Westat | Improving Lives Through Research®

Statistical Assessment of the NEISS and NEISS-AIP Samples:

Final Technical Report

Authors

David Marker
Jim Green
Frost Hubbard
Richard Valliant

September 24, 2020

Prepared for:
Consumer Product Safety Commission
4330 East West Highway
Bethesda, MD 20814

Prepared by:
Westat
An Employee-Owned Research Corporation®
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

Table of Contents

Table of Contents (continued)

1. Introduction

The United States Consumer Product Safety Commission (CPSC) requires an independent statistical assessment of the validity of the current National Electronic Injury Surveillance System (NEISS) and NEISS-All Injury Program (NEISS-AIP) samples of hospital emergency departments. This includes analyzing the pros and cons of keeping the current samples, expanding the sample, or resampling hospital emergency departments. As part of the assessment the CPSC also requests guidance on alternative methods for adjusting for replacement hospitals that can cause significant changes to annual estimates; alternative methods for estimating injuries for incidents with high variability or geographical concentration; and information on suppression criteria for determining which estimates are not of sufficient accuracy to be published.

This report summarizes Westat's research into updating the NEISS samples. Section 2 describes the current sample, selected in 1996, examines how the sampled hospitals have changed over time, and compares them with the current universe of hospital emergency departments. Section 3 summarizes our research into optimal allocations for a sample considering the current universe, cost constraints, and the desire to maintain a time series of injuries. Section 4 describes a series of sample maintenance recommendations, including a bridge sample overlapping old and new samples, replacing hospitals, improved methods for adjusting for monthly nonresponse, weighting and variance estimation, and geographically-concentrated estimates. Section 5 summarizes suppression criteria currently used by a wide range of government agencies, across the United States and internationally.

Updating the sample will better reflect the current hospital population, which is quite different than it was 25 years ago. The major recommendations in this report are:

- **NEISS sample** allocation Using the same strata as in the current design, we recommend in Section 3.4 a budget constrained allocation of 100 hospitals whose total cost is $2 million per year and which minimizes the average of the relvariances of the estimated totals over a set of six key statistics.

- **NEISS-AIP sample allocation**. Based on analyses of four injury types of special interest in NEISS-AIP, we recommend an allocation of either 59 hospitals that respects an annual budget constraint of $2.5 million; or, if budget permits, the CPSC can use the NEISS allocation with a subsample of four of the eight Children's hospitals in the

NEISS allocation, giving a total number of sample hospitals of 96 with an approximate cost of $3.5 million per year. This larger sample will produce much more accurate estimates of the number of injuries. See section 3.5.

- **Transition to new sample.** To facilitate the transition to a updated sample, the new sample should be selected to maximize the overlap with the current sample. We estimate that about 74 percent of the current sample will be retained. Ideally a one year bridge will be used when the new sample is phased in. See section 4.1.

- **Substitution.** We recommend an improved substitution methodology that will minimize breaks in time series when hospitals must be replaced (Section 4.2).

- **Nonresponse adjustments.** We have proposed improved methodology for adjusting for monthly nonresponding hospitals that reflects the recent history of the hospital, even if it has changed size over time. We also recommend that annual, rather than monthly nonresponse adjustments be used (Sections 4.3, Appendix C.3).

- **Weighting and variance estimation.** Improved weighting and variance methodology will better track the overall population of hospital emergency rooms as they continue to change over time. To reflect the complexity of the NEISS estimators of descriptive statistics, we recommend that a replication variance estimator be used. By providing public use files with replicate weights, the CPSC will give analysts the ability to compute correct standard error estimates using readily available software for complex survey analysis (Sections 4.4, Appendix C).

2. Current Sample

2.1 The 1996 Sample

The current NEISS and NEISS-AIP samples were selected by Westat in 1996 from the SMG marketing group's 1995 data file(s). The target population included all hospitals with six or more beds with an emergency department, excluding psychiatric and penal institutions. The population and sample were divided into five strata – four strata determined by size (total emergency room visits (ERVs)) and a fifth stratum for children's hospitals. A total of 102 hospitals were selected across the five strata, of which 76 were selected in the previous NEISS and were retained through overlap (Keyfitz, 1951) control methods.

The hospital population and sample distribution by stratum, along with the year 1994 ERVs, by sampling stratum are provided in Table 2-1.

Table 2-1 Hospital population and sample distribution by stratum

Stratum	1994 ERVs – range	Hospitals – 1995 population	Hospitals – 1995 sample	1994 ERVs from 1995 frame
Small	1 – 16,830	3,179	48	23,255,912
Medium	16,831 – 28,150	1,059	14	23,238,999
Large	28,151 – 41,130	674	9	22,977,640
Very Large	41,131+	426	23	24,880,758
Children's	Various	50	8	1,994,190
Total		5,388	102	96,347,499

The NEISS sample design was a stratified, single-stage sample of hospitals. Within each stratum, the hospitals were sorted geographically before applying and selecting a systematic sample (after identification and selection of hospitals with conditional probabilities of one due to the Keyfitz approach.) Hospitals were sorted by Census division, state, metropolitan and non-metropolitan status within state, and ZIP code.

2.2 Stratum Migration

As noted in Schroeder (2019), the eligible universe decreased from 5,388 hospitals in 1995 to 4,809 in 2017. Although the size of the universe has been fairly stable since 2015, this is partly due to new ineligibles being balanced by new eligibles. For example, 24 hospitals on the 2016 frame were not eligible in 2017; 26 hospitals were newly eligible in 2017.

Table 2-1 in Schroeder (2019) also shows that there has been a noticeable amount of "stratum jumping" between 1995 and 2017. For example, there were 3,179 hospitals in the Small stratum in 1995. Of those, 197 (6.2%) had increased in size and jumped into one of the other three size strata by 2017.

Tables 2-2 and 2-3 below show the current cross-classification of 1995 sampling stratum and the appropriate 2018 stratum to which a hospital would be assigned based on current ERVs, for the NEISS and NEISS-AIP samples respectively. Twenty-six of the 88 responding NEISS hospitals have changed strata, whereas 17 of the 59 responding NEISS-AIP hospitals have changed strata. Those that have not changed stratum are shown in red.

Table 2-2 2018 NEISS responding hospitals: 1995 stratum design by 2018 stratum design

	2018 Strata Definitions					
1995 Design Strata Definitions	Small	Medium	Large	Very Large	Children's	Total
Small	38	0	0	1	0	39
Medium	3	2	6	0	0	11
Large	1	0	5	3	0	9
Very Large	1	5	6	9	0	21
Children's	0	0	0	0	8	8
Total	43	7	17	13	8	88

Table 2-3 2018 NEISS-AIP responding hospitals: 1995 stratum design by 2018 stratum design

	2018 Strata Definitions					
1995 Design Strata Definitions	Small	Medium	Large	Very Large	Children's	Total
Small	27	0	0	0	0	27
Medium	2	1	4	0	0	7
Large	1	0	3	2	0	6
Very Large	0	5	3	6	0	14
Children's	0	0	0	0	5	5
Total	30	6	10	8	5	59

2.3 Consequences of the 1996 Sample Being Out-of-Date

Since the NEISS sample was selected 24 years ago, its efficiency and ability to represent the current universe of eligible hospital has gradually degraded. This has a number of consequences:

- Eligible hospitals have changed size strata; and
- Some hospitals have become nonrespondents; replacement hospitals are selected based on the nonresponding hospital's original stratum.

This leads to inefficiencies in the weighting and discontinuities in estimates. Redesigning the sample will increase precision of estimates as illustrated in sections 3 and 4. In addition, designing and selecting a new sample, albeit with overlap control, also brings the additional benefit of addressing stratum migration.

3. Optimal Sample

3.1 Constraints and Objectives

Sample design optimization balances cost and precision, and as such, objectives and constraints based on a realistic cost model, precision targets, and desired sample size are required. We have generated three multivariate NEISS sample design optimization options whose objectives are to maximize the precision for six key estimates constrained to cost, overall sample size, or both. As described further in section 3.3.2, the objective function was the average of the relative variances for six key estimates, rather than focusing on a single statistic as was done for the 1995 NEISS redesign.

In terms of costs, NEISS pays roughly $4.5 million to hospitals for 750,000 records annually. This includes the CPSC cases and the Centers for Disease Control and Prevention (CDC) cases (about half of the records meet the CPSC's eligibility requirements). The current costs are split into about $2 million, which the CPSC pays hospitals for NEISS cases, and $2.5 million paid to the hospitals by the CDC for NEISS-AIP cases. With the knowledge that the NEISS and NEISS-AIP budgets may not be increasing, we set the overall NEISS and overall NEISS-AIP data collection cost constraints to their current costs of $2 million and $2.5 million, respectively.

In constraining the desired sample size, we recognize that an overall allocation larger than 100 selected hospitals would not likely be feasible due to the increase in staff and project costs this would entail. Thus, we constrained the overall sample size to 100 hospitals for two of the three optimal allocations presented in section 3.3. We removed the overall sample size constraint while keeping the NEISS cost constraint of $2 million dollars for one option in order to present the gains in precision that would be achieved if such an option were feasible.

A final constraint applied to all three options was the Children's hospital stratum must have an allocation of at least eight hospitals. Given that there were only 63 hospitals within the Children's hospital stratum on the 2018 NEISS sampling frame this stratum contributes the least to the overall variance of each of the injury statistics. Given this, the nonlinear programming algorithm will assign fewer than eight hospitals to the Children's stratum unless a constraint is applied. We understand that estimates of the number of children's injuries is one of NEISS' key outcomes and thus wanted to ensure that the allocation size of the Children's hospital stratum in any redesign option at least matched the allocation size of 8 from the 1995 NEISS sample design. In Section 3.5 for one of the NEISS-AIP options we loosen this constraint, because children's injuries are less of a focus for NEISS-AIP.

3.2 Key Estimates

Based on discussion with the CPSC, we produced three alternative allocations around the following list of six key estimates:

- Total injuries
- Sports-related concussions
- Elderly hospitalizations
- Children's injuries
- Firearms injuries
- Power saw injuries

Product groupings (e.g., children's injuries) were used because they were the most reliable to deal with – not individual product-level estimates (e.g., cribs or car seats) for which there are fewer observations.

3.3 Alternative Allocations

This section presents sample size, estimated total per-case data collection costs, and estimated precision for three NEISS sample allocations. Each option focuses on minimizing the overall coefficient of variation (CV) but differing on being constrained by a target sample size of 100 hospitals, an overall per-case data collection cost of $2 million dollars for the CPSC cases, or both. The three sample allocations are:

- Option 1: Budget and Sample Size Constrained Average Relvariance:
  - A budget ($2 million) and sample size (*n* = 100) constrained approach that minimizes the average expected relvariance across a set of six estimates, by choosing one set of stratum-specific hospital sample sizes.

