Document title:	Bayesian Estimation of State-Level Crime Victimization Rates
Author:	Emily Berg, PhD, Iowa State University
BJS Project Manager(s):	Heather Brotsos, Deputy Director of Statistical Programs Rachel E. Morgan, PhD, Chief, Victimization Statistics Unit
Document No.:	NCJ 308785
Publication Date:	November 25, 2025

Abstract:

This report was produced under an Intergovernmental Personnel Act (IPA) agreement, authored by Dr. Emily Berg of Iowa State University. The report describes the feasibility of using model-based techniques to produce estimates of state-level crime victimization rates. Specifically, a Bayesian multivariate lognormal model was applied. Direct estimates from the National Crime Victimization Survey served as model response variables. Covariate information was derived from the FBI's Uniform Crime Reporting Program. The model incorporates multiple crime types and data for two aggregate time periods. This report presents results from applying the model to produce state-level estimates of crime victimization rates.

Disclaimer

This page is intentionally left blank.

Bayesian Estimation of State-Level Crime Victimization Rates

Emily Berg, PhD
Iowa State University

November 2025

Prepared under an Intergovernmental Personnel Act Agreement with:

Bureau of Justice
Statistics 999 N. Capitol
St., NE Washington, DC
20531

Contents

Executive Summary

The Bureau of Justice Statistics (BJS) regularly publishes direct statistical estimates of nonfatal violent and property victimization in the United States based on data from the National Crime Victimization Survey (NCVS). Due to small sample sizes, direct estimates may not be possible or appropriate for smaller geographic levels, such as states. However, state-level estimates are important for research and policy purposes. To meet the demands of data users, the NCVS sample was boosted and redesigned in 2016. The redesign permitted publication of direct estimates for the 22 largest states (Kena and Morgan, 2023). The use of model-based procedures for producing state-level estimates is of interest for smaller states due to the instability of direct estimates. The operation of using model-based estimates to improve upon direct estimates is generally referred to as small area estimation (Rao and Molina, 2015).

This document examines the feasibility of using small area estimation techniques to produce estimates of crime victimization rates for all 50 states and the District of Columbia.[1] Estimates are produced separately for violent crime and property crime. The types of violent crime included are simple assault and violent crime excluding simple assault. In this case, violent crimes include crimes of rape or sexual assault, robbery, and aggravated assault. Similarly, property crime is divided into burglary/trespassing and all remaining theft which includes motor vehicle theft and other types of household theft. The temporal reference periods are the two aggregate time frames of 2017–2019 and 2020–2022.

The specific model employed is called a multivariate lognormal model. The multivariate component enables production of unified estimates for multiple types of crime victimization rates during the two aggregate time frames of interest. The log transformation improves the fit of the model and benchmarking ensures the aggregation of state-level estimates is compatible with the direct estimate of the national-level crime rate. Bayesian inference procedures are used. The Bayesian paradigm easily accommodates the log transformation and reflects the effect of benchmarking on measures of uncertainty.

The multivariate lognormal model uses two sources of input data: response variables and covariates. The response variables are direct estimates from the NCVS. The covariates are derived from the FBI's Uniform Crime Reporting (UCR) Program. By synthesizing these two types of information, the model can produce estimates that are more reliable than what can be deduced from the NCVS alone.

This report highlights the results from applying the multivariate lognormal model using NCVS and UCR data to produce state-level estimates of criminal victimization. The state-level estimates vary considerably around the U.S. estimate for both violent crime and property crime. Significant declines in the property crime rate at the state level are observed between the 2017–2019 and 2020–2022 aggregate time frames. Coefficients of variation (CVs) for the state-level estimates are generally in the range of 10% to 30%. These CVs are usually below the CVs of the direct estimators.

The adequacy of the model fit is evaluated using three criteria: standardized residuals, posterior predictive p-values, and benchmarking ratio adjustment factors. The standardized residuals display no systematic trends as a function of the fitted values. Posterior predictive p-values (a Bayesian measure of

[1]The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release. Approval #CBDRB-FY24-POP001-0021.

goodness of fit) are close to 50%, indicating that the data generated from the model are compatible with the observed data. The raking ratio adjustment factors are exceedingly close to one. None of the model diagnostics raises concern about the adequacy of the model for the NCVS data.

Despite the positive qualities of the proposed model, several challenges emerge from the analysis. The most important issue is that the UCR data provide only weak covariate information, as many of the regression coefficients do not differ significantly from zero. Also, direct estimates of zero for some states or time periods present a problem that is not fully resolved by the model. Finally, evidence of spatial dependence exists for simple assault.

The primary recommendation based on this analysis is to consider modeling a 6-year average. This would lead to a natural estimation cycle. Direct estimates would be published for the 22 states and 3-year aggregate time frames, as in Kena and Morgan (2023). These estimates would be paired with model-based estimates for all states over a 6-year time frame.

Introduction

The importance and demand for subnational estimates of crime have been well documented by the research community, policymakers, and data users. The challenge in producing subnational estimates is that standard survey estimators (i.e., direct estimators) can suffer from instability due to small sample sizes. One way to overcome the limitations associated with direct estimates is to use model-based estimates when producing subnational statistics. Model-based estimates incorporate auxiliary information and stronger assumptions. Under the model assumptions, model-based estimates are more efficient than direct estimates. This report develops and evaluates estimates of state-level crime rates based on a multivariate lognormal model. For more information about the development of this model, see the *Background on formulation of multivariate lognormal model* textbox.

Model-based estimates for violent crime and property crime are provided for two 3-year aggregate time frames: 2017–2019 and 2020–2022. Violent crime is disaggregated as simple assault and violent crime excluding simple assault, where violent crime excluding simple assault includes robbery, aggravated assault, and rape/sexual assault. Property crime is separated into burglary/trespassing and all remaining theft, which includes motor vehicle theft and other types of household theft.

- State-level estimates vary considerably around the national estimate for both violent crime and property crime.
- Overall property crime and burglary/trespassing decline significantly for several states between the 2017–2019 and 2020–2022 time frames. In contrast, few significant changes occur for state-level violent crime between the two time frames.
- The model typically provides coefficients of variation in the range of about 10% to 30%.

Development of the multivariate lognormal model is a nuanced task. However, rather than focus on model formulation, this report emphasizes the results. First, state-level estimates of violent crime and property crime rates are presented, followed by an evaluation of the model, and a discussion of the challenges. The report also provides a summary and recommendations, along with a discussion of the technical details of the model.

Data sources

The results in this report are based on data from the National Crime Victimization Survey (NCVS). The NCVS is a probability-based survey sponsored by the Bureau of Justice Statistics (BJS) and administered by the U.S. Census Bureau. The NCVS collects data from persons ages 12 or older from a nationally representative sample of U.S. households. The NCVS provides direct estimates for many types of nonfatal personal crime (i.e., rape or sexual assault, robbery, aggravated assault, simple assault, and personal larceny) and household property crime (i.e., burglary/trespassing, motor vehicle theft, and other types of household theft). See the BJS NCVS webpage for more information at https://bjs.ojp.gov/programs/ncvs and the NCVS Dashboard Terms & Definitions webpage at https://ncvs.bjs.ojp.gov/terms.

In addition to the NCVS data, the model integrates data from the FBI's Uniform Crime Reporting (UCR) Program. Data from the UCR provide estimates of crime rates based on information submitted to the FBI by states and local law enforcement agencies. The UCR data have potential to aid in estimating crime rates if the correlation between the NCVS and the UCR crime rates is sufficiently high. More information about the UCR Program is available at https://www.fbi.gov/how-we-can-help-you/more-fbi-services-and-information/ucr.

Background on the formulation of the multivariate lognormal model

The Bureau of Justice Statistics (BJS) publishes direct estimates based on the National Crime Victimization Survey (NCVS). A direct estimate is a weighted sum of sampled units where the weights reflect the selection probabilities and nonresponse adjustments. Historically, these direct estimates have been at the national level. This practice restricted the utility of the survey for policymakers and researchers who often request subnational statistics.

To meet the demand for subnational data, the NCVS sample was boosted and reallocated in 2016 to improve efficiency for state-level estimates. This facilitated the publication of direct estimates for the 22 largest states in Kena and Morgan (2023). The work of Kena and Morgan (2023) was pioneering because it marked the first publication of direct estimates at the state level by BJS.

Despite the sample redesign, many issues for subnational estimation remain. In particular, estimates are desired for the 28 states that did not receive the sample boost. Sample sizes in those states are too small to support reliable direct estimates.

To overcome the challenges with direct estimates and meet the demand for subnational data, the National Academy of Sciences recommended that BJS "investigate the use of modeling NCVS data to construct and disseminate subnational estimates of major crime types" (National Research Council, 2008). This led BJS to invest in considerable research on how to best produce subnational estimates (See the BJS NCVS Subnational Estimates Program webpage at https://bjs.ojp.gov/subnational-estimates-program).

Building on past efforts, this report describes work conducted to develop and evaluate model-based estimates of state-level crime rates. The research for this report has been completed through an Intergovernmental Personnel Act (IPA) agreement with Iowa State University. As this research is ongoing, the specific estimates provided in this report are subject to change.

The model used for this analysis is a multivariate lognormal model. The model uses direct NCVS estimators as response variables. The FBI's Uniform Crime Reporting (UCR) Program data are incorporated as covariates (i.e., predictor variables that explain variation in the response). The multivariate component means that the model has multiple response variables. Using this method, estimates can be produced for multiple types of crime in a unified fashion. The lognormal component means that the model is fit in the log scale. An analysis of posterior predictive p-values indicated that the model fit the data better in the log scale than in the original scale. Bayesian methods are used for inference, and benchmarking ensures compatibility with direct national estimates.

Two separate models are used: one for violent crime and one for property crime. The response variables in the model for violent crime are overall violent crime and simple assault. An estimate for violent crime excluding simple assault is deduced from the estimates for the other two categories. Property crime, burglary/trespassing, and all remaining theft are modeled directly, and an estimate of overall property crime is subsequently constructed.

This report provides model-based estimates for all states, including the 22 states presented in Kena and Morgan (2023). The model-based estimates in this report differ from the direct estimates provided in Kena and Morgan (2023). This is to be expected, as the estimates are constructed under different paradigms. Data users are advised to synthesize all available information, recognizing that the estimates in this report are based on an ongoing research project.

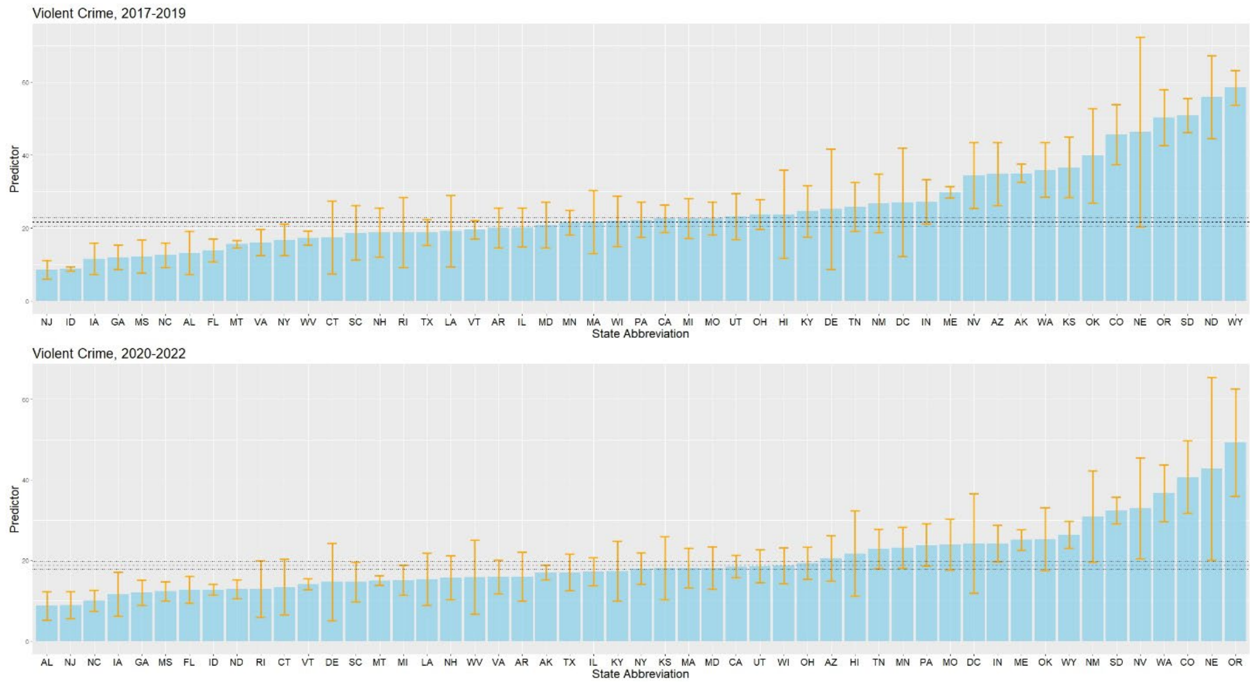
Violent Crime

Comparisons of state estimates to the U.S. average

State estimates of the victimization rates for overall violent crime, simple assault, and violent crime excluding simple assault are shown in Figures 1 to 3 below. The estimates are ordered by magnitude, with estimates for 2017–2019 presented in the top panel of each figure and estimates for 2020–2022 in the bottom panel. The state estimates are compared to the overall U.S. rate, represented by the dashed horizontal line. An estimate is regarded as significantly below the U.S. rate if the upper interval endpoint for the state estimate is below the lower interval endpoint for the U.S. estimate. Likewise, an estimate is considered significantly above the U.S. rate if the lower interval endpoint for the state estimate is above the upper interval endpoint for the U.S. estimate. The comparison of confidence intervals has an interpretation similar to a statistical significance test.

