Survey of Sexual Victimization Sample Redesign Research

Yeng Xiong, Laura Bechtel, and Nicole Czaplicki

Statistical Methods and Sample Design Staff

Economic Statistical Methods Division

October 2023

1

Executive Summary

The Survey of Sexual Victimization (SSV) collects data on the incidence of sexual victimization in correctional facilities under the Prison Rape Elimination Act of 2003. This executive summary and research report document the research, key findings, and recommendations from the analysis conducted to evaluate and redesign the SSV sampling procedure.

Limitations:

1. State and federal prisons are assumed to be censuses with no missing data, and their data are reported as a single, aggregated observation for each frame. Results from current analysis are not applicable to these facilities.
2. Analysis does not evaluate allocation methods to determine sample size for each sampling frame but instead uses the sample sizes as requested by the Bureau of Justice.
3. The only auxiliary variable that has data available for every unit across the adult sampling frames is the average daily population, which is used as the measure of size variable. Because the public and private jail sampling frames are the only ones that have other auxiliary variables that are complete, validation results can only be obtained for public and private jails.
4. The outcome, total substantiated cases, is too rare of a characteristic in the tribal jail sample, making it difficult to get replicable and reliable results.

Key Findings:

1. There is a weak correlation between the current measure of size and total substantiated cases. However, a better alternative variable for the measure of size is not available.
2. Even with sequential sorting, some regions may not be represented in the private jail, tribal jail, and juvenile samples, sometimes for a period of years.
3. Large proportion of the private prison and private jail samples are certainty units (i.e., are deemed self-representing).
4. For units in private prison, private jail, and tribal jail frames that have a measure of size of zero, one is arbitrarily added to get a positive inclusion probability in order to acquire a probability proportional to size sample.

Recommendations:

1. Request that the state and federal prisons report the disaggregated data.
2. Make the largest unit in each region a certainty to ensure that all regions are represented in the private jail, tribal jail, and juvenile samples.
3. Have a consistent and automated sample design across all the frames.
4. Use a data-driven approach to define the certainty units.
5. Consider taking a census of the private jails.
6. Convert the sample designs of the private prison, private jail (if not a census), tribal jail and juvenile frames into stratified sequential random sampling from probability proportional to size sampling. Public jail is already a stratified sequential random sample design.
7. Estimate variances with the standard stratified sampling variance formula. If additional auxiliary variables become available, re-investigate variance estimator.
8. Create one overall frame of public jails, private prisons, private jails, and tribal jails and analytically determine sample size allocation to each component.
9. Consolidate all SSV documentation into one with a table summarizing the sampling procedure for each sampling frame.

Table of Contents

1) Survey of Sexual Victimization Background 4

1.1) Outcome and Frame Variables 4

1.2) Sample Design for Components 6

2) Research Objective 11

3) Representativeness of Sample 11

3.1) Comparison by Gender for Adult Facilities (excluding Tribal Component) 12

3.2) Comparison by Region for Adult Facilities 13

3.3) Comparisons by Offender Status for Juvenile Facilities 15

3.4) Comparison by Region for Juvenile Facilities 16

3.5) Highlights from Analysis on Representativeness of Sample 18

4) Correlation Analysis 18

4.1) Relationship between Total Substantiated and Current MOS 18

4.2) Relationship between Total Substantiated and Alternative MOS Variables 21

4.3) Highlights from Correlation Analysis 22

5) Analysis on Certainty Units Selection 22

5.1) Distribution of Current Certainty Units 23

5.2) Alternative Data-Driven Methods to Select Certainty-due-to-Size Units 24

6) Accounting for Size in Sampling (STRS vs. πPS) 30

7) Number of Strata 31

8) Sample Allocation Evaluation 34

8.1) Simulation 35

8.2) Adult Results 35

8.3) Juvenile Results 37

9) Recommendations 39

Appendix 43

1) Survey of Sexual Victimization Background

The Survey of Sexual Victimization (SSV) collects data on the incidence of inmate-on-inmate sexual victimization—defined as nonconsensual sexual acts, abusive sexual contact, and sexual harassment[1]—and staff-on-inmate sexual victimization – defined as sexual misconduct and sexual harassment- in correctional facilities under the Prison Rape Elimination Act of 2003. Determination of sexual victimization is based on allegations reported to authorities, and these allegations are investigated and then categorized as substantiated, unsubstantiated, unfounded, or investigation ongoing. The SSV publishes estimates on the number of allegations and the number of substantiated incidents.

The mandate requires that data be collected from *all* correctional facilities in the state prison systems, state juvenile correctional systems, the federal prison system, US Immigration and Customs Enforcement (ICE), and the U.S. military. Most of these report one aggregated response (rather than having one response per facility). ICE respondents report at the facility-level. Without the disaggregated data, these facilities cannot be analyzed and are excluded from this research. Heretofore, any reference to the SSV sample excludes these units.

The mandate also requires that data be collected from *a random sample* consisting of public jail jurisdictions, privately operated adult prisons and jails, facilities in Indian country, and local and private juvenile justice facilities. The overall sample must cover at least 10 percent of each facility type and have at least one facility from each state. The sample design for these facilities is described in Section 1.2.

The SSV publishes separate data tables on incidence of sexual victimization for adult facilities and juvenile facilities. For the adult facilities, estimates are produced by facility type using the broader categories of prisons (with additional break down into public prisons at the federal and state levels and private prisons), jails, and other adult facilities (includes military prisons, ICE facilities, and tribal jail facilities).

1.1) Outcome and Frame Variables

Counts of allegations and substantiated incidents are the primary survey variables for the SSV. Substantiated cases is of more interest, and thus, is the outcome variable for this research. Analyzing this variable is challenging because it is a count variable rather than a continuous one and it is sensitive. Thus, it is not surprising that substantiated cases has a high percentage of zeros (but it is unlikely that all of these are true zeros) and missing observations as shown in Table 1. Measuring the extent of response and nonresponse biases that may be present is beyond the scope of this project but should be conducted. Note also that the distribution of total substantiated cases, given in Figure 1, is not highly skewed, especially in the smaller components.

Table 1. Percentages of Zeros and Nonresponse in Number of Total Substantiated Cases

Year	% of Zero Cases in Adult Facilities	% of Missing Cases in Adult Facilities	% of Zero Cases in Juvenile Facilities	% of Missing Cases in Juvenile Facilities

[1] For specific definitions of these terms, see 2019 SSV Questionnaire.

2015	76.3%	1.9%	81.6%	4%
2016	72.4%	5.9%	79.0%	10.5%
2017	73.0%	15.7%	77.5%	21.6%
2018	70.3%	19%	74.6%	20.3%
2019	73.0%	11.4%	77.2%	13.7%

Source: 2015-2019 Survey of Sexual Victimization.

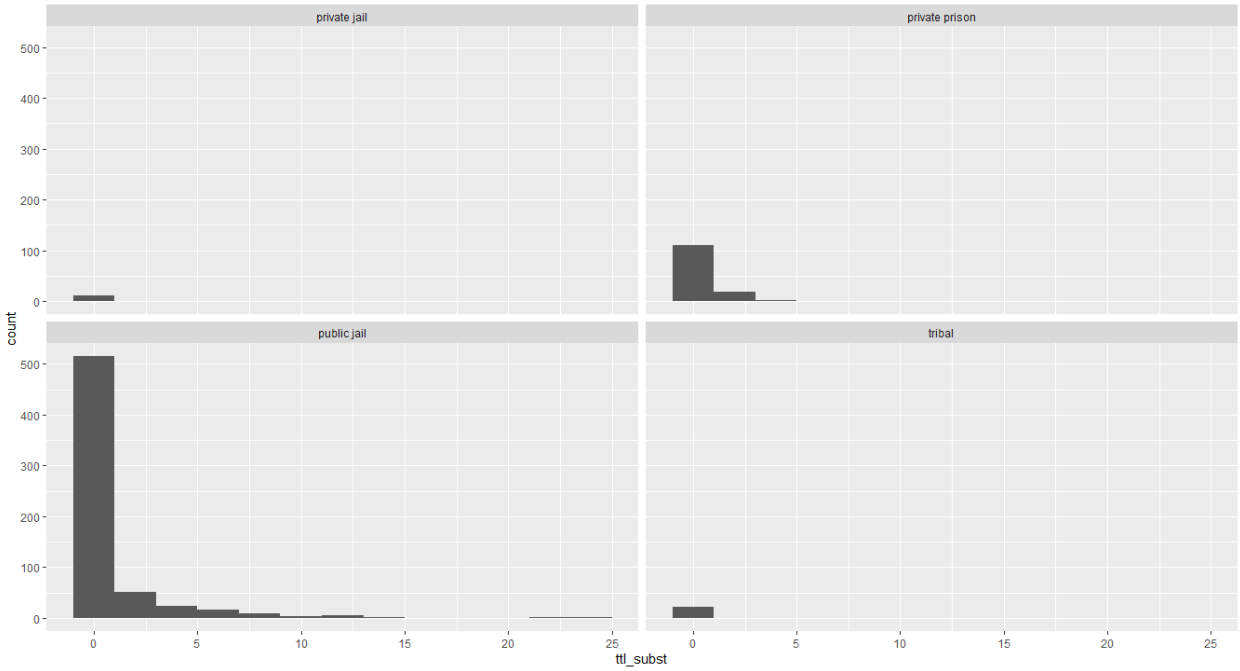


Figure 1. Distributions of total substantiated cases by facility type. Source: 2019 Survey of Sexual Victimization.

In addition to the outcome variable, there are some administrative frame variables available for analysis. All variables are listed in Table 2. Aside from the currently used measure of size (MOS) variable, average daily population (ADP), the other frame variables are not readily available across all the adult frames.

Table 2. List of administrative SSV variables available for analysis

Variables	Availability in Frame	Description
Ttl_subst	All but for sampled units only	Outcome - number of substantiated cases of sexual victimization
region	All	Census region – Midwest, Northeast, South, West
state	All	US state
ADP/adp_sum	All adult	Average daily population; measure of size; sum of adpfemale and adpmale
adpfemale	Public and private jail and private prison	Average daily female population

adpmale	Public and private jail and private prison	Average daily male population
admis	Public and private jail	Number of newly admitted inmates; sum of admisfemale and admismale
admisfemale	Public and private jail	Number of newly admitted female inmates
admismale	Public and private jail	Number of newly admitted male inmates
adultf	Public, private, and tribal jail	Number of females confined at midyear
adultm	Public, private, and tribal jail	Number of males confined at midyear
totgender	Public, private, and tribal jail	Number of confined inmates at midyear; sum of adultf and adultm
pa2bd	Juvenile	Persons assigned to beds; measure of size
num_offenders	Juvenile	Number of offenders
num_nonoffenders	Juvenile	Number of non-offenders

1.2) Sample Design for Components

The SSV sample includes a random sample of public jail jurisdictions, privately operated adult prisons and jails, facilities in Indian country, and local and private juvenile justice facilities. Administratively, these are separate and independent components within the correctional network. As such, each component is maintained as a separate frame—referred to as public jail, private prison, private jail, tribal jail, and juvenile—and a random sample is independently obtained from each.

Table 3 summarizes each component's sample design, including the sample size, criteria for certainty strata, sampling method, and the variables used for MOS and sorting. The certainty strata contain the facilities that have a sample weight of 1. Units are classified into a certainty stratum based primarily on size. Prior to sampling, noncertainty units are serpentine sorted such that units are ascendingly ordered by the first variable and descendingly by the second variable and again ascendingly by the third variable (if a third variable is present). This provides an implicit stratification.

Table 3. Summary of Sampling Procedures for SSV Frame Components

Frame Component	Sample Size	MOS	Certainty Strata	Sampling Method	Serpentine Sort on
Public Jail	700	ADP	• Largest in the state; • MOS > 1000	Stratified sequential random sample (STRS) with 3 noncertainty size strata	Region, state, MOS

Private Prison	155	ADP	• MOS ≥ cutoff[2]	Systematic probability-proportional-to-size (πPS/PPS) random sample	Region, state, MOS
Private Jail	15	ADP	• MOS ≥ cutoff	Systematic πPS random sample	Region, state, MOS
Tribal Jail	25	ADP	• MOS ≥ cutoff	Systematic πPS random sample	State, MOS
Juvenile	918 - 950	pa2bd	• Tribal facilities; • Largest in the state for local and private facilities; • MOS ≥ cutoff	Stratified systematic πPS random sample	Region, state, MOS

Source: 2015-2019 Survey of Sexual Victimization.

Public Jail

The public jail component is stratified into two certainty strata and three noncertainty strata. To meet the requirement from the Prison Rape Elimination Act of 2003 stating that at least one facility is from each state, the largest public jail from each state is placed into a certainty stratum, which will be referred to as largest-in-state certainty. The facilities with an ADP of at least 1000 are considered to be significantly large and are placed into a second certainty stratum, which will be referred to as certainty-due-to-size units. The 1000 cutoff is in line with the methodology implemented in the Census of Jails. The noncertainty strata uses Delanius and Hodges' cumulative square root of frequency (cum-root-f) method to form three size classes of ADP, and their sample size allocation is calculated with optimal allocation based on the number of confined persons (as of 12/31 in a given year). Facilities in the noncertainty strata are serpentine sorted by region, state, and ADP before sampling. In total, 700 facilities are sampled from this frame. The sample accounts for at least 24% of the public jail population, satisfying the 10% coverage mandated.

Table 4. Population and Sample from Public Jail Frame

	Frame	Population size	Sample size (% of pop)	# of largest-in-state certainty[3]	# of size certainty (MOS ≥ 1000)	# of sampled noncertainty
2015	Death in Custody	2904	700 (24.1)	46	111	543
2016	Death in Custody	2884	700 (24.3)	46	108	546

[2] Size cutoff varies from year to year but are given in Table 4 to Table 7.

[3] There are public jails in 45 states and the District of Columbia. Connecticut, Delaware, Hawaii, Rhode Island, and Vermont do not have public jails.

2017	Mortality in Correctional Institutions	2874	700 (24.4)	46	103	551
2018	Death in Custody	2841	700 (24.6)	46	106	548
2019	Census of Jails	2882	700 (24.3)	46	105	549

Source: 2015-2019 Survey of Sexual Victimization.

Private Prison

The facilities that are large, in terms of ADP, compared to the other private prisons are sampled with certainty. The cutoff ADP value varies from year to year, but how it is derived is not clearly documented. The rest of the facilities are serpentine sorted by region, state, and ADP and then sampled with probability proportional to size (πPS/PPS). For any units that have an ADP of zero, one is added to obtain a positive inclusion probability. In total, 155 facilities are sampled from this frame. The sample accounts for at least 24% of the private prison population, satisfying the 10% coverage mandated.

Table 5. Population and Sample from Private Prison Frame

	Frame	Population size	Sample size (% of pop)	Certainty MOS cutoff	# of size certainty (MOS ≥ cutoff)	# of sampled noncertainty
2015	2012 Prison Census	462	126 (27.3)	450	111	15
2016	2012 Prison Census	530	155 (29.2)	628	78	77
2017	2016 frame	518	155 (29.9)	587	81	74
2018	2017 frame	518	155 (24.6)	587	81	74
2019	2019 Census of State and Federal Adult Correctional Facilities	406	155 (38.2)	321	75	80

Source: 2015-2019 Survey of Sexual Victimization.

Private Jail

The facilities that are large, in terms of ADP, compared to the other private jails are sampled with certainty. In 2015-2017, the ADP cutoff is 900 but increases to 1000 in 2018-2019 to be in line with methodology implemented in Census of Jails. The rest of the facilities are serpentine sorted by region, state, and ADP and then sampled with πps design. For any units that have an ADP of zero, one is added to obtain a positive inclusion probability. In total, 15 facilities are sampled from this frame. The sample accounts for at least 38% of the private jail population, satisfying the 10% coverage mandated.