- Option 2: Budget Constrained Average Relvariance:
  - A budget ($2 million) constrained approach that minimizes the average expected relvariance across a set of six estimates, by choosing one set of stratum-specific hospital sample sizes. No constraint on total sample size was used.
- Option 3: Sample Size Constrained Average Neyman Allocation across 6 Key Statistics:
  - A sample size (n = 100) constrained Neyman allocation approach that uses the average of the Neyman allocations across each of six estimates.

As noted in Section 3.1, for all three of these allocations, we required that at least eight sample hospitals be allocated to the Children's stratum.

To obtain these results, unweighted stratum-level population variances S_h^2 were estimated using NEISS data from 2015-2018. Rather than using each hospital's stratum as assigned when the sample was selected, we have reassigned the 1996 sample of hospitals to strata based on the 2018 values in the frame file provided to us by the CPSC[1]. As noted in section 2.2, there has been some "stratum jumping" since 1996. Reassigning hospitals to strata based on 2018 data provides more up-to-date estimates of the variance components needed for allocation. One effect of reassigning strata is that the hospitals within a stratum have sizes that are more similar to each other than if the 1996 strata had been used. This, in turn, produced population stratum variance estimates that were smaller than they would be using the 1996 strata. Population variances for each of the six injury types were estimated and averaged across the four years. The average cost per hospital in each reassigned stratum was also computed from hospital-level cost data for the 2020 fiscal year provided by the CPSC. The details of the formulas used to produce the average expected relvariances (square of the CV) and allocations per stratum (n_h) for each option are in Appendix A.

Table 3-1 presents the allocations for the three alternatives listed above.

---

[1] We reassigned hospitals to strata based on 2018 data rather than 2019 data because the frame file provided by the CPSC ended for data collection year 2018 and did not have stratum values for 2019. This was also the reason our estimates of S_h^2 used NEISS data through 2018 even though the 2019 data were available online at the time we calculated the alternative allocations.

Table 3-1 NEISS optimal allocations by stratum based on NEISS 2015-2018

Stratum	Current Design: 2018 Frame		Current Design: 1996 Sample Selected Hospitals[a]	Current Design: Average CPSC Cost per NEISS Hospital – Using 2018 Design Strata	Current Design: 2018 Responding Hospitals[b]		Redesign Allocation Options: Option 1: Budget and Sample Size Constrained Average Relvariance; Constraints: n = 100, Total Cost ≤ $2 million		Option 2: Budget Constrained Average Relvariance Constraint: Total Cost ≤ $2 million		Option 3: Sample Size Constrained Neyman Allocation Average across 6 Key Stats; Constraint: n = 100	
	N	%			n	Total Cost[c]	n	Total Cost[d]	n	Total Cost	n	Total Cost
Small	3,002	63%	48	$8,599	43	$369,763	43	$367,655	47	$402,262	36	$306,115
Medium	868	18%	14	$12,088	7	$84,615	26	$316,329	27	$327,577	22	$271,786
Large	536	11%	9	$31,083	17	$528,403	12	$383,909	12	$365,582	16	$502,416
Very Large	295	6%	23	$51,064	13	$663,837	11	$556,799	10	$529,340	18	$916,063
Children's	63	1%	8	$46,905	8	$375,240	8	$375,271	8	$375,240	8	$375,240
Total	4,764	100%	102	$23,923	88	$2,021,857	100	$1,999,963	104	$2,000,000	100	$2,371,620

[a] By 2018, hospitals in four Small stratum PSUs had closed without a replacement. By 2019, this had increased to five Small stratum hospitals.

[b] The responding hospitals are assigned to the strata based on their number of ERVs recorded on the 2018 American Hospital Association (AHA) database.

[c] $2 million is the annual CPSC data collection cost; the CDC currently funds an additional $2.5 million for the NEISS-AIP sample that is not considered in the allocations in Table 3-1.

[d] The Total Cost calculations for Options 1, 2 and 3 are based on expected allocation rather than the whole number allocation (i.e., n) displayed in the Table 3-1. Thus, the Total Cost does not equal the product of the value in the 'Average CPSC cost per NEISS Hospital – Using the 2018 Design Strata' and the whole number allocation provided in the table. For example, the expected allocation for the Small stratum in Option 1 is 42.75486 (compared to the whole number 43 reported in the table) and the average cost for the Small stratum is $8,599.13, producing a Total Cost of $367,655.

Option 1 optimally allocates a sample of 100 hospitals with a cost not to exceed $2 million. Option 2 only constrains the total cost and results in a total sample of 104 hospitals. Option 3 computes an optimal allocation (Neyman) of 100 hospitals separately for each of the six key statistics, and then computes the average of these six allocations in each stratum. In all cases, the optimization maximizes precision of national estimates by minimizing the relvariance (square of the CV) of the estimates.

We recognize that Option 2 with 104 hospitals will incur additional costs for recruiting and maintaining a larger group of hospitals but maintains the $2 million per year variable costs associated with coding data. Having more than 100 hospitals would have some advantages, one of which would be the ability to have wider geographic representation.

Table 3-2 provides the expected precision (CVs) for each of the sample allocation options presented in Table 3-1. The first row contains the estimated CVs for the current participating sample of hospitals, with the new allocation options in the following rows. Note that the CVs based on the 2018 respondents in the first row use their design strata from 1995 and not their reassigned strata, as the three options do. This row also uses the 2018 weights and gives the CVs that an analyst of NEISS data would calculate. The CVs for Options 1-3 were computed as the square root of the relvariance in equation (1) shown in Appendix A.

Table 3-2 CVs by key statistic and NEISS allocation method, followed by improved precision for each method

Allocation Method[2]	Total Injuries	Sports-Related Concussions	Elderly Hospitalizations	Children's Injuries	Firearms Injuries	Power Saw Injuries
2018 Responding Hospitals – Using 1995 Design Strata	8.5%	11.7%	13.2%	12.7%	10.3%	8.4%
Option 1: Budget and Sample Size Constrained Average Relvariance	6.6%	8.9%	11.7%	8.1%	8.7%	7.0%
Option 2: Budget Constrained Average Relvariance	6.5%	8.9%	11.7%	8.1%	8.6%	6.9%
Option 3: Sample Size Constrained Neyman Allocation Average across 6 Key Stats	6.3%	8.0%	10.6%	7.7%	8.5%	6.7%
Option 1 CV / 1995 Design Strata CV	77.4%	75.9%	88.5%	64.2%	83.7%	83.0.%
Option 2 CV / 1995 Design Strata CV	77.0%	76.4%	89.0%	64.0%	82.7%	81.6%
Option 3 CV / 1995 Design Strata CV	74.5%	68.0%	80.2%	61.1%	82.0%	79.2%

Options 1 through 3 all yield smaller CVs than the “2018 Responding Hospitals” allocation. Part of the reduction is because each of Options 1 through 3 uses at least 100 hospitals while the “2018 Responding Hospitals” allocation has only 88. Most of the reductions, however, are due to the options having allocations that are more efficient. In particular, the stratum of Medium hospitals receives a much larger allocation in each option in the 2018 sample (see Table 3-1). There are only seven Medium stratum hospitals among the 2018 respondents, which is inefficiently low compared to Options 1-3.

3.4 Recommendation for NEISS Sample

Beyond the differences in total sample size, or total overall per-case data collection costs, when not constrained, the differences in sample sizes by stratum are largely limited to the Small and Medium hospital strata. In Options 1-3 the Children’s hospital stratum is always set to a minimum of eight,

[2] The CVs in the first row of Table 3-2 were calculated using the weights provided on the public use file and the SAS PROC SURVEYMEANS procedure. Options 1 through 3 were calculated using unweighted stratum-level population variances S_h^2 from 2015-2018 NEISS data. The S_h^2 values were calculated unweighted to avoid issues of increased variation in the weights due to stratum jumpers because these options use the 2018 design strata and not the 1995 design strata. All hospitals within a stratum in these three options were treated as having equal weights.

even though it contains less than 2 percent of all ERVs. With respect to precision, the results of all three approaches are similar, and more alike one another than for the current allocation (i.e., 2018 responding hospitals). By comparison, the CVs for any estimate of any of the three options are always an improvement over the current sample. The ratio of the CV for the options in Table 3-2 to the CV for the 2018 respondents (shaded rows in the table) range from 61.1% (Option 3 for children's injuries) to 89.0% (Option 2 for elderly hospitalizations). The improvements are substantial given that we required that the samples respect the current NEISS budget of $2 million. Thus, an improved, updated allocation of a sample of 100 or 104 hospitals is estimated to cost the same for data collection as the current sample of 88 respondents and yield estimates that are more precise.

In addition to the relative increase in precision (CVs) offered by each of the allocation options, designing and selecting a new sample, albeit with overlap control, also brings the additional, but less-tangible benefits of being contemporary (i.e., allowing participation from all existing hospitals) and addressing stratum migration. Stratum jumping is likely to continue in the future and will inevitably lead to some loss in efficiency over time. However, the calibration estimation described in section 4.4 will help offset stratum migration by adjusting weights to respect annually updated population counts.

Given that an overall sample of 104 hospitals presented in Option 2 would incur extra recruiting costs that are not practical for the CPSC at this time and that the CVs for any estimate of Options 1 and 2 are nearly equivalent, we recommend Option 1 as the final sample allocation in a NEISS sample redesign.

3.5 Recommendation for NEISS-AIP Sample

With Option 1 as the recommended final sample allocation for a NEISS sample redesign, we generated the following three options to examine how the NEISS sample redesign will affect CDC costs for the NEISS-AIP sample as well as the CVs of key estimates. The CPSC provided data on four injury types of interest to CDC NEISS-AIP: firearm injuries, work injuries, adverse drug effects, and self-inflicted injuries. The three options are:

- **Budget Constrained Average Relvariance Option:**
  - A budget ($2.5 million) constrained approach that selects a subsample of the Option 1 100 hospitals which minimizes the average expected relvariance across the four CDC estimates, by choosing one set of stratum-specific hospital sample sizes. Each stratum was required to have at least eight hospitals allocated.
- Full NEISS Redesign Option:
  - Have all of the 100 Option 1 NEISS sample hospitals be a part of the NEISS-AIP sample.
- NEISS redesign with subsample of children's hospitals
  - The same sample as above, but retain only 4 of the 8 children's hospitals.

