

Impact of Influential Observations on Enumeration and Variance Estimation in the National Crime Victimization Survey

FCSM Conference
November 2013

Michael Planty
Lynn Langton
U.S. Department of Justice
Bureau of Justice Statistics

Statement of Problem

- Outliers and influential observations on estimates from samples
- Enumeration of crime victimizations: Small percent of victims report a very large number of incidents
- Examine the impact of various enumeration strategies for extreme observations on estimates

Statement of Problem

Outliers

- Outlying observation is one that appears to deviate markedly from other members of the sample in which it occurs (Eltinge & Cantwell 2006; Chambers, 1985; Dixon, 1950; Grubbs 1969)

Type of outlier (Chambers 1985)

- Measurement/Gross Error
- Non-Representative
- Representative

Statement of Problem

Outlier: key component is influence

Including outliers results in extreme weights and increased volatility:

- Large shifts in point estimates
- Observation accounts for large proportion of estimate
- Large increases in variance, reduction of power
- Large shifts in point estimates over time

Statement of Problem

Outlier detection/identification and treatment

- What is an outlier?
 - Various techniques that identify outliers
 - Standard deviations from the mean
 - Visual inspection
 - Statistical techniques/distance/index of influence or leverage
 - Examine for influential data points
- What should you do with an outlier?
 - Treatment types
 - Include
 - Eliminate/Trim
 - Downweighting/Winsorization/Imputation strategies
 - Model-based alternatives (Elliott and Little, 2000)

Statement of Problem

National Crime Victimization Survey (NCVS)

Household survey

- Sample of persons 12 or older
- Rotating panel design, interviewed every 6 months for 3 years
- 160,000 interviews (in-person or phone)
- 87% HH and 87% person response rate

NCVS estimation:

- Annual totals counts and rates of crime
- Annual change

Counts and rates for key subdomains:

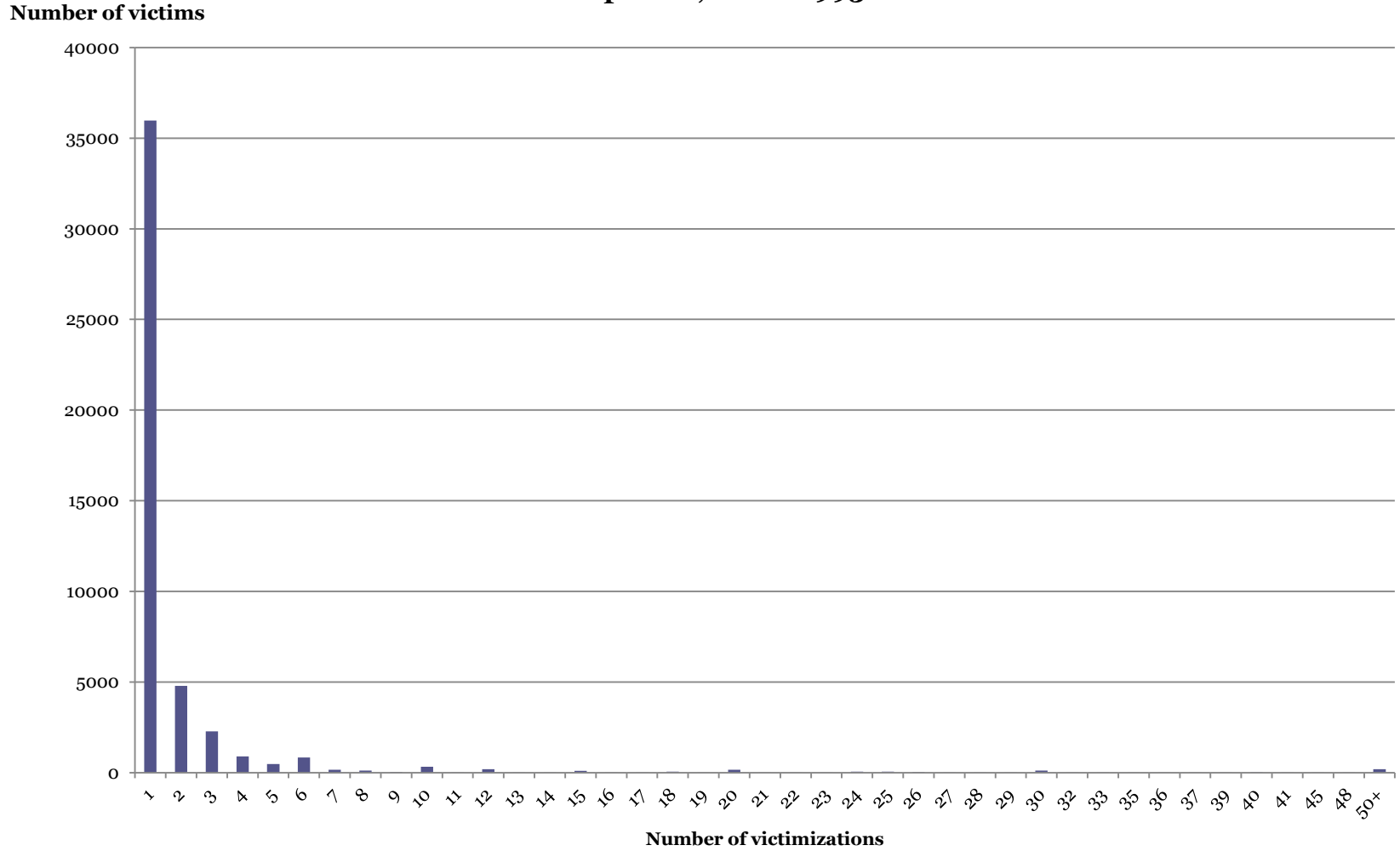
- Type of crime- violent and property crimes
- Victim characteristics- gender, age, race/ethnicity, etc.
- Incident characteristics- reported to police, injury, weapon use, etc.