Table 6. Population and Sample from Private Jail Frame

	Frame	Population size	Sample size (% of pop)	Certainty MOS cutoff	# of size certainty (MOS ≥ cutoff)	# of sampled noncertainty

2015	Deaths in Custody	39	15 (38.5)	900	7	8
2016	Deaths in Custody	39	15 (38.5)	900	7	8
2017	Mortality in Correctional Institutions	36	15 (41.7)	900	6	9
2018	Deaths in Custody	37	15 (40.5)	1000	6	9
2019	Census of Jails	29	15 (51.7)	1000	4	11

Source: 2015-2019 Survey of Sexual Victimization.

Tribal Jail

The facilities that are large, in terms of ADP, compared to the other tribal jails are sampled with certainty. The cutoff ADP value varies from year to year, but how it is derived is not clearly documented. The rest of the facilities are serpentine sorted by state and ADP and then sampled with πps design. For any units that have an ADP of zero, one is added to obtain a positive inclusion probability. In total, 25 facilities are sampled from this frame. The sample accounts for at least 38% of the tribal jail population, satisfying the 10% coverage mandated.

Table 7. Population and Sample from Tribal Jail Frame

	Frame	Population size	Sample size (% of pop)	Certainty MOS cutoff	# of size certainty (MOS ≥ cutoff)	# of sampled noncertainty
2015	Jails in Indian Country	57	25 (38.5)	83	7	18
2016	Jails in Indian Country	60	25 (38.5)	78	7	18
2017	Jails in Indian Country	63	25 (41.7)	107	6	19
2018	Jails in Indian Country	61	25 (40.5)	104	7	18
2019	Census of Jails	60	25 (51.7)	104	7	18

Source: 2015-2019 Survey of Sexual Victimization.

Juvenile

The tribal and nontribal juvenile frames are created from different files. Tribal juvenile facilities are from the Jails in Indian Country (JIC), and nontribal juvenile facilities are either from the Juvenile Residential Facility Census (JRFC) or the Census of Jails in Residential Placement (CJRP). With so few tribal juvenile facilities, these units are placed into a separate stratum and included into the sample with certainty.

Table 8. Population and Sample from Juvenile Frame

	Frame	Population size	Sample size (% of pop)	# of tribal certainty	# of initial certainty[4]	# of size certainty (MOS ≥ cutoff)	# of sampled noncertainty
2015	2014 JRFC	1890	950 (50.3)	19	485	122	324
2016	2015 CJRP	1847	918 (49.7)	19	453	114	332
2017	2016 JRFC	1752	918 (52.4)	21	455	130	333
2018	2017 CJRP	1793	922 (51.4)	21	456	130	315

Source: 2015-2018 Survey of Sexual Victimization.

Out of the nontribal juvenile facilities, state central reporters and facilities that report separately are in the sample with certainty. DC facilities are considered as state facilities. The rest are grouped by the facility type: detention, local, and private. The local and private facilities that are largest in each state are also in the sample with certainty. Within each of the facility types, large facilities are separated into a corresponding take-all stratum. A large facility is defined as one with a MOS greater than the Hansen-Hurwitz cutoff

$$c = K \cdot Y/n$$

where Y is the total MOS for the facility group, n is the sample size in the facility group, and $K = 0.85$. The remaining noncertainty facilities are divided into 7 strata. See Table 8 for a summary of the stratum design for the nontribal juvenile facilities. Sample size for each noncertainty stratum is determined with proportional allocation. Noncertainty facilities are serpentine sorted by region (where applicable), state, collapsed facility type[5], and MOS within stratum and then sampled with a systematic πps design. In total, over 900 facilities are sampled from this frame. The sample accounts for at least 49% of the juvenile jail population, satisfying the 10% coverage mandated.

Table 9. Strata Description and Sample Size from Juvenile Frame

Stratum Label	Description	Sample Size
10	State central reporters and facilities that report separately	371
15	Tribal juvenile facilities	21
20	Largest locally-operated facility in each state	37
30	Largest privately-operated facility in each state	48
40A	Midwest detention centers	41
40B	Northeast detention centers	11
40C	South detention centers	37
40D	West detention centers	29
40E	Detention facility certainties due to size	47
51	Local non-commitment	5
52	Local commitment	28
53	Local facility certainties due to size	31
60	Private facilities	164

[4] The initial certainty count includes state facilities, largest locally-operated facilities in each state, and largest privately-operated facilities in each state.

[5] A collapsed facility means that a facility has been assigned to a new type and stratum number since there was not enough respondents in the stratum or the unit is too big and is converted to a certainty.

63	Private facility certainties due to size	52

Source: 2015-2019 Survey of Sexual Victimization.

2) Research Objective

The population for the SSV is all correctional facilities in the U.S., but since only the public jails, private prisons, private jails, tribal jails, and juvenile facilities are sampled, they are the target population in scope for our sample redesign research. Currently, they operate as separate populations with independent frames and sample designs. For the adult facilities, we highly recommend having just one sample design that can select facilities from each component but within one frame. This requires, though, having a consistent methodology that can be applied across the components and automating the process (to the greatest extent possible).

In designing a consistent sample methodology, we need to consider what features of the population should be preserved in the sample. Since the mandate requires a coverage of each state and each facility type, these should be accounted for in the sample design. Additionally, as is typical for establishment surveys, larger units should be retained as much as possible in the sample to produce estimates that are reasonably close to the truth. This is because larger units tend to contribute more to estimates in establishment surveys. The current design already accounts for much of these features. A sample is selected from each component, which corresponds to facility type, and facilities are geographically sorted by region (except for the tribal jail component) and state prior to selection.

Size is also already incorporated in the current design but in an inconsistent manner. In the public jail component, size is used to create strata whereas, in the other smaller components, inclusion probabilities are calculated proportionally to size. Also, while all the components include a selection of the largest facilities with certainty into their samples, determining the size threshold to identify these units differs across the components. Our research then concentrates on establishing a sample design accounting for size, including developing an approach to designate certainty units, that can be uniformly applied to all the components.

We first review the current sampling methodology by:

- assessing how adequately the current sample represent the target population (Section 3),
- determining if the current MOS is correlated with the outcome (Section 4.1),
- exploring if other administrative variables are more correlated with the outcome and could be used as an alternate MOS variable (Section 4.2),
- and comparing other possible certainty selection methods to the current one (Section 5.2).

From this, we find that a stratified sequential random sample (STRS) may be more appropriate functionally than a probability proportional to size sample. Under such a design, we discuss how many strata should be used in Section 7, and we investigate allocation methods through a simulation study in Section 8. The concluding section contains our recommendations (Section 9).

Some additional notes to bear in mind regarding our research, analyses are completed component by component. For ease of reading, only a subset of analysis plots are included in this report, but all analysis plots may be found here.

3) Representativeness of Sample

A probability sample (e.g., STRS sample, πps random sample) typically resembles its target population, but it is possible that it might not fully capture a subgroup, especially a small one. By comparing the sample to the population within a covariate subgroup, we can determine if the sample is imbalanced and correct for it in the sample design.

For the adult components, we compare the weighted MOS distribution between the sample and the population within gender (female/male) and region (South, West, Midwest, and Northeast); for the juvenile component, comparisons are within offender status (offender/nonoffender) and region. Region is included as a sorting variable in the current sample design for all but the tribal component, so it is controlled for but not the other subgroups. Therefore, if the sample and the population are similar within the subgroups, this is an indication that the sample is representative of the population with respect to the particular subgroup. On the other hand, if there is a large discrepancy in a subgroup, then the sample design may need to directly control for it to get better representation of it.

3.1) Comparison by Gender for Adult Facilities (excluding Tribal Component)

The data for the adult facilities (public jail, private prison, and private jail components) breaks down the MOS of average daily population (ADP) by gender. The tribal component did not have this information and is excluded from this analysis. We calculate the sample percentage of the total ADP that is male and female as

$$\hat{\mathrm{p}}_{\mathrm{M}}^{ADP} = 100 \cdot \frac{\sum_i^n w_i x_{M,i}}{\sum_i^n w_i x_i} \text{ and } \hat{\mathrm{p}}_{\mathrm{F}}^{\mathrm{ADP}} = 100 \cdot \frac{\sum_i^n w_i x_{F,i}}{\sum_i^n w_i x_i}$$

where w_i is facility i's sample weight, $x_{M,i}$ is a facility's male ADP, $x_{F,i}$ is a facility's female ADP and x_i is a facility's ADP. The population percentage is similarly computed but with the sum over all the facilities in the component population and $w_i = 1$ for all facilities.

Figure 1 plots the percentages for each component from 2015 to 2019. The red line represents the population percentage; blue is the sample percentage. Overall, the sample seems to represent the population with respect to gender quite well since the two lines are generally overlapping. This holds for all the components.

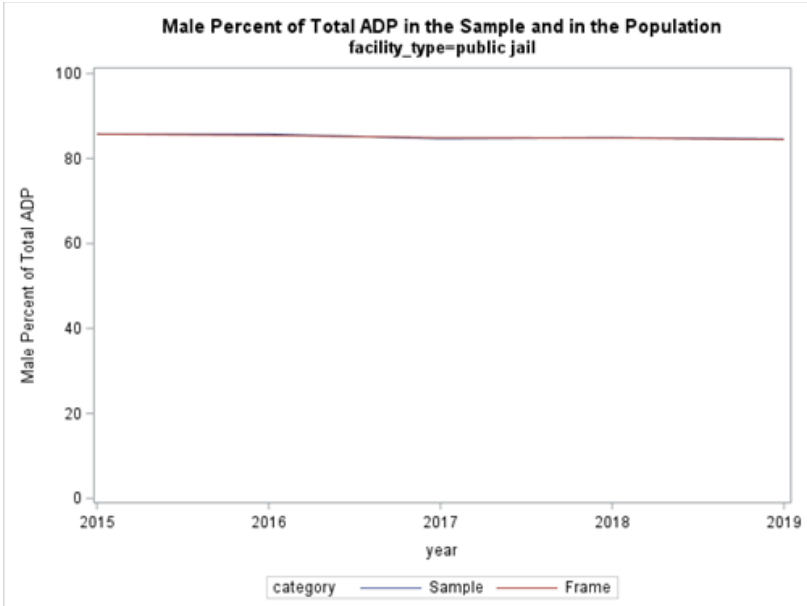


Figure 2. Comparison of percentage of total ADP that is male between the sample (blue) and population (red) for public jail component. Source: 2015-2019 Survey of Sexual Victimization.

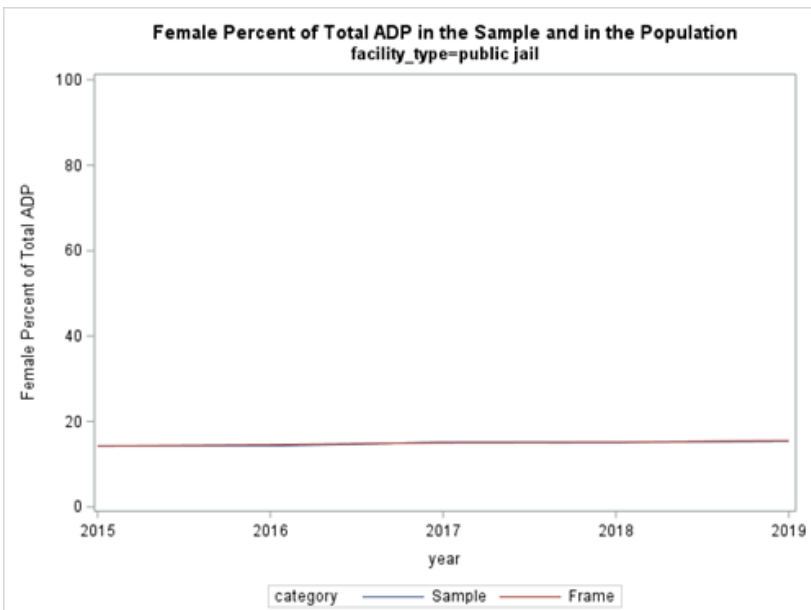


Figure 3. Comparison of percentage of total ADP that is female between the sample (blue) and population (red) for public jail component. Source: 2015-2019 Survey of Sexual Victimization.

3.2) Comparison by Region for Adult Facilities

We calculate the sample percentage of the total ADP within a region as

$$\hat{\mathrm{p}}_{\mathrm{R}}^{ADP} = 100 \cdot \frac{\sum_i^n w_i x_i I_{R,i}}{\sum_i^n w_i x_i}$$

where w_i is facility i's sample weight, x_i is a facility's ADP and $I_{R,i}$ is a region indicator ($I_{R,i} = 1$ if a facility is in region $R = \{N\text{ortheast}, \text{Midwest}, \text{West}, \text{South}\}$ and $I_{R,i} = 0$ otherwise). The population

percentage is similarly computed but with the sum over all the facilities in the component population and $w_i = 1$ for all facilities.

In the public jail and private prison components, the sample closely follows the population across the different region. However, in the private jail and tribal jail components, there is a notable difference between the sample and population. See Figure 3 for private jail and Figure 4 for tribal jail.

It should be noted that in 2018 and 2019, only one private jail facility was selected from the Northeast, and it was a certainty unit. Same in 2017 from the Midwest. In 2016 and 2018, no private jail facilities were selected from the West. The private jail component is the smallest with only 40 facilities in the population. Even though the private jail component includes region as a sorting variable, with such a small frame, the sample may miss some of the regions. Given this, it seems reasonable to collect data from all private jail facilities if possible.

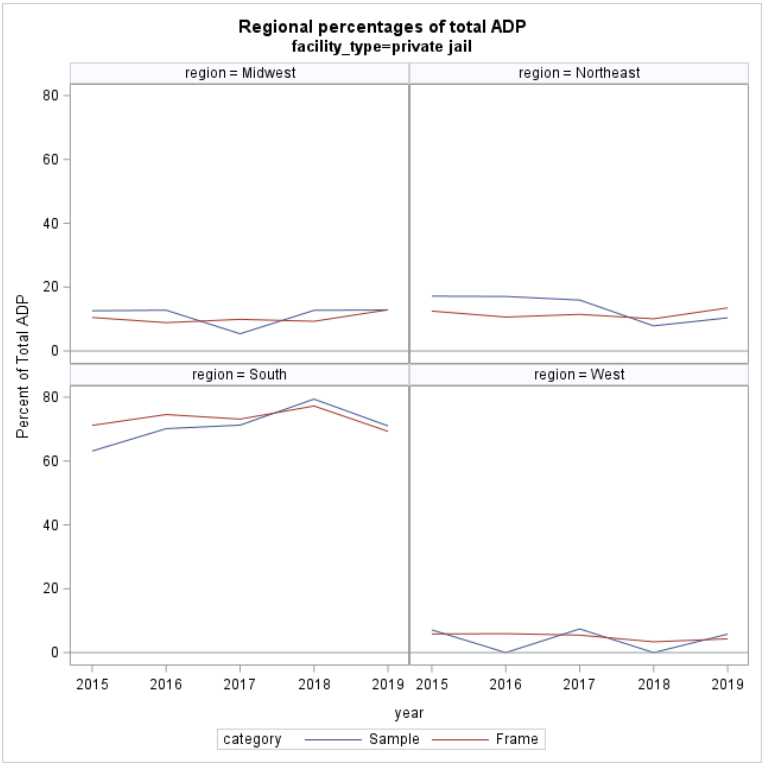


Figure 4. Comparison of percentage of total ADP within region between sample (blue) and population (red) for private jail component. Source: 2015-2019 Survey of Sexual Victimization.

The sample for the tribal jail component seems to be similar to the population in the Midwest but has some deviations in the South and West. The tribal jail component does not include region as a sorting variable. Therefore, adding it may help improve the sample imbalance in the West. However, in the South, there is only one or two facilities in the population. These units could be made into a certainty even though it is not a large facility to ensure that every region is represented in the sample.