To estimate the CDC costs for these three options, we calculated the average CDC cost per NEISS-AIP hospital based on the 2018 design strata using the NEISS fiscal year 2020 contract funding files provided by the CPSC. We then estimated the total cost by stratum by multiplying the average cost by the allocated sample in each option. While the CDC costs contained in Table 3-3 are our best estimate using current information, future costs for a new sample of hospitals may be potentially different from what is listed here.

As Table 3-3 presents below, the "Budget Constrained Average Relvariance" option is consistent with the current funding level of $2.5 million to collect NEISS-AIP. While we recognize the "Full" option well exceeds the current level of funding, it serves as a useful comparison point to the "Budget Constrained" option on costs. In addition, the "Full" option shows the increases in precision (provided in Table 3-4) that could be achieved if the CDC has the available funds and desire to increase their NEISS-AIP budget beyond its current level. Alternatives between these two options could be created to match intermediate funding levels; an example is given in Option 3 where the number of children's hospitals has been cut in half.

.

Table 3-3 NEISS-AIP optimal allocations by stratum based on NEISS-AIP 2015-2018

Stratum	2018 NEISS-AIP Responding Hospitals – 1995 design strata	Average CDC Cost per NEISS-AIP Hospital – 2018 design strata	2018 NEISS-AIP Responding Hospitals – 2018 design strata		<u>Budget Constrained Average Relvariance Option:</u> <u>Constraints</u>: Total Cost ≤ $2.5 million; $n_h \ge 8$		<u>Full NEISS Redesign Option:</u> All 100 NEISS-AIP Hospitals using Option 1 Allocation		<u>NEISS Redesign w/ Children's Hospital Subsample Option:</u> NEISS-AIP Hospitals using Option 1 and Children subsample Allocation	
			n	Total Cost	n	Total Cost	N	Total Cost	n	Total Cost
Small	27	$14,612	30	$438,360	19	$270,913	43	$628,316	43	$628,316
Medium	7	$29,752	6	$178,512	15	$450,431	26	$773,552	26	$773,552
Large	6	$50,835	10	$508,350	8	$406,680	12	$610,020	12	$610,020
Very Large	14	$128,338	8	$1,026,704	8	$1,026,704	11	$1,411,718	11	$1,411,718
Children's	5	$43,159	5	$215,796	8	$345,272	8	$345,272	4	$172,636
Total	59	$42,793	59	$2,367,718	59	$2,500,000	100	$3,768,878	96	$3,596,242

Table 3-4 CVs by key statistic and NEISS-AIP allocation method, followed by percent change in precision for each method

Allocation Method[3]	Firearm Injuries	Work Related Injuries	Adverse Drug Effect Injuries	Self-Inflicted Injuries
2018 Responding NEISS-AIP Hospitals – 2018 Design Strata	26.1%	15.8%	13.5%	10.8%
(1) Budget Constrained Average Relvariance	20.7%	9.8%	12.6%	11.2%
(2) Full Sample: 100 NEISS-AIP Hospitals using NEISS Option 1 Allocation	16.7%	7.2%	9.9%	8.3%
(3) NEISS Option 1 Allocation with Children's hospital Subsample	16.7%	7.2%	9.9%	8.5%
(1) CV / 2018 Design Strata CV	79.1%	62.0%	93.5%	103.4%
(2) CV/ 2018 Design Strata CV	63.9%	45.5%	73.9%	76.9%
(3) CV/ 2018 Design Strata CV	63.9%	45.5%	73.9%	78.3%

All three options both yield smaller CVs than the "2018 Responding Hospitals" allocation, with the exception of self-inflicted injuries for the "Budget Constrained" option. This is expected given that the options optimize the allocation to minimize precision for a given cost. As was the case for the NEISS allocation, while a large part of the reduction for the "Full Sample" option for NEISS-AIP is that it uses 100 hospitals, reductions are also due to the options having more efficient allocations. For example, compared to the "2018 Responding Hospitals" baseline option, the "Budget Constrained" option increased the Medium hospital stratum allocation by nine even though the overall allocation had the same number of hospitals (see Table 3-3).

There is very little difference in CVs for Options 2 and 3. This is because few of these types of injuries are treated in Children's hospitals, so including 4 versus 8 such hospitals has little impact on the estimates. Given that Option 3 has $170,000 less in annual data collection costs, it is preferable to Option 2.

We recommend either the "Budget Constrained" or Option 3 be implemented, with the latter preferable if the funding is available. Option 3 clearly provides the highest level of precision due to

[3] The CVs in the first four rows of Table 3-4 were calculated using unweighted stratum-level population variances S_h^2 from 2015-2018 NEISS data.

the substantial increase in overall sample size. If the $2.5 million cost cannot be increased, the "Budget Constrained" option provides the best balance of precision and cost.

4. Sample Maintenance

4.1 Implementation, Overlap, and Bridge Period

Recruitment will be part of implementing the new sample, and given the costs and level of effort associated with recruitment, retaining hospitals previously in the sample is advantageous. Overlap control procedures have been used previously in NEISS, and we would recommend their continued use with a new sample.

The expected overlap between a new NEISS sample and the 1996 sample, assuming NEISS sample allocation Option 1, is shown in Table 4-1.

Table 4-1 Expected overlap between 1996 and 2020 NEISS samples

	1996				2020			
Stratum	Pop.	Sample	Sampling Rate	Sample – 2018 Respondents	Pop.	Sample	Sampling Rate	Expected Overlap[4]
Small	3,179	48	0.0151	39	3,002	43	0.0143	37
Medium	1,059	14	0.0132	11	868	26	0.0300	11
Large	674	9	0.0134	9	536	12	0.0224	9
Very Large	426	23	0.0540	21	295	11	0.0373	11
Children's	50	8	0.1600	8	63	8	0.1270	6
Total	5,388	102		88	4,764	100		74

With an expectation that around 14 hospitals (88-74) will be dropped from the current sample and 26 others added, it will be important to include a bridge period when data are collected from both old and new hospitals. If this can be maintained for a full year, it will help smooth discontinuities over time with the updated sample. This is particularly important for relatively rare types of injuries that might be more common in a small percentage of all hospitals. The cost of the bridge sample will be the cost of keeping the 14 hospitals in for the additional year.

[4] These are rough estimates only, reflect rounding, and cannot anticipate the effect hospitals new to the sampling frame at the time of the new sample selection will have on these results.

4.2 Nonresponse, Substitution, and Implications/Procedures for Trends

Historically, two or three sampled and participating hospitals become ineligible (e.g., close) or become nonrespondents and are replaced via substitution each year., on average. It can take some time to successfully recruit substitutes for such hospitals.

Ongoing unit nonresponse has been and can continue to be treated via substitution. Substitution, especially at the institutional or establishment level, is one remedy for unit nonresponse and has both some precedent in survey sampling and consideration in survey literature. Substitute schools have previously been used in the National Center for Education Statistics (NCES)' National Assessment of Education Progress (NAEP).[5] In NAEP, nonresponding original schools and their substitutes were matched via a distance measure on frame characteristics, including enrollment demographics, size, and school characteristics. Literature concerning the practice of substitution includes the following: Lynn 2004; Rubin and Zanutto 2002; and Vehovar 1999. The Substance Abuse and Mental Health Services Administration's (SAMHSA) Drug Abuse Warning Network (DAWN) is also currently using substitute hospitals within explicit and implicit strata in light of nonresponse and the need to meet rigid participation numbers.

Previously, NEISS has reached out to sampled hospitals and attempted to gain cooperation in advance of fielding, and a substitute is activated only when all efforts to obtain cooperation fail. Several alternates are provided, and are worked in their designated turn. Given the NEISS sample design (stratified, equal probability within stratum), substitute hospitals from the same design stratum have the same probabilities and weights of original selections. This is an advantage both from the point-of-view of clarifying how to assign weights to the substitutes and of explaining methodology to data users.

However, there are two drawbacks to this approach. First, those weights assume the set of hospitals from which the replacement is drawn is the same as that when the original sample was selected, which excludes newer information available from updated American Hospital Association (AHA)

---

[5] https://nces.ed.gov/nationsreportcard/tdw/sample_design/2002_2003/sampdsgn_2002_main_subst.aspx

frames. Second, it assumes that hospitals are still of similar size to when they were originally selected. Both of these are problematic.

We recommend modifying the procedure used to select substitutes. At the time that a substitute is needed, the most current AHA frame should be used as a basis for selection. This implies stratifying the frame into four size categories (plus the Children's hospital stratum) based on the previous year's ERVs. All NEISS hospitals, including the no longer participating hospital should be assigned to sampling strata based on that frame. A replacement hospital should then be selected from nonsampled hospitals in the same stratum as the non-participant (i.e., not based on its stratification in the 1996 survey design, but the stratification based on the most recent AHA frame data). All sampled hospitals then receive an updated base weight, defined as the number of hospitals in their frame stratum divided by the number of sampled hospitals in their stratum. This procedure properly allows newer hospitals to be eligible to be selected as a replacement and updates all survey weights to properly reflect stratum migration since the last time the sample was updated.

The choice of the replacement hospital should be based on a multivariate distance measure using variables on the updated AHA frame, and be allowed to cross state lines (previously this was discouraged). A variant of this is predictive mean matching (Little 1988), in which the number of injuries for a set of key products is predicted for all hospitals on the frame based on a regression model. A substitute would be selected from those whose multivariate distance of the predictions is close to the predicted (or observed) numbers of injuries for the original. Substitute hospitals should be identified for the new sample when it is selected because it is likely that some newly sampled hospitals may refuse to participate, and can be identified subsequently in an annual process when new AHA information becomes available. The distance measure should be developed as part of the actual updating of the NEISS sample under a future contract.

Discontinuities in injury time series. Replacement hospitals can also be a source of a change in trend. Substitute hospitals can behave very differently from the original hospitals that they replace, at least for some subsets of estimates, due to natural differences in hospital emergency department catchment areas, purposes and specialties, and geographic location and proximity to unique or rare causes of particular injuries. There are several possible treatments, including:

- Implementing a method of smoothing or wedging-in change.

- Obtaining previous months injuries for a replacement hospital to assess differences and make adjustments
- Utilizing disclaimers and cautionary language in publications

The third treatment requires no work outside of the writing of reports, but does not make attempts to dampen the effects of change in estimates attributable to the differences in injury frequencies between the original and sampled hospital. If dampening the effects of change is preferable to the CPSC, we would recommend an approach that, for one reporting cycle (i.e., a year), combines the remaining treatments listed above. Such an approach could obtain previous months for a replacement hospital (by reviewing historical electronic health records), adjust the reported case frequencies for a select subset of estimates, and ERVs, for the replacement hospital to those of the original hospital for a period of up to 12 months, and thereby smooth or wedge-in the change for a one year period.