Figure 1 contains the ordered estimates of the overall violent crime victimization rates. Estimates of violent crime rates for 2017–2019 (in units of violent crime victimizations per 1,000 persons ages 12 and older) range from 8.5 violent crimes per 1,000 persons (SE 1.3) in New Jersey to 58.4 crimes per 1,000 persons (SE 2.4) in Wyoming. For comparison, the overall national estimate of the violent crime rate during 2017–2019 is 21.6 crimes per 1,000 persons. Eleven states (Alabama, Florida, Georgia, Idaho, Iowa, Mississippi, Montana, New Jersey, North Carolina, Virginia, and West Virginia) are significantly below the U.S. estimate. In contrast, 12 states (Alaska, Arizona, Colorado, Kansas, Maine, Nevada, North Dakota, Oklahoma, Oregon, South Dakota, Washington, and Wyoming) have estimates that are significantly above the U.S. estimate. Confidence intervals for remaining states overlap with the confidence interval for the U.S. estimated rate.

Figure 1: Ordered state estimates of victimization rates for overall violent crime with corresponding confidence intervals, 2017–2019 and 2020–2022



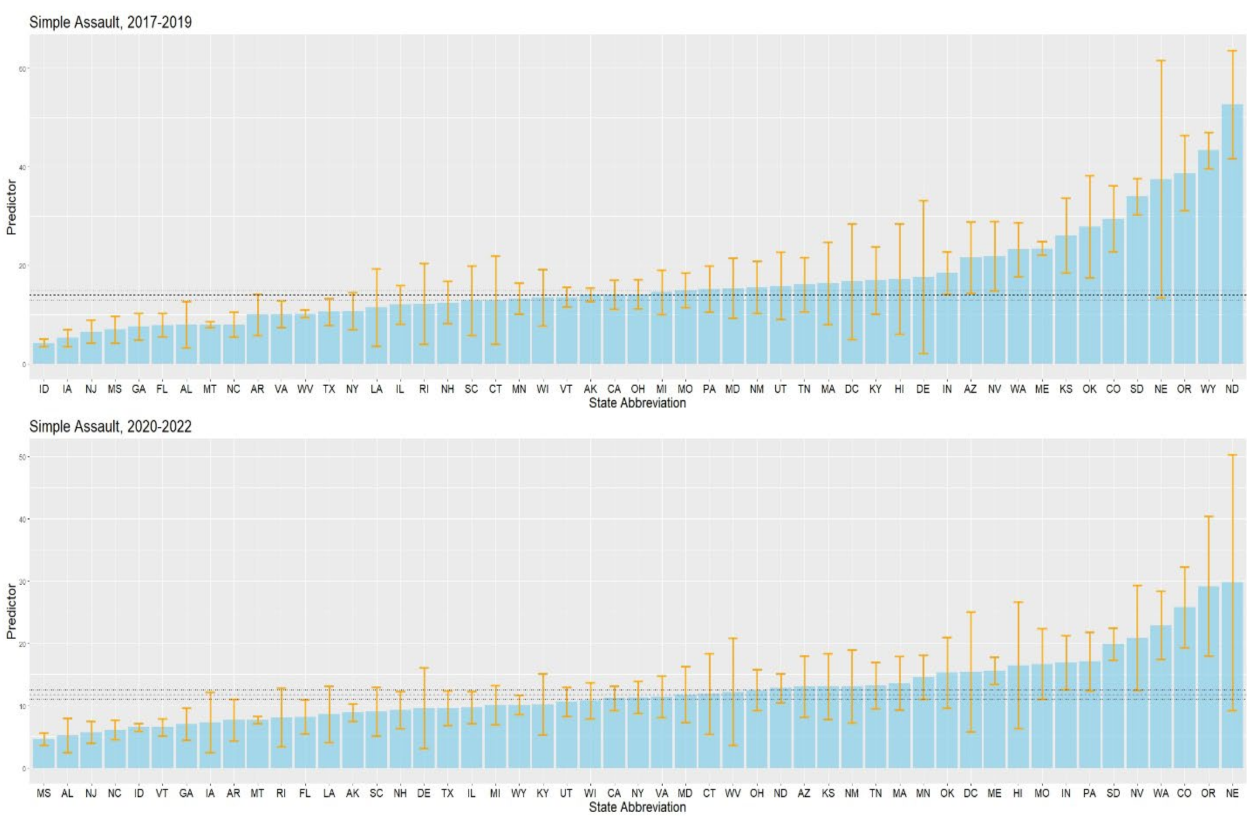
Note: Dashed line is U.S. estimate. Violent victimization includes rape or sexual assault, robbery, aggravated assault, and simple assault. Estimates include 95% confidence intervals. The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

The patterns observed from 2017–2019 are fairly similar to the patterns for the 2020–2022 time frame. During 2020–2022, the estimates of violent crime range from 8.7 crimes per 1,000 persons (SE 1.8) in Alabama to 49.3 crimes per 1,000 persons (SE 6.8) in Oregon. The U.S. rate during 2020–2022 is 18.8 crimes per 1,000 persons (SE 0.5). Similar to 2017–2019, 11 states (Alabama, Florida, Georgia, Idaho, Iowa, Mississippi, Montana, New Jersey, North Carolina, North Dakota, Vermont) are significantly below the U.S. estimate during 2020–2022. Among these states, North Dakota is the only one with a changing relationship compared to the U.S. rate. The change for North Dakota should be interpreted with caution because the 2020–2022 estimate for violent crime excluding simple assault in North Dakota is highly unstable. Eight states (Colorado, Maine, Nebraska, Nevada, Oregon, South Dakota, Washington, and Wyoming) are significantly above the national estimate during 2020–2022. For the remaining states, the confidence intervals overlap the confidence interval for the national rate.

The simple assault victimization rates are presented in Figure 2. Simple assault is the largest contributor to violent crime and the simple assault estimates follow similar patterns to the overall violent victimization estimates. The state with the highest simple assault rate during the aggregate period of 2017–2019 is North Dakota with 52.7 crimes per 1,000 persons (SE 5.6) and the state with the lowest simple assault rate is Idaho with 4.2 crimes per 1,000 persons (SE 0.4). For 2017–2019, the national estimate of the simple assault rate is 13.9 crimes per 1,000 persons (SE 0.5). Most of the same states with violent crime rates below the U.S. rate during 2017–2019 have simple assault rates below the corresponding rate during 2017–2019. These states are Alabama, Florida, Georgia, Idaho, Iowa, Mississippi, Montana, New Jersey, North Carolina, Virginia, and West Virginia. The 9 states with simple assault rates above the national rate (Colorado, Kansas, Maine, North Dakota, Oklahoma, Oregon, South Dakota, Washington, and Wyoming) overlap heavily with the 12 states that have violent crime rates above the corresponding U.S. rate during 2017–2019. The remaining states are not different from the overall U.S. estimate in that they have overlapping confidence intervals.

The patterns for simple assault during the aggregate period of 2020–2022 mirror the patterns observed during 2017–2019. The simple assault estimates during 2020–2022 range from 4.6 crimes per 1,000 persons in Mississippi (SE 0.5) to 29.8 crimes per 1,000 persons in Nebraska (SE 10.5). Ten states (Alabama, Alaska, Florida, Georgia, Idaho, Mississippi, Montana, New Jersey, North Carolina, and Vermont) have estimates significantly below the U.S. rate of 11.8 crimes per 1,000 persons (SE 0.5). Six states (Colorado, Indiana, Maine, Oregon, South Dakota, and Washington) have estimates significantly above the U.S. rate.

Figure 2: Ordered state estimates of victimization rates for simple assault with corresponding confidence intervals, 2017–2019 and 2020–2022



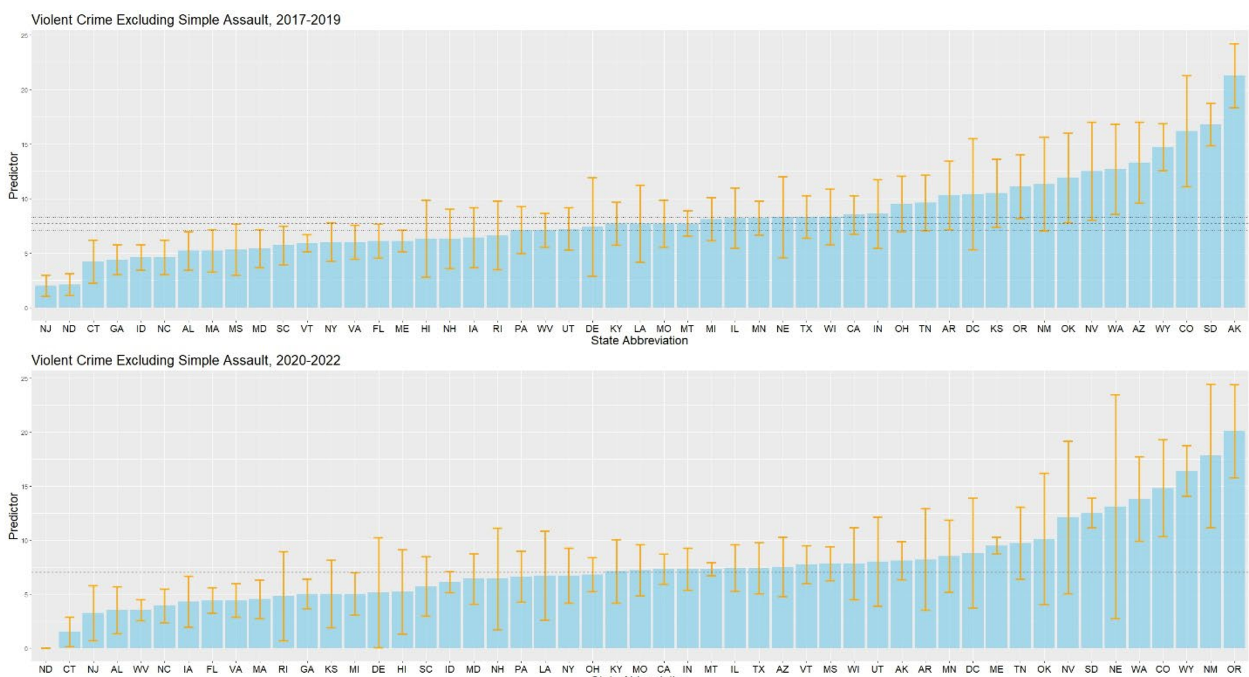
Note: Dashed line is U.S. estimate. Estimates include 95% confidence intervals. The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).

Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

Violent crime excluding simple assault is depicted in Figure 3. The estimates during 2017–2019 range from 2.0 crimes per 1,000 persons in New Jersey (SE 0.5) to 21.3 crimes per 1,000 persons in Alaska (SE 1.5). The corresponding U.S. estimate is 7.7 crimes per 1,000 persons (SE 0.3). The nine states with estimates significantly below the U.S. rate are Alabama, Connecticut, Georgia, Idaho, Maine, New Jersey, North Carolina, North Dakota, and Vermont. The six states with estimates significantly above the national estimate are Alaska, Arizona, Colorado, South Dakota, Washington, and Wyoming.

The U.S. estimate for violent crime excluding simple assault during 2020–2022 is 7.0 crimes per 1,000 persons (SE 0.2). The state with the highest estimate is Oregon, having a rate of 20.1 crimes per 1,000 persons (SE 2.2). Seventeen states have confidence intervals during 2020–2022 that do not overlap with the confidence interval for the national estimate. Of these, 10 states (Alabama, Connecticut, Florida, Georgia, Massachusetts, New Jersey, North Carolina, North Dakota, Virginia, and West Virginia) have rates below the U.S. rate, and 7 states (Colorado, Maine, New Mexico, Oregon, South Dakota, Washington, and Wyoming) have rates above the U.S. rate.

Figure 3: Ordered state estimates of victimization rates for violent crime excluding simple assault with corresponding confidence intervals, 2017–2019 and 2020–2022



Note: Dashed line is U.S. estimate. Violent crime excluding simple assault includes rape or sexual assault, robbery, and aggravated assault. Note that the direct estimate near zero for violent crime excluding simple assault for North Dakota in 2020–2022 is considered unreliable because the direct estimate is zero and the coefficient of variation for the model-based estimate is extreme. Estimates include 95% confidence intervals. The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).

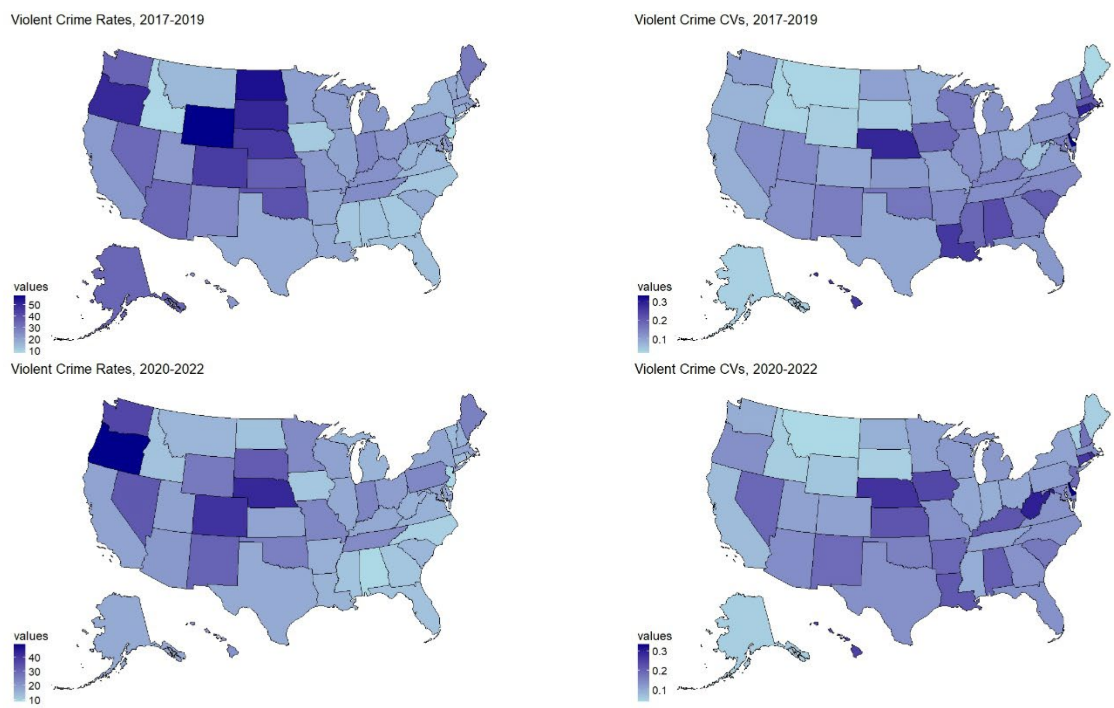
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020-–2022).

