summary Reporting System data that Andrew has mentioned that they were using. As time goes on, I think both are going to be able to be used together more. So NIBRS is making a real push to have better coverage.

As Andrew said, a few years ago most states just had like 70% coverage or better. That is improving each year now as more and more agencies are transitioning. And I think once that happens, I think the two will be able to complement each other very nicely in terms of what they measure, how they measure. For example, the NCVS measures a lot more in terms of services provided and things like that. So that information can be used to augment some information that's in the UCR and the NIBRS data that isn't in the NCVS. So I think the two can work together and I think that's how it's probably best to be able to be used. And the last thing I will say on that is that they can be used to complement each other and, I think, also to validate each other. As far as Andrew showed, they are both trying to measure the same basic thing, and so they can help you gauge whether or not the estimates you're providing are reasonable and accurate in

terms of what you're trying to estimate and present and any sort of research that you're doing.

GRACE KENA: Thank you. There are a couple of questions on whether there are plans to produce estimates for the largest cities or counties and then for any updates on the small area work. I can speak to that briefly. Thank you to both of the requestors for that information--for those questions rather. So, yes, BJS is continuing to do work in this area. Heather Brotsos would be the best person to talk more about that continued analysis and work to build out this program. And then also just to add a little bit to my earlier comments about accessing the state-level data, as I mentioned earlier, there are other types of subnational data that can be accessed, including some of the small area work that don't require accessing the restricted-use data. So please do look at the subnational website to see if there are other types of data that might meet particular research needs that wouldn't require accessing the restricted-use data. So just a plug for those other pieces, because the founders of this program knew that every type of data or analysis wouldn't meet every person's needs, and so the idea was to form different strategies to provide different options for people depending on needs and resources, etc.

There is another question about understanding how the estimates affect budgets for the NCVS to assist victims. I'm not sure if this question is referencing the Crime Victims Fund. If so, the Office for Victims of Crime oversees that fund, it isn't something that BJS oversees. So, hopefully, I've addressed the question there. And then there is another question on why Washington appears to have the highest crime rate in the new NCVS analysis but they're much lower if you look at state rankings using UCR and NIBRS. I think we've touched on some of that, but another point that I just want to add is that, you'll note that we were very deliberate in our report not to compare the states against each other and that we also included a confidence interval just to show that, you know, there's a range of what the true estimate could be. And so that's something that we really want to keep in mind when we're looking at these state-level analyses. But, again, also keep in mind the differences between the two collections and why we might expect to see some of those things. There is a question about ResearchDataGov and where do you find the subnational data to apply for access? Rachel, do you want to take that one?

DR. RACHEL MORGAN: Yes. So you would select the NCVS in there. It's just the core NCVS. And then once you get approved and get access, you would have the state FIPS codes as Andrew was talking about. Right now, it's only available through 2015 and so, you know, if you want to send us an email, we can give you a better timeline for when the more recent data will be available, but, yeah, it's just accessing the core data. That's what you would request.

GRACE KENA: Okay. Thank you. Any other questions from the audience? Tammy just placed also in the chat some of the general resources on analyzing NCVS data. So, hopefully, those of you who are not familiar with those will find them helpful. If there is nothing else then we can give everyone 10 minutes back in their day. Thank you so

much for joining us today. We really appreciate it and I hope that you find this helpful. Please, again, reach out to any of us with any questions. The slides and video for this webinar will be posted on the BJS website in the coming weeks. Thank you.

DARYL FOX: Great. Thanks so much. So on behalf of the Bureau of Justice Statistics and our panelists, we want to thank you for joining today's webinar. This will end today's presentation.