Appendix B describes estimation options that can be used to handle discontinuities in time series of injury estimates. The first is wedging (see section B.1), which is discussed below with two examples. Wedging is probably the simplest to implement but does assume that a discontinuity induced by a replacement hospital is a better reflection of the truth than was the previous level. A second option in section B.2 is poststratified or calibrated estimation, in which weights are adjusted based on updated frame totals of a set of covariates (e.g., ERVs, strata counts, or hospital counts by urbanicity). This option will only indirectly smooth out discontinuities by forcing weighted sums of covariates to match totals in the updated population for every time period.

A third option (see section B.3) is Winsorization, in which the injury reports for a hospital are compared to those of the rest of the sample to determine whether the reports are outliers. This method is injury-specific and can be applied at the editing stage. The last option in Appendix B is weight trimming, which is based on the thinking that an outlying hospital is not representative of other hospitals and, thus, should not be weighted-up to represent others. Weight trimming has the disadvantage of applying to all injuries reported by a hospital regardless of whether a report is extreme or not. Either wedging or Winsorization are most likely to be useful for NEISS.

Consider the following two examples as potential changes-in-trend (identified using data provided by the CPSC) and how they might be addressed by wedging-in, at an annual basis, over a two year period (such that $p = 2$ in Appendix B, and the below applies to the first year or $u = 1$ of the two year wedging-in period).

Example 1: PSU 55 – Amusement park injuries

In Table 4-2, for PSU 55, are the last 12 months for the responding original hospital (hospital A) and the first 12 months for its responding substitute hospital (hospital B). Unfortunately, we do not have data for the exact same period, so we have to compare different 12-month periods.

Table 4-2 Amusement park injuries for 12 months before original hospital was inactive and 12 months after substitute hospital became active

Hospital	12 Months Before Original was Inactive and 12 Months After Substitute was Active												Total
A	Jan-00	Feb-00	Mar-00	Apr-00	May-00	Jun-00	Jul-00	Aug-00	Sep-00	Oct-00	Nov-00	Dec-00	46
	1	0	0	2	3	9	11	14	4	2	0	0	
B	Oct-01	Nov-01	Dec-01	Jan-02	Feb-02	Mar-02	Apr-02	May-02	Jun-02	Jul-02	Aug-02	Sep-02	1
	0	0	0	0	0	1	0	0	0	0	0	0	

Hospital A's 12 month total was 46, whereas hospitals B's 12 month total was only 1. One possible wedging-in approach for a two year period would be for the first year reporting to include hospital B's data and inflate the reported annual total to 43.5 (reported 1 plus (46 – 1)/2). This would limit the decrease in the estimate in the first year of hospital B's reporting.

Example 2: PSU 41 – Firearms injuries

Table 4-3, for PSU 41, are the last 12 months for the responding original hospital (hospital A) and the first 12 months for its responding substitute hospital (hospital B). We should note, however, that this original-substitute pair is an instance where the original hospital migrated strata, and with the procedure recommended above, its substitute would be from the same "current" stratum as the original, and thus the differences in the monthly reported values and annual averages should be less extreme. Nonetheless, it is a useful example for when the discrepancy in counts between hospital A and B is larger.

Table 4-3 **Firearm injuries for 12 months before original hospital was inactive and 12 months after substitute hospital became active**

Hospital	12 Months Before Original was Inactive and 12 Months After Substitute was Active												Total
A	Nov-09	Dec-09	Jan-10	Feb-10	Mar-10	Apr-10	May-10	Jun-10	Jul-10	Aug-10	Sep-10	Oct-10	2
	1	0	0	0	0	0	1	0	0	0	0	0	
B	Jun-12	Jul-12	Aug-12	Sep-12	Oct-12	Nov-12	Dec-12	Jan-13	Feb-13	Mar-13	Apr-13	May-13	794
	87	95	87	92	82	0	0	74	47	62	74	94	

Hospital A's annual total was only 2 firearm injuries, whereas Hospitals B's annual total was 794. (It may be that hospital B did not report any data to NEISS in November and December 2012, in which case the annual total number of firearm injuries is probably even bigger.) Clearly, the substitute hospital in this case is very different from the original one, and the change doesn't represent an increase in firearm injuries. Because hospital A had become a Medium-sized hospital it would have been much better to replace it with another hospital in the Medium stratum, rather than hospital B, which was in the Very Large stratum using the more recent AHA ERVs. But to demonstrate the wedging-in approach for the first year reporting including hospital B's data, the reported annual total is reduced to reflect an annual total of 398 (2 plus (794-2)/2), limiting the increase in the estimate in the first year of hospital B's reporting. When the difference is this big it may be desirable to spread the wedging across more than two years, but hopefully the revised substitution methodology will greatly reduce incidences with such large differences.

4.3 Monthly Nonresponse Adjustment

In addition to refusals and substitution, hospitals that are otherwise responding sometimes fail to provide data for one or more months. See Table 4-4 for an accounting of the current sample and the number of hospitals by months missing data based on 2017-2018 data (i.e., a 24 month period).

Table 4-4 **Number of months missing and count of sampled hospitals**

Months Missing	Count
0	69
1 to 3	6
4 to 6	4
7 to 9	3
10+	8
All	12
Total	102

In the case of the monthly adjustment for nonresponse, the variance may actually be greater between hospitals (within a stratum) than within hospitals across months. Under such circumstances, an adjustment across months within a hospital may be preferable. See Piesse and Green (2004) for more details on similar findings and an approach from DAWN.

The most effective nonresponse adjustment to implement for the occasional, limited, monthly nonresponse of otherwise participating hospitals depends on the relationship between the NEISS case counts for the missing hospital across month and across hospitals. Currently, the monthly nonresponse adjustment inflates the monthly weight of responding hospitals by the ratio of the number of in-scope sample hospitals to responding hospitals in that month, within original sampling stratum. This adjustment is equivalent to imputing the mean of the respondents for every product-related injury within each original stratum to each nonrespondent.

Analyzing and reviewing the variance of reported NEISS case counts overall, between hospitals within stratum, and between months within hospital can help the CPSC choose among the above alternatives. To this end, we used an analysis of variance (ANOVA) model to analyze the hospital data for the six estimates (total injuries, injuries involving children, elderly hospitalizations, power saw injuries, sports-related concussions, and firearms injuries) and using a panel of all 61 consistently reporting hospitals (i.e., no missing data across 4 years 2015-2018). Table 4-5 provides the mean square errors (MSEs) for each type of injury. The hospital MSE is between hospitals within stratum, the month MSE is within hospital across months of the year, and the residual MSE is the remaining variability within hospital/month across the four years. The magnitude of the MSEs is not important, it is a function of how many injuries of that type are in the file. What is important is the relative size of the MSE columns for a given type of injury. The ANOVA results show that for all six NEISS estimates the variance is predominantly between hospitals within stratum. For all injury types, the hospital MSE is 10 to 40 times the residual MSE.

Table 4-5 ANOVA results for 61 NEISS hospitals and six injury types

Injury	Hospital MSE	Month MSE	Residual MSE
Total Injuries	4,094,459	15,801	98,848
Children’s	1,054,041	8,925	70,276
Elderly hospitalizations	15,836	33	645
Power saws	68	2	6
Sports-related concussions	1,409	120	141
Firearms	6.5	0.5	0.6

Based on the results in Table 4-5, we recommend the CPSC consider an alternative monthly nonresponse adjustment approach that uses historical information from the same hospital rather than across hospitals, for example the value from the last reported month. This inclusion of imputed data in a public file is currently done by the National Science Foundation's Survey of Science and Engineering Research Facilities https://www.nsf.gov/statistics/srvyfacilities/data/2017-facilities-puf-user-guide.pdf (the Energy Information Administration also has used this in its surveys, but we were unable to confirm if this is still produced). In the case of some NEISS estimates where injuries are seasonal (e.g., fireworks), it would be preferable to not use data from the preceding month, but rather use data from the previous year.

There is another advantage to imputing missing data over the current weight adjustment. One of the causes for disruptions in data series when a hospital stops reporting and is then replaced can be attributed to the weighting adjustment. By adjusting for nonresponse by increasing the weights of responding hospitals from the same original stratum, NEISS is assuming that the nonresponding hospital is typical for that stratum. But if the hospital has migrated strata that assumption is not true. By using historical data from the nonresponding hospital, the imputed values will be consistent with the recently reported data from that hospital.

Although we recommend the CPSC consider the above approach for intermittent nonresponse, we should note that the calibration estimators presented in section 4.4 can be viewed as a way of both reducing standard errors and correcting for nonresponse bias, all in a single procedure. Those estimators control the weights to the updated frame totals of variables on the AHA file. The resulting weight adjustments implicitly correct nonresponse bias, assuming that any nonresponding hospitals are missing at random and that the calibration model holds for both responding and nonresponding hospitals. The calibration estimator, however, is making a common adjustment for all injuries, while the above approach allows for injury-specific adjustments and therefore is preferable.

When intermittent nonresponse becomes longer term (say, more than a year since the hospital last participated in NEISS), the above recommendation may no longer be ideal. The relationship between the currently missing monthly injuries and the last reported data will diminish over time. Ideally in such cases a substitute hospital will be in place, but if that has not occurred, the CPSC may need to develop ad hoc weighting adjustment based on a review of the nonresponding hospital's injury levels relative to those still reporting.

In summary, selecting imputation, weight adjustment, or calibration to correct for nonresponse in NEISS deserves more research that we have been able to give it here. We have made recommendations but recommend that a more extensive study be made using previous NEISS data. For example, artificial nonresponse might first be created from the complete data samples over time. The artificial nonresponse should reflect the patterns of non-reporting actually seen in NEISS throughout its history. Alternative weight and imputation strategies could then be applied to evaluate how the alternatives perform over time. Beginning with complete data and controlling the patterns and degree of missingness will allow biases to be estimated for the alternative procedures.

Finally, we note that using monthly nonresponse adjustments that can vary over the year substantially complicates variance estimation. Appendix C discusses this in detail, but the nub of the problem is reflecting the randomness of adjustments that may differ across months. Appendix C.3 describes how to make a single, annual nonresponse adjustment, which we recommend, that can be used in conjunction with the ratio adjustment to total ERVs or with the calibration estimation discussed below in Section 4.4. A replication variance estimator is also sketched in Appendix C.3 that will correctly reflect the complexity of the NEISS estimator and can be implemented by analysts using available software.