Victimization: NCVS estimation

Data collection

- Respondent asked: Recall, date, and enumerate every criminal incident within reference period (6 months)
- Most report 0 or 1 incidents, but a small percentage report a very high number (a finding consistent across victimization surveys, various populations, and crime types)

Figure 1. Number of victimization per victim reported in the 6-month reference period, NCVS 1993-2012



Victimization: NCVS estimation

Data collection

- Respondent asked: Recall, date, and describe every criminal incident within reference period (6 months)
- Most report 0 or 1 incidents, but a small percentage report a very high number (a finding consistent across victimization surveys, various populations, and crime types)
- High volume repeat victims- high burden to recall, date, and describe accurately every incident
- Tension between burden and accuracy

Victimization: series incidents

NCVS data collection procedure: cap incidents collected in field

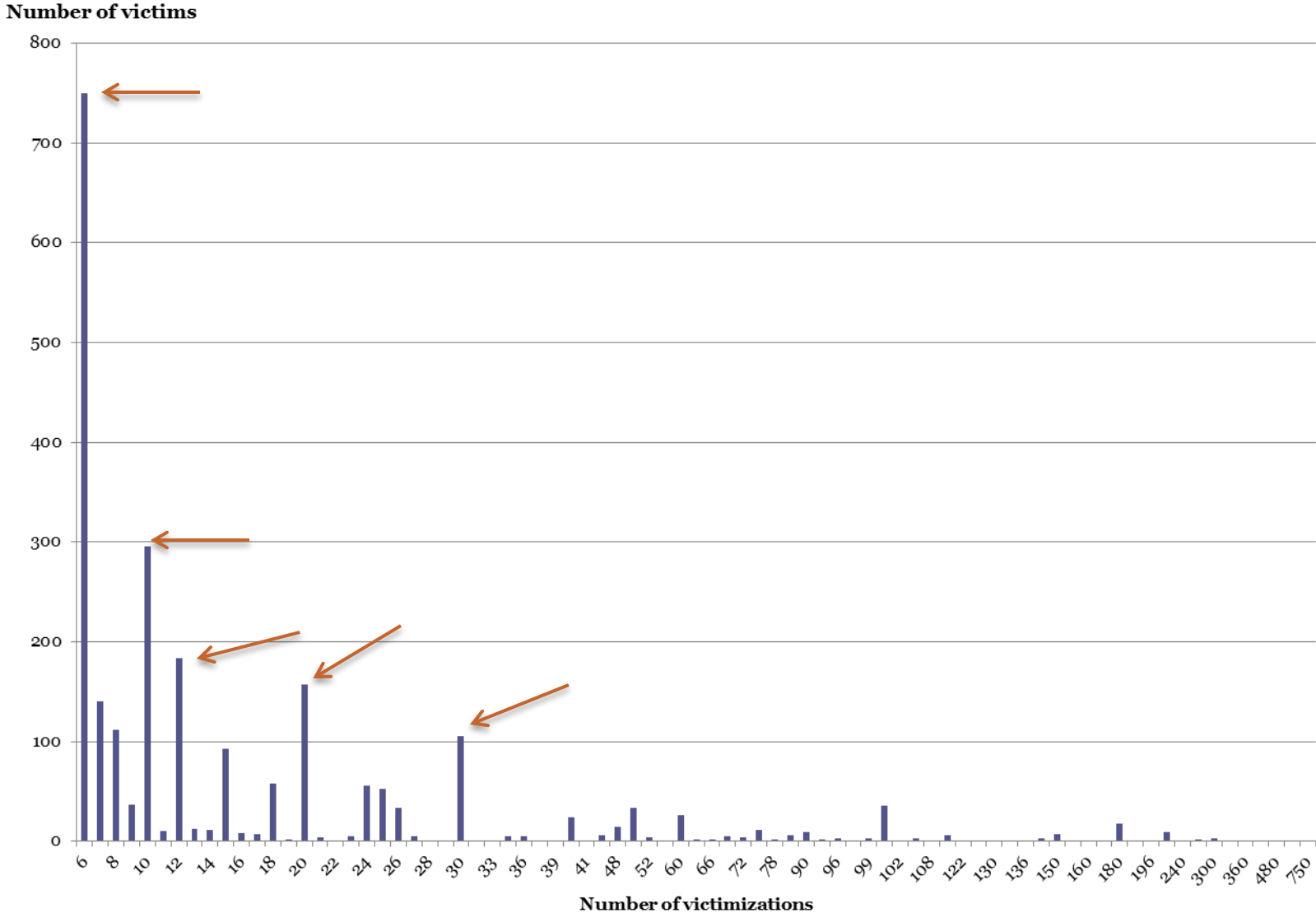
Most victimizations surveys use a capping method to data collection and estimation:

- Crime Survey for England and Wales (formerly British Crime Survey) – cap at 5 incidents
- Mexico's Encuestas Nacionales Sobre Iseguridad (ENSI) - cap at 5 incidents
- Canadian Victimization Survey (GSS Cycle 18)- up to 20 but estimation capped at 3

NCVS “series incident” protocol

1. Record number of times incident occurred
2. Record all incidents that can be recalled in separate detail
3. 6 or more incidents that are similar in detail are grouped and details of the last incident are recorded (Series incident)