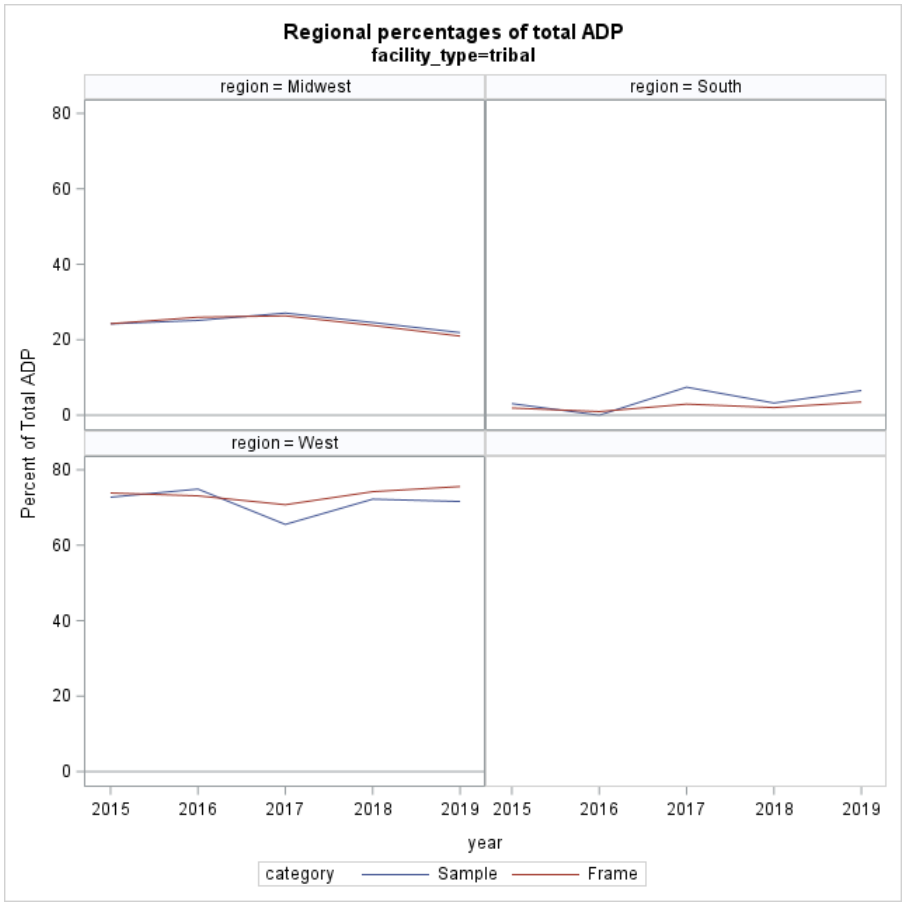


Figure 5. Comparison of percentage of total ADP within region between sample (blue) and population (red) for tribal jail component. Source: 2015-2019 Survey of Sexual Victimization.

3.3) Comparisons by Offender Status for Juvenile Facilities

The data for the juvenile facilities breaks down the MOS of persons assigned to bed (pa2bd) by offender status. We calculate the sample percentage of the pa2bd that is offender and nonoffender within stratum h as

$$\hat{\mathrm{p}}_{\mathrm{O,h}}^{PA2BD} = 100 \cdot \frac{\sum_{i \in h}^{n_h} w_i z_{O,i}}{\sum_{i \in h}^{n_h} w_i z_i} \text{ and } \hat{\mathrm{p}}_{N,h}^{\mathrm{PA2BD}} = 100 \cdot \frac{\sum_{i \in h}^{n_h} w_i z_{N,i}}{\sum_{i \in h}^{n_h} w_i z_i}$$

where w_i is facility i's sample weight, $z_{O,i}$ is the pa2bd that are offenders, $z_{N,i}$ is the pa2bd that are nonoffender and z_i is a facility's pa2bd. The population percentage is similarly computed but with the sum over all the facilities in the component population and $w_i = 1$ for all facilities. Certainties are not included in the percentage calculation because they are always sampled. The juvenile sample tends to be representative of the population. The only stratum that differs is the local non-commitment, but this has a small sample size.

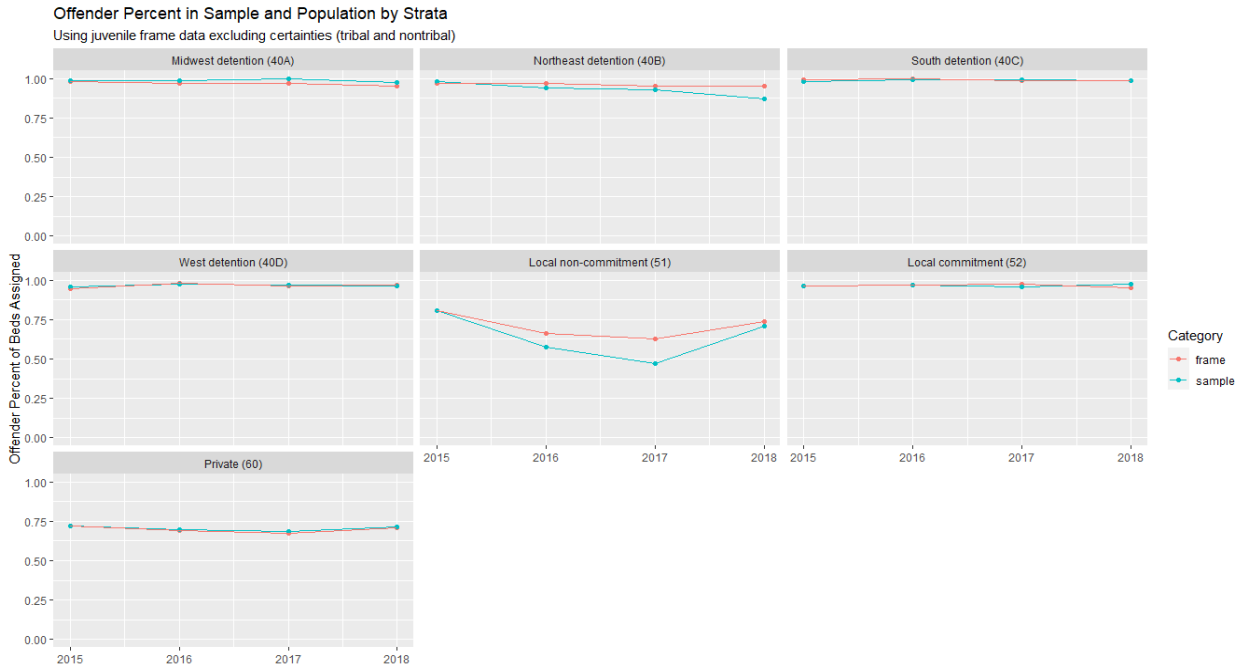


Figure 6. Comparison of percentage of total persons assigned to beds that is offender between the sample (blue) and population (red) for strata within the juvenile component. Source: 2015-2018 Survey of Sexual Victimization.

3.4) Comparison by Region for Juvenile Facilities

We calculate the sample percentage of the pa2bd within a region for a stratum as

$$\hat{\mathrm{p}}_{\mathrm{R,h}}^{PA2BD} = 100 \cdot \frac{\sum_{i \in h}^{n_h} w_i z_{\,i} I_{R,i}}{\sum_{i \in h}^{n_h} w_i z_i}$$

where w_i is facility i's sample weight, z_i is a facility's pa2bd and $I_{R,i}$ is a region indicator ($I_{R,i} = 1$ if a facility is in region $R = \{\mathrm{Northeast}, \mathrm{Midwest}, \mathrm{West}, \mathrm{South}\}$ and $I_{R,i} = 0$ otherwise). The population percentage is similarly computed but with the sum over all the facilities in the component population and $w_i = 1$ for all facilities. Certainties are not included in the percentage calculation because they are always sampled. The juvenile stratum samples seem to be fairly balanced across the regions. The exceptions are local commitment stratum in the Northeast and the local noncommitment stratum. This, again, is a result of the small number of facilities that are in these populations, especially for the local noncommitment facilities in the West where there is, in total, 0-1 facility (in 2016, there was none) and for the local commitment facilities in the Northeast where there are, in total, 0-3 facilities (in 2015, there was none). With so few facilities in these subpopulations, they are true self-representing units and should be sampled with certainty.

Figure 7. Comparison of percentage of total persons assigned to beds within region between the sample (blue) and population (red) for local noncommitment stratum within the juvenile component. Source: 2015-2018 Survey of Sexual Victimization.

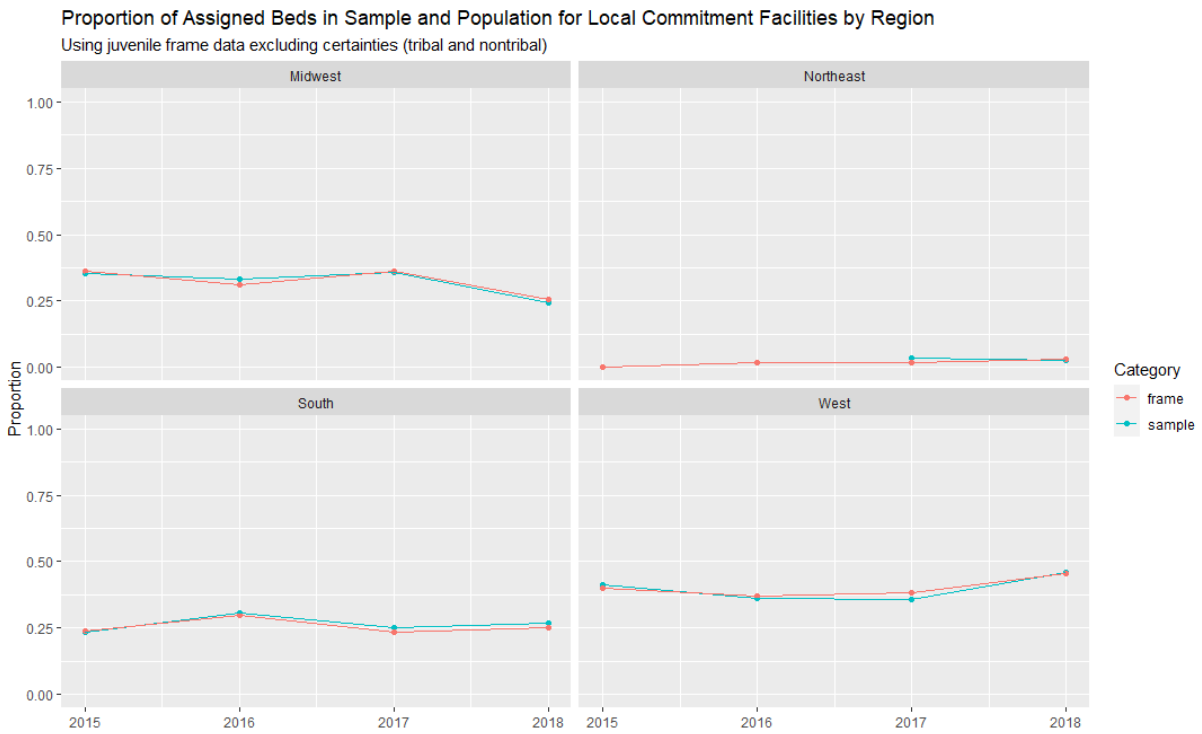


Figure 8. Comparison of percentage of total persons assigned to beds within region between the sample (blue) and population (red) for local commitment stratum within the juvenile component. Source: 2015-2018 Survey of Sexual Victimization.

3.5) Highlights from Analysis on Representativeness of Sample

- The samples are generally of the same composition as the adult component frames with respect to the proportion of male and female ADP and region.
- In some years, not all regions are represented in the public jail and tribal jail samples.
- The juvenile sample is generally of the same composition as the juvenile frame with respect to the proportion of offender and nonoffender ADP and region.
- Within a facility type of the juvenile sample, a region may fail to be represented, sometimes consecutively for years. Local noncommitment facilities in the West have not been sampled at all from 2015-2018 (but there were no facilities to be sampled in 2016).
- If there is a region with only one or two facilities, this region's facilities should be selected with certainty as they are definitely self-representing facilities.

4) Correlation Analysis

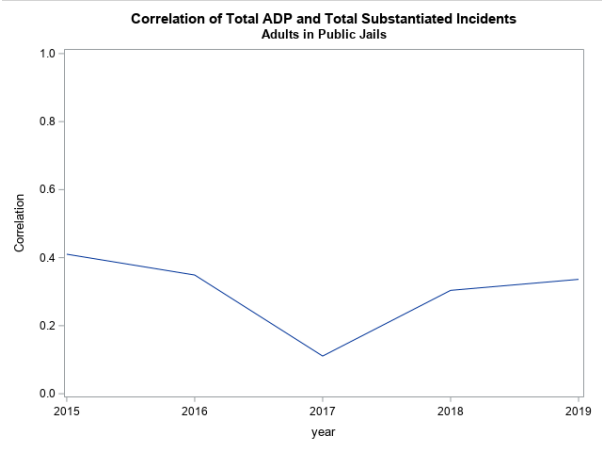
All the sample designs for the component frames set up inclusion probability as a function of MOS. The public jail component frame does so by dividing facilities into strata based on MOS, and the other smaller component frames by using MOS to directly determine inclusion probability. This helps obtain an efficient sample. However, the gain in efficiency largely depends on the assumption that the MOS is highly correlated with the outcome. In this section, an unweighted correlation analysis is performed to test this assumption and to see if there is another administrative variable that may be better suited as the size variable.

4.1) Relationship between Total Substantiated and Current MOS

From 2015-2019, the Pearson correlation for public jail is between 0.11 and 0.41; private prison is between 0.22 and 0.51; private jail is between -0.09 and 0.77; tribal jail is -0.07 in 2015 and -0.15 in 2018; and juvenile is between 0.14 and 0.53. See the first plot of Figure 9-Figure 13. From these results, the relationship between total substantiated cases and the current MOS tends to range from a nonexistent one to a weakly positive one. Private jail does have a surprisingly high correlation value of 0.77 in 2015, but it should be noted that it is a small frame, having less than 40 facilities. Therefore, its results are more varied, as seen in Figure 11a, and more susceptible to outliers, as is the case in 2015. Also of note, the correlation for tribal jail cannot be calculated in 2016, 2017, and 2019 because no substantiated case was found in those years. Even in 2015 and 2018, only one or two substantiated cases are discovered.. With so few cases, though, obtaining valid and generalizable results is difficult.

Scatterplots of total substantiated cases versus current MOS are also produced to visually inspect the relationship between the two variables. See the second plot of Figure 9-Figure 13, in which 2019 data is used, but the pattern is also consistent for other years; see Correlation Analysis Plots. The scatterplots agree with the correlation conclusion of a generally weak relationship. In addition, the scatterplots for private prison and juvenile facilities illustrates that, aside from a few (possibly outlying) cases, total substantiated cases are uniformly distributed across MOS despite having a correlation around 0.4 in 2019. Given these results, ADP (for adult facilities) and persons assigned to beds (for juvenile facilities) may not be the most appropriate choice for MOS.