4.4 Weighting and Variance Estimation

The current weights for NEISS are based on inverse selection probabilities that are ratio adjusted based on the number of ERVs on a recently updated frame of eligible hospitals (Schroeder, 2019). The ratio adjustment helps account for the fact that the hospital population does not remain static over time. Hospitals close, merge, and open, as well as change in the volume of ERVs. Ratio adjusting is an example of a calibration estimator using the single covariate, ERVs. As illustrated later in this section, the ERV ratio adjustment also importantly reduces standard errors for estimated total number of injuries for the products we studied.

4.4.1 Estimating Totals

The redesigned NEISS again will be a stratified simple random sample (*stsrs*). The base weight of sample hospital *i* in stratum *h* will be $w_{hi} = N_h / n_h$ where N_h is the number of hospitals in the frame in stratum *h*, and n_h is the number of sample hospitals selected from that stratum. More generally, w_{hi} would be the inverse of the selection probability including a nonresponse adjustment and the ERV adjustment. The π - or Horvitz-Thompson estimator of a total number of injuries is

$$\hat{t} = \sum_{h=1}^{H} \sum_{i \in s_h} w_{hi} y_{hi}$$
$$= \sum_{h=1}^{H} N_h \bar{y}_h$$

where y_{hi} is the number of injuries reported by hospital *hi*, s_h is the set of sample hospitals in *h,* and $\bar{y}_h = \sum_{i \in s_h} y_{hi} / n_h$. The estimator $\hat{t}$ is unbiased in repeated stratified simple random sampling, assuming that any nonresponse adjustment corrects for nonresponse bias.

The precision of survey estimators can often be improved by using additional covariates that are predictive of the number of injuries treated in hospital emergency departments. For example, hospital location (urban, suburban, or rural) is related to the occurrence of certain types of injuries. The goal is to identify a single set of covariates that will be used for all estimates. This will result in a single set of weights at each time period that will be used for all estimates as is now done in NEISS. The method we describe below uses control totals for covariates (e.g., total ERVs) that would be updated annually using the AHA file. Updated controls will help adjust for changes in the hospital universe. Software is also available that will allow the CPSC to implement the weighting method (e.g., see An, 2020; Valliant, Dever, and Kreuter, 2018, Chapter 14).[6]

Calibrating estimators to a set of population totals of covariates also can also be viewed as a single-step adjustment to account for hospital nonresponse in a particular time period and for the change of the composition of the eligible hospital population, similar to the methods currently used for NEISS that are described by Schroeder (2019). Consequently, a calibrated estimator could be used

[6] An (2020) reports on a SAS macro for calibration that may fit well with the NEISS processing system. However, based on internal testing by Westat, the current version of the macro fails for some problems and currently needs revision; SAS has been notified of the problems.

to accomplish the same purpose as the nonresponse adjustment discussed in section 4.3, although it will not be injury-type specific like that adjustment.

One example of a calibration estimator is a general regression estimator (GREG). Suppose that *p* auxiliaries are used, $\mathbf{x}_{hi}$ is the vector of *p* auxiliaries for hospital *hi*, $\mathbf{t}_x = \left(t_{x1}, \ldots, t_{xp}\right)^T$ is the vector of the *p* population totals of the auxiliaries, and $\hat{\mathbf{t}}_x = \sum_h \sum_{s_h} w_{hi}\mathbf{x}_{hi}$ is the estimator of those totals. The GREG is defined as

$$\begin{aligned}\hat{t}_G &= \hat{t} + \left(\mathbf{t}_x - \hat{\mathbf{t}}_x\right)^T \hat{\mathbf{B}} \\ &= \sum_h \sum_{i \in s_h} \underbrace{\left[1 + \left(\mathbf{t}_x - \hat{\mathbf{t}}_x\right)^T \left(\mathbf{X}^T \mathbf{W}\mathbf{X}\right)^{-1} \mathbf{x}_{hi}\right]}_{g_{hi}} w_{hi} y_{hi}\end{aligned}$$

where $\hat{\mathbf{B}} = \left(\mathbf{X}^T \mathbf{W}\mathbf{X}\right)^{-1} \sum_h \sum_{i \in s_h} w_{hi}\mathbf{x}_{hi} y_{hi}$, $\mathbf{X}$ is the $n \times p$ matrix of auxiliaries for the *n* sample hospitals, and $\mathbf{W}$ is the $n \times n$ diagonal matrix of the w_{hi}'s for the sample hospitals. These estimators have been well developed theoretically (Deville and Särndal, 1992), and are in common use by government agencies (Estevao, Hidiroglou, and Särndal, 1995; European Union, 2014).

The calibrated weight for hospital *hi* is

$$w_{Ghi} = w_{hi} g_{hi}$$

with g_{hi} defined above. A related estimator would be one that rakes the w_{hi} weights to the same set of *x*-totals. An estimated total using the GREG weights is computed as

$$\hat{t} = \sum_{h=1}^{H} \sum_{i \in s_h} w_{Ghi} y_{hi} .$$

Thus, analysts use the calibrated weights in the same way as the ones that NEISS currently supplies. Westat made a preliminary investigation of how to use ERVs and other covariates to determine whether reductions of standard errors can be made. One of the advantages of calibration is that covariates can be used in estimation that are not explicitly or implicitly part of the sample design. Both quantitative and qualitative covariates are possibilities. There are many potential predictors that are available on the AHA file, including control or ownership (government, non-Federal; nongovernment not-for-profit; investor owned, for profit; government Federal), type of services offered (general medical and surgical; surgical, psychiatric, etc.), whether a hospital/system has a health plan license, total Medicare and Medicaid inpatient discharges, net patient revenue, and many other variables that are available at the individual hospital level.

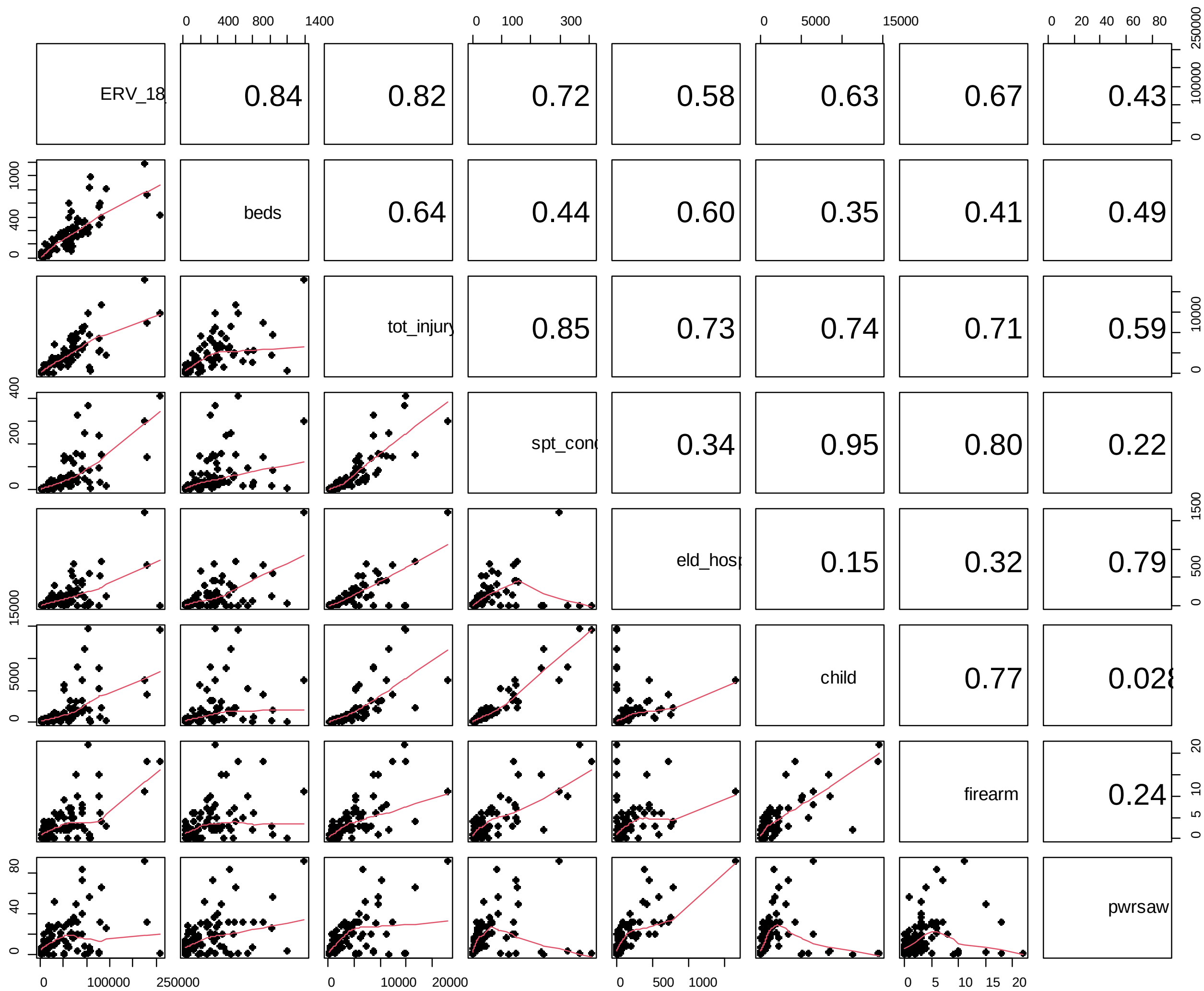
In the time available for this project, we have investigated a few of the possibilities for covariates, and it seems clear that improvements in precision can be made. However, we recommend that the results below be treated only as a proof-of-concept to be confirmed once the updated sample is selected and new data are collected. Further work may identify other covariates from the AHA frame that may be produce more efficient estimators.

We used the 2018 NEISS sample of 88 responding hospitals and considered the variables from the AHA frame as potential auxiliaries:

- Number of ERVs
- Number of inpatient beds
- 2018 stratum membership
- Census region and division
- Metropolitan statistical area (MSA coded as metro, micro, rural)

Figure 4-1 is a scatterplot matrix of ERVs, beds, and the six injury types shown in earlier tables (total injuries, sports-related concussions, elderly hospitalizations, children's, firearms, and power saws) using the 2018 data for the 88 responding hospitals. Correlations between variables are printed in the upper panels. The correlation of ERVs with the six injury types are in the first row and range from a low of 0.43 for power saw injuries to a high of 0.82 for total injuries. The correlations of inpatient beds with types of injuries are also reasonably high, but beds itself has a correlation of 0.84 with ERVs, implying that beds and ERVs are redundant as predictors.