Figure 2. Number of victimizations by series victim, NCVS 1993-2012



Number of series incidents

Figure 3. Number of series incidents reported per year, NCVS, 1993-2012

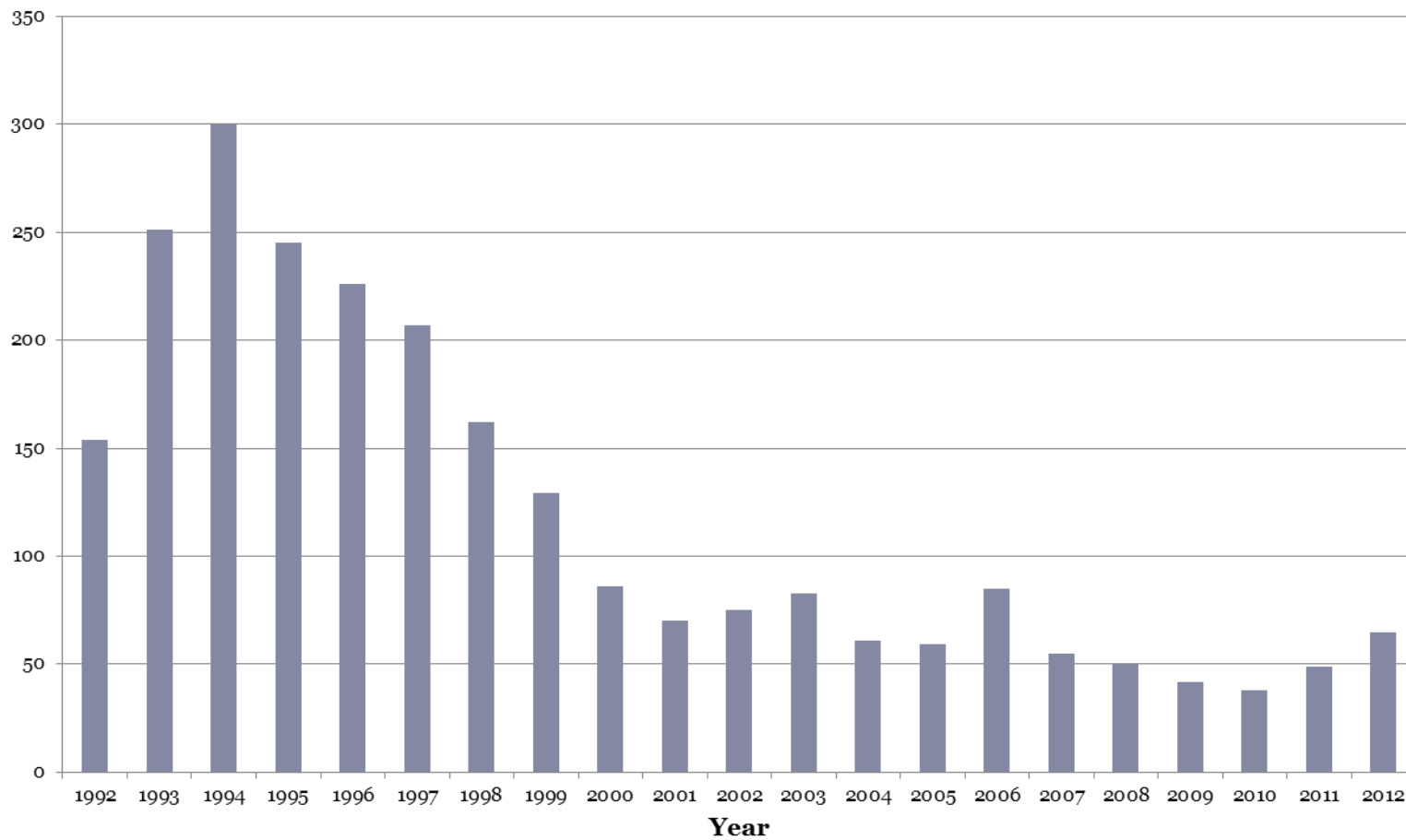
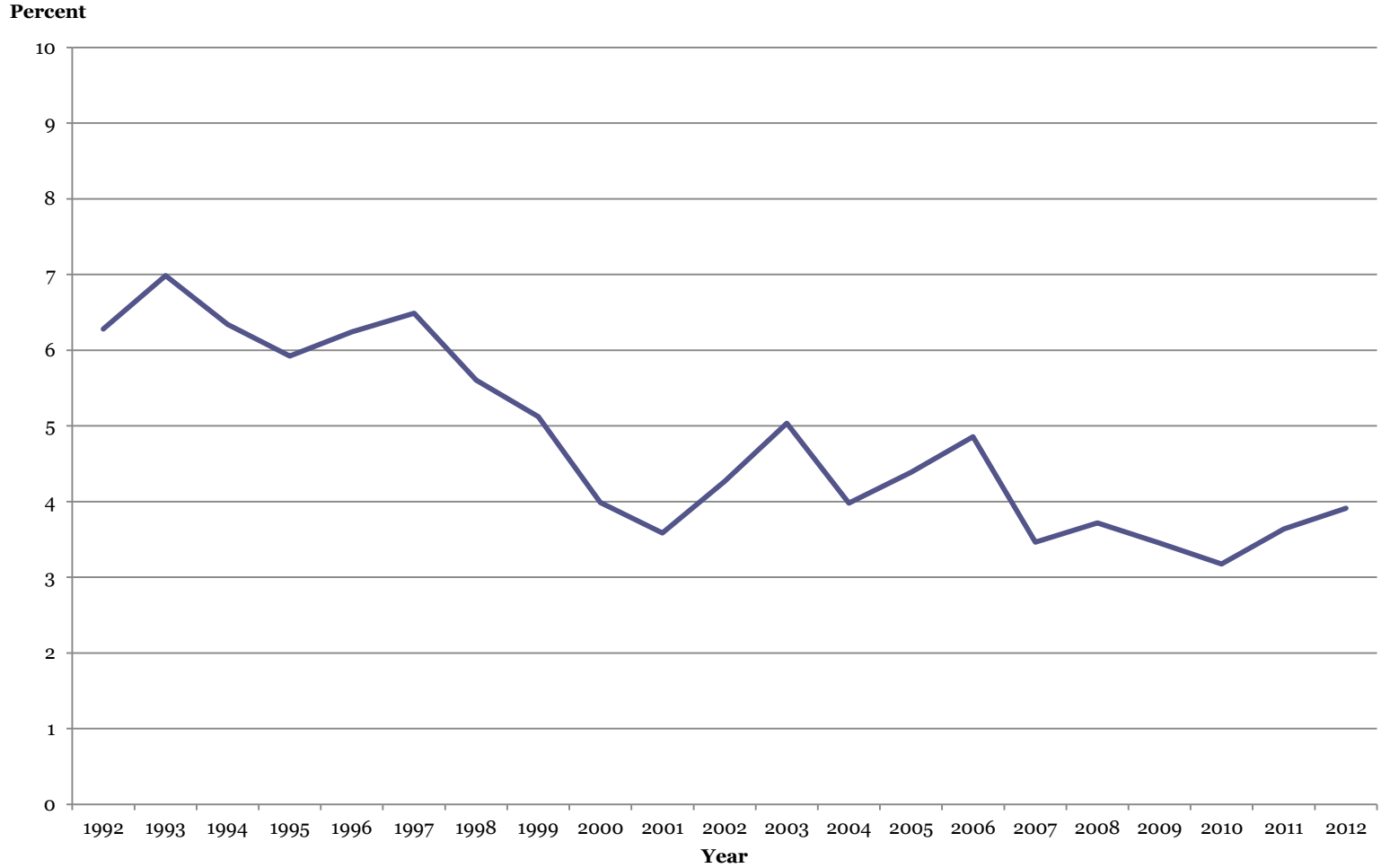