Public Jail Correlation Plots



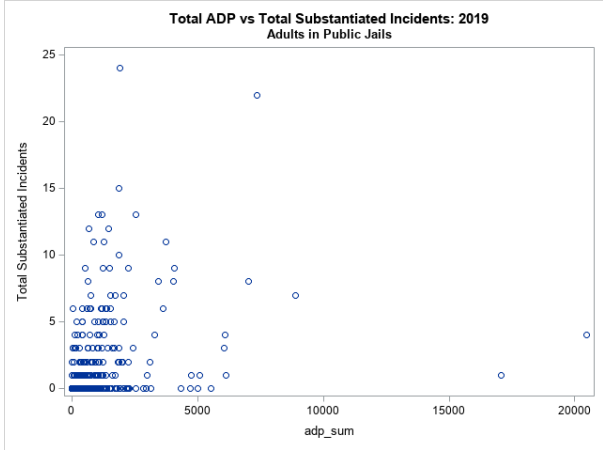
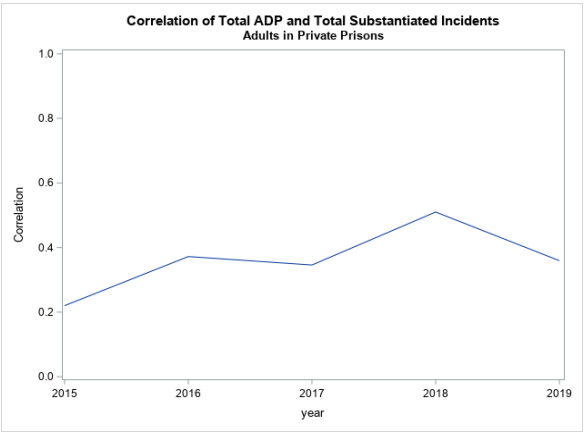


Figure 9. Figure 9a (left) plots the Pearson correlation between ADP and total substantiated cases for public jails from 2015-2019. Figure 9b (right) plots the 2019 public jail data of total substantiated cases against ADP. Source: 2015-2016 Survey of Sexual Victimization.

Private Prison Correlation Plots



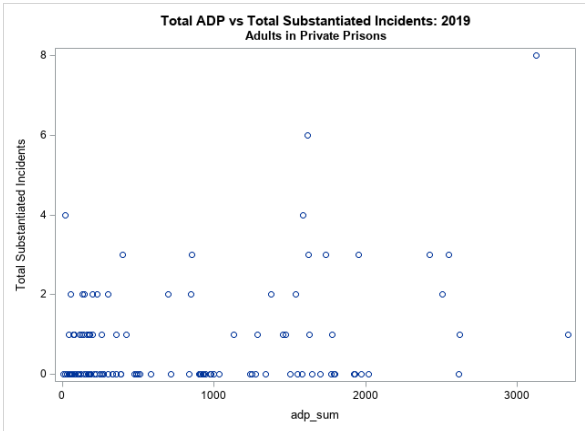
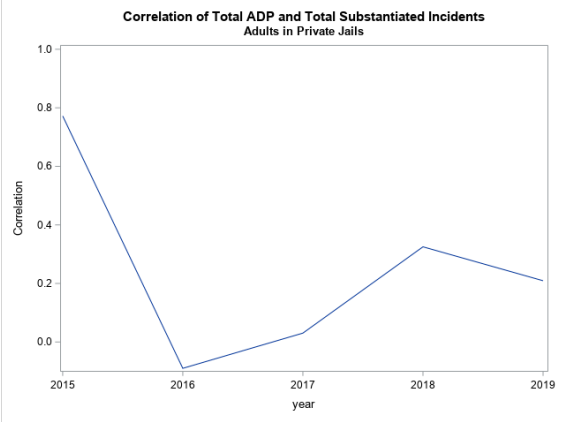


Figure 10. Figure 10a (left) plots the Pearson correlation between ADP and total substantiated cases for private prisons from 2015-2019. Figure 10b (right) plots the 2019 private prison data of total substantiated cases against ADP. Source: 2015-2016 Survey of Sexual Victimization.

Private Jail Correlation Plots



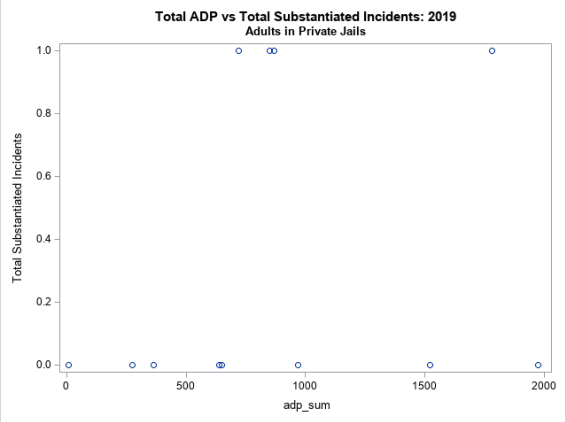
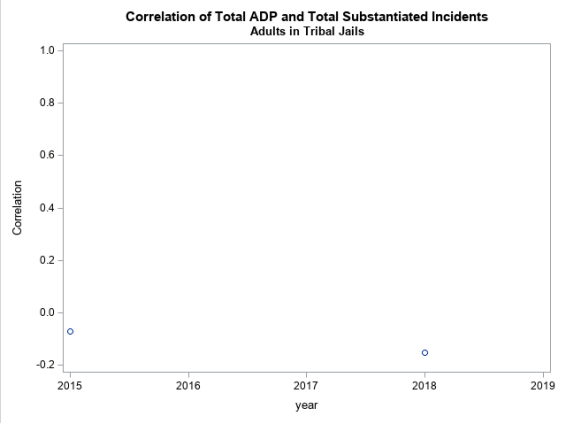


Figure 11. Figure 11a (left) plots the Pearson correlation between ADP and total substantiated cases for private jails from 2015-2019. Figure 11b (right) plots the 2019 private jail data of total substantiated cases against ADP. Source: 2015-2016 Survey of Sexual Victimization.

Tribal Jail Correlation Plots



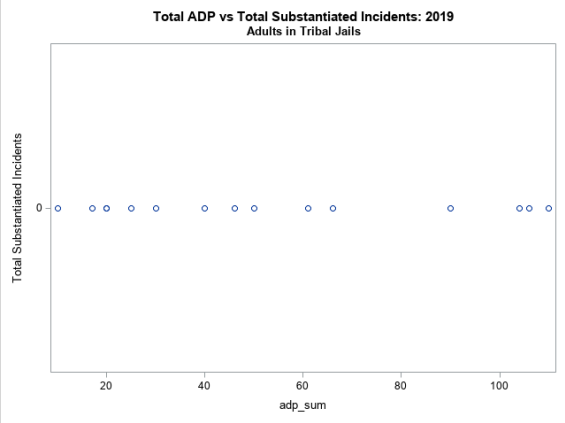
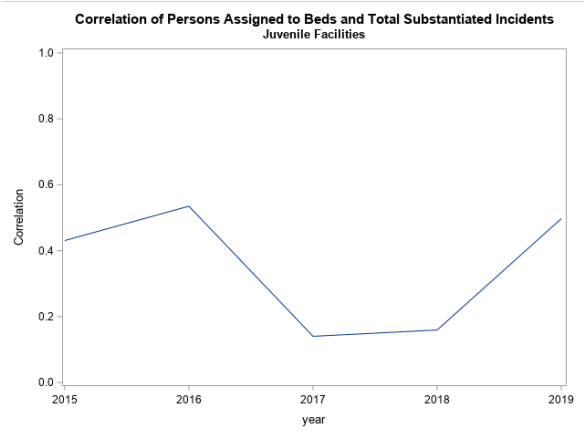


Figure 12. Figure 12a (left) plots the Pearson correlation between ADP and total substantiated cases for tribal jails from 2015-2019. Figure 12b (right) plots the 2019 tribal jail data of total substantiated cases against ADP. Source: 2015-2016 Survey of Sexual Victimization.

Juvenile Correlation Plots



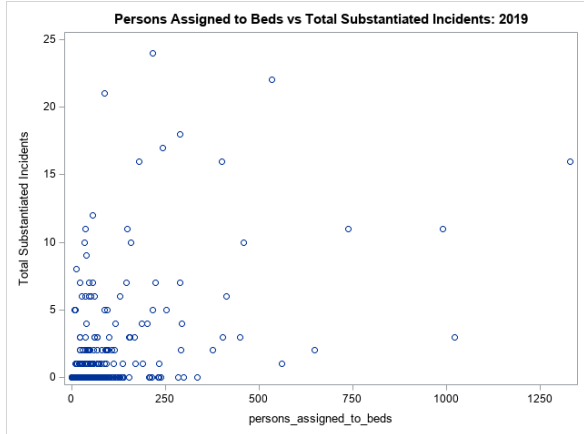


Figure 13. Figure 13a (left) plots the Pearson correlation between persons assigned to beds and total substantiated cases for juvenile jails from 2015-2019. Figure 13b (right) plots the 2019 juvenile jail data of total substantiated cases against persons assigned to beds. Source: 2015-2016 Survey of Sexual Victimization.

4.2) Relationship between Total Substantiated and Alternative MOS Variables

For the adult facilities, there are two possible alternative MOS variables: number of confined inmates and number of admitted inmates. These variables are also divided into males and females, but because these are highly correlated with number of confined inmates and number of admitted inmates, the correlation results are not included here. For juvenile facilities, possible alternative MOS variables include the number of held persons, the number of held males, and the number of held persons aged 18-20. The Pearson correlations are calculated between total substantiated cases and these alternative MOS variables in the public jail component and are plotted in Figure 14. For other components, see the plots here. To clarify, the correlation could not be calculated in 2016 and 2019 for the tribal jail component because all units have zero substantiated cases, and so these points are nonexistent in its scatterplot. In general, the correlation results are similar to those with the current MOS across the components. Since the alternative MOS variables do not necessarily have a better relationship with total substantiated and, unlike the current MOS, are not available for all units across all the frames, they are no longer considered.

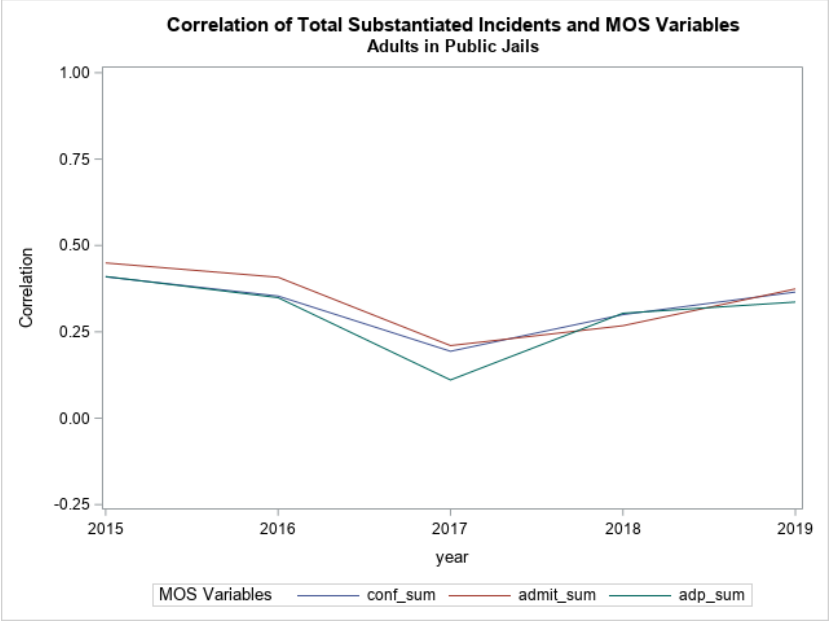


Figure 14. Correlation between total substantiated cases versus number of confined persons (blue line), number of admitted persons (red line), and the current MOS of ADP (green line) for public jails from 2015-2019. Source: 2015-2019 Survey of Sexual Victimization.

4.3) Highlights from Correlation Analysis

- The current MOS (ADP for adult facilities and persons assigned to beds for juvenile facilities) are weakly correlated with total substantiated cases.
- Other size variables are not any more related to total substantiated cases and are not necessarily available for all units across all frames.
- Results are less conclusive for private jails because the frame is too small and for tribal jails because the outcome is hardly observed.

5) Analysis on Certainty Units Selection

A unit's sample weight estimates how many other units it represents if sampled. For example, a unit with a weight of 3 indicates that there are two other units similar to it in the population. Sample weight is the inverse of the inclusion probability. In the SSV component samples, since a facility's inclusion probability is a function of its size, the sample weight is also related to size. However, some facilities are much larger than other facilities and should be distinguished as such. These units are given a sample weight of 1, signifying that they are self-representing and are included in the sample for certain. Hence, they are called certainty units.

There are two types of certainty units in the SSV component samples. Any facility with a MOS larger than a size threshold would be designated as a certainty. The threshold varies by component (see Section 1.2). Additionally, in the public jail and the juvenile components, the largest facility in each state is automatically a certainty even if it does not meet the size threshold. This ensures that the largest facility in each state is sampled. We refer to these as largest-in-state certainty (cert-largest) whereas the other certainty units are referred to as certainty-due-to-size (cert-size).

Determination of the size threshold is not clearly documented, but instead of setting a hard, possibly arbitrary, threshold, we investigate three data-driven methods to select the certainty units. If adopted, this has the advantage of being adaptable to each component's unique constitution and can be consistently administered and automated across the components. This will only affect the certainty-due-to-size units. The largest-in-state certainty units will still be identified and separated into their own take-all stratum.

5.1) Distribution of Current Certainty Units

Figure 15 shows the percentage of sampled units that are certainty units for each component sample. Less than a third of the sample for the public jail and tribal jail components are certainty units. However, the private jail and private prison components have a disproportional number of certainty units with around half of the sample being a certainty. By definition, a certainty unit is self-representing in terms of size, but looking at the scatterplot in Figure 16, the certainty units are not necessarily distinctively different from the noncertainty units in both size and outcome. Therefore, the number of certainty units could be reduced in the private jail and private prison component samples. Given that MOS is not highly related to the outcome, having this many certainty units may not be necessary. Because certainty units are always in sample, it can be burdensome to be designated as a certainty, especially for the smaller facilities in the private prison and private jail components.

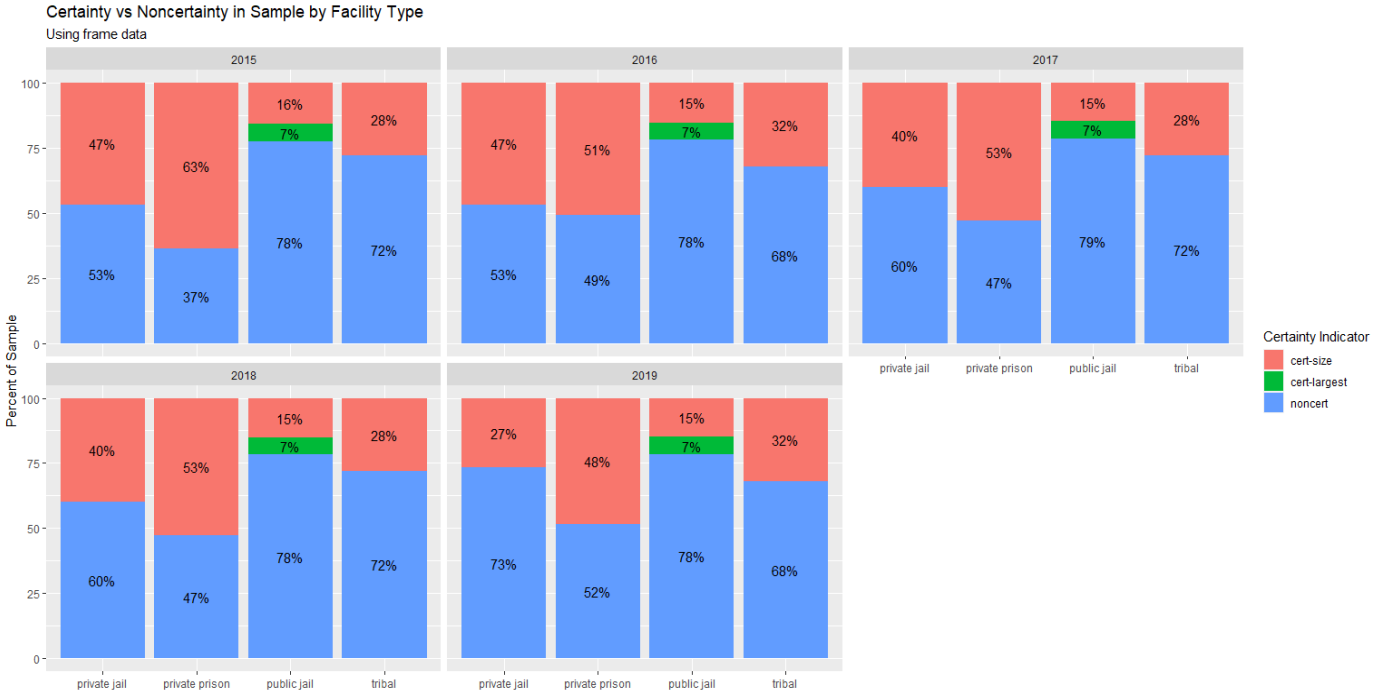


Figure 15. Distribution of certainty and noncertainty units in the sample by component. Source: 2015-2019 Survey of Sexual Victimization.