Figure 4-1 **Scatterplot matrix of ERVs, beds, and the six injury types based on 2018 NEISS data. Correlations are in the upper panels. The line in each lower panel is a nonparametric smoother that shows the underlying pattern (if any)**



To identify useful covariates, we also fit survey-weighted models for the six injury types. The variables above were used as main effects. In addition, we examined the interaction of ERVs with several of the qualitative variables. The results for those models showed that ERVs, 2018 stratum membership, Census region, MSA type, ERVs within 2018 strata, and ERVs within Census region might all be effective auxiliaries in a GREG or raking estimator.

However, with only 88 respondents, fitting a model with more auxiliaries than are feasible to use in calibration is a danger. Thus, we selected four models for comparison:

(1) A stratified ratio estimator in which the weighted sum of total ERVs within collapsed strata (Children's, Small + Medium, Large + Very Large) are forced to equal the 2018 frame totals. This is similar to the current NEISS estimator, except that the ERV adjustment is made for annual data—not monthly. This estimator is labeled Ratio in tables.

(2) Poststratification (denoted PS in tables) in which the design strata themselves were used as poststrata. Hospitals were assigned to poststrata based on their 2018 ERVs, and the population controls were NEISS frame counts;

(3) A GREG estimator with the quantitative variable, ERVs, and the categorical variables, MSA type and 2018 stratum membership. These were some of the more important predictors of numbers of injuries and were chosen to illustrate whether gains in precision are possible. This estimator adds two covariates (ERVs and MSA type) to 2018 stratum membership in (1), and, can be viewed as a step up in complexity from poststratification. We also bounded the resulting GREG weights so that each was at least 0.4 of its input w_{hi} weight and at most 4 times that weight (although the upper bound was never approached). This estimator is denoted by GREG1 in tables and figures.

(4) A GREG estimator that includes only ERVs and MSA type. Poststratifying to 2018 stratum membership adds some undesirable variability to the weights, so this alternative was included as a way of reducing that variation. This estimator is denoted by GREG2 in tables and figures.

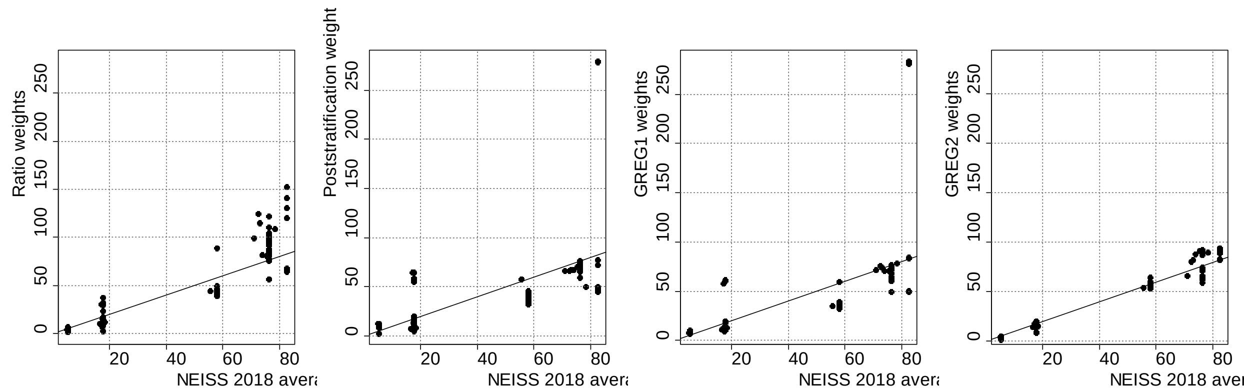
For analyses below, we used the NEISS weights on the 2018 public use file and treat them as inverse probability weights for computing the estimated totals $\hat{t}$. Since the NEISS hospitals did not all respond in all months of 2018, for each hospital we used its average weight across the months in which it did respond. The sum of the average 2018 NEISS weights is 4,782 compared to 4,764 in the frame. Thus, the NEISS estimate of total hospitals is very close to the population count. Estimates based on these average weights are labeled as HT in tables.

Figure 4-2 shows a plot of the ratio, poststratification, and GREG weights versus the NEISS average weights. Poststratification and GREG1 have two large weights of about 280 which occur in the Medium stratum. Ratio and GREG2, which does not include 2018 stratum membership as a covariate, do not have this issue. One measure of weight variability is the Kish design effect, $deff_K$; for the current design and four sets of weights, they are shown in Table 4-6.

Table 4-6 Design effects from differential weighting

Weights	$deff_K$
2018 NEISS	1.29
Ratio	1.41
PS	1.64
GREG1	1.64
GREG2	1.32

Figure 4-2 Plot of poststratification and GREG weights vs. NEISS 2018 average weight for the sample hospitals. A line is drawn at 45°



Thus, the GREG2 weights are comparable to the NEISS weights in variability, but, as shown in Table 4-7 below, GREG2 generally produces more precise estimates than the latter. Table 4-7 compares the estimated totals, standard errors, and CVs for estimated totals using the HT, ratio poststratification, and calibrated GREG weights. The GREG1 weights were controlled so that the sum of the weights was equal to the 2018 frame count of hospitals as shown in Table 3-1, the weighted sum of ERVs equaled the total of ERVs on the frame, and the estimated counts of hospitals by MSA type and by 2018 stratum equaled those in the frame.

As shown in Table 4-7, the estimated totals for the six injury types are similar, although somewhat lower than HT for almost all products. The Ratio estimates are closer to the HT estimates than PS. The GREG2 totals are from 2 to 5% smaller than the HT estimates, but the differences are all less than one GREG standard error. The standard errors of the Ratio, PS, and GREG estimators are uniformly less than those of the HT estimators with the one exception of Ratio for power saws. The last column of the table shows the ratios of the CVs for the alternative estimators to the CV of the HT estimator. These range from 0.55 for sports concussions with GREG1 to 1.09 for power saw injuries with Ratio. The Ratio estimator reduces CVs by 2% to 35% compared to HT, except for power saws where it is 9% worse than HT. Poststratification would reduce CVs by 15% to 42%, depending on the type of injury. GREG1 reduces CVs 19% to 45%. GREG2 shows reductions in CVs of 5% to 37%.

As mentioned earlier, the Ratio is the estimator most like the one now being used by NEISS. The PS, GREG1, and GREG2 estimators all are more precise than Ratio for the products we studied. These results imply that gains in precision can be achieved by using an estimator that is only slightly more complicated than NEISS' current estimator. However, we have by no means exhausted all the possibilities for improved estimators. Further investigation would be worthwhile on a selection of covariates and on whether a raking estimator may be an improvement over the GREG.

Another approach to identifying a set of effective covariates is to fit classification and regression trees (CART) separately to the six injury types studied in this report. CART and other machine learning algorithms are especially good at identifying interactions that may be worth considering when calibrating. It will be important to obtain a limited set of covariates whose predictive power is consistent across time and is effective for a range of products and hazards. Having a set that is too elaborate risks the creation of unstable estimators that do not always reduce variances and correct for changes in the hospital population.

Table 4-7 Estimated totals, standard errors, and coefficient of variation (CVs) with NEISS and GREG weights based on NEISS 2018 data. (HT = estimates with 2018 NEISS weights; PS = poststratification weights)

Injury Type	Estimator	Estimated total	Standard Error	CV	Ratio $\frac{CV(\text{alternative})}{CV(HT)}$
Total injuries	HT	13,960,492	1,181,528	0.085	
Total injuries	Ratio	13,456,359	856,249	0.064	0.73
Total injuries	PS	12,559,612	713,636	0.057	0.60
Total injuries	GREG1	12,761,320	663,195	0.052	0.56
Total injuries	GREG2	13,518,180	785,256	0.058	0.67
Sport concussions	HT	148,894	17,380	0.117	
Sport concussions	Ratio	135,624	10,563	0.078	0.61
Sport concussions	PS	129,626	10,044	0.077	0.58
Sport concussions	GREG1	132,036	9,629	0.073	0.55
Sport concussions	GREG2	142,173	11,654	0.082	0.67
Elderly hospitalizations	HT	616,669	81,629	0.132	
Elderly hospitalizations	Ratio	560,881	60,892	0.109	0.75
Elderly hospitalizations	PS	506,174	51,140	0.101	0.63
Elderly hospitalizations	GREG1	521,309	50,075	0.096	0.61
Elderly hospitalizations	GREG2	592,008	66,867	0.113	0.82
Children	HT	4,378,963	556,340	0.127	
Children	Ratio	4,147,546	361,978	0.087	0.65
Children	PS	3,980,323	343,407	0.086	0.62
Children	GREG1	4,022,247	333,887	0.083	0.60
Children	GREG2	4,180,245	394,004	0.094	0.71
Firearms	HT	12,256	1,268	0.103	
Firearms	Ratio	12,251	1,246	0.102	0.98
Firearms	PS	11,487	957	0.083	0.76
Firearms	GREG1	11,611	948	0.082	0.75
Firearms	GREG2	11,822	1,115	0.094	0.88
Power saws	HT	76,499	6,399	0.084	
Power saws	Ratio	79,011	6,961	0.088	1.09
Power saws	PS	71,340	5,422	0.076	0.85
Power saws	GREG1	71,882	5,194	0.072	0.81
Power saws	GREG2	75,109	6,067	0.081	0.95

4.4.2 Variance Estimation

In the absence of nonresponse, a design-unbiased estimator of the variance under stratified, simple random sampling of the estimated total $\hat{t}$ with inverse-probability weights, defined in section 4.4.1, is

$$v(\hat{t}) = \sum_{h=1}^{H} \frac{N_h^2}{n_h}\left(1 - \frac{n_h}{N_h}\right) s_h^2$$
$$= \sum_{h=1}^{H}\left(1 - \frac{n_h}{N_h}\right)\frac{n_h}{n_h - 1}\sum_{i \in s_h}\left(w_{hi} y_{hi} - \frac{1}{n_h}\sum_{j \in s_h} w_{hj} y_{hj}\right)^2$$

The second line above is the form that statistical software (like SAS, Stata, SUDAAN, or R survey) will compute if the default option of a with-replacement variance estimator is requested with an *ad hoc* finite population correction factor included. When a GREG estimator is used, this variance estimator is modified to use a residual based on the regression model used in calibration. Stata and R survey include the modified variance estimation options for a GREG, while SAS does not. Properly accounting for the calibration is critical in order to reflect the effects of controlling to a set of population totals.