Figure 4. Percent of violent crime that are series incidents, NCVS, 1993-2012



Victimization: NCVS series incidents

Series incidents: substantively important victims

3 general categories of assaults:

- Domestic violence: intimate partners or other family members
- School violence
- Workplace violence (e.g., police officers/security guards, medical workers).

Series protocol addresses the recall and burden issues for the respondent, but at the cost of less information.

Richard Dodge (1984) “unwanted offspring of the NCS”

How should these cases be handled?

Estimation strategies- reduce bias and contain volatility/variance

Count series:

- 0 (exclude)
- 1
- cap to median (6-10)
- actual respondent estimate (6-750)

Examine impact:

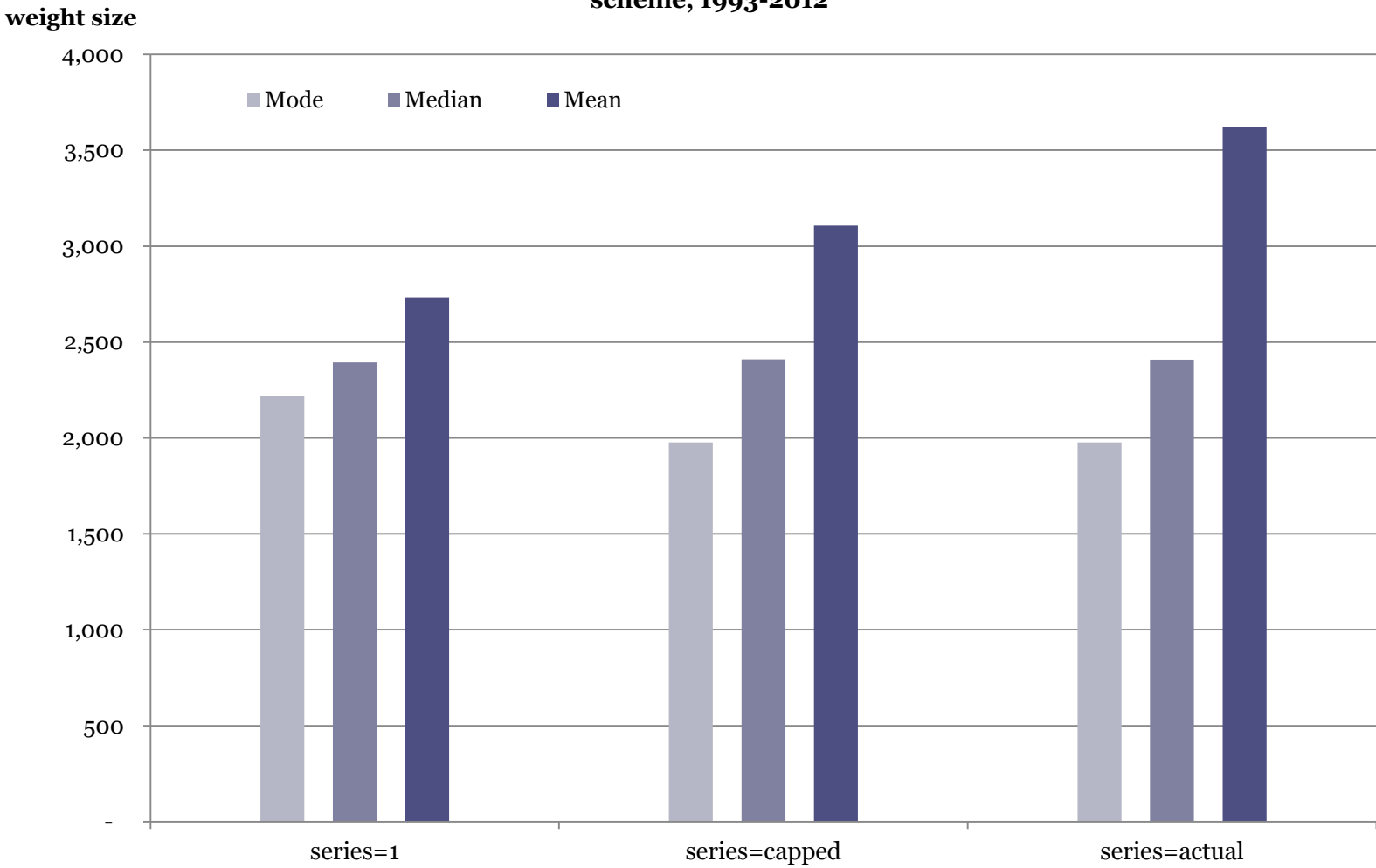
- Point estimate
- Standard error
- Coefficient of variation