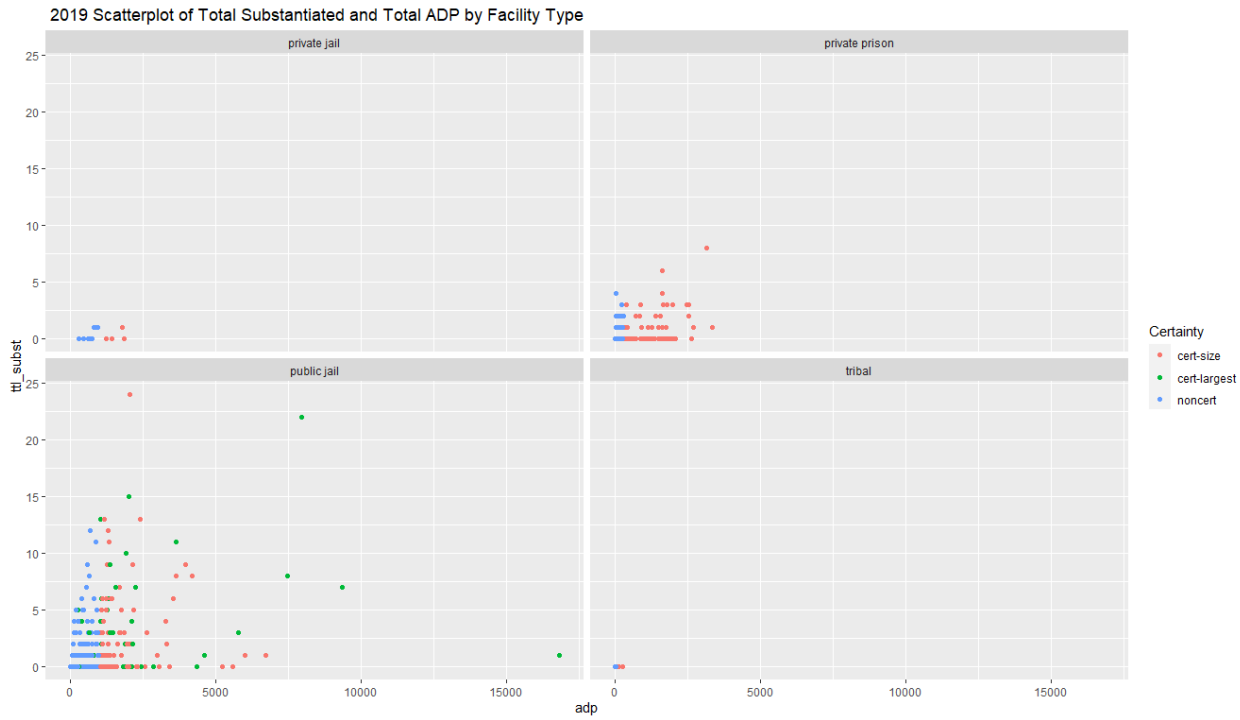


Figure 16. Survey Outcome versus MOS with certainty status. Source: 2019 Survey of Sexual Victimization.

5.2) Alternative Data-Driven Methods to Select Certainty-due-to-Size Units

We explore three data-driven methods to determine a certainty threshold. Any unit with a MOS greater than or equal to the threshold would be designated as a certainty unit. The three methods are k-means clustering, cum-root-f clustering (a heuristic method we derived), and Lavallee-Hidiroglou (LH) stratification. A fourth method is also included in which inclusion probability is calculated proportionally to the MOS and any unit with an inclusion probability greater than or equal to one is designated as a certainty unit. This is how inclusion probabilities are set in a πPS sample, and so we call this the PPS-cert method. It is less data-driven than the other methods and is only appropriate and should only be considered if a πPS sample is selected. To have access to the full MOS range, all units undergo this procedure (i.e., the largest-in-state units have not yet been removed into their own stratum).

The k-means clustering method (SAS/STAT PROC FASTCLUS Procedure) applies the popular machine learning method to form two clusters using the MOS. The total within-cluster variability, i.e., the total sum of squares (SSE), is calculated as

$$SSE = \sum_{k=1}^{2} \sum_{i=1}^{N_k} (x_i - \mu_k)^2$$

where k is indexing a cluster with N_k units, x_i is the MOS, and μ_k is the mean MOS within cluster k. The two clusters are formed so that the total within-cluster variability is minimized, and the between-clusters variability is maximized. In other words, the facilities within a cluster should have similar MOS values, and the facilities in different clusters should have different MOS values. The facilities that fall in the cluster containing the larger MOS values will be the certainty units. The minimum ADP for these units is the certainty threshold.

To approximate the k-means clustering, we develop the cum-root-f clustering method which uses Delanius and Hodges' cumulative root of the frequency (cum-root-f) to stratify the MOS variables into initial groups and then pool them to form two clusters (Delanius and Hodges, Jr., 1959). The cum-root-f requires specifying the number of initial groups, but since we do not know which will heuristically yield the two clusters such that the total within variability is minimized, we iteratively increase the number and calculate the total within variability. The algorithm goes through these steps:

1. Divides the range of MOS values into 100 bins of equal intervals ($b = 1, \dots, 100$)
2. Count the frequency in each bin (n_b)
3. Square root the frequency: $\sqrt{n_b}$
4. Calculate a running total of the square rooted frequency $\left(q_t = \sum_{b=1}^{t} \sqrt{n_b}\,, t = 1,2, \dots, 100\right)$
5. Set the number of initial groups K (start with $K = 2$)
6. Assign each bin to an initial group by computing: $k = \left\lceil K\left(\frac{q_t}{\sum_{t=1}^{100} q_t}\right)\right\rceil$
7. Identify the initial group with the largest value of MOS as the certainty cluster: $k_{\text{cert}} \coloneqq \arg\max_{k}\left(\max_{i \in k} x_i\right)$
8. All other groups are pooled into a noncertainty cluster: $k_{\text{noncert}} \coloneqq (\cup_{c=1}^{K} k) - k_{\text{cert}}$
9. Calculate the minimum MOS in the certainty cluster: $\tau_K = \min_{i \in k_{cert}} x_i$
10. Calculate the total within-cluster variability: $SSE_K = \sum_{c=1}^{K} \sum_{i=1}^{N_c} (x_i - \mu_c)^2$
11. Increase the number of initial groups by 1 ($K = K + 1$) and repeat steps 6—10
12. Stop once the minimum MOS and total within variability is calculated for $K = 25$
13. The minimum MOS that aligns with the smallest total within-cluster variability is selected as the certainty threshold: $\tau = \tau_{K^*}$ where $K^* = \arg\min_{K} SSE_K$

We can apply this method to all the component frames. Note that when applied to the public jail component, we removed outlying facilities whose MOS are clearly significantly larger than the rest (ADP > 15000).

The LH stratification method divides the MOS values with the objective of minimizing the sample size for a given coefficient of variation (Lavallee and Hidiroglou, 1988). This approach is specifically derived for skewed population in the sense that it produces a "take-all" stratum. The minimum MOS for this stratum is the certainty threshold.

The derived certainty thresholds are plotted against the current certainty threshold. The threshold from the k-means method does help reduce the number of certainty units, but it is only the public jail and private prison components are large enough (have at least 100 facilities) for the method to be applicable (SAS/STAT PROC FASTCLUS Procedure). The LH method gives the most reduction in the number of certainty units in the public jail and private prison components, but it may be too much. On the other hand, it increases the number of certainty units in the private jail and tribal jail components. We suspect this is happening because the MOS variable is less skewed for these components.

The cum-root-f clustering method that we set up gives us a good compromise. Its thresholds are similar to the k-means in the public jail and private prison components, so it is selecting less certainty units but not overly so. In the smaller components, it does not change the number of certainty units from the current approach. The PPS-cert method helps reduce the number of certainty units in the private prison

component without changing the number of certainty units in the other components, but again, this method is only applicable if a πPS sample is selected.

We also consider including a certainty stratum containing facilities with a high count of substantiated cases from previous years. We suspect that some facility may consistently have a high number of cases and so should be in the sample even though they may not necessarily be a large facility. Our preliminary analysis show that this is not the case. Results from this preliminary analysis can be found here.

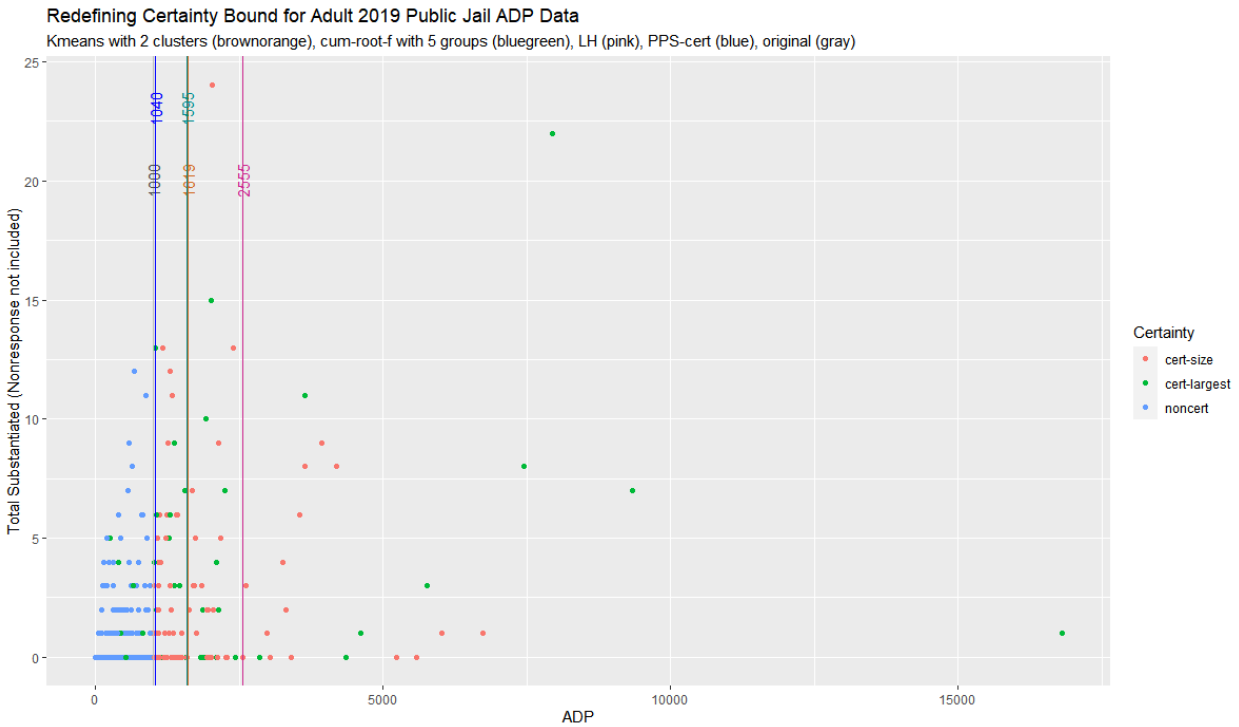


Figure 17. Size thresholds in 2019 public jail component. Source: 2019 Survey of Sexual Victimization.

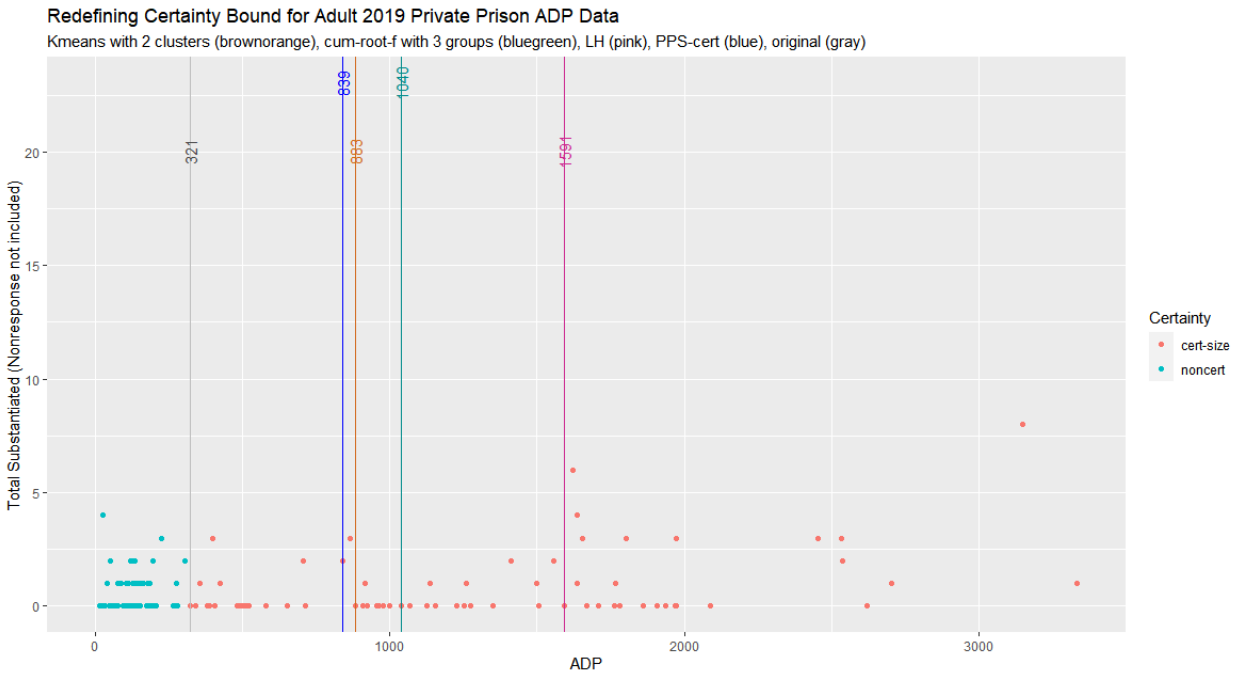


Figure 18. Size thresholds for 2019 private prison component. Source: 2019 Survey of Sexual Victimization.