Geographic distribution of the estimates and coefficients of variation

The predictions for violent crime, simple assault, and violent crime excluding simple assault are presented in Figures 4, 5, and 6. The left panel has the predicted crime rates, and the right panel has the associated coefficients of variation (CV). The top panel in each figure represents the 2017–2019 period and the bottom panel represents the 2020–2022 period. In each plot, the magnitude of a value increases as the coloring darkens.

Figure 4 contains the estimates and associated CVs for violent crime. The estimated crime victimization rates range from approximately 10 to 50 crimes per 1,000 persons and the corresponding CVs range from approximately 10% to 30%. Relatively high crime rates in several states in the northwest and plains regions emerge. Lighter colors in the southeast region of the country suggest relatively lower crime rates. The estimates in the left panels of Figure 4 should be interpreted relative to the CVs provided in the right panels. For example, Louisiana has a relatively small estimated crime rate, but it also has a relatively high CV. The CVs exhibit an inconsistent association to the crime rates. Some states with relatively low crime rates have somewhat higher CVs, while the converse is true in other states.

Figure 4: Estimates of overall violent crime rates (left) and coefficients of variation (right), 2017–2019 and 2020–2022

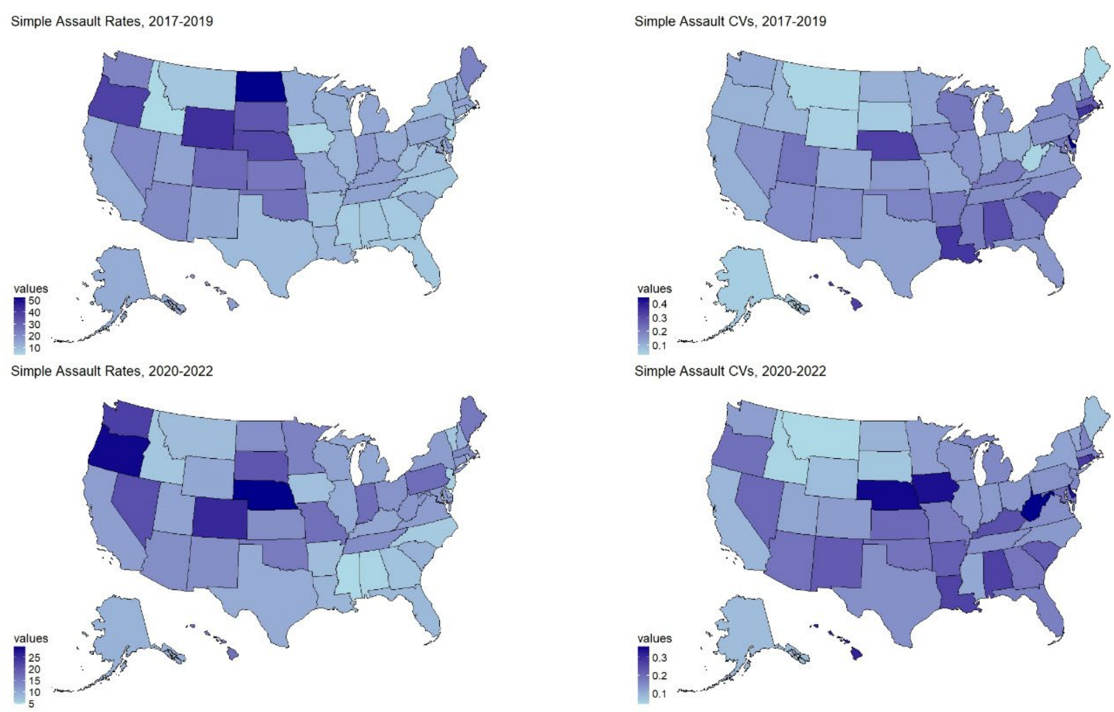


Note: Violent crime includes rape or sexual assault, robbery, aggravated assault, and simple assault. The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).

Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

Figure 5 depicts the estimates and associated CVs for simple assault. The geographic patterns for simple assault largely mirror the geographic patterns for violent crime, as shown in Figure 4. Several states in the plains region and coastal northwest have relatively high simple assault rates. A cluster of states in the southeast have relatively low estimated simple assault rates; however, these states also have relatively high CVs. Note that Nebraska has a relatively high estimate for simple assault in 2020–2022, but this state also has a fairly high CV.

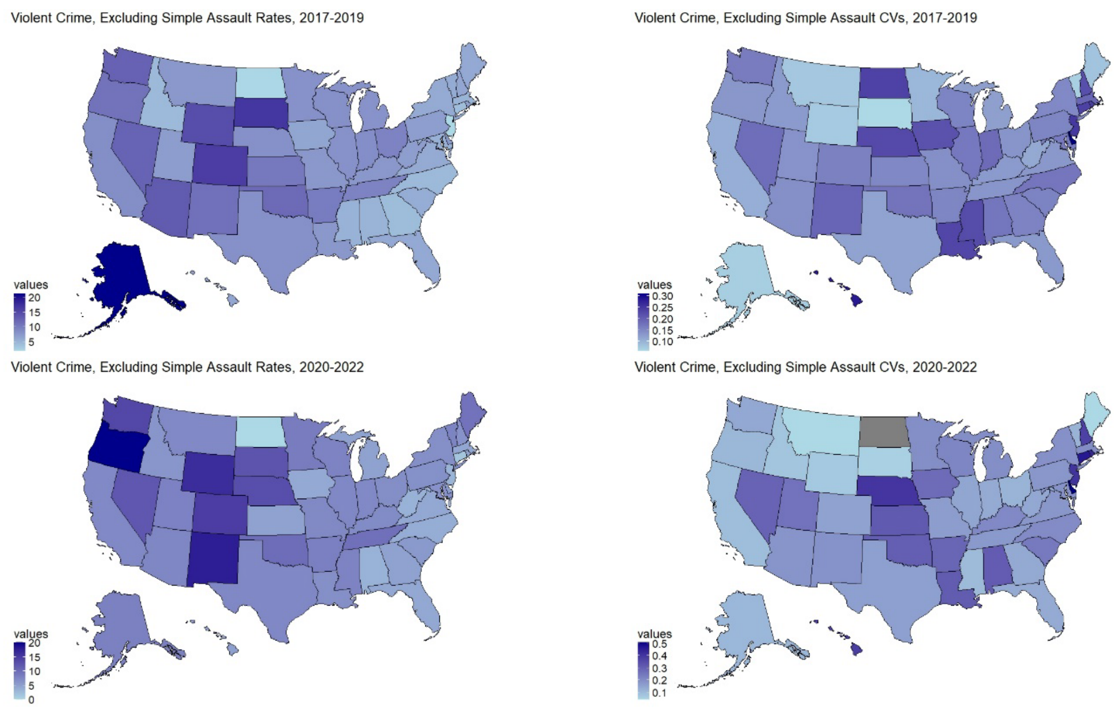
Figure 5: Estimates of simple assault rates (left) and coefficients of variation (right), 2017–2019 and 2020–2022



Note: The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

The estimates for violent crime excluding simple assault in Figure 6 also show relatively high crime rates in the plains and northwest coastal regions. The CV for violent crime excluding simple assault in North Dakota is suppressed as indicated by the olive color in the figure. This estimate is highly unstable due to the direct estimate of zero. Including the CV in the maps would distort the geographic depiction of the estimates for the remaining states. Some states in the northeast region (e.g., Rhode Island and New Jersey) that have relatively low estimated crime rates also have relatively high CVs.

Figure 6. Estimates of violent crime excluding simple assault rates (left) and coefficients of variation (right), 2017–2019 and 2020–2022



Note: Violent crime excluding simple assault includes rape or sexual assault, robbery, and aggravated assault. The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

Change over time for violent crime

Figure 7 presents the state-level differences between the estimates for the 3-year period of 2020–2022 and the estimates for 3-year period of 2017–2019. The states are listed in increasing order by the predicted change.

Most states exhibit no significant change in the crime rate between the two time periods. The confidence intervals for most states overlap zero heavily. This suggests that a slight increase in crime in 2022 has been off-set by a slight decline in the crime rate during the COVID-19 pandemic.

A few states appear to have significant changes (i.e., confidence intervals do not contain zero). For violent crime, significant changes are observed in Arizona, Arkansas, Kansas, Maine, Michigan, North Dakota, Oklahoma, South Dakota, Vermont, and Wyoming. For simple assault, significant changes are observed in Arkansas, Delaware, Kansas, North Dakota, Oklahoma, South Dakota, Vermont, and Wyoming. For violent crime excluding simple assault, only Arkansas changed significantly.

North Dakota shows a significant decline in the crime rates for overall violent crime and for simple assault during both the 2017–2019 and 2020–2022 time frames. This result should be interpreted with caution. For the 2020–2022 time frame, the direct estimate of violent crime excluding simple assault in North Dakota is zero, and consequently the model-based estimate is of poor quality. This can impact the results for the other crime types as all of the estimates are related. It is possible that the reported confidence intervals may understate the true degree of uncertainty for North Dakota. The reported result does not necessarily mean that North Dakota experienced a decline in violent crime between these two time frames.

Figure 7: Estimates of change over time with corresponding prediction intervals, 2017–2019 and 2020–2022

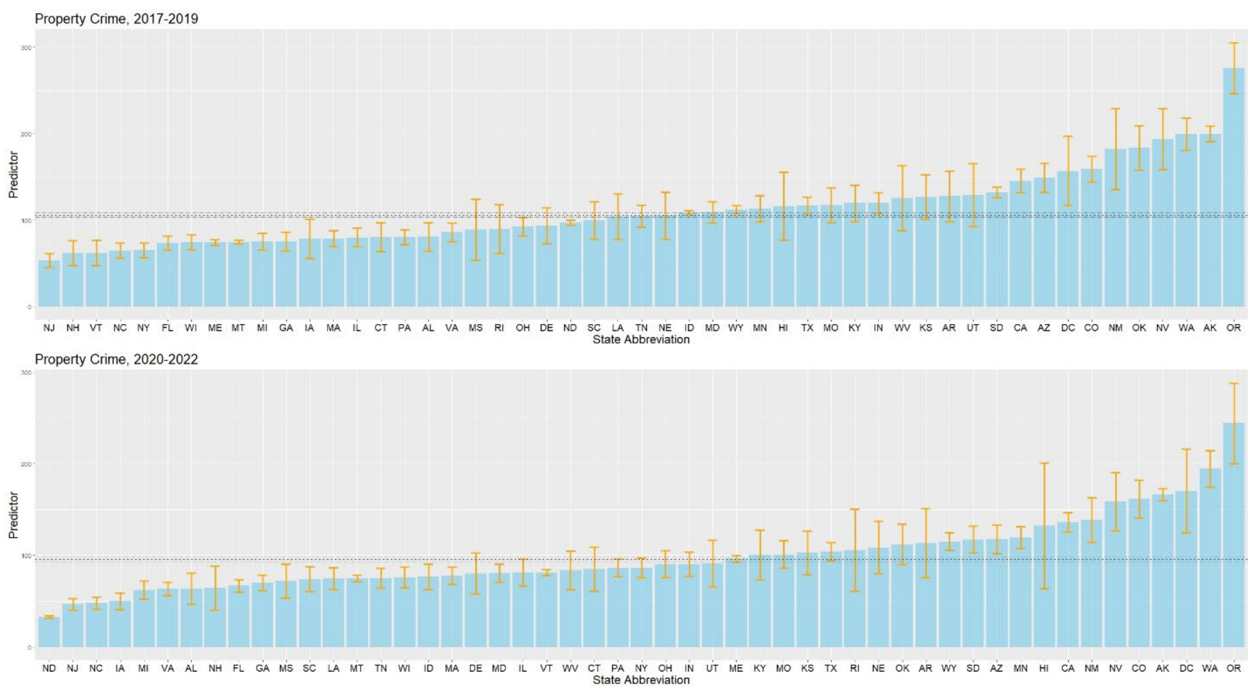
Note: Violent crime includes rape or sexual assault, robbery, aggravated assault, and simple assault. When interpreting the confidence intervals in Figure 7, the reader should recognize that Figure 7 includes numerous comparisons and the reported confidence intervals do not incorporate a multiple comparisons adjustment. In any collection of many comparisons, one expects a small number of spurious results that appear significant that may not actually differ significantly from zero. The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

Property Crime

Comparisons of state estimates to the U.S. average

Figure 8 contains the overall property crime victimization rates with corresponding confidence intervals for the aggregate periods of 2017–2019 (top) and 2020–2022 (bottom). During 2017–2019, estimates for property crime rates range from 53.0 crimes per 1,000 households (SE 4.0) in New Jersey to 276.1 crimes per 1,000 households (SE 15.1) in Oregon. In comparison, the U.S. estimate of the overall property crime rate is 105.9 crimes per 1,000 households (SE 1.3). Twenty states have property crime rates significantly below the U.S. estimate: Alabama, Connecticut, Florida, Georgia, Illinois, Iowa, Maine, Massachusetts, Michigan, Montana, New Hampshire, New Jersey, New York, North Carolina, North Dakota, Ohio, Pennsylvania, Vermont, Virginia, and Wisconsin. Eleven states have rates significantly above the U.S. rate: Alaska, Arizona, California, Colorado, the District of Columbia, Nevada, New Mexico, Oklahoma, Oregon, South Dakota, and Washington. The remaining states do not have property crime rates that differ significantly from the U.S. rate.