Violent crime counts:

- Total violence
- Rape/sexual assault
- Robbery
- Aggravated assault
- Simple assault

Enumeration strategies- weights

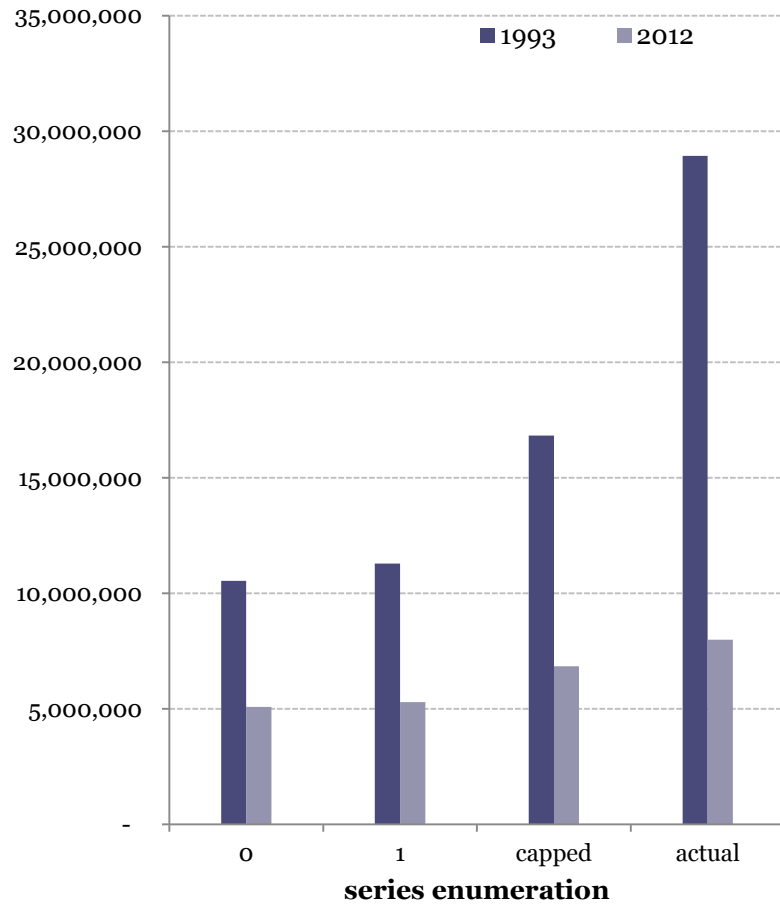
Figure 5. Measures of central tendency for victimization weights by series enumeration scheme, 1993-2012



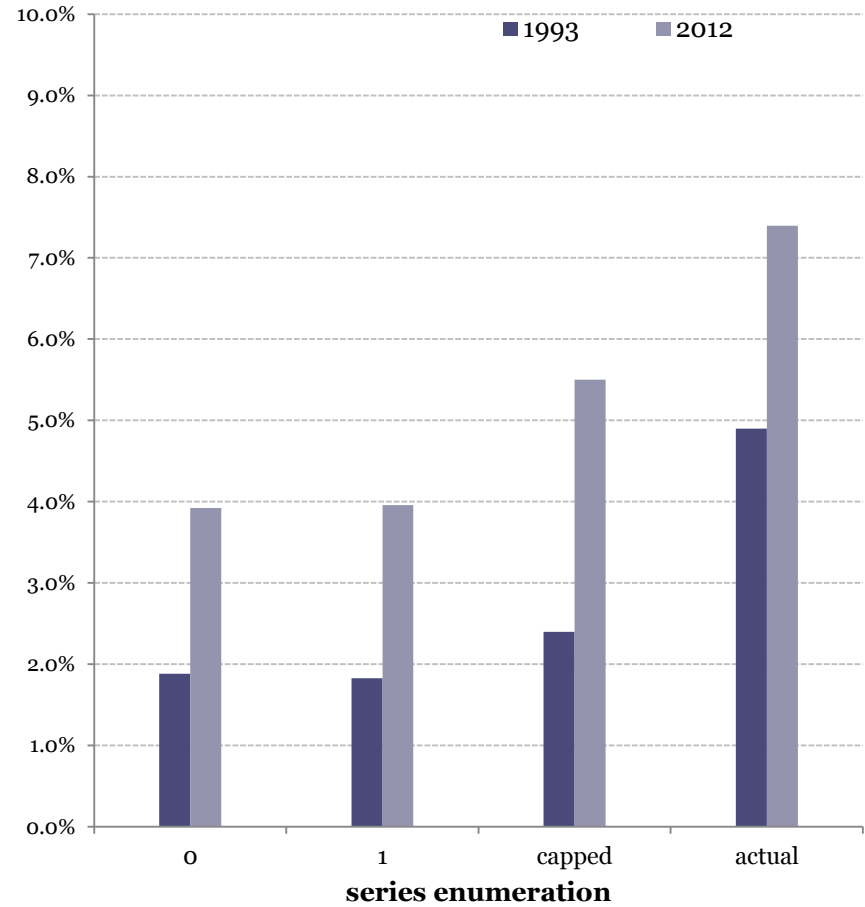
Enumeration strategies

Figure 6. Total violence, point estimates and coefficient of variation, 1993, 2012