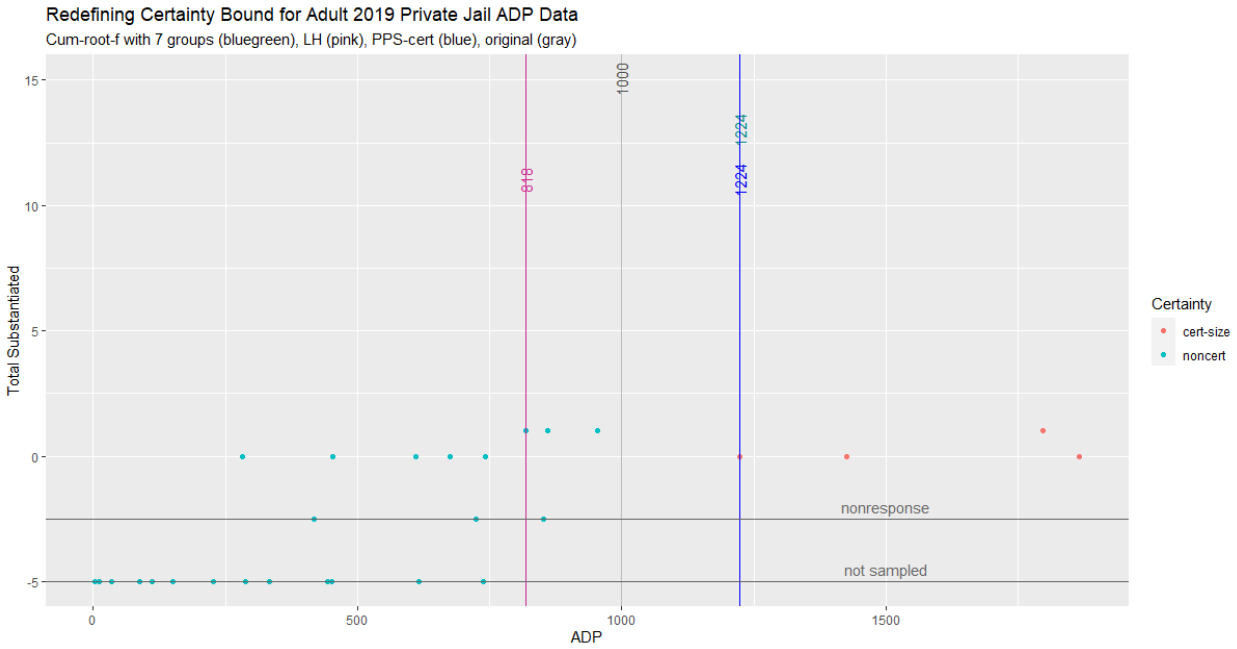


Figure 19. Size thresholds for 2019 private jail component. Source: 2019 Survey of Sexual Victimization.

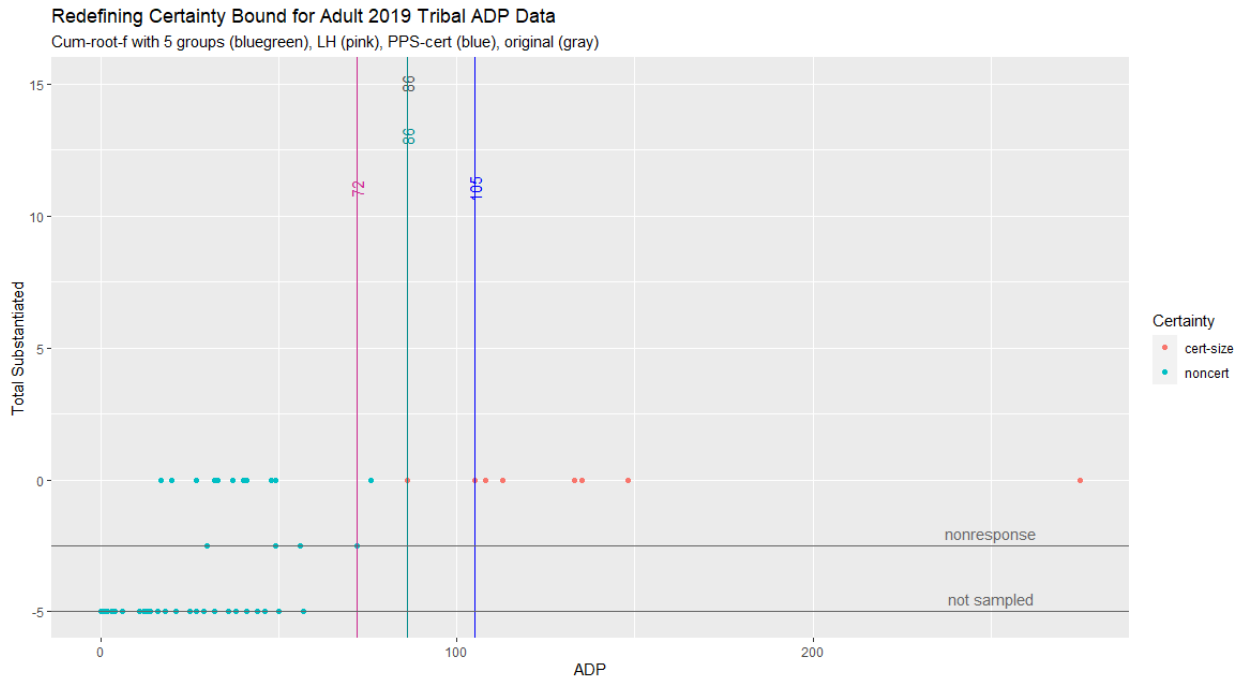


Figure 20. Size thresholds for 2019 tribal jail component. Source: 2019 Survey of Sexual Victimization.

After the certainty-due-to-size units are identified (and stratification of noncertainty units), the largest-in-state certainty units in the public jail component and the juvenile component may be separated into their own take-all stratum. Some of these may have already been designated as certainty units because their MOS is larger than the size threshold. In this case, these would be moved from the certainty-due-to-size stratum to the largest-in-state certainty stratum. This has no effect on a facility's certainty probability but serves as a guarantee that all states have been accounted for per the Prison Rape Elimination Act of 2003. Hence, we choose to identify the largest-in-state certainty units secondary because they are not truly self-representing with respect to size or the number of substantiated cases. On the contrary, removing them prior to determining the size threshold would confound the results. If we remove the largest-in-state certainty units first, the size threshold would likely be a lower value than what would have been determined if they are not removed, leading to more of the "smaller-medium" units being misclassified and held to the same reporting responsibility as the truly large facilities.

We apply the proposed cum-root-f clustering method to data from previous year and a summary of the results for the adult population is given in Table 10. This includes the size threshold found and a count of certainty units versus noncertainty units with the new threshold. Juvenile results are given in the **Appendix**.

Table 10. Size threshold for delineating certainty facilities in adult population under the proposed cum-root-f clustering method

	Year	Size Threshold	# of Certainty	# of Noncertainty
Public Jail	2019	1595	60	2822
	2018	1836	50	2791
	2017	1858	50	2824
	2016	1861	49	2834
	2015	1754	60	2843
Private Prison	2019	1040	42	364
	2018	1200	47	470
	2017	1200	47	470
	2016	1137	50	479
	2015	1029	50	412
Private Jail	2019	1224	4	25
	2018	581	19	18
	2017	580	18	18
	2016	540	19	19
	2015	621	13	24
Tribal Jail	2019	86	8	52
	2018	104	7	54
	2017	71	10	52
	2016	106	5	56
	2015	114	6	51

Source: 2015-2019 Survey of Sexual Victimization

6) Accounting for Size in Sampling (STRS vs. πPS)

Size can be accounted for in the sample design by stratification of facilities, allocation of sample size, and/or assignment of inclusion probability. A πPS random sample directly accounts for size by assigning a facility's inclusion probability as the product of the sample size and its proportion of total MOS. This results in unique inclusion probabilities for each MOS value. In practice, when the MOS is weakly correlated with the survey outcome, it may be hard to sample the smaller facilities, but if they are sampled, their large weight could contribute to a large sampling variance. Furthermore, adjustment may be needed to be sure that inclusion probabilities are strictly positive and at most one. For example, an arbitrary "small" value is typically added to units with zero MOS to get a positive inclusion probability, but this is more ad-hoc rather than analytically justified.

A STRS will create strata using the MOS and then allocate sample size to the strata (possibly with respect to size). Units within a stratum are sampled with equal probability. With this sampling method, inclusion probability is still a function of size, but the equal probability (within stratum) protects against arbitrarily assigning a factitious size to those facilities with a zero MOS and gives small units a more probable chance of being sampled.

Table 11 shows the percentages of facilities in the population with MOS less than the smallest sampled facility. Note that the samples from the components with a πPS design continually misses the smallest units from year to year whereas the public jail component, which has a stratified design, does not have this issue. As an illustration, we simulate 1000 samples from the private prison component using a stratified design (this follows the design we are proposing) and the current πPS design in Figure 21. The red distribution is the stratified, and the blue is the current design. The two distributions tend to overlap on the right tail, so the main difference between them occurs towards the lower values of MOS. Therefore, the stratified samples are getting a better representation of the smaller units but are still able to capture the large facilities just like the πPS samples, especially since the stratified design has a take-all stratum containing the largest units in the component. The same holds for the private jail and tribal jail components.

Table 11. Percentage of Population Units with MOS less than Smallest Sampled Unit

Facility Type	2015	2016	2017	2018	2019
Private jails	45.2	15.8	25	24.3	27.6
Private prisons	6.9	10.8	4.4	2.7	3.7
Public jails	0.3	0	0.5	0	0
Tribal	24.6	11.5	25.8	29.5	33.3

Source: 2015-2019 Survey of Sexual Victimization

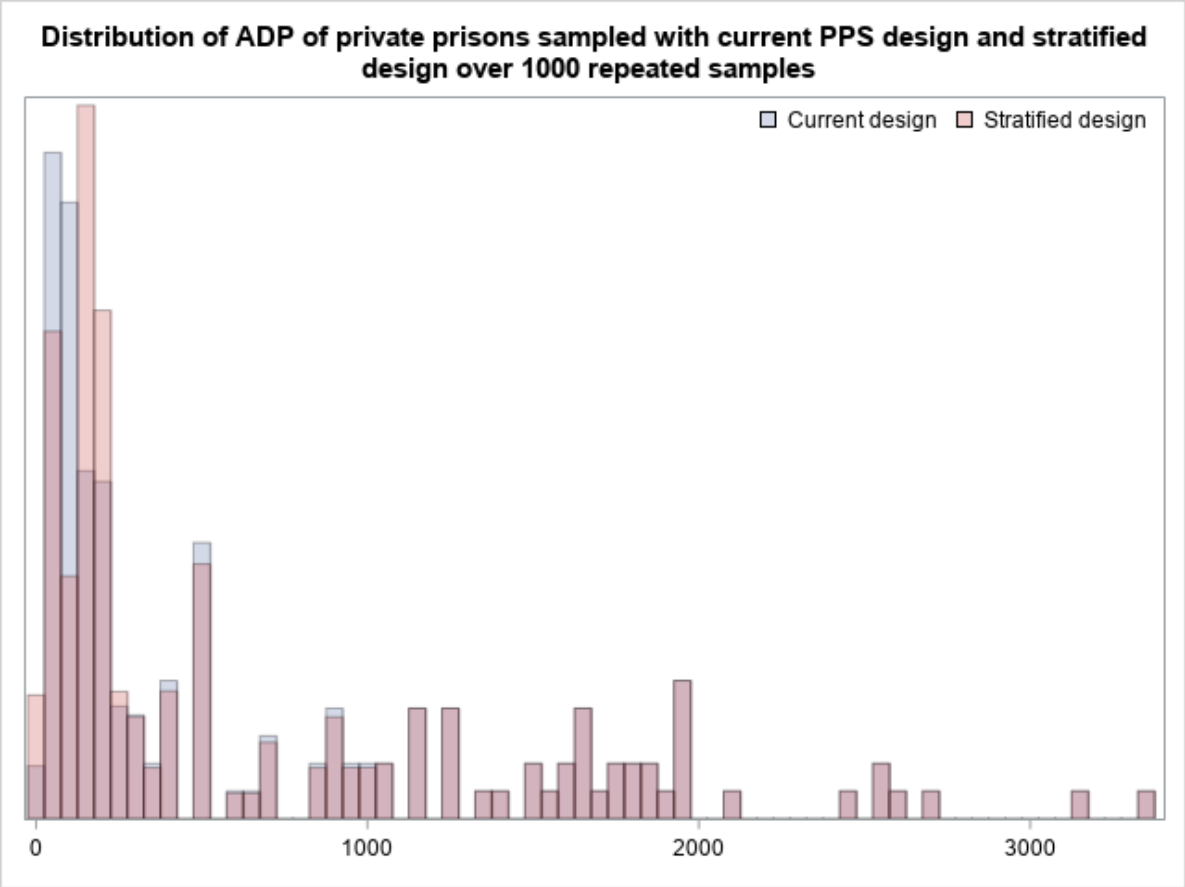


Figure 21. Distribution of ADP from 1000 repeated samples of the private prison component. Source: 2019 Survey of Sexual Victimization

Given this illustration and the weak correlation we found between the MOS and substantiated cases, the STRS seems to be a more suitable design than a πPS sample. Furthermore, in their simulation study, Thompson et. al. (2023) observes that, when the survey outcome is rare and informative covariates are unavailable, a πPS design by Pareto sampling produces more biased results compared to a STRS, which "provide[s] some robustness against misspecification of a statistical relationship between unit size and the rare characteristic."

7) Number of Strata

Unlike the adult components, the juvenile component has strata that are primarily based on facility type and geography (but only for the detention centers). It also has some take-all strata, three of which house certainty-due-to-size units. In our proposed sample design for this component, we are maintaining the same strata definitions but will use the cum-root-f clustering approach to determine the certainty-due-to-size units. The rest of this section focuses on determining the number of strata for use in the adult components under a stratified design.

Because the MOS is a continuous variable, it needs to be discretized into size strata. We will apply Delanius and Hodges' cum-root-f method (Delanius and Hodges, Jr., 1959) to create the strata, but we need to first decide on the number of strata. Stratification is most ideal when groups of similar units are

formed so that the variability is reduced within strata, leading to an overall variability reduction. Thus, the means between strata should be fairly different (Lohr 2010). The more strata there are, the more similar the units would be within a stratum, but the number of units within a stratum would also decrease. In choosing the number of strata, we need to have a balance in which the stratum means are most different but there should still be a sufficient number of units in the strata to be sampled. As a rule of thumb, there should be at least 15 units in a stratum, allowing for some nonresponse when data is collected.

After identifying the certainty-due-to-size units with the cum-root-f clustering approach (see Section 5.2), we remove them into a take-all stratum. The noncertainty population, including the largest-in-state certainty unit that are smaller than the size threshold, will then be stratified using the cum-root-f method with the number of strata varying from 1 (no stratification) to 5 strata. The largest-in-state certainty units will be placed in their own take-all stratum after the strata have been formed. This follows the same rationale as when determining the size threshold. They are not inherently unique, but removing these facilities beforehand would change the noncertainty population and distort the strata definition. Since identification of the largest-in-state certainty units only pertains to the public jail and juvenile components, removing them after defining the strata would also make the methodology for these components more consistent with the other components that do not have a take-all stratum of the largest-in-state certainty units.

Table 12. Count of facilities with size less than certainty size threshold (this includes the largest-in-state certainty units) in a stratum (top number), number of largest-in-state certainty (in parentheses), and stratum mean MOS (italicized bottom number) with different specification of number of strata for the 2019 public jail component.