Figure 8: Ordered state estimates of victimization rates for overall property crime with corresponding confidence intervals, 2017–2019 and 2020–2022

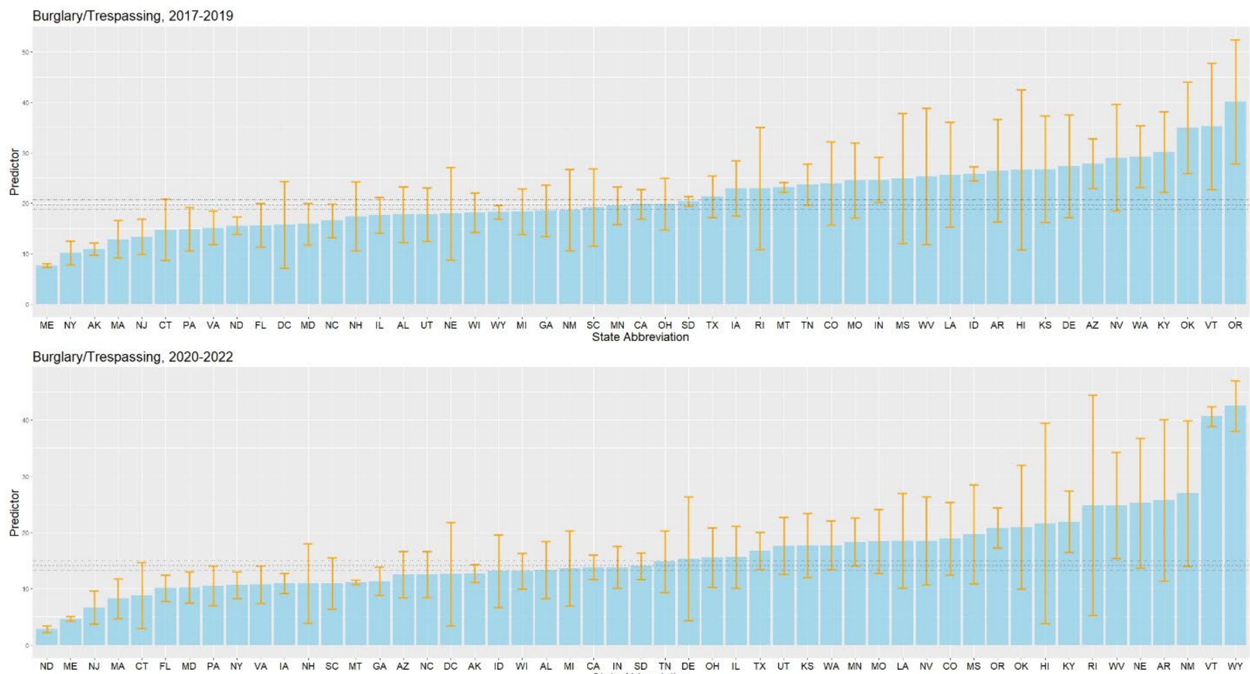


Note: Dashed line is U.S. estimate. Property crime includes burglary/trespassing, motor vehicle theft, and other theft. Estimates include 95% confidence intervals. The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

Property crime rates from 2020–2022 range from 32.5 crimes per 1,000 households in North Dakota (SE 0.7) to 243.9 crimes per 1,000 households in Oregon (SE 22.5). The national level estimate of the property crime rate for that period is 95.6 crimes per 1,000 households (SE 1.2). Similar to the 2017–2019 time frame, 20 states (Alabama, Florida, Georgia, Idaho, Iowa, Louisiana, Maryland, Massachusetts, Michigan, Mississippi, Montana, New Hampshire, New Jersey, North Carolina, North Dakota, South Carolina, Tennessee, Vermont, Virginia, and Wisconsin) have rates significantly below the U.S. estimate, and 12 states (Alaska, Arizona, California, Colorado, District of Columbia, Minnesota, Nevada, New Mexico, Oregon, South Dakota, Washington, and Wyoming) have rates above the U.S. estimate in the 2020–2022 period.

Burglary/trespassing, depicted in Figure 9, constitutes an important component of property crime. During 2017–2019, the estimated rates for burglary/trespassing range from 7.6 crimes per 1,000 households in Maine (SE 0.2) to 40.1 crimes per 1,000 households in Oregon (SE 6.3). In this time frame, seven states (Alaska, Maine, Massachusetts, New Jersey, New York, North Dakota, and Virginia) have burglary/trespassing rates significantly below the U.S. rate of 19.7 crimes per 1,000 households (SE 0.5), and eight states (Arizona, Idaho, Kentucky, Montana, Oklahoma, Oregon, Vermont, and Washington) have rates significantly above.

Figure 9: Ordered state estimates of victimization rates for burglary/trespassing with corresponding confidence intervals, 2017–2019 and 2020–2022



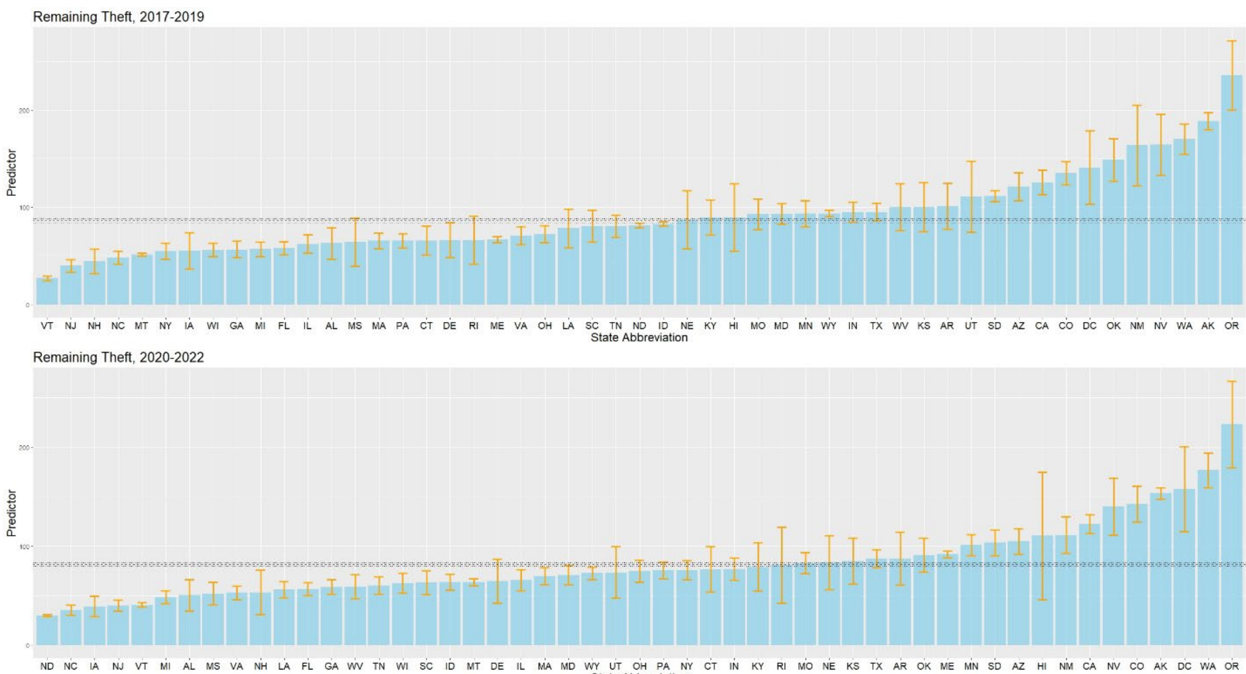
Note: Dashed line is U.S. estimate. Estimates include 95% confidence intervals. The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

During the aggregate period of 2020–2022, the lowest burglary/trespassing rate is in North Dakota with 2.8 crimes per 1,000 households (SE 0.3) and the highest burglary/trespassing rate is in Wyoming with

42.5 crimes per 1,000 households (SE 2.3). The national-level burglary/trespassing rate during this time frame is 14.1 crimes per 1,000 households (SE 0.4). The states with estimates significantly below the national rate are Florida, Iowa, Maine, Maryland, Massachusetts, Montana, New Jersey, New York, and North Dakota. The states with a burglary/trespassing rate significantly above the U.S. estimate are Kentucky, Oregon, Vermont, West Virginia, and Wyoming. Burglary/trespassing rates do not differ significantly from the U.S. rate for remaining states.

Figure 10 displays property crime victimization rates for all remaining theft, including motor vehicle theft and other types of household theft. During 2017–2019, the estimates range from 26.8 crimes per 1,000 households (SE 1.3) in Vermont to 235.9 crimes per 1,000 households (SE 18.3) in Oregon. The national estimate of the remaining theft rate is 86.3 crimes per 1,000 households (SE 1.1) during 2017–2019. The 21 states with estimates significantly below the U.S. estimate are Alabama, Connecticut, Delaware, Florida, Georgia, Illinois, Iowa, Maine, Massachusetts, Michigan, Montana, New Hampshire, New Jersey, New York, North Carolina, North Dakota, Ohio, Pennsylvania, Vermont, Virginia, and Wisconsin. The 12 states with remaining theft rates above the national rate are Alaska, Arizona, California, Colorado, the District of Columbia, Nevada, New Mexico, Oklahoma, Oregon, South Dakota, Washington, and Wyoming.

Figure 10: Ordered state estimates of household victimization rates for all remaining theft with corresponding confidence intervals, 2017–2019 and 2020–2022



Note: Dashed line is U.S. estimate. Remaining theft includes motor vehicle theft and other theft. Estimates include 95% confidence intervals. The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

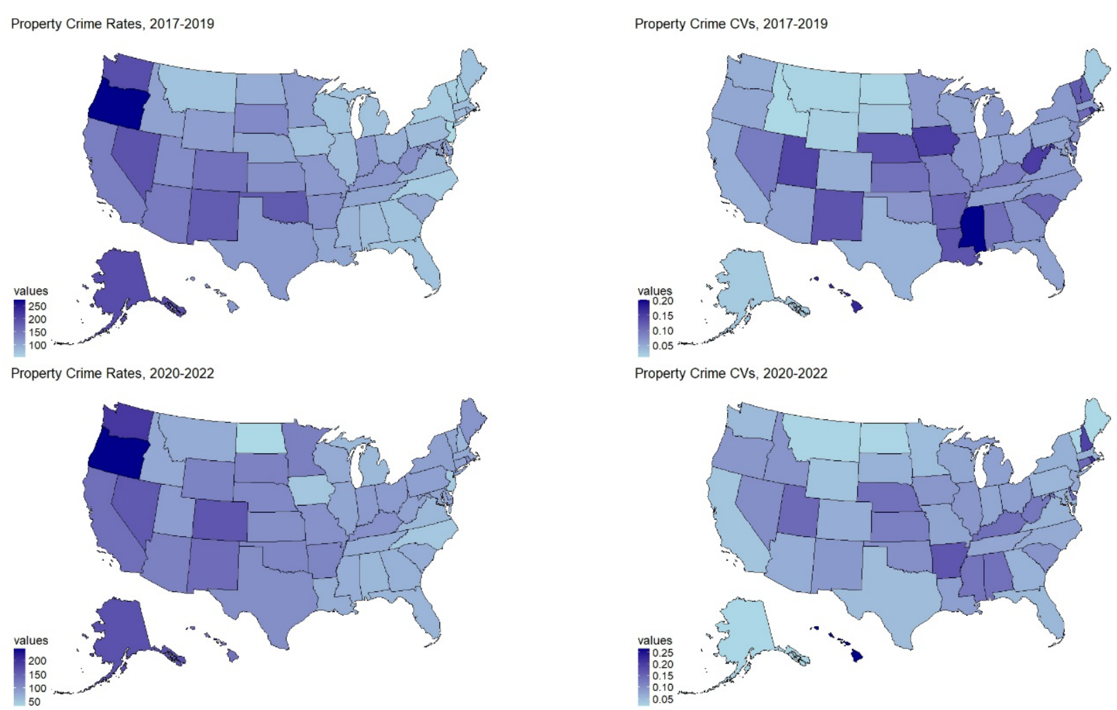
During 2020–2022, the estimates for remaining theft (including motor vehicle theft and other types of household theft) range from 29.7 crimes per 1,000 households in North Dakota (SE 0.6) to 223.1 crimes per 1,000 households in Oregon (SE 22.4). For comparison, the U.S. estimate of the remaining theft rate for 2020–2022 is 81.5 crimes per 1,000 households (SE 1.0). Similar to the 2017–2019 time frame, 22 states (Alabama, Florida, Georgia, Idaho, Illinois, Iowa, Louisiana, Massachusetts, Michigan, Mississippi, Montana, New Hampshire, New Jersey, North Carolina, North Dakota, South Carolina, Tennessee, Vermont, Virginia, West Virginia, Wisconsin, and Wyoming) have rates significantly below the national rate, and 12 states (Alaska, Arizona, California, Colorado, District of Columbia, Maine, Minnesota, Nevada, New Mexico, Oregon, South Dakota, and Washington) have rates above the national rate.

Geographic distribution of state-level property crime estimates

Estimated state-level rates with the associated CVs for property crime, burglary/trespassing, and remaining theft are shown in Figures 11, 12, and 13. In each figure, the estimated rates are presented at left with the corresponding CVs at right. The top half of each figure is for the 2017–2019 period and the bottom half is for 2020–2022. The colors darken as the numeric values for the rates or CVs increase.

Estimated property crime rates range from about 50 to 250 crimes per 1,000 households, generally decreasing from west to east (Figure 11). The estimates of property crime rates are fairly reliable, with CVs ranging from about 5% to 25%. The CVs do not exhibit a consistent trend with the estimates. Some states with very high CVs have low estimates and the reverse is also true. A cluster of states in the northwest region seem to have relatively small CVs, close to 5%.