Counts

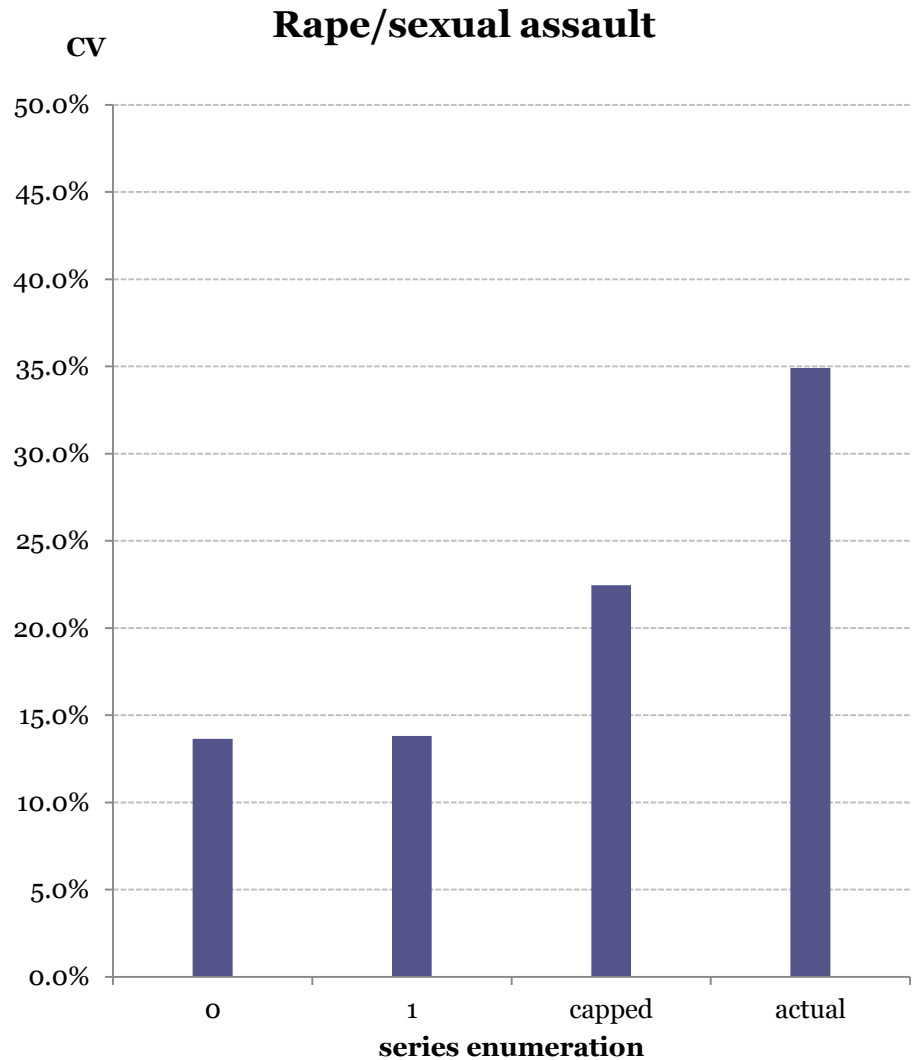
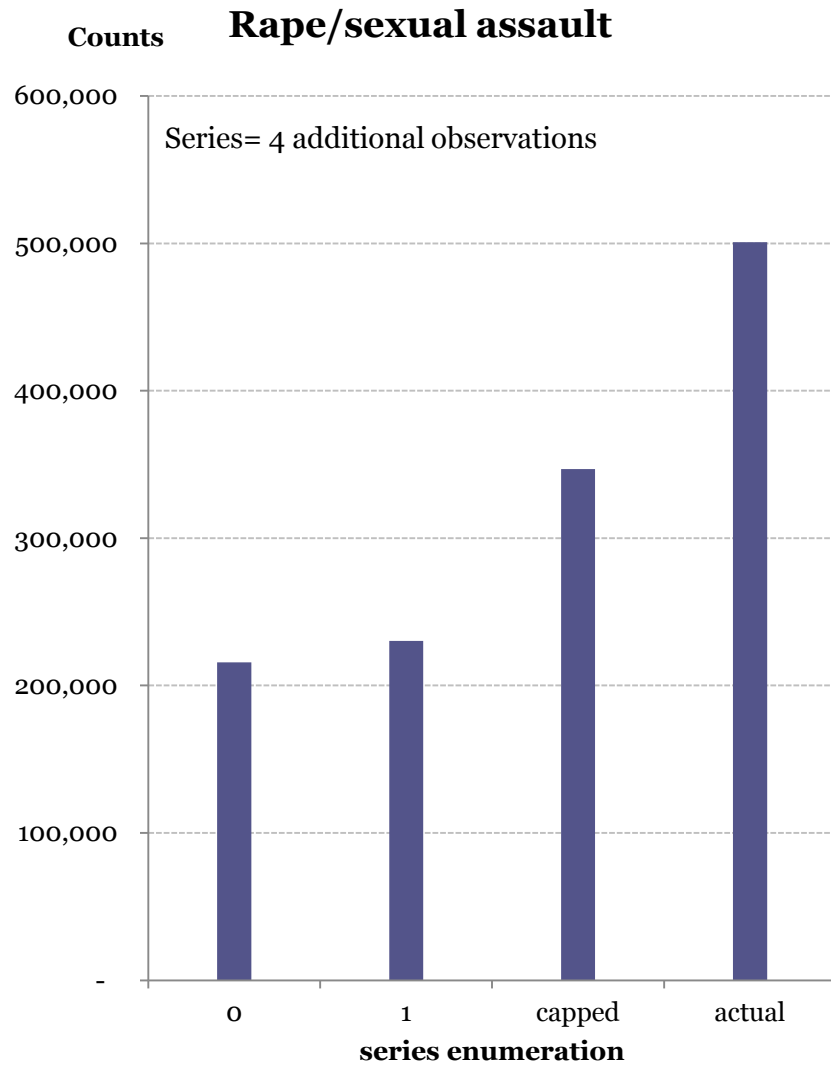