Use of a nonresponse adjustment and/or inclusion of auxiliary data to reduce standard errors complicates the variance estimation. The most straightforward way of handling such complications is replication variance estimation. The stratified jackknife variance estimator for a total is

$$v_J(\hat{t}) = \sum_{h=1}^{H} \frac{n_h - 1}{n_h} \sum_{i \in s_h}\left(\hat{t}_{(i)} - \hat{t}\right)^2$$

where $\hat{t}_{(i)}$ is referred to as a replicate estimate and is calculated after deleting sample hospital *hi*. As written above, the number of replicate estimates to calculate is $n = \sum_h n_h$, the full sample size. To reduce computation, it is also legitimate to create random groups of sample hospitals within each stratum or across strata, although reducing the number of replicates may be unnecessary given NEISS' small sample size.

The key to having the jackknife reflect the nonresponse and any other adjustments is treat repeat the nonresponse, calibration, or any other adjustments separately for each replicate. A replicate estimate can be written as

$$\hat{t}_{(i)} = \sum_{\substack{j \in s_h \\ j \neq i}} w_{hj(i)} y_{hj}$$

where $w_{hj(i)}$ is the weight for hospital *hj* after hospital *hi* is dropped from the sample. Note that when a hospital is dropped from a particular stratum, all sample hospitals are retained in the other strata.

However, the monthly nonresponse adjustments used in the current NEISS make replication impractical. The details are discussed in Appendix C. That appendix also describes a simpler annual nonresponse adjustment, which we recommend, instead of the monthly adjustments. The annual nonresponse adjustments could be computed separately for each variance estimation replicate as would any calibration adjustments that are used.

Analysts would be provided with a file that includes a full sample weight and a set of replicate weights for each hospital, as is done for many other government data series. The file would contain annual records only—not monthly records as now done. By specifying software that a replication estimator like the jackknife should use, the estimated standard errors will then reflect the effects of repeated sampling, nonresponse adjustment, and calibration. Analysts who use software designed to handle complex surveys should all have access to routines that will correctly calculate jackknife standard errors. For example, SAS, Stata, SUDAAN, the R survey package, and WesVar can all handle jackknife replicate weights.

4.5 Geographically-Concentrated Injuries

When injuries are most likely to occur in specific locations, the resulting estimates from NEISS can be unstable. Locations can refer to parts of the country (e.g. downhill skiing in mountainous areas in the north and west) or limited locations where an activity occurs (e.g., horseracing injuries). With a sample of 100 hospitals, NEISS is only likely to have a few hospitals whose natural catchment area encompasses such injuries. Because only those few hospitals contribute to NEISS estimates, they are subject to much greater variability than for injuries found across the country.

There are two common methods for stabilizing such estimates, modelling or combining multiple years data. Statistical models would make use of techniques commonly referred to as small area estimation (Rao and Molina, 2015), where data from other data sources are used to "borrow strength" and improve the estimates. As described in the final report of Westat's report on Small Area Estimation of Product-Related Injuries (Marker et al., 2018) there are very little data on product-related injuries available outside of NEISS that can be used for this purpose. The Agency for Healthcare Research and Quality's (AHRQ) Healthcare Cost and Utilization Project (HCUP) collects emergency room data from a larger group of hospitals, but the coding schemes aren't always consistent with NEISS and the data are not available in a consistent fashion across the country.

The best method available to the CPSC would be to combine estimates across multiple years. This will definitely stabilize estimates, but at the expense of picking up timely changes in injury patterns.

At its simplest, multiple years of NEISS could be combined to produce the current estimate. For example, data from 2016-2019 could be combined to produce the best 2019 estimate. This would contain four times as many cases from which to produce the estimates, although it would not change the number of contributing hospitals. The data reported as best estimates would be more consistent, but it would be slower to reflect real changes. As an example, if a new, but injury-prone, product became popular in, say, January 2018, this 4-year average for 2019 would only reflect this in half of its data (2018 and 2019, but not 2016 or 2017).

One potential improvement would be to use an exponentially-weighted moving average, rather than a simple equally-weighted average. This puts the most weight on the newest data, with weight decreasing for each past year. For example, a weight of 0.5 could be applied to the most recent year, 0.25 ($=.5^2$) to the preceding year, 0.125 ($=.5^3$) to the year before that, and so on. In the example above, 75 percent of the weight would be given to data since the introduction of the new product rather than only 50 percent.

The general formula for an exponentially-weighted moving average estimator for year *t* would be

$$\hat{y}_t = \alpha * y_t + \alpha^2 * y_{(t-1)} + \alpha^3 * y_{(t-2)} + \alpha^4 * y_{(t-3)} + \ldots$$

where a = the weighting factor (0.5 in the above example)

y_t = the injury total collected by NEISS for year *t*.

5. Suppression Criteria for Inaccurate Estimates

Both the CPSC and the CDC regularly make national estimates from the NEISS. It would be appropriate if both organizations used the same rules for determining when estimates are sufficiently accurate to be published.

To guide the CPSC and the CDC in their decision making, we have collated the rules used by a wide range of U.S. Federal agencies and foreign statistical offices. Table 5.1 shows the rules used by:

- CPSC
- CDC
- Other agencies within the U.S. Department of Health and Human Services (Agency for Healthcare Research and Quality (AHRQ), Substance Abuse and Mental Health Services Administration (SAMHSA), Centers for Medicare and Medicaid Services (CMS), and the National Cancer Institute (NCI))
- U.S. Census Bureau
- U.S. Department of Education's National Center for Education Statistics (NCES)
- U.S. Department of Agriculture's National Agricultural Statistical Service (NASS)
- U.S. Energy Information Administration (EIA)
- Statistics Canada
- Israeli Central Bureau of Statistics
- British Office of National Statistics
- Australian Bureau of Statistics
- Statistics New Zealand

Many organizations also suppress data to avoid disclosing individual data. We did not collect those rules for two reasons. First, there is a great deal of flux at the moment as agencies determine how to protect not only from disclosing individuals directly from their tables, but also through intruders potentially combining their data with other publically available data. Second, NEISS data provide unique disclosure control concerns, because at times rare injuries happen to high profile people, providing unique opportunities for disclosure.

We focus here on decisions to suppress data because of their statistical precision. These typically take the form of rules on how big is the CV (also referred to as the relative standard error (RSE)), minimum sample sizes, and response rates.

5.1 General Patterns

CV/RSE. Most U.S. agencies either don't publish estimates with a CV > 30%, or indicate that those between 30% and 50% should be used with caution while refusing to publish those with a CV > 50%. International offices often have tighter restrictions, but many do not have firm rules applied consistently across all data products. Israel puts a warning on estimates with CVs between 15% and 30%, while the British do not publish with a CV > 20%. Australia puts warnings on totals with 10% < CV < 50% while New Zealand suppresses agricultural or economic estimates with CV > 30% (experimental series can be produced up to 50%). The Swedish Central Statistical Bureau does not have rules (thus not in Table 5.1), but in general they warn against using data with a CV > 50%.

Three agencies have special rules related to the width of a confidence interval. These rules are meant to avoid problems with very small proportional estimates. The National Center for Health Statistics (NCHS) suppresses estimates with a confidence interval width of 0.30; if the width is between 0.05 and 0.30, they suppress if the relative width exceeds 130% (corresponding to a CV (on p or (1-p)) of 33%). The CDC's National Immunization Survey (NIS) attaches a warning if a state's confidence interval half width is greater than 10%. Australia requires the standard error to be no bigger than 5.1% (so that a confidence interval on a percentage is plus or minus 10%) and suppresses any percentage for which a confidence interval would go below 0% or above 100%.

Sample size. The requirement of a minimum of 20 sampled cases that both the CPSC and the CDC use for NEISS is similar to the minimum of 16 used by NCI for cancer surveillance. The Swedish Central Statistical Bureau provides general guidance against using estimates with less than 20 respondents. A more common minimum cell size of 30 is used by CDC at NCHS and for the NIS. The U.S. Department for Health and Human Services (HHS) requires a minimum of 50 for the