Figure 11. Estimates of property crime rates in units of crimes per 1,000 households (left) and coefficients of variation (right), 2017–2019 and 2020–2022

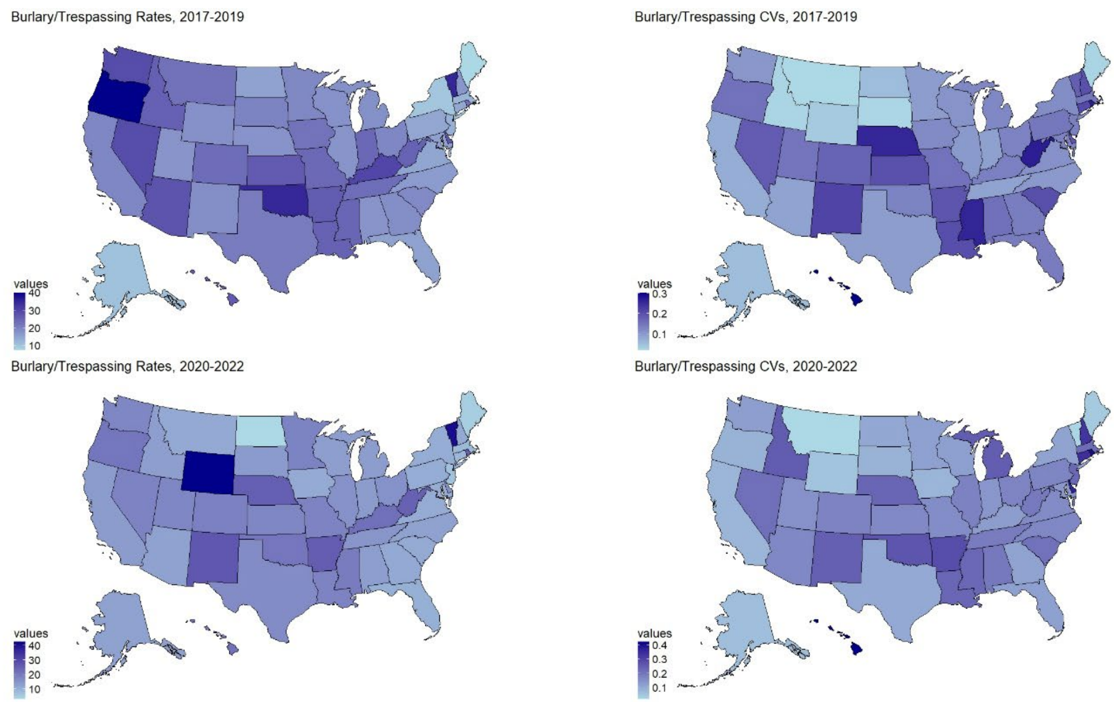


Note: Property crime includes burglary/trespassing, motor vehicle theft, and other theft. The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to

this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

The burglary/trespassing rates (Figure 12) are below the overall property crime rates and range from about 10 to 40 crimes per 1,000 households. As this is a smaller category, the CVs are slightly higher than those for overall property crime ranging from about 10% to 30%. No obvious geographic trends emerge in the estimates for burglary/trespassing.

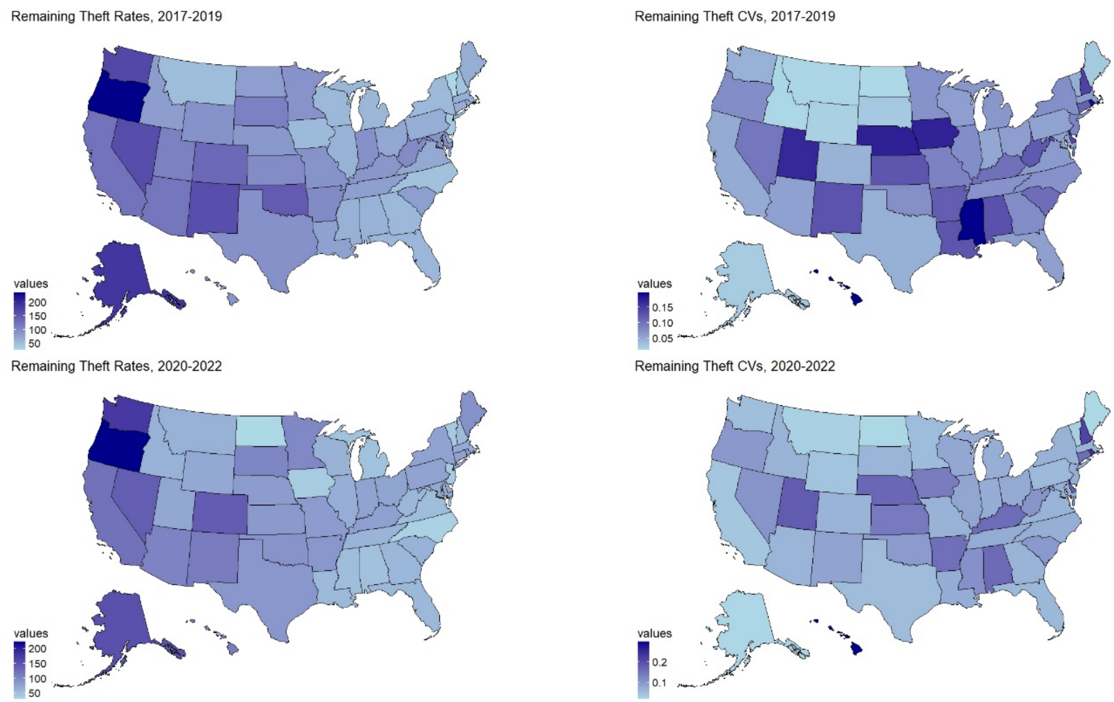
Figure 12. Estimates of burglary/trespassing rates in units of crimes per 1,000 households (left) and coefficients of variation (right), 2017–2019 and 2020–2022



Note: The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

Figure 13 shows estimated rates and CVs for the remaining theft category, which includes motor vehicle theft and other types of household theft. As the largest contributor to overall property crime, the remaining theft estimates mirror the estimates for overall property crime. The rates exhibit a decreasing east to west trend, similar to overall property crime rates. The rates range from about 50 to 200 crimes per 1,000 households. The estimates for this category are fairly reliable, with estimated CVs in the range of 5% to 20%.

Figure 13. Estimates of remaining theft rates (left) in units of crimes per 1,000 households and coefficients of variation (right), 2017–2019 and 2020–2022

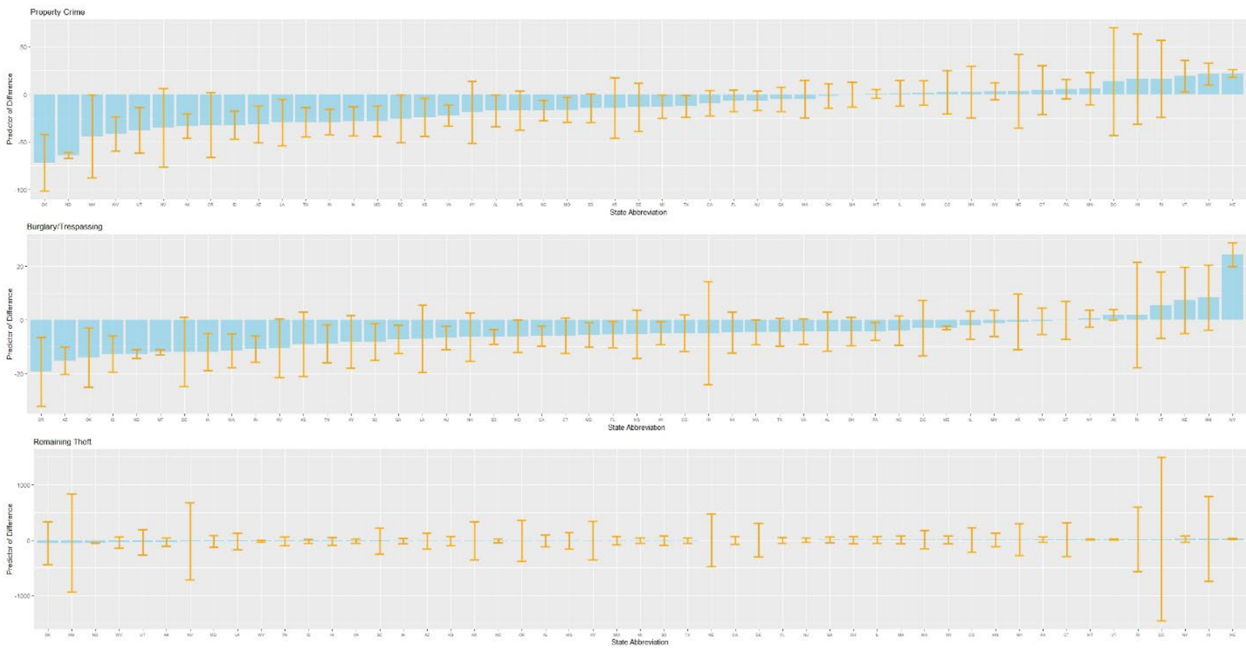


Note: Remaining theft includes motor vehicle theft and other theft. The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

Change over time for property crime

Figure 14 contains the estimates of change over time for property crime with corresponding confidence intervals. The change is the difference between the 2020–2022 estimate and the 2017–2019 estimate. The estimated changes are sorted from smallest (i.e., biggest declines in crime rates) to largest (i.e., biggest increases in crime rates).

Figure 14. State estimates of change over time for property crime, with corresponding confidence intervals, 2017–2019 and 2020–2022



Note: Property crime includes burglary/trespassing, motor vehicle theft, and other theft. The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

The properties of the estimates of change depend on the type of crime. The overall U.S. property crime rate declines from 106.0 crimes per 1,000 households (SE 1.3) during 2017–2019 to 95.6 crimes per 1,000 households (SE 1.2) during 2020–2022, and this decrease is reflected in the state-level estimates. The overall property crime rate between the two time frames decreases significantly for several states: Alaska, Arizona, Idaho, Indiana, Iowa, Kansas, Louisiana, Maryland, Missouri, North Carolina, North Dakota, Oklahoma, Tennessee, Utah, Virginia, and West Virginia. The pattern for overall property crime is reflected in the estimated changes for the burglary/trespassing rates.

The national estimate for the burglary/trespassing rate decreases from 19.7 crimes per 1,000 households during 2017–2019 (SE 0.5) to 14.1 crimes per 1,000 households during 2020–2022 (SE 1.2). This decrease is also observed at the state level. Many states exhibit a significant decline in the burglary/trespassing rate between the two time periods, including Arizona, Indiana, Iowa, Montana, North Dakota, Oklahoma, Oregon, and Washington, among others.

The patterns for the remaining theft category differ slightly from the nature of the changes for overall property crime and burglary/trespassing. The national estimate for remaining theft decreases from 86.3 crimes per 1,000 households (SE 1.1) during 2017–2019 to 81.5 crimes per 1,000 households (SE 1.0) during 2020–2022. This modest decrease, however, is not reflected at the state level. Essentially no states have a significant change in the remaining theft rate between the two time periods.

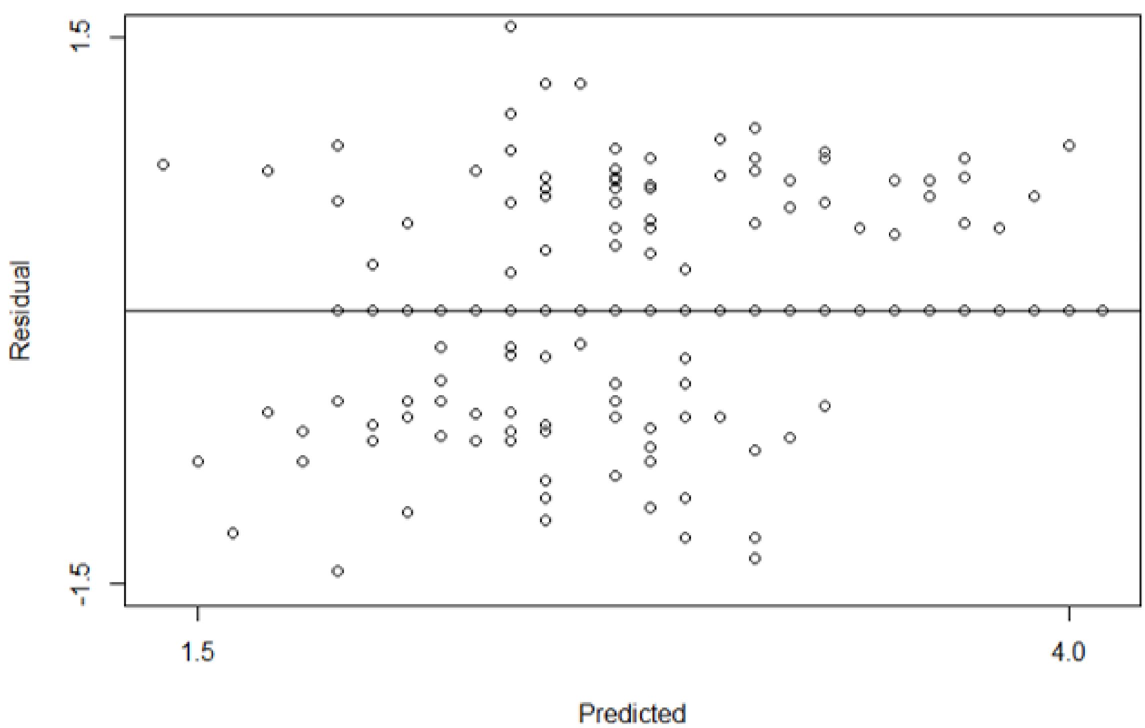
Diagnosing Model Goodness of Fit

The model-based results implicitly assume that the model provides an adequate fit to the data. Therefore, it is important to diagnose the goodness of fit of the model. Below, model goodness of fit is evaluated using standardized residuals, posterior predictive p-values, and benchmarking ratio adjustment factors.

Standardized residuals

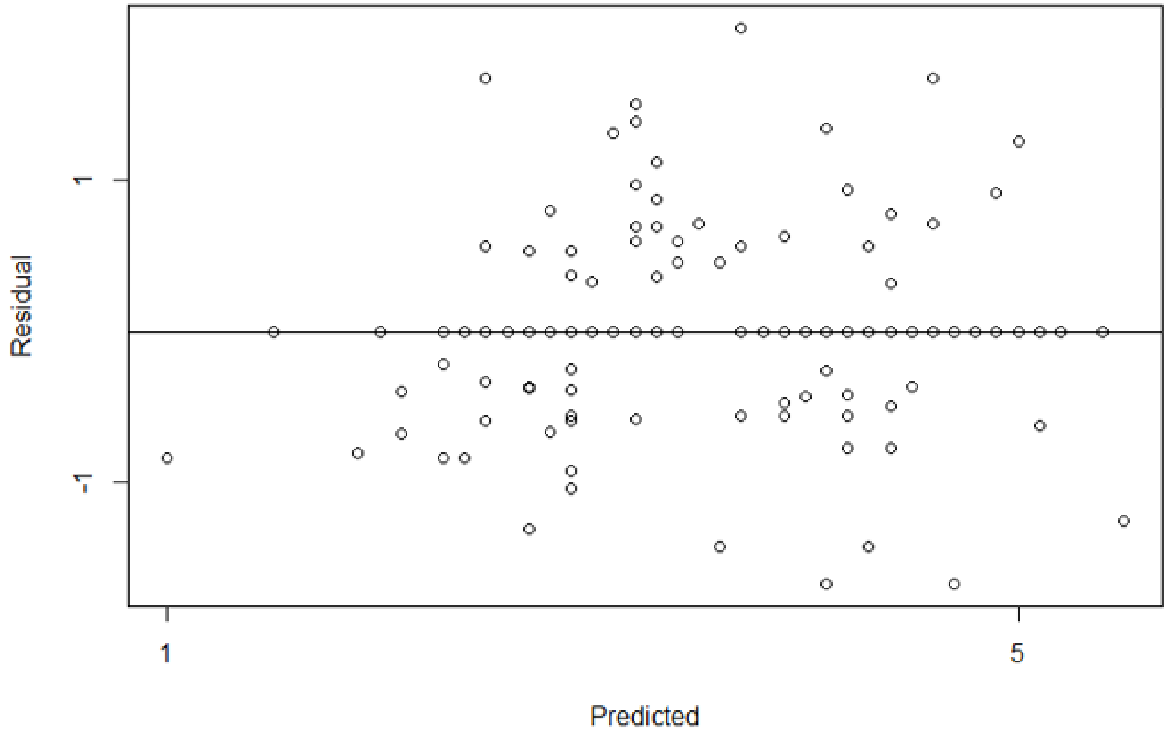
A standardized residual is defined as the ratio of the difference between the model response variable and predicted value to the standard error of the direct estimator. If the model fits the data, then the residuals should be evenly scattered between -2 and 2 (approximately) and exhibit no systematic trends as a function of the predicted values. Figure 15 has the standardized residuals for violent crime and Figure 16 has the standardized residuals for property crime. The residuals have no extreme values and do not exhibit clear patterns as a function of the predicted values. The plots of standardized residuals lead to no cause for concern about the adequacy of the model assumptions for the NCVS data for violent crime or property crime during the aggregate time periods of 2017–2019 and 2020–2022.

Figure 15: Standardized residuals for violent crime model (in log scale)



Note: The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

Figure 16: Standardized residuals for property crime model (in log scale)



Note: The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

Posterior predictive p-values

The posterior predictive p-value is a measure of goodness of fit that is recommended in Gelman et al. (2013) and is appropriate for Bayesian models. The posterior predictive p-value compares a statistic generated based on the model to the corresponding statistic based on the observed data. A posterior predictive p-value that is close to 50% and is not close to 0 or 1 indicates that the model fits the data well. Gelman et al. (2013) consider p-values in the range of 0.05 to 0.95 to be reasonable.

The posterior predictive p-value is calculated for three statistics: the mean, variance, and skewness. Table 1 contains the posterior predictive p-values for violent crime and property crime. All posterior predictive p-values are considered close enough to 0.5 for reasonable statistical standards. The analysis of the posterior predictive p-values indicates that the model provides a very good fit to the data for both violent crime and property crime.

Table 1: Posterior predictive p-values

Statistic	Violent crime	Property crime
Mean	0.411	0.527
Variance	0.797	0.615
Skewness	0.322	0.680

Note: The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

Benchmarking ratio adjustments

Finally, the benchmarking ratio adjustments are examined. A ratio adjustment factor closer to 1 indicates a better fitting model. The benchmarking ratios for violent crime and property crime are given in Table 2. The benchmarking ratio adjustments are extremely close to 1. The effect of benchmarking on the model-based predictors is negligible.

Table 2: Benchmarking ratio adjustment factors

	2017–2019	2020–2022
Property crime	0.992	0.999
Burglary/trespassing	0.994	0.999
Remaining theft	0.992	0.999
Violent crime	1.008	1.015
Simple assault	0.988	1.015
Violent crime excluding simple assault	1.048	1.014

Note: The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

Comparison of Model-Based Predictors to Direct Estimators

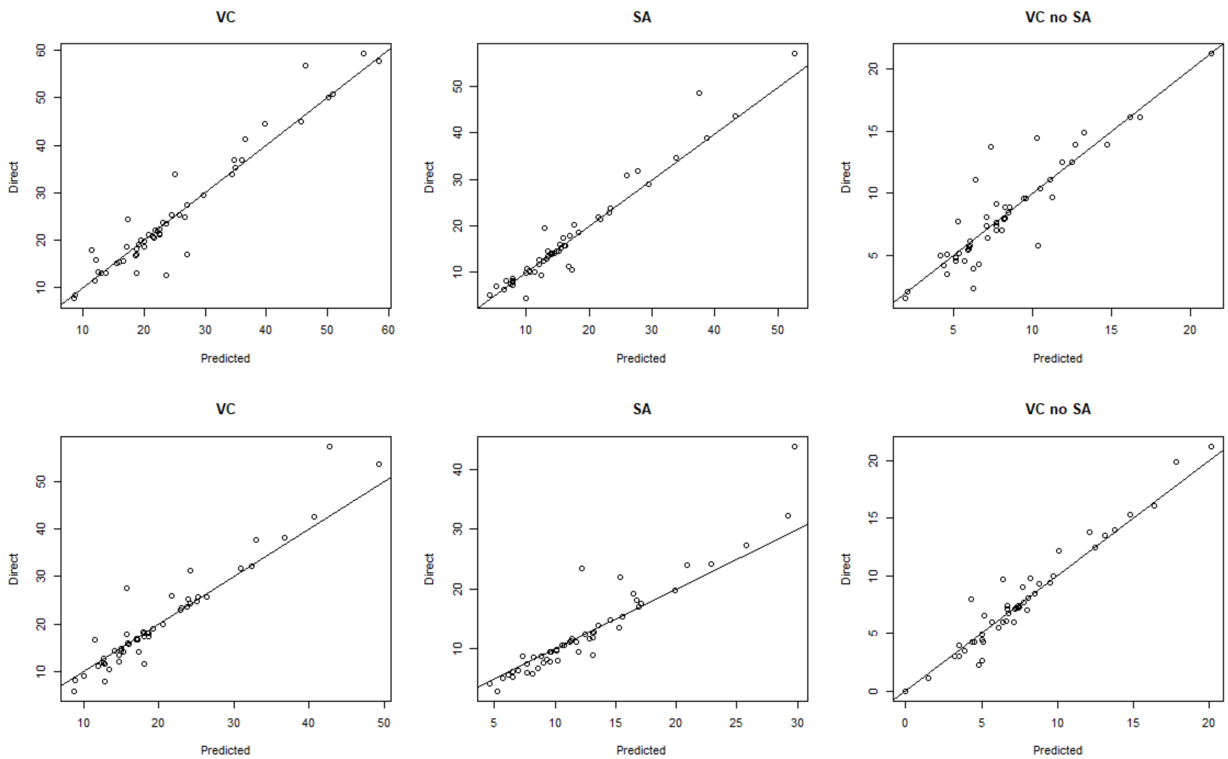
To gain further insight into the properties of the model, the model-based predictors are compared to the direct estimators. First, the model-based predictors of the rates are compared to the corresponding direct estimators. Subsequently, the CVs for the two types of estimators are considered.

Comparison of model-based predictors of crime rates to direct estimators of crime rates

In Figures 17 and 18, the direct estimators of the crime rates are plotted on the vertical axis with the model-based predictors of crime rates on the horizontal axis. Top panel is for 2017-2019 and bottom panel is for 2020-2022. Figure 17 is for violent crime and Figure 18 is for property crime.

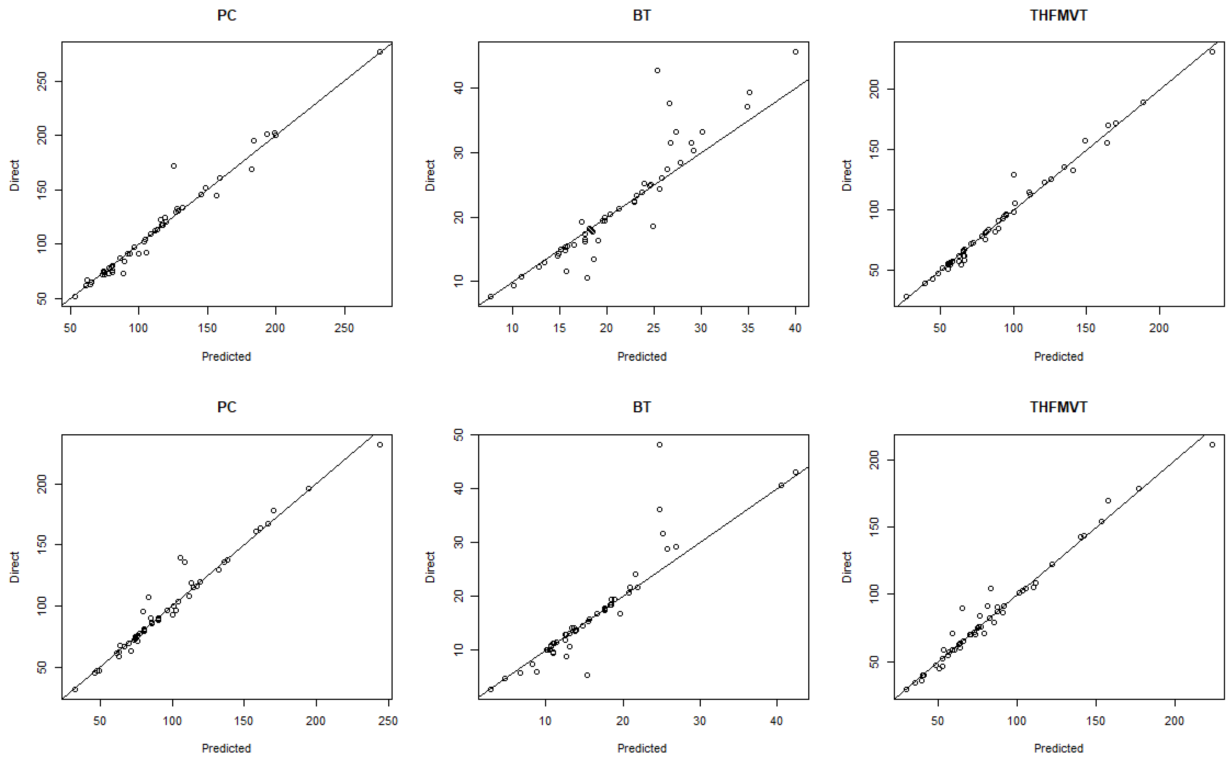
Figures 17 and 18 indicate that the model makes only modest changes to the direct estimators, in general. The predictors and direct estimators are tightly clustered around the 45-degree line. Deviations from this relationship are not systematic. The absence of a systematic deviation from the 45-degree line in Figures 17 and 18 indicates that the model does not introduce systematic biases. Figures 17 and 18 reveal a handful of outliers that are not unusual relative to the model assumptions (according to the residual plots presented in Figures 15 and 16).