2019 Public Jail	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5	Certainty Stratum
With 1 noncert stratum	2822 (25) *188.88*					59 (20) *2999.54*
With 2 noncert strata	2220 (2) *78.62*	602 (23) *595.51*				59 (20) *2999.54*
With 3 noncert strata	1779 (1) *49.76*	756 (4) *265.80*	287 (20) *848.67*			59 (20) *2999.54*
With 4 noncert strata	1476 (1) *36.05*	744 (1) *163.06*	409 (5) *413.86*	193 (18) *980.45*		59 (20) *2999.54*
With 5 noncert strata	1177 (1) *25.33*	823 (0) *115.52*	429 (2) *273.85*	245 (5) *547.84*	148 (17) *1057.05*	59 (20) *2999.54*

Source: 2019 Survey of Sexual Victimization

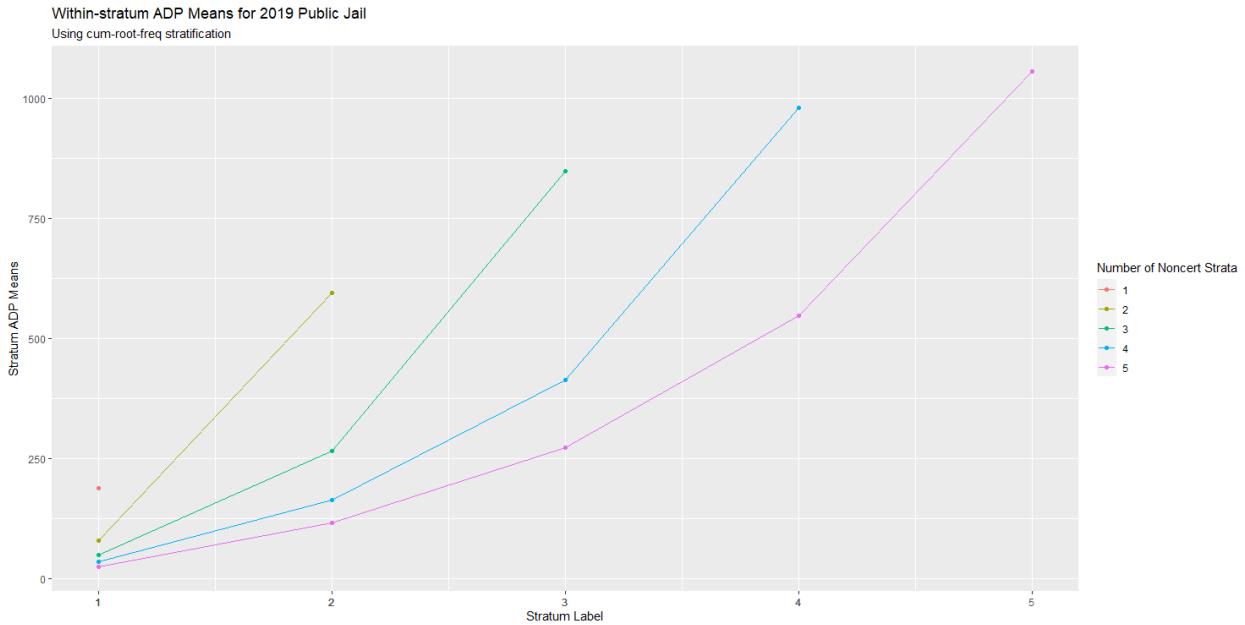


Figure 22. Comparison of stratum mean MOS under different specification of number of strata for 2019 public jail component. Source: 2019 Survey of Sexual Victimization.

Once the strata are formed, we calculate the stratum mean MOS and the stratum size. Table 11 provides this information for the public jail component. The top number is the number of facilities with size smaller than the certainty size threshold including the number of largest-in-state certainty units. A separate count of the largest-in-state certainty units is provided in parentheses. The bottom italicized number is the stratum mean MOS. For the public jail component, there are enough facilities that we can have five noncertainty strata without concern about the stratum size, but we also need to consider the difference in stratum means. This is best seen in a line plot as in Figure 22. Strata with the most different means will have steeper slopes between the points and a less curved line. Comparing the olive-colored line (where we have only two strata) to the green-colored line (where we have three strata), their shape is clearly different indicating that adding an extra stratum improves the classification. Looking just at the green-colored line, the slopes between the points are also of different steepness, so the stratum MOS means are widely dissimilar. On the contrary, the blue-colored line (where we have four strata) is similar in shape to the purple-colored line (where we have five strata), and thus, adding the extra stratum here minimally helps with the classification. With this logic, having three or four strata seems to be the most fitting. Arguably, there is a slight improvement with having four strata, but to be on the conservative side, we recommend three strata.

To evaluate this choice, we graph the MOS against the survey outcome for the respondents and color-code the strata in Figure 23. We see that the strata we defined using MOS are able to form groups of facilities with similar values of total substantiated cases too unlike with the current strata definition, which is in Figure 24. The stratum boundaries are less aligned with the outcome values.

We conduct the same analysis for the private prison, private jail, and tribal jail components. Their tables are listed in the **Appendix** and graphs are given here. For the private prison component, two or three noncertainty strata seems appropriate. With the private jail and tribal jail components being so small, they should only have one noncertainty stratum.

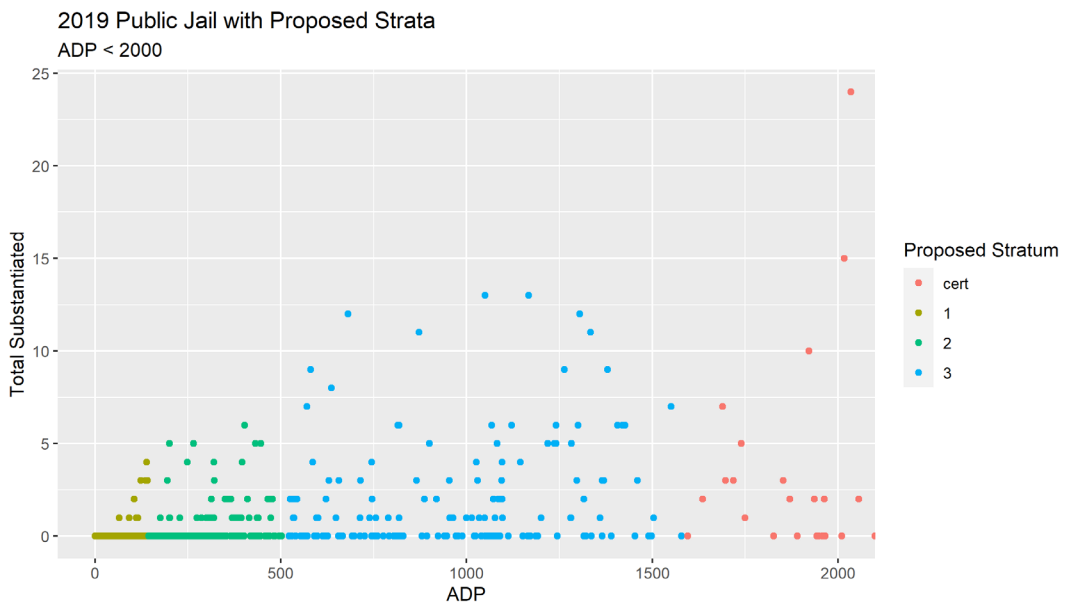


Figure 23. Total substantiated versus ADP with proposed strata definitions for the 2019 public jail component. Source: 2019 Survey of Sexual Victimization.

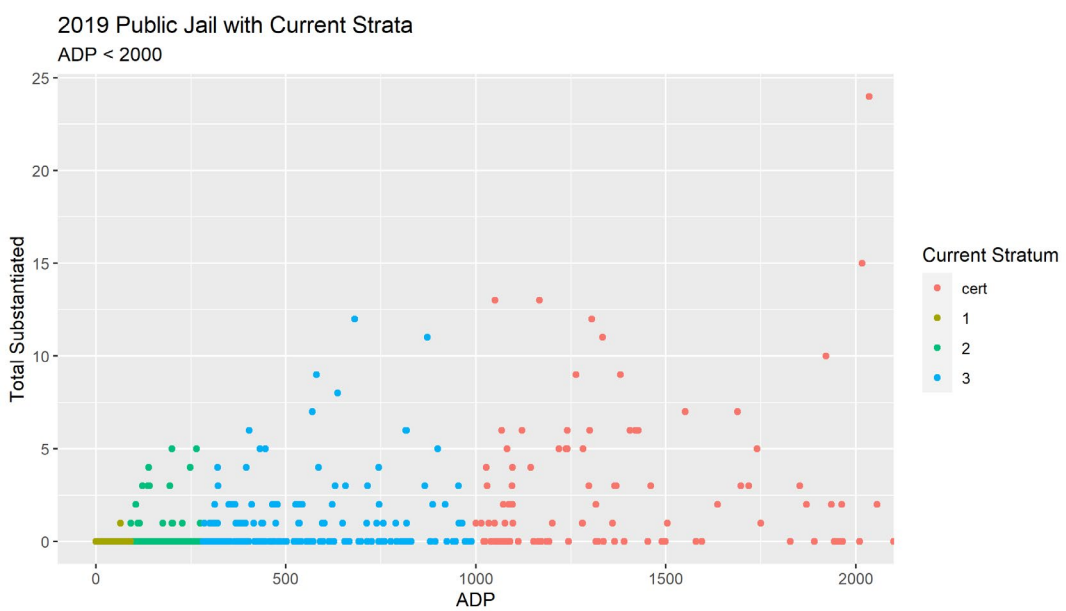


Figure 24. Total substantiated versus ADP with current strata definitions for the 2019 public jail component. Source: 2019 Survey of Sexual Victimization.

8) Sample Allocation Evaluation

With the strata and sample design selected, the next step is choosing how to allocate the sample to the strata such that facility size is accounted for without underrepresenting facilities of smaller MOS values. With this in mind, three allocation procedures are considered: optimal allocation, proportional allocation, and $\sqrt{MOS}$ - proportional allocation. Optimal allocation allocates sample to the strata by minimizing the variance of the variable of interest subject to the constraint that the strata sample sizes

sum to the sample size of the component (i.e., $n = \sum_{h=1}^{H} n_h$, where h is the stratum, and H is the number of strata). The formula used for optimal allocation (Neyman allocation, Neyman 1934) is given by $n_h = n\left(\frac{N_h S_h}{\sum_{j=1}^{H} N_j S_j}\right)$, where S_h is an estimate of variability for stratum *h.* Proportional allocation allocates the sample proportionally to the strata either based on the sample size or some auxiliary information. In our case, we use MOS to proportionally allocate the sample, so $n_h = n\left(\frac{\sum_{i=1}^{N_h} x_i}{\sum_{j=1}^{H}\sum_{i=1}^{N_j} x_i}\right)$. This will give the strata that have the most proportion of the total MOS more allocation, which can result in small allocation for the strata contributing less to the total MOS. This has an effect akin to assigning unique inclusion probabilities in a πPS sample design. The $\sqrt{MOS}$ - proportional allocation (a power allocation procedure as proposed in Bankier 1988) helps mitigate this situation by compressing the MOS range using the power function. This allocation procedure uses the following formula to allocate the sample to each stratum h: $n_h = n\left(\frac{\sqrt{\sum_{i=1}^{N_h} x_i}}{\sum_{j=1}^{H}\sqrt{\sum_{i=1}^{N_j} x_i}}\right)$.

8.1) Simulation

A simulation study is conducted to compare the three considered allocation methods but only for the public jail and juvenile components. These are the only components with sufficient auxiliary variables beside the MOS variable. For the public jail component, the variables Admitted (Total, Female, and Male) and Confined (Total, Female, and Male) are used for our analysis; for the juvenile component, the variables offenders, nonoffenders, and persons assigned to beds.

For each allocation method, we selected 5,000 stratified sequential samples and obtained empirical estimates for bias, variance, and mean squared error (MSE). We calculated relative bias as $\left[\frac{1}{5{,}000}\sum_s(\hat{Z}_{l,s}^m - Z_l)\right] \Big/ Z_l$ where Z_l is the true total of auxiliary variable l and $\hat{Z}_{l,s}^m$ is the total estimate of auxiliary variable l for sample s using the allocation method m. We calculated variance as $\frac{1}{5{,}000}\sum_s(\hat{Z}_{l,s}^m - \hat{\bar{Z}}_l^m)^2$, where $\hat{\bar{Z}}_l^m = \frac{\sum_{s=1}^{5{,}000}\hat{Z}_{l,s}^m}{5{,}000}$ (an average of total estimates). We calculated the MSE as $\frac{1}{5{,}000}\sum_s(\hat{Z}_{l,s}^m - \hat{\bar{Z}}_l^m)^2 + (\frac{1}{5{,}000}\sum_s \hat{Z}_{l,s}^m - Z_l)^2$. We compared these statistics among the three considered allocation methods using line plots. The method with the smallest absolute relative bias, variance, and MSE would be considered the best performing method.

8.2) Adult Results

For brevity's sake, we present the relative bias and variance results for only the variables admitted adults, admitted females, and admitted males. The confined adults, confined females, and confined males results are very similar. The MSE plots are not given because the bias is so small that the MSE results look nearly identical to the variance results.

When considering the relative bias results, we want an allocation method that on average has a relative bias value that is closest to zero. Looking at the graphs in Figure 25, no method seems to be "closer" to zero than any other. As a matter of fact, all the relative biases are all very close to zero.

Public Jails: Relative Bias

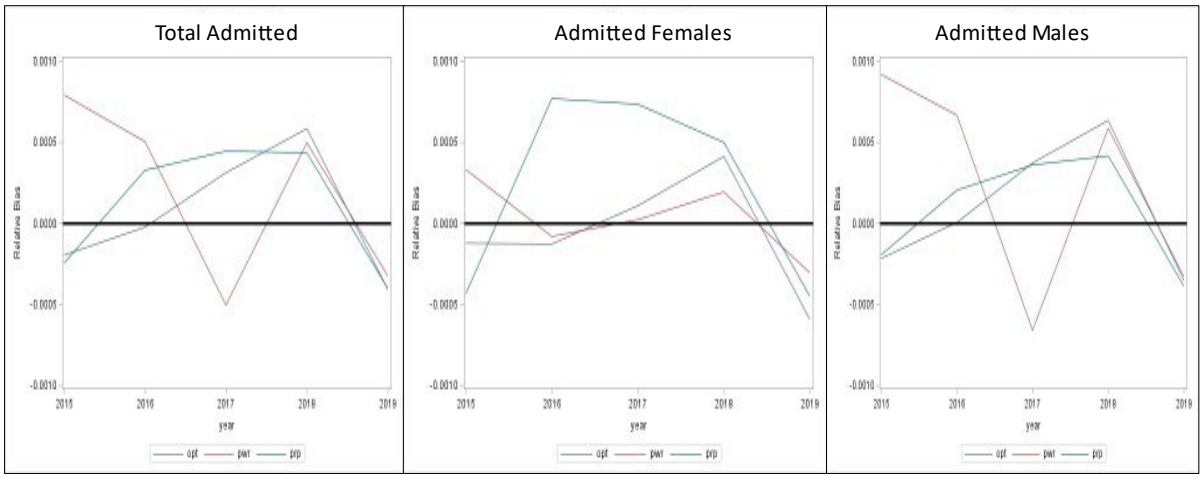


Figure 25. Relative Bias Results for Public Jails Total Admitted, Admitted Females, and Admitted Males. The dark blue line represents for optimal allocation (opt), the red line represents $\sqrt{MOS}$ – proportional allocation (pwr), and the green/teal line represents proportional allocation (prp). Source: 2015-2019 Survey of Sexual Victimization.

When considering the variance estimates presented in Figure 26, we are looking for the allocation method that results in the smallest variances. Here there is a clear pattern, and the optimal allocation method always results in the smallest variance estimates. Since all have zero relative bias but the optimal allocation has the smallest variance, we recommend it as the strata allocation method for the adult components of the SSV.

Public Jails: Variance

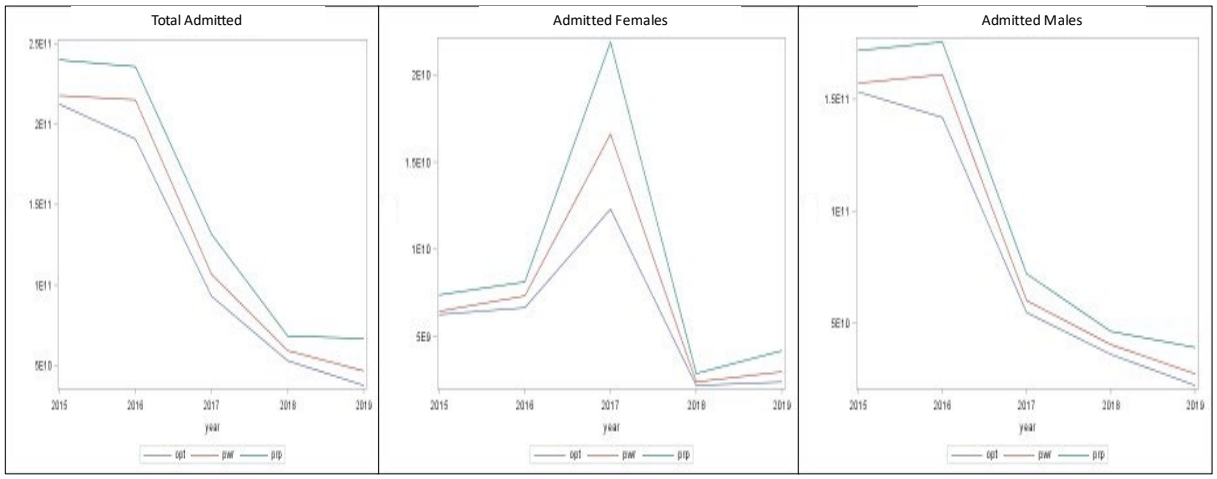


Figure 26. Variance Results for Public Jails Total Admitted, Admitted Females, and Admitted Males. The dark blue line represents for optimal allocation (opt), the red line represents $\sqrt{MOS}$ – proportional allocation (pwr), and the green/teal line represents proportional allocation (prp). Source: 2015-2019 Survey of Sexual Victimization.