CV



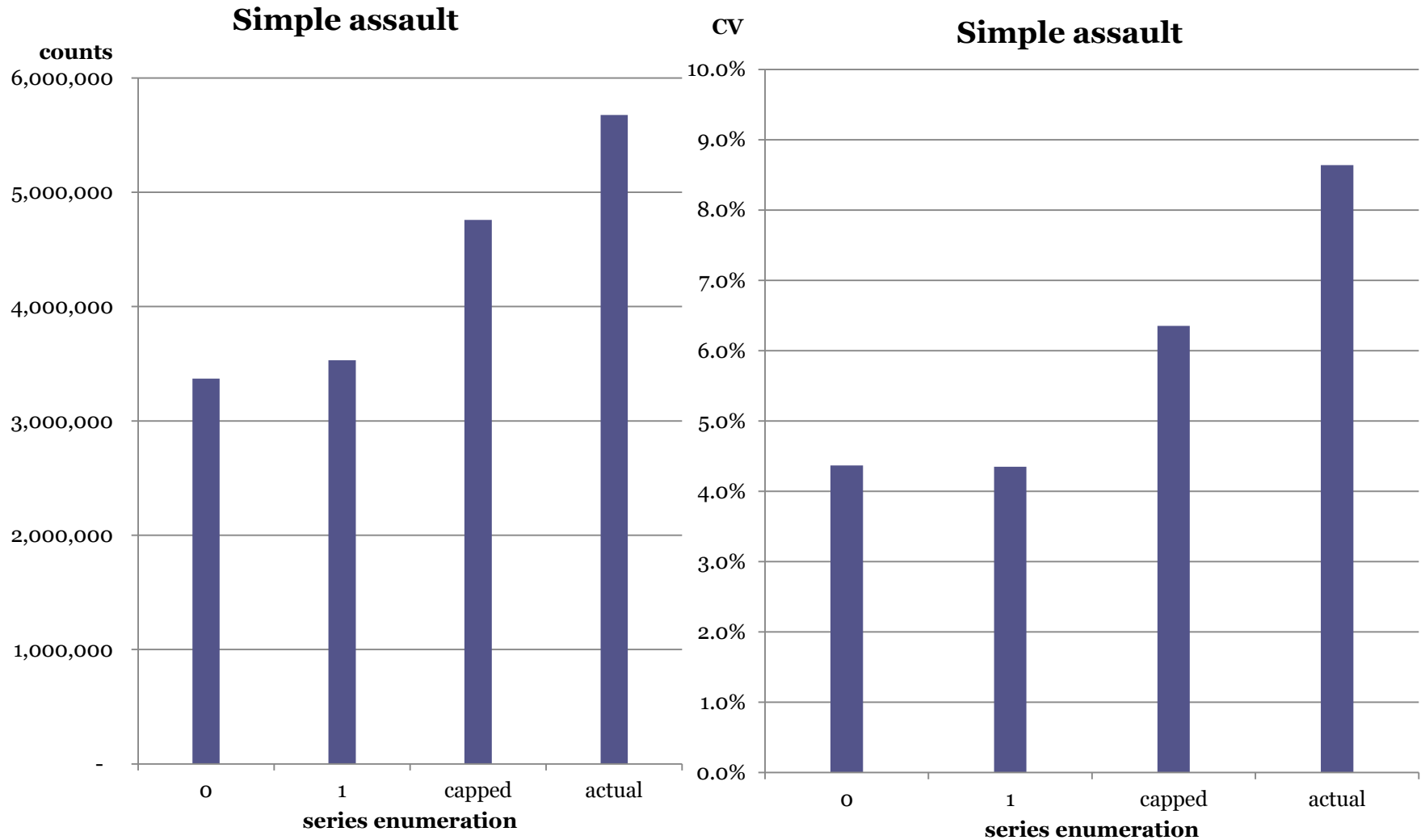
Enumeration strategies

Figure 7. Total rape/sexual assault, point estimates and coefficient of variation, 2012



Enumeration strategies

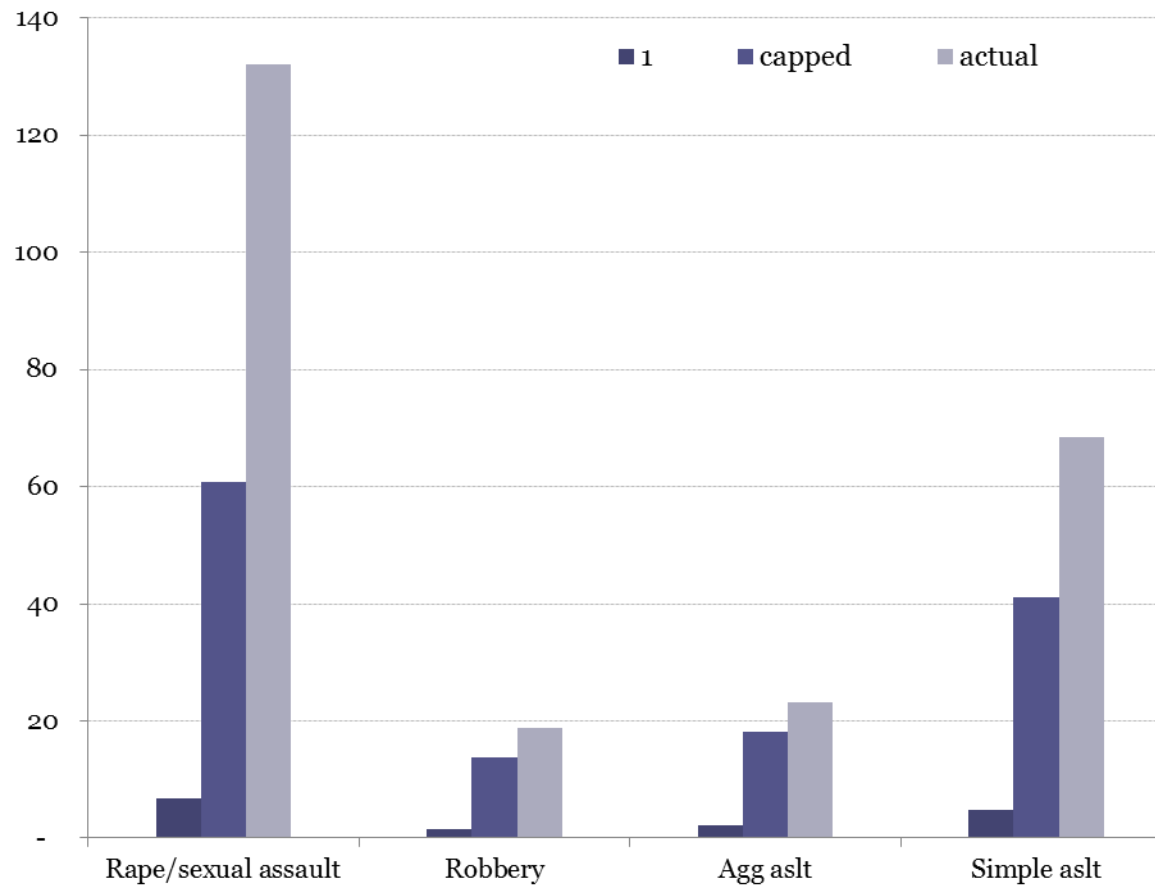
Figure 8. Total simple assault, point estimates and coefficient of variation, 2012