Figure 17: Direct estimators against predictors for violent crime, 2017–2019 and 2020–2022



Note: Top panel is for 2017–2019 and bottom panel is for 2020–2022. VC = violent crime, SA = simple assault, VC no SA = violent crime excluding simple assault. The solid lines in the plots are 45-degree lines through the origin. The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

Figure 18: Direct estimators against predictors for property crime, 2017–2019 and 2020–2022



Note: Top panel is for 2017–2019 and bottom panel is for 2020–2022. PC = property crime, BT = burglary/trespassing, and THFMVT = remaining theft. Note that "remaining theft" is the combination of the NCVS variables of "other theft" and "motor vehicle theft." The solid lines in the plots are 45-degree lines through the origin. The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

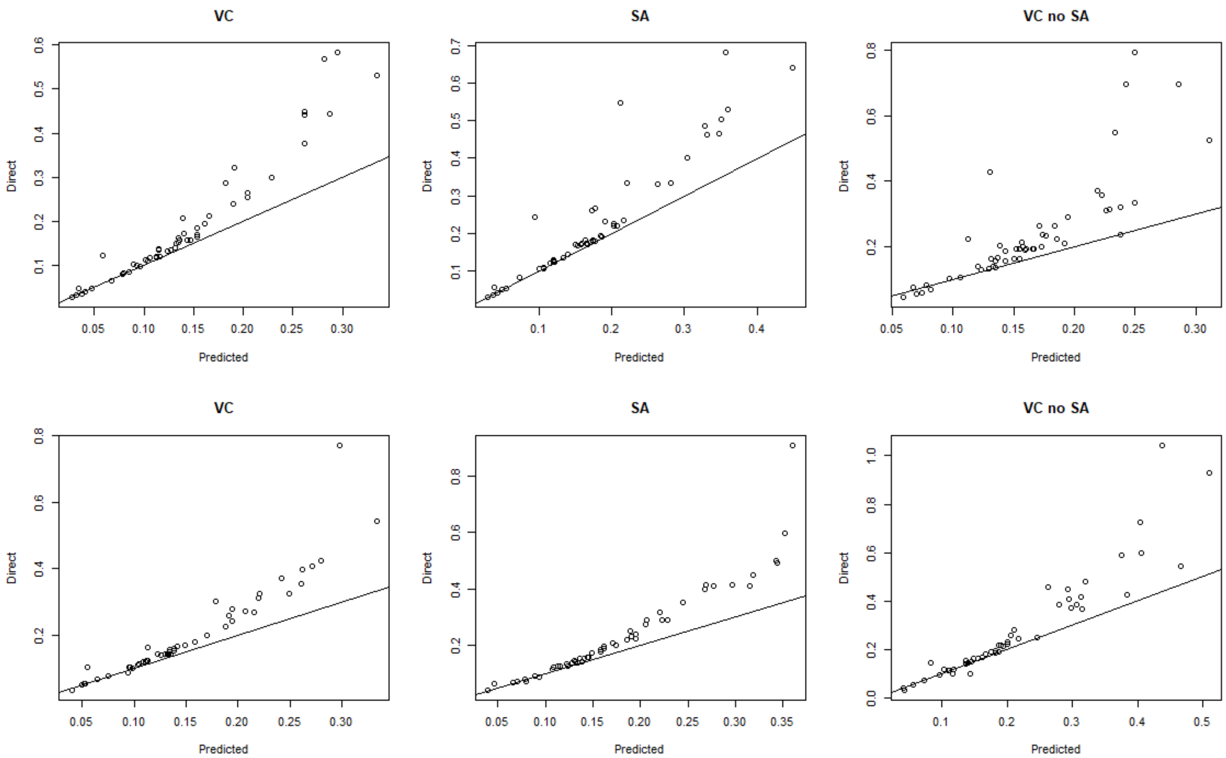
Comparison of CVs of predictors to CVs of direct estimators

Figures 19 and 20 compare the CVs of the predictors to the CVs of the direct estimators. The CV of the direct estimator is plotted on the vertical axis with the CV of the predictor on the horizontal axis. The structure of these figures is analogous to the structure of Figures 17 and 18.

The CVs of the predictors are generally below or equal to the CVs of the direct estimators. This positive finding indicates that the model-based procedure garners an efficiency gain relative to the direct estimators. When the CV of the direct estimator is small, the model-based and direct CVs are similar. In contrast, when the CV of the direct estimator is large, the model-based procedure can render a substantial reduction in CV relative to the direct estimator. This pattern seems reasonable. When the direct estimators are reliable the model makes only a modest change. In contrast, when the direct estimators are unreliable the model has greater impact.

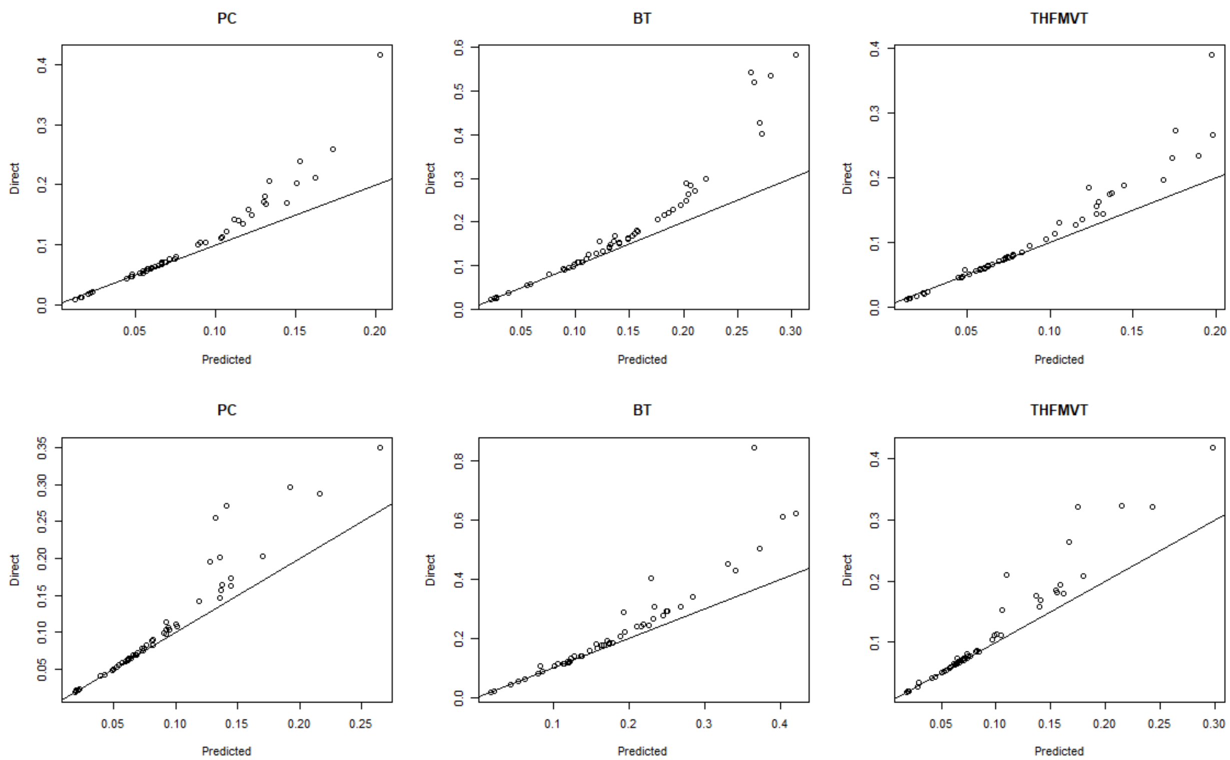
The CVs for violent crime excluding simple assault during 2020–2022 excludes the estimate for North Dakota. The direct plot of estimate for this state is nearly zero. As a result, the model-based estimate is also nearly zero and the corresponding CV is extreme.

Figure 19: Coefficients of variation (CV) of direct estimators against CVs of predictors for violent crime, 2017–2019 and 2020–2022



Note: Top panel is for 2017–2019 and bottom panel is for 2020–2022. VC = violent crime, SA = simple assault, VC no SA = violent crime excluding simple assault. The solid lines in the plots are 45-degree lines through the origin. The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

Figure 20: Coefficients of variation (CV) of direct estimators against CVs of predictors for property crimes, 2017–2019 and 2020–2022



Note: Top panel is for 2017–2019 and bottom panel is for 2020–2022. PC = property crime, BT = burglary/trespassing, and THFMVT = remaining theft. The solid lines in the plots are 45-degree lines through the origin. The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

Challenges, Limitations, and Areas for Future Research

Three main challenges are encountered when developing model-based estimates of state-level crime rates. First, direct estimates at the state level can equal zero. Second, data from the UCR Program are weak covariates for many types of crime. Third, evidence of spatial dependence exists in the model residuals. This section elaborates on these issues and discusses possible ways to address them through future research.

Zero direct estimates

The first challenge is that direct estimates at the state level can equal zero for 3-year time intervals. The most prominent example of this is the estimate for violent crime excluding simple assault for North Dakota in 2020–2022. To circumvent the difficulties associated with the estimate of zero, the model is fit to the other two violent crime categories, and then the estimate for violent crime excluding simple assault is deduced. Despite this workaround, the model-based estimate for violent crime excluding simple assault for North Dakota remains highly unreliable during the 2020–2022 period. The model-based estimate for this domain is essentially zero and as a result the CV far exceeds any reasonable standard.

A second example of a zero direct estimate occurs for the motor vehicle theft category. Several states had estimates of motor vehicle theft that are nearly zero. To overcome this issue, motor vehicle theft is aggregated with the other household theft category creating a remaining theft category. While this improves the reliability of estimates for remaining theft, it results in a loss of granularity and prevents construction of state-level estimates for motor vehicle theft.

Future work may explore improved avenues to handle the zero estimates. One approach is to model the zero values directly. Modifications to the variance-covariance matrices to handle the zero estimates would be needed. Future work may entail modifying the approach proposed in Berg and Fuller (2012) for the NCVS context. Modeling the zeros directly may enable production of nonzero model-based estimates even if the direct estimate is equal to zero.

UCR data are weak covariates

A second and important issue is that the UCR crime rates provide only a weak covariate for modeling the NCVS crime rates. The correlation between the NCVS and UCR crime rates is low. As a result, many of the regression slopes in the model do not differ significantly from zero.

This problem is illustrated in Tables 3 and 4. These tables contain 95% confidence intervals for the regression slopes for the different model response variables. Table 3 is for violent crime and Table 4 is for property crime. A regression slope differs significantly from zero if the lower interval endpoint (2.5%) is above zero. For overall violent crime during the aggregate period of 2017–2019, the regression slope differs significantly from zero. However, the regression slope does not differ significantly from zero for overall violent crime during 2020–2022. For simple assault, the regression slope does not differ significantly from zero for either period. The association to the UCR covariate is somewhat stronger for property crime than for violent crime. The regression slopes for remaining theft differ significantly from zero in both periods. The regression slope for overall property crime only differs significantly from zero during 2017–2019.

Table 3: 95% confidence intervals for regression slopes for the violent crime model

	2017–2019		**2020–2022**	
	Violent crime	**Simple assault**	**Violent crime**	**Simple assault**
Lower level (2.5%)	0.225	-0.058	-0.029	-0.190
Upper level (97.5%)	0.882	0.797	0.570	0.500

Note: The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

Table 4: 95% confidence intervals for regression slopes for the property crime model

	2017–2019		2020–2022	
	Burglary/ trespassing	**Remaining theft**	**Burglary/ trespassing**	**Remaining theft**
Lower level (2.5%)	0.074	0.389	-0.066	0.196
Upper level (97.5%)	0.630	0.711	0.681	0.605

Note: The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB-FY24-POP001-0021).
Source: Bureau of Justice Statistics, National Crime Victimization Survey, Restricted-use data, 2017–2022; and the FBI's Uniform Crime Reporting Program, Summary Reporting System (2017–2019) and National Incident-Based Reporting System (2020–2022).

As illustrated in Table 3, the UCR covariate is particularly weak for simple assault where the regression slopes for both aggregate periods do not differ significantly from zero. Modeling violent crime excluding simple assault instead of simple assault may strengthen the association to covariates. This approach depends on finding a better way to handle the zero estimate for North Dakota.

For property crime, the coefficient for burglary/trespassing is only significant during the 2017–2019 period. It is possible that modeling overall property crime and remaining theft together and subsequently deducing burglary/trespassing would lead to a stronger model. This is an approach to consider in possible future research.

Tables 3 and 4 demonstrate that the association between UCR and NCVS data is modest at best. This means that the UCR data are weak covariates, and the model has low predictive power. Altering the model form by changing the model response variables is one possible way to improve the association between the response and covariate. Research into alternative sources of auxiliary information is another possible direction for further study.

Evidence of spatial dependence

Given that the data for states are spatially structured, it is natural to examine the data for spatial dependence. The Moran's I test is used to evaluate if the predicted random effects have spatial dependence. Evidence of spatial structure is found for simple assault, but not for overall violent crime. Failure to appropriately model spatial dependence can lead to problems with the predictors and confidence intervals. A region effect is added to the model, but this does not lead to improvements. Modifying the model further to account for spatial dependence is an area for future research.

Summary and Recommendations

Estimation of crime rates at the state level is challenging. Small sample sizes at the state level can lead to unstable estimates with high CVs. As a result of small sample sizes, direct estimates at the state level are often regarded as unreliable.

This work endeavors to improve upon the direct estimates through the aid of statistical modeling techniques. A multivariate lognormal model has been developed to obtain state-level estimates. The model incorporates multiple crime types and data for two aggregate time frames in a unified fashion. Bayesian methods are used for inference.

The model is applied to violent crime and property crime separately. The model generally produces CVs in the range of roughly 10% to 30% for most states and crime types. Violent crime rates exhibited no significant changes at the state level between the two time frames considered. For property crime, significant declines in the crime rate at the state level are observed between 2017–2019 and 2020–2022. For both property crime and violent crime, the state-level crime rates vary significantly around the U.S. estimate.

To gain further insight into the properties of the model, the correspondence between the model-based estimates and the direct estimates is examined. The model-based and direct estimators are clustered tightly around a straight line through the origin. The CVs of the direct estimators are generally higher than the CVs of the model-based predictors. This indicates that the model improves upon the efficiency of the direct estimators, without introducing systematic biases.

Standard statistical tools are employed to examine the goodness of fit of the model. The residuals exhibit no trends as a function of the predicted values. The posterior predictive p-values are close to 50%. Benchmarking ratio adjustment factors are close to one. These diagnostics indicate that the model fits the data well.

Despite the positive properties of the model, many areas for improvement remain. One issue is that the model does not handle zero direct estimates very well. This is a particular problem for violent crime excluding simple assault in North Dakota, where the model-based estimate remains highly unstable. The second issue is that the UCR data provide only weak covariate information. The regression coefficients for many crime types and years do not differ significantly from zero. The final problem is that evidence exists of spatial dependence in the random effects for simple assault.