8.3) Juvenile Results

When looking at the juvenile results, it should be noted that we only had the MOS (persons assigned to beds) and breakdowns of the measure of size (offenders and nonoffenders) for evaluation. Ideally, we would not be conducting our analysis on the very variable we used to allocate the sample. However, these were the only variables available. The relative bias results for the juvenile component are presented below in Figure 27. Here again we are looking for relative biases that are nearly zero. Much like the adult component of SSV we find that all three allocation methods resulted in relative bias values very close to zero.

Juvenile
Relative Bias

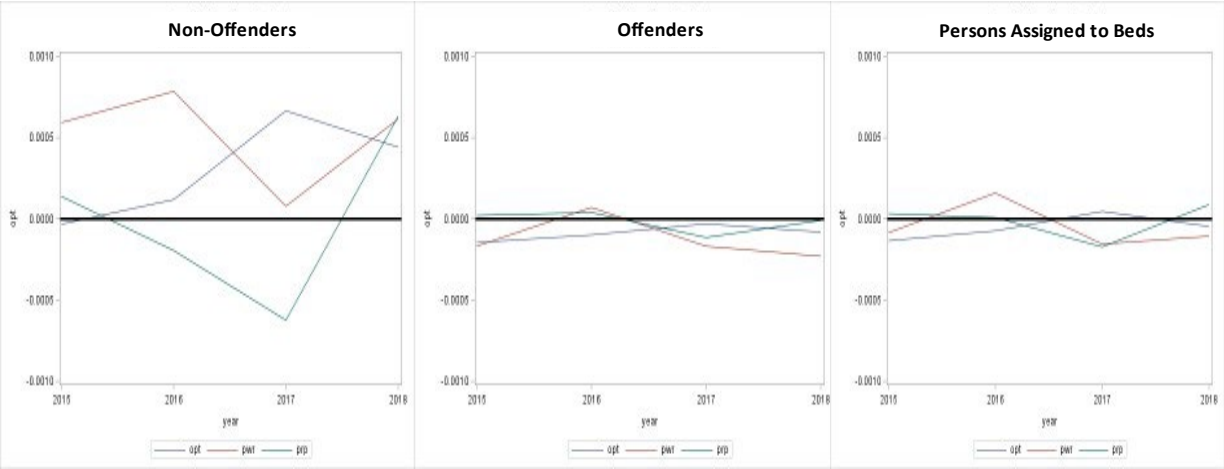


Figure 27. Juvenile Relative Bias Results for Non-Offenders, Offenders, and Persons Assigned to Beds. The dark blue line represents for optimal allocation (opt), the red line represents $\sqrt{MOS}$ – proportional allocation (pwr), and the green/teal line represents proportional allocation (prp). Source: 2015-2018 Survey of Sexual Victimization.

However, when we look at the variances presented in Figure 28, we do see a pattern emerge. First, the $\sqrt{MOS}$ - proportional allocation method seems to consistently result in the largest variances. On the other hand, it appears that both proportional and optimal allocation are very close to each as the allocation methods that result in the smallest variances. From these results, it would seem that either proportional or optimal allocation methods would work well for the juvenile component of SSV. However, one of our goals in this research is to find ways to streamline the sampling methodology so that the approaches are as consistent as possible for the different populations that make up SSV. Consequently, we recommend optimal allocation to align with the recommendations for the adult components of SSV.

Juvenile Variance

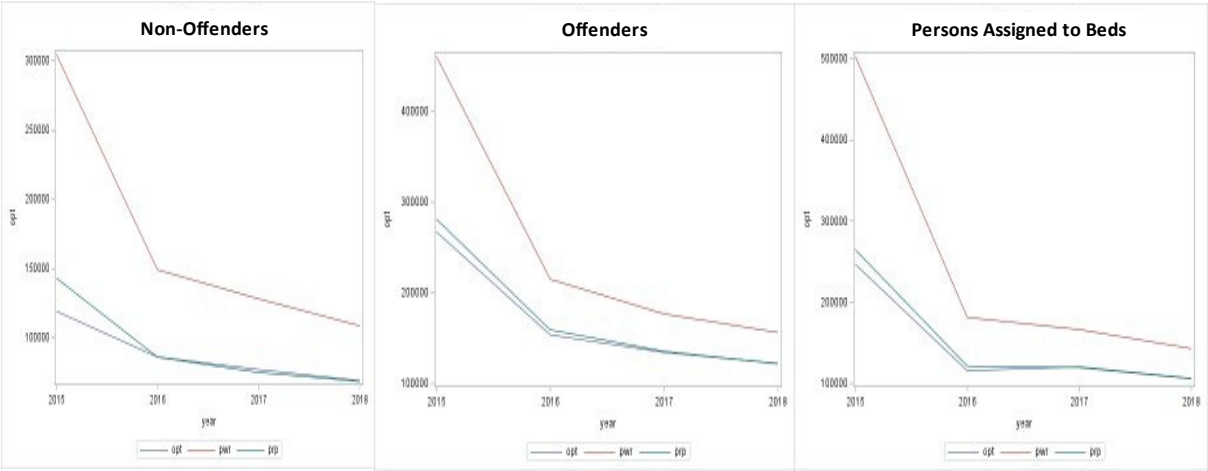


Figure 28. Juvenile Variance Results for Non-Offenders, Offenders, and Persons Assigned to Beds. The dark blue line represents for optimal allocation (opt), the red line represents $\sqrt{MOS}$ – proportional allocation (pwr), and the green/teal line represents proportional allocation (prp). Source: 2015-2018 Survey of Sexual Victimization.

9) Recommendations

1. Request that the state and federal jails and prisons report the disaggregated data.

   State jails report one value for total substantiated cases, aggregating all the states' data. This is the same for state prisons and for facilities at the federal level. Presumably, then, the micro data for individual facility existed and could be reported. Though, a census of these units is taken, their micro data could provide additional insights into the behavior of the outcome and how it may relate to other variables. This could help improve the sample design and estimation even more. Additionally, a thorough nonresponse analysis could also be conducted to assess nonsampling error.

2. Make the largest unit in each region a certainty to ensure that all regions are represented private jail, tribal jail, and juvenile samples.

   Although geography is controlled for in the sample design, we find that not all regions are represented in the smaller components, sometimes consecutively for years. Making the largest unit in each region a certainty will solve this and aligns with the methodology to make the largest facility in a state a certainty. However, this could be burdensome on the units, so an alternative would be to create a stopping rule for sampling that checks if a sample contains at least one unit from each region.

3. Have a consistent and automated sample design across all the component frames.

The components may be typologically different, but in terms of the outcome and MOS, they tend to behave similarly, so they do not need to have different designs. Having a general design among them means that it can be automated, streamlining the programming and maintenance of the sample.

4. Use a data-driven approach to define the certainty units.

   The current size thresholds for demarcating the certainty units tend to produce an overabundance of certainty units in the sample, especially in the smaller components. We find that many of these units are not necessarily unique with respect to both the outcome and the MOS. Hence, we develop an algorithm using the cum-root-f method to form clusters of exchangeable units to set the size threshold. The methodology is transparent and data driven.

5. Consider taking a census of the private jail component.

   The private jail component has a small population of only 30 jails or less. With the outcome being such a rare occurrence, taking a census of the private jails may give a more accurate count. Burden may be an issue, but having to always report may help them build a routine and ease this burden (Willimack and Nichols, 2010).

6. Convert the sample designs of the private prison, private jail (if not a census), tribal jail and juvenile frames into stratified sequential random sampling from probability proportional to size sampling. Public jail is already a stratified sequential random sample design.

Proposed Sample Designs							
	Facility Type	**# of requested units**	**Types of certainty units**	**How certainty units are defined**	**Stratified sampling?**	**# of strata***	**How strata are defined**
Adult	Public jails	700	Certainty by size Largest in State	Cum root f clustering method Facility with largest MOS in the state	Stratified	4	Cum root f
Adult	Private prisons	155	Certainty by size	Cum root f clustering method	Stratified	3	Cum root f
Adult	Tribal jails	25	Certainty by size	Cum root f clustering method	Stratified	2	Cum root f
Adult	Private jails	29	n/a	n/a	Census	n/a	n/a

Juvenile	Local and private facilities	922	Certainty by size	Cum root f clustering method	Stratified	2	Cum root f
Juvenile	Tribal facilities	21	n/a	n/a	Census	n/a	n/a

* Counts include one stratum containing all certainty units in each component.

The efficiency gains from a πPS sample design hinge on having a MOS that is perfectly related to the outcome. However, our research finds that there is a small to nonexistent relationship between the MOS and the outcome. Thus, the advantages stemming from such a design are minimal. On the other hand, a πPS sample design yields some real complications that need to be resolved such as handling zero MOS and inclusion probability greater than one and potentially having a large sampling variability from varying weights. Combined with the fact that the outcome does not have a highly skewed distribution, this prompts us to propose using a stratified sequential sample design across the board for all components and thus streamlining the processes. This design still assumes a relationship between MOS and the outcome but also safeguards against it if it is not true.

7. Estimate variances with the standard stratified sampling variance formula. If additional auxiliary variables become available, re-investigate variance estimator.

   Without sufficient auxiliary data, we did not have the means to properly propose and evaluate other variance estimators. Further research into variance estimation is recommended.

8. Create one overall frame of public jails, private prisons, private jails, and tribal jails and analytically determine sample size allocation to each component.

   The choice to have independent designs for the components makes it hard to optimize the survey as a whole. For example, both the public jail component and the private jail component contribute to the jail estimate and its variance estimate. It is difficult to optimally minimize this variance estimate because the two components are sampled differently. Optimally designing each component does not guarantee an optimal estimate when they are pooled. However, if the adult facilities are grouped into one frame with the components acting as strata, then the sample size for each component could be calculated such that the sample variance is minimized. It may also help keep the data more consistent across the components.

9. Consolidate all SSV documentation into one with a table summarizing the sampling procedure for each sampling frame.

   Going along with Recommendation 8, because the components collectively contribute to estimates, it is important to consolidate their documentation, making it easier for data users to understand, especially if the component sample designs are kept independent.

References

Bankier, M. (1988). Power Allocations: Determining Sample Sizes for Subnational Areas. The American Statistician, 42(3), 174-177.

Dalenius, T. and Hodges Jr, J. L. (1959), "Minimum Variance Stratification," *Journal of the American Statistical Association*, 54, pp. 88-101.

Lavallee, P. and Hidiroglou, M. A. (1988), "On the Stratification of Skewed Population," *Survey Methodology*, 14 (1), pp. 33-43.

Lohr, S. (2010) *Sampling: Design and Analysis.* Boston: Brooks and Cole.

Neyman, J. (1934). "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection", Journal of the Royal Statistical Society, Vol. 97, 558-606.

SAS/STAT® 13.2 User's Guide. "The FASTCLUS Procedure," retrieved on 3/24/2022 from The FASTCLUS Procedure (sas.com).

Thompson, K. J. and et. al. (2023), "Using Partially Synthetic Frames to Evaluate Alternative Sample Designs for Estimating a Rare Business Characteristic," Working Manuscript.

Willimack, D. K. and Nichols, E. (2010), "A Hybrid Response Process Model for Business Surveys," *Journal of Official Statistics*, 26(1), pp. 3-24.

Appendix

Table 13 provides the size threshold when the cum-root-f clustering method is applied to the juvenile population and gives the count of certainty and noncertainty facilities.

Table 13. Size threshold for delineating certainty facilities in juvenile population.

Year	Stratum	Size threshold	# of Certainty	# of Noncertainty
2018	Detention	62	28	429
	Local	94	16	185
	Private	88	34	652
2017	Detention	69	22	439
	Local	62	44	153
	Private	165	8	735
2016	Detention	65	32	432
	Local	72	26	171
	Private	161	10	788
2015	Detention	86	24	438
	Local	137	7	190
	Private	209	6	786

Source: 2015-2018 Survey of Sexual Victimization.

To determine how many strata may be best if the sample design of each component is converted to a STRS, we compare the change in stratum ADP means as the number of strata is increased. In the following tables, the top number is the count of facilities, and the second italicized number is the mean ADP within the stratum. The last column gives the same statistics for the certainty stratum[6], which is found using the cum-root-f clustering algorithm that we proposed in Section 5.2.

Table 14. Count of facilities with size less than certainty size threshold (this includes the largest-in-state certainty units) in a stratum (top number), number of largest-in-state certainty (in parentheses), and stratum mean MOS (italicized bottom number) with different specification of number of strata for the 2019 private component.

2019 Private Prison	**Stratum 1**	**Stratum 2**	**Stratum 3**	**Stratum 4**	**Stratum 5**	**Certainty Stratum**
With 1 noncertainty stratum	364 *125.78*					42 *1772.83*
With 2 noncertainty strata	274 *55.15*	90 *340.82*				42 *1772.83*

[6] These components do not have largest-in-state certainty units, so this contains only the certainty-due-to-size units.

With 3 noncertainty strata	201 *40.44*	118 *119.36*	45 *523.78*			42 *1772.83*
With 4 noncertainty strata	145 *32.00*	129 *81.16*	58 *179.79*	32 *632.69*		42 *1772.83*
With 5 noncertainty strata	115 *27.80*	119 *63.87*	69 *121.88*	35 *245.63*	26 *691.50*	42 *1772.83*

Source: 2019 Survey of Sexual Victimization.

Table 15. Count of facilities with size less than certainty size threshold (this includes the largest-in-state certainty units) in a stratum (top number), number of largest-in-state certainty (in parentheses), and stratum mean MOS (italicized bottom number) with different specification of number of strata for the 2019 private jail component.

2019 Private Jail	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5	Certainty Stratum
With 1 noncert stratum	25 *440.04*					4 *1578.00*
With 2 noncert strata	12 *172.00*	13 *687.46*				4 *1578.00*
With 3 noncertainty strata	8 *92.50*	8 *410.50*	9 *775.22*			4 *1578.00*
With 4 noncertainty strata	6 *60.50*	6 *283.50*	6 *541.83*	7 *812.29*		4 *1578.00*
With 5 noncertainty strata	4 *34.75*	6 *195.50*	5 *420.40*	4 *656.25*	6 *827.00*	4 *1578.00*

Source: 2019 Survey of Sexual Victimization.

Table 16. Count of facilities with size less than certainty size threshold (this includes the largest-in-state certainty units) in a stratum (top number), number of largest-in-state certainty (in parentheses), and stratum mean MOS (italicized bottom number) with different specification of number of strata for the 2019 tribal jail component.

2019 Tribal Jail	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5	Certainty Stratum
With 1 noncertainty stratum	52 *26.31*					8 *137.88*
With 2 noncertainty strata	26 *9.81*	26 *42.81*				8 *137.88*
With 3 noncertainty strata	18 *5.78*	17 *24.94*	17 *49.41*			8 *137.88*

With 4 noncertainty strata	14 *3.71*	12 *16.92*	12 *32.33*	14 *51.79*		8 *137.88*
With 5 noncertainty strata	11 *2.18*	10 *12.90*	10 *24.20*	10 *37.40*	11 *54.45*	8 *137.88*

Source: 2019 Survey of Sexual Victimization.