Table 5-1 Suppression rules by agency

				Statistically Unreliable				
Agency	**Center**	**Data System**	**Example Statistic**	**RSE**	**Sample Size**	**Response Rate**	**Other**	**References**
CPSC		NEISS		CV≥33%	n<20		National estimate less than 1,200	
CDC	NCIPC	Web-based Injury Statistics Query and Reporting System (WISQARS)	Firearm injuries from NEISS	CV≥30%	n<20		National estimate less than 1,200	
CDC	NIOSH	NEISS-Work	Work-related injuries	CV>50% suppressed. 30%<CV<50% are published with note that "data are statistically unreliable"	Minimum weighted and unweighted counts. Also minimum number of hospitals.			The required minimum sizes are not publicly reported.
CDC	NCHS			CI Width>0.30; if 0.05<CI Width<0.30 and relative width>130% (corresponds to approximately RSE>33%)	n<30		If Degrees of Freedom <8 estimates must be reviewed (not applied to Vital Statistics). Estimates of 0 must also be reviewed.	Parker JD, Talih M, Malec DJ, et al. National Center for Health Statistics Data Presentation Standards for Proportions. National Center for Health Statistics. Vital Health Stat 2(175). 2017. https://www.cdc.gov/nchs/data/series/sr_02/sr02_175.pdf
CDC	NCHS	Vital Statistics		CI Width>0.30; if 0.05<CI Width<0.30 and relative width>130% (corresponds to approximately RSE>33%)	n<20 in numerator for rates		Less than 85% of a states births(death) in a 12-month period available. Less than 80% of records for the variable are complete.	Technical Notes for Hamilton BE, Martin JA, Osterman MJK, Rossen LM. Births: Provisional data for 2018. Vital Statistics Rapid Release; no 7. Hyattsville, MD: National Center for Health Statistics. May 2019. https://www.cdc.gov/nchs/data/vsrr/vsrr-007-508.pdf
CDC	NCCDPHP	Youth Risk Behavior Surveillance System (YRBSS)			n<100			Methodology of the YRBSS - 2013 Centers for Disease Control and Prevention. MMWR 2013. https://www.cdc.gov/mmwr/pdf/rr/rr6201.pdf
CDC	NCIRD	National Immunization Survey (NIS)		RSE>30%	n<30		State estimates with CI half width >10% are marked with a footnote	Technical Notes for NIS Surveillance Tables. National Center for Immunization and Respiratory Diseases, 2016. https://www.cdc.gov/vaccines/imz-managers/coverage/nis/child/tech-notes.html
HHS	AHRQ	Medical Expenditure Panel Survey (MEPS)		RSE>30%	n<100			2018 NATIONAL HEALTHCARE QUALITY AND DISPARITIES REPORT 2018 National Healthcare Quality and Disparities Report Detailed Methods for the Medical Expenditure Panel Survey. AHRQ Publication No. 19-0070-1-EF September 2019. https://www.ahrq.gov/sites/default/files/wysiwyg/research/findings/nhqrdr/2018qdr-mepsmethods.pdf
HHS	SAMHSA	National Survey on Drug Abuse and Health (NSDUH)		RSE>50% for year at first use	n<100 (or effective sample size <68)		RSE(p)/(-ln(p)) >17.5%, if p≤0.5; RSE(1-p)/(-ln(1-p)) >17.5%, if p>0.5; p<0.00005 or p>0.99995	"The rule prior to 1991 [RSE>50%] imposed a very stringent application for suppressing estimates when is small but imposed a very lax application for large ." Substance Abuse and Mental Health Services Administration. (2019). 2018 National Survey on Drug Use and Health: Methodological summary and definitions. Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Agency	Center	Data System	Example Statistic	Statistically Unreliable				References
				RSE	Sample Size	Response Rate	Other	
HHS	CMS	Medicare Current Beneficiary Survey (MCBS)		Proportions: CI Width>0.30; if 0.05<CI Width<0.30 and relative width>130% (corresponds to approximately RSE>33%). Other estimates RSE>30%	n<50			Technical notes for MCBS 2016 Chartbook. https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Data-Tables-Items/2016Chartbook
HHS	NIH/NCI	Surveillance, Epidemiology, and End Results Program (SEER)			n<16			Assumes cancer incidence is Poisson distributed, n<16 is equivalent to a RSE<25%
Commerce	Census Bureau	Demographic surveys		Serious data quality issues related to sampling error occur when the CV>30%. 5-year average tables suppress cells if Median<MOE (CV>.61). Otherwise an entire table is suppressed if the median CV>61%	Medians with n<3.	Unit response rates for surveys or censuses, or cumulative unit response rates for panel or longitudinal surveys, are below 60%. Item response rates or total quantity response rates on key items are below 70%. Coverage ratios for population groups associated with key estimates are below 70%. Combined rates for key estimates (e.g., computed as unit response item response coverage) are below 50%	This requirement does not apply to secondary estimates. For example, if the estimated month-to-month change is the key estimate, and the monthly estimates are secondary, the requirement applies only to the estimated month-to-month change. Thresholds are provided because bias is often associated with low response rates or with low coverage ratios. If nonresponse bias analyses or other studies show that the bias associated with nonresponse is at an acceptable level, or that steps taken to mitigate nonresponse bias or coverage error are effective, these thresholds do not apply.	Data that exceed these cut offs can be published with approval and identification of cells that do not meet the standards. 2013 Quality Standards (to be updated in 2021) https://www.census.gov/about/policies/quality/standards.html . Nonresponse bias analyses must be conducted when unit, item, or total quantity response rates for the total sample or important subpopulations fall below the following thresholds. The threshold for unit response rates is 80 percent. The threshold for item response rates of key items is 70 percent. The threshold for total quantity response rates is 70 percent. (Thresholds 1 and 2 do not apply for surveys that use total quantity response rates.)
	Census Bureau	Economic Surveys		CV>30%	Surveys that select the largest units in the target universe and do not attempt to collect data from the small units in the universe. These surveys have serious data quality issues if the responding units comprise <70% of the target universe, based on the unit response rate or the total quantity response rate, as appropriate.	Total quantity response rate TQRR<50%. If 50%<TQRR<70% reported with a warning.		

				Statistically Unreliable				
Agency	Center	Data System	Example Statistic	RSE	Sample Size	Response Rate	Other	References
Education	NCES			30%<CV<50% publish with warning; CV>50% suppress		RR<50% needs approval from Senior Staff; RR<90% for universe surveys		
USDA	NASS	Local Food Marketing Practice Survey		Lowest level of table (e.g. state) must have majority of CV<30%. 15%<CV<30% are coded yellow. CV>30% "Caution should be used when using this estimate in any form"				
USDA	NASS		County-level estimates		n<30; unless (n≥3 and ≥25% coverage) or (n≥10 and ≥10% coverage)			
Energy Information Administration (EIA)		Energy Consumption Surveys and Monthly Electricity Surveys		CV>50%			Energy Consumption surveys if <10 households; or < 20 buildings	
Energy Information Administration (EIA)		Weekly Motor Gasoline Survey; Heating Oil and Propane Survey		CV>5%				These are price estimates (e.g., average price per gallon of gasoline) that are ratios of two correlated estimated totals.
Energy Information Administration (EIA)		Weekly Petroleum and Natural Gas Surveys				Volume weighted <90% (proposed rule will change this to <80%)		
Statistics Canada				CV>30% or 35%, depending on the program			A working group is developing recommendations, including for proportions near 0% or 100%. Minimum levels of geography required.	Some programs use a letter grade for cell CV, and don't publish above a certain CV; goal of Quality Secretariat is to make this more standardized.
Israeli Central Bureau of Statistics			Proportions	15%<CV≤30% publish with warning; CV>30% not published	Complex rules governing effective sample size of both numerator and denominator			2015 memo communicated by Dr. Danny Pfeffermann, Director of the Israeli CBS
Israeli Central Bureau of Statistics			Totals	15%<CV≤30% publish with warning; CV>30% not published	n≤11 not published; 12≤n≤39 published with warning if CV≤30%			2015 memo communicated by Dr. Danny Pfeffermann, Director of the Israeli CBS
British Office of National Statistics		Labour Force Survey (LFS) and Annual Population Surveys (APS)		Parliamentary questions on the LFS do not publish if CV>20%	Public Policy Analysis Division (PPAD) Policy Evidence & Analysis Team publishs City estimates with shading if 2<n<30. Others shade up to 25 or 27.		LFS do not publish if weighted n<4,500. PPAD publishes with shading if 4,500<weighted n<6,000	https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/measuringandreportingreliabilityoflabourforcesurveyandannualpopulationsurveyestimates

				Statistically Unreliable				
Agency	Center	Data System	Example Statistic	RSE	Sample Size	Response Rate	Other	References
British Office of National Statistics		Annual Survey of Hours and Earnings (ASHE)		ASHE publishes with warnings. It shades any cells with CV>5%, with CV>20% results are "considered unreliable for practical purposes"				Decisions on other surveys are made separately for each output and should reflect the known and likely uses. The Crime Survey for England and Wales has a lower limit on the number of incidents with rare crime types. The ASHE uses coefficient of variation directly, including shading cells in the table according to precision band and suppressing those with large CVs. Summary document at https://gss.civilservice.gov.uk/policy-store/privacy-and-data-confidentiality-methods-a-national-statisticians-quality-review-nsqr/
Australian Bureau of Statistics (ABS)			Totals (e.g., counts and value)	10%<RSE<25% ^; 25%<RSE<50% *; RSE>50% **				These estimates should be used with caution as they are subject to sampling variability too high for some purposes, most practical purposes, or too unreliaible for general use, respectively.
Australian Bureau of Statistics (ABS)			Proportions	MOE>10% (Std err >5.1%); or Estimate < MOE; or Est+MOE>1			For derived statistics, e.g. the ratio of two estimates, the annotation is based on the standard error of the derived statistic.	Annotated with #, labelled "Proportion has a high margin of error and should be used with caution"
Stats NZ		Agricultural Production Survey	Livestock, farming, horticulture and forestry statistics	RSE>30% are suppressed				
Stats NZ		Annual Enterprise Survey (AES)	Financial statistics covering the entire economy by industry and sector groups	RSE>30% are suppressed			Where actual values are suppressed, seasonally adjusted and trend values are also suppressed.	
Stats NZ		AES - Business Performance Benchmark	Experimental series that give lower level financial information than AES	RSE>50% are suppressed				

Medical Expenditure Panel Survey, while AHRQ and SAMHSA both require 100. Israel and Britain both publish with less than 30 sampled cases but use warning messages.

Agencies that survey highly skewed populations often have rules that tie together number of sampled cases and percent of the population covered by those cases. For example, while the NASS requires a sample size of 30, it will also publish with sample sizes as small as 10 if they cover 10% of the farms/product, or even 3 cases, if they cover at least 25% of farms/product.

Response rates. Response rates are not used to suppress in most agencies. For economic surveys the Census Bureau will not publish if less than half of the total volume is represented among the respondents and they publish with a warning if between 50% and 70% is included in the sample. For demographic surveys they require unit response rates of 60%, item response rates of 70%, and the combined rate of at least 50%. NCES requires senior leadership approval to publish if the response rate is less than 50%.

Other. Like the CPSC and the CDC, the British do not publish cells with less than a pre-set weighted number of respondents. Multiple agencies will publish trend and ratio estimates that meet these standards, even if some of the underlying statistics need to be suppressed.

References

An, Tony (2020). A SAS Macro for Calibration of Survey Weights. SAS Global Forum, Paper 4284-2020.

Beaumont, J-F. & Rivest, L-P. (2009). Dealing with Outliers in Survey Data. Chapter 11 in *Sample Surveys: Design, Methods and Applications*, Vol. 29A, *Handbook of Statistics* 29. Eds. C.R. Rao and D. Pfeffermann. Amsterdam: Elsevier.

Deville J. C., Särndal C-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87(418), 376–382.

Estevao, V., Hidiroglou, M.A., and Särndal C-E. (1995). Methodological Principles for a Generalized Estimation System at Statistics Canada. *Journal of Official Statistics*, 11(2), 181-204.

European Union (2014). Method: Generalised Regression Estimator. *Memobust Handbook on Methodology of Modern Business Statistics.* EUROSTAT, https://ec.europa.eu/eurostat/cros/content/weighting-and-estimation-generalised-regression-estimator-pdf-file_en.

Lynn, P. (2004). The use of substitution in surveys. *The Survey Statistician*. 49, 14-16.

Little, R. J. A. (1988). Missing-Data Adjustments in Large Surveys (with Discussion). *Journal of Business Economics and Statistics* 6(3), 287–301.

Mance, S. (2016). How hard is it to count? Reconciling monthly survey and administrative data: experiences from the CES program, *International Conference on Establishment Surveys*, Geneva, Switzerland, November 2016.

Marker, D.A., Shlomo, N., Baskin, R., and Campione, J. (2018). Small Area Estimation of Consumer Product-Related Emergency Department-Treated Injuries. Submitted to the U.S. Consumer Product Safety Commission.

Piesse, A.R., and Green, J.L. (2004). Calibrated imputation for the Drug Abuse Warning Network (DAWN). Proceedings of the Section