

Welcome to the National Crime Victimization Survey or NCVS Knowledge Corner. This short video is focused on weighting the NCVS data. This video briefly discusses sample weights and focuses primarily on NCVS-specific weights. Self-study is encouraged for data users that need further knowledge or information about weighting sample data in general.

A sample weight is an adjustment that is applied to sample data to make the data representative of a specific population of interest. For the NCVS, sample weights are applied to make the sample data representative of the U.S. population of households or persons ages 12 or older.

Sample data that are not weighted, or unweighted data, can be biased as a result of who completed and did not complete the survey. Unweighted data are not representative and cannot be used to generalize about specific populations of interest. BJS does not report out on unweighted sample data. Statisticians examine unweighted data to understand sample characteristics and assess sample sizes for estimate reliability. All NCVS statistical estimates that are published in BJS reports are weighted to be representative of the U.S. population. The NCVS has four types of weights: household, person, incident, and victimization. And knowing when to use each type of weight is critical for your analysis.

Each NCVS annual data file is comprised of two 6-month data files. So, the NACJD, the National Archive of Criminal Justice Data, who processes and publishes the NCVS data files, receives two 6-month incident files, two 6-month household files, and two 6-month person files and then combines them together to create annual data files. Because these 6-month files are standalone files and include persons and household weights that are representative of the entire population, the person and household weights must be adjusted when combined, or they would be doubling the population when they're combined. So the first household weight on the file is V2116, and this adjusts each 6-month period to a total count of U.S. households. However, in order to avoid doubling the household population, we have to divide V2116 by 2 to get an annual count of U.S. households. So we use this weight, WGTHHCY, which means weight, household, collection year, to generate an annual account of U.S. households when calculating property crime rates. So an example from the 2020 NCVS household file, you can see that for the first row V2116, the weight value is 2178, but we divide that in half to get 1089 for WGTHHCY.

The process is similar for the NCVS person weights that are included on the person data file. The original weight on the file, V3080, adjusts each 6-month period to U.S. population counts of persons ages 12 or older. But in order to not double that, when we combine to make an annual file, we adjust that weight dividing it in half. And the resulting weight is WGTPERCY, we call it -- WGT, PER for person, and then collection year. So we use this weight to generate the annual count of U.S. persons ages 12 or older when calculating violent crime or personal larceny rates. So an example from the 2020 NCVS person file, we can see in row four the weight value for 3080 is 26355 and dividing that in half we get 13177 for WGTPERCY.

So before I get into the details on the differences between the incident and victimization weights, I must explain the difference between these two crime measures. An incident is the number of specific criminal acts involving one or more victims. A victimization is the total number of times that people or households were victimized by crime. So if every incident had one victim, the number of incidents would be equal to the number of victimizations. However we know that's not the case, not every crime incident has one victim. So for example, let's say someone was walking down the street with a friend and they were robbed. That would be one robbery incident because they were together when the crime occurred, but two victimizations because there were two victims involved in that incident. So in this case, there was more than one victim so the incident weight is the victimization weight divided by the number of victims in the incident, which is two. So we don't want the incident to be reported multiple times by each of the victims and then overcounted, so we have to adjust.

The number of victimizations is the number of victims of a violent or personal larceny crime, and then for property crimes each crime has a single victim, which is the affected household since we're counting property crimes at the household level. Only respondents that reported an incident or a victimization are included in the incident data file. And please watch the NCVS Knowledge Corner video on data file structure for more information on the NCVS data files. Data users must decide if they are interested in analyzing incidents or victimizations and then apply the appropriate weight. You could use either, it really just depends on your research question and what you're interested in learning with the NCVS data.

So there are two victimization weights in the incident data file, WGTVICCY and then SERIES\_WEIGHT. So, WGTVICCY provides a total count of victimization but does not account for series crimes. Series crimes are 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 each event in detail. For example, intimate partner violence or domestic violence, which unfortunately can occur multiple times in a short period of time and a victim may not be able to describe each event separately. The series weight accounts for these high-frequency repeat victimizations and then caps them at a maximum of 10 in order to avoid extreme outliers in the data. So, series crimes don't make up a large percentage of crimes in 2020. They were less than 3% of violent victimizations, were considered series victimization. And if the crime was not a series victimization the weight values for WGTVICCY, and SERIES\_WEIGHT would be equal. So we can see here as an example from the 2020 data file, a nonseries weight would be 585. Because it wasn't a series crime, the series weight would be the same. And then if it was a series crime, so this particular person experienced 12 victimizations, but we capped it at 10 so their original weight was 750, and then multiplied by 10 would be 7506.

So there are also two incident weights on the incident data file, which is similar to the victimization weights that I just discussed. One doesn't have an adjustment for series crimes, this one is V4527, provides the total count of incidents. And then one is adjusted for series crimes, this is SERIES\_IWEIGHT for the I -- meaning incident -- whereas the one I just talked about before the victimization series weight was just

SERIES\_WEIGHT. So this one accounts for high-frequency repeat incidents and caps it at 10 just like I talked about before to avoid the extreme outlier effect. Again, if it's not a series incident, these weight values would be equal in the data. Remember that the difference between victimizations and incidents is the number of victims involved. So, for here for the nonseries weight 4527, 585, the same for SERIES\_IWEIGHT, and then again this series person experienced 12 incidents, but it was capped at 10 to avoid being an outlier in the data. And so the original weight was 375, and then multiplied by 10 would be 3753.

So in 2020, the household weights ranged from 61 to about 10,000, the person weights ranged from 72 to 17,000, and the series victimization weight ranged from 151 to 77,560. So in other words, a weighted count of about 77,000 victimizations could be based on an n of 1, so one person or hundreds of people really depending on the characteristics of the victim in order to make these sample data representative of the U.S. population. So it's always important to check unweighted counts to know if you're dealing with an n of 1, an n of 2, or you have sufficient sample size, and BJS statisticians ensure that all the published estimates that we put out meet data quality standards. So, in certain cases we may have to flag estimates or suppress them and not show them if they don't meet a minimum sample size or coefficient of variation that is above a certain threshold.

So, in closing if there are further questions please see the NCVS page on the BJS website. We have an NCVS technical documentation that's also available that has more information about the methodology, specifically the weights that are included on the data files. You're welcome to email [askbjs@usdoj.gov](mailto:askbjs@usdoj.gov) with additional questions and those emails are forwarded to BJS statisticians that have topical expertise. Make sure to mention the NCVS so it comes to myself or one of my colleagues. And then the NCVS public use data files are available at the National Archive of Criminal Justice Data. And there's a lot more information on the NCVS and NCVS methodology in the annual code books.

Thank you.