These issues suggest two main directions for future work. For operational purposes, use of a 6-year average instead of two 3-year averages may be advisable. This has potential to alleviate problems with zero direct estimates and stabilize associations to covariates. BJS has the capability of producing direct estimates for the 22 largest states based on a 3-year average. These direct estimates could naturally be paired with model-based estimates for all states based on a 6-year average. While this would reduce the temporal granularity, it would likely improve efficiency at the state level. The operation of publishing direct estimates for the 22 largest states based on a 3-year average combined with model-based estimates for all states based on a 6-year average may also lead to a natural estimation cycle.

The second future research direction would be to conduct further research aimed at improving upon the current model for the 3-year intervals. This would entail an investigation of spatial dependence, efforts to try to strengthen the covariates, and research into improved methods for handling zero direct estimates.

Understanding model-based estimation

Why use a model?

Model-based estimates have the potential to improve upon direct estimates. Model-based procedures garner efficiency gains in two main ways. First, models can incorporate auxiliary variables. These are covariates that are external to the National Crime Victimization Survey (NCVS) data. An ideal auxiliary variable is correlated with the NCVS direct estimates. By taking advantage of this correlation, the model-based estimates can be more precise than the direct estimates. Second, models use statistical assumptions that the estimates for different states share common distributional properties. This supports procedures that "borrow information" from all states when constructing the estimates for the individual states. If the statistical assumptions are correct, then model-based estimates are more efficient than the direct estimates.

Interpreting the model-based results

The direct estimates for the 22 states reported in Kena and Morgan (2023) differ from the model-based estimates in this report. Users should understand that the two sets of estimates are constructed under different frameworks. Data users are encouraged to synthesize all available information when drawing conclusions based on the NCVS results.

Uncertainty in the model is measured through the root mean square error (MSE). The MSE is a measure of uncertainty that combines bias and variability. For simplicity, the model-based root MSEs are referred to as standard errors (SE). Coefficients of variation are defined as ratios of the root MSE to the model-based estimate. Normal theory confidence intervals are obtained by adding and subtracting 1.96 times the SE to the model-based prediction. Note that estimates based on the model are technically "predictions" because they are predictors of random variables in the model. We use the terms "predictions" and "estimates" interchangeably throughout.

The model form

The model form is a multivariate lognormal model. The multivariate component means that the model encompasses multiple crime types and two aggregate time periods (2017–2019 and 2020–2022) in a unified fashion. Lognormal means that the model is fit in the log scale. The log transformation is only used when fitting the model. The results are presented in the original scale.

Two separate models are used for violent crime and property crime. Each model has four response variables (i.e., is a multivariate model). The response variables for violent crime are overall violent crime and simple assault in each of the two time frames. The estimate for violent crime excluding simple assault is then deduced by subtracting the estimate of simple assault from the estimate for overall violent crime. This approach circumvents difficulties associated with a zero direct estimate for violent crime excluding simple assault in North Dakota for the period 2020–2022. For property crime, motor vehicle theft is aggregated with other types of household theft to form a category called remaining theft. This operation avoids problems with zero estimates for motor vehicle theft. Then, burglary/trespassing and all remaining theft are modeled together for the two time frames. Subsequently, the estimate for property crime is deduced by adding the estimates for the more detailed categories.

The model incorporates auxiliary information from the Uniform Crime Reporting (UCR) Program. The UCR and NCVS have a complex relationship, as discussed in detail in Fay and Li (2011). The UCR is an administrative database that uses different definitions and data collection procedures than the NCVS. As a result, the NCVS and UCR measures of crime have an imperfect association. An exploratory analysis guides which UCR variables to include as model covariates. Correlations and multiple R-squared values from regressions lend insight into the associations between the UCR crime rates and the NCVS crime rates. This analysis suggests using the UCR rape category as the covariate for the violent crime model, the UCR burglary category as the covariate for NCVS burglary/trespassing, and the UCR motor vehicle theft category as the covariate for all remaining theft.

The benchmarking procedure of You et al. (2002) is used to ensure that the model-based estimates are consistent with the overall U.S. estimates. This benchmarking procedure applies a simple raking ratio adjustment to benchmark the predictors. The benchmarking adjustments had a negligible effect on the predictors. The ratios of the benchmarked to non-benchmarked predictors of the overall U.S. crime rate are in the range of 0.98–1.05 for all variables.

Bayesian inference procedures are used throughout. The Bayesian framework can easily accommodate the log transformation. The Bayesian framework also naturally reflects the slight increase in mean square error that results from the benchmarking operation. A further benefit of the Bayesian paradigm is that it naturally permits the use of unstructured covariance matrices for the multivariate component of the model. Diffuse priors for the fixed model parameters are used, and in our assessment, these priors have negligible impact on the model results. Gibbs sampling is used to approximate the posterior distributions.

Technical details of the model

Model development entails several stages. To select the covariates, correlations between the NCVS and UCR data are examined, and adjusted R-squared values are calculated. Then, a model is fit in the original scale. The posterior predictive p-values based on the model in the original scale are very close to zero or 1, indicating lack of fit. This motivates consideration of a log transformation. After the log transformation, the model appears to fit adequately.

In this section, the log-scale model is defined in technical terms. Let Y_{ikt} be the log of the direct estimator of the crime rate for type k and year t, where $k = 1,2$ and $t = 1,2$. Then, collect the direct estimators into a vector, denoted as Y_i, where $Y_i = (Y_{i11}, Y_{i21}, Y_{i12}, Y_{i22})'$. A multivariate Fay-Herriot model is specified for Y_i. Assume $Y_i = \theta_i + e_i$, where $e_i \sim N(0, \Psi_i)$, Ψ_i is the direct estimate of the design variance of Y_i, and Ψ_i is treated as fixed in the model. Then, assume that $\theta_i = B_0 + X_i B_1 + u_i$, where $u_i \sim N(0, \Sigma_{uu})$, Σ_{uu} is an unstructured 4x4 covariance matrix, B_0 = $(B_{01}, B_{02}, B_{03}, B_{04})'$, and B_1 = $(B_{11}, B_{12}, B_{13}, B_{14})'$. Parametrize X_i as $diag(X_{i1}, X_{i2}, X_{i3}, X_{i4})$.

The specific forms for X_i and Y_i depend on whether the model is for violent crime or property crime. For both models, $t = 1, 2$ are for 2017–2019, 2020–2022, respectively. For violent crime, $k = 1$ corresponds to overall violent crime, and $k = 2$ corresponds to simple assault. For property crime, $k = 1$ and $k = 2$ are for burglary/trespassing and all remaining theft, respectively. For violent crime, X_{i1} and X_{i2} are both the log of the average of UCR (revised) rape estimate across the years 2017–2019, and X_{i3} and X_{i4} are both the log of the average of UCR (revised) rape estimate across 2020–2021. For property crime, X_{i1} is the log of the average of UCR burglary estimate across the years 2017–2019, X_{i2} is the log of the average of UCR motor vehicle theft estimate across the years 2017–2019, X_{i3} is the log of the average of UCR burglary estimate across 2020–2021, and X_{i4} is the log of the average of UCR motor vehicle theft estimate across 2020–2021. UCR data for 2022 are excluded due to lack of data availability at the time of development of this report.

The Bayesian approach requires prior distributions for fixed model parameters. Improper uniform priors are specified for B_0 and B_1, where the improper uniform prior is basically a flat line at 1. The prior for Σ_{uu} is defined as $\Sigma_{uu} \sim Inverse - Wishart(.001, I_4.001)$.

Gibbs sampling is used to simulate from the posterior distribution. All full conditional distributions have known forms, enabling a straightforward Gibbs sampling procedure. The Gibbs sampling procedure begins with initial values $B_0^{(0)}$, $B_1^{(0)}$, and $\Sigma_{uu}^{(0)}$. For iterations $t = 1, \ldots, T$, the Gibbs sampling procedure involves repeating the following steps:

- Generate $\theta_i^{(t)} \sim N\left(B_0^{(t-1)} + X_i B_1^{(t-1)} + \Sigma_{uu}^{(t-1)}\left(\Sigma_{uu}^{(t-1)} + \Psi_i\right)^{-1}\left(Y_i - \left(B_0^{(t-1)} + X_i B_1^{(t)}\right)\right), \Sigma_{uu}^{(t-1)} - \Sigma_{uu}^{(t-1)}\left(\Sigma_{uu}^{(t-1)} + \Psi_i\right)^{-1}\Sigma_{uu}^{(t-1)}\right)$, and set $\theta^{(t)} = \left(\left(\theta_1^{(t)}\right)', \ldots, \left(\theta_D^{(t)}\right)'\right)'$ and $u^{(t)} = \left(\left(u_1^{(t)}\right)', \ldots, \left(u_D^{(t)}\right)'\right)'$, where $u_i^{(t)} = \theta_i^{(t)} - B_0^{(t-1)} - X_i B_1^{(t-1)}$.
- Define $\beta = (B_0', B_1')'$, and generate $\beta^{(t)} \sim N(\mu_\beta, V_\beta)$, where $V_\beta = (Z'R^{-1}Z)^{-\mathbf{1}}$, $Z = (1_D \otimes I_4, X)$, $X = (X_1', \ldots, X_D')'$, $R = diag(D) \otimes \Sigma_{uu}^{(t-1)}$, and $\mu_\beta = V_\beta Z' R^{-1} \theta^{(t)}$.
- Generate $\Sigma_{uu}^{(t)} \sim Inverse - Wishart(D + .001, S + I_4.001)$, where $S = \Sigma_{i=1}^{D} u_i^{(t)}(u_i^{(t)})'$.

This results in samples $\{\theta_i^{(t)}: t = 1, \ldots, T\}$.

Three chains of the Gibbs sampler are run with dispersed starting values. Each chain has length 2100, and the first 100 iterations are discarded as burn-in. The 2000 samples from the three chains are then pooled together. The results are based on a total of 6000 iterations. Trace plots and scale reduction factors give no cause for concern about convergence of the Markov chain.

The parameters are of the form $\ell' \exp(\theta_i)$, where $\exp(\theta_i)$ is the vector of the exponentials of the elements of θ_i. For predicting simple assault, $\ell' = (1, -1)$, and for predicting property crime, $\ell' = (1, 1)$. For the other crime types, ℓ' is either (1, 0) or (0, 1). The distribution of $\{\ell' \exp(\theta_i^{(t)}): t = 1, \ldots, T\}$ approximates the distribution of $\ell' \exp(\theta_i)$. The optimal model-based predictors are the posterior means of $\{\ell' \exp(\theta_i^{(t)}): t = 1, \ldots, T\}$.

The optimal model-based predictors do not preserve the direct estimators of crime rates at the national level. The benchmarking procedure of You et al. (2002) is implemented to ensure that appropriately weighted sums of final predictors equal the direct, national estimates. This benchmarking procedure applies a simple ratio adjustment to the model-optimal predictors. The ratios obtained are provided in Table 2 of this report. You et al. (2002) also define a procedure to obtain the mean square error of the benchmarked predictors. They demonstrate that the mean square error of the benchmarked predictor is the sum of the posterior variance and the squared difference between the benchmarked and non-benchmarked predictors. The method of You et al. (2002) is used to obtain the posterior mean square error of all predictors in this document.

Final estimates are rounded to one decimal place to comply with the standards of the U.S. Census Bureau disclosure review board.

Relationship to past work on model-based estimation for the NCVS

BJS has invested considerable resources in model-based subnational estimation over roughly the last decade, developing the multivariate dynamic model (Fay et al., 2013; Fay and Diallo, 2015; Fay, 2021). The multivariate dynamic model has both similarities and differences with the model used for this report. Both models are multivariate, meaning multiple crime types serve as model responses. Both models use UCR data as covariates. The key difference between the two models is that the multivariate dynamic model uses a lengthy time series, while the model for this report uses only two time frames. Another substantive difference is that the multivariate dynamic model is specified in the original scale, while the model used for this report operates in the log scale. The multivariate dynamic model uses frequentist inference procedures instead of Bayesian methods. A benefit of the Bayesian paradigm is that it can easily reflect the effect of benchmarking on the MSE, while the multivariate dynamic model ignores the impact of benchmarking on the measures of uncertainty. The differences between the two models indicate that the two approaches should be viewed as complementary (i.e., not in competition with each other). Again, data users should integrate all sources when basing decisions on NCVS data.

References

Berg, E. J., & Fuller, W. A. (2012). Estimators of error covariance matrices for small area prediction. *Computational Statistics & Data Analysis*, *56*(10), 2949–2962.

Fay, R. E., & Li, J. (2011). Predicting violent crime rates for the 2010 redesign of the National Crime Victimization Survey (NCVS). *Section on Survey*, *10*, 1663–1676.

Fay, R. E., Planty, M., & Diallo, M. S. (2013). Small area estimates from the National Crime Victimization Survey. In *Proceedings of the Section on Survey Research Methods.* American Statistical Association, 1544–1557.

Fay, R. E., & Diallo, M. (2015). Developmental Estimates of Subnational Crime Rates Based on the National Crime Victimization Survey. Available at: https://bjs.ojp.gov/library/publications/developmental-estimates-subnational-crime-rates-based-national-crime-0.

Fay, R. E. (2021). Constructing and Disseminating Small Area Estimates from the National Crime Victimization Survey, 2007–2018. Available at: https://www.ojp.gov/pdffiles1/bjs/grants/300603.pdf.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

Kena, G., & Morgan, R. E. (2023). *Criminal Victimization in the 22 Largest U.S. States, 2017–2019*. NCJ 305402. Bureau of Justice Statistics. Available at: https://bjs.ojp.gov/library/publications/criminal-victimization-22-largest-us-states-2017–2019.

National Research Council. (2008). *Surveying victims: options for conducting the National Crime Victimization Survey.* Washington, DC: The National Academies Press. https://doi.org/10.17226/12090.

Rao, J. N. K., & Molina, I. (2015). *Small area estimation*. John Wiley & Sons.

You, Y., Rao, J. N. K., & Dick, J. (2002, May). Benchmarking hierarchical Bayes small area estimators with application in census undercoverage estimation. In *Proceedings of the Survey Methods Section*, 86–90.