Enumeration strategies

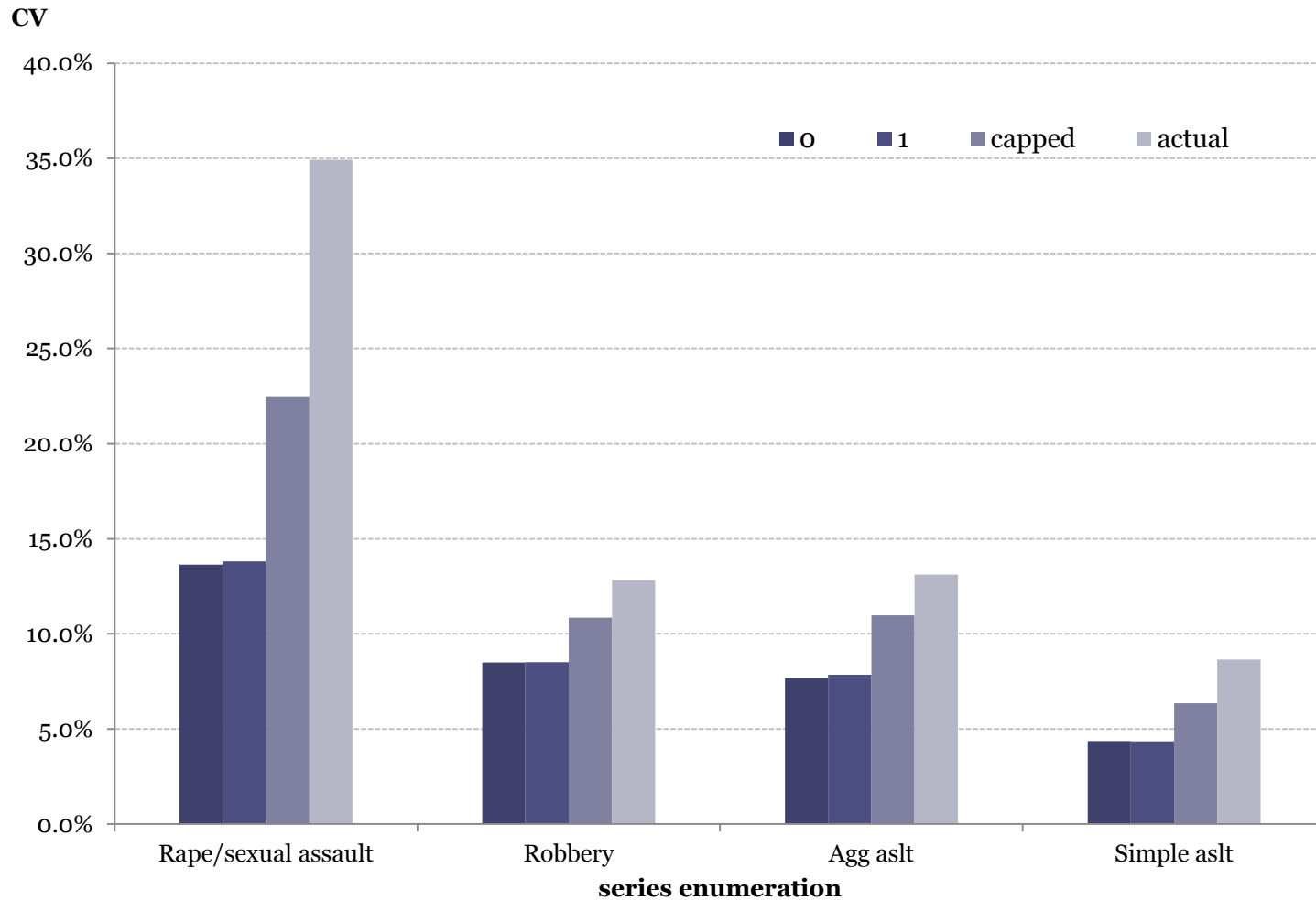
Figure 9. Type of crime, percent increase point estimates (victimization counts), 2012

Percent increase



Enumeration strategies

Figure 10. Crime type, coefficient of variation, 2012



Summary/Conclusions

- High volume repeat victims pose challenges to data collection and estimation
- Reporting actual respondent counts produces extreme and influential weights resulting in:
 - Large increases in point estimates
 - Large shifts in annual change
 - Large increases in variance, reduction in statistical power
 - Relatively rare crimes such as rape/sexual assault experience large increases in volatility, other crime types less susceptible (e.g., robbery)
 - Relatively common incidents such as simple assault account for a large number of series, but less prone to large shifts because of the large number of total observations.
- Capping series incidents with the median estimate reduces the impact on the estimate and variance while balancing substantive accuracy.
- Concerns: Inliers for subdomains and small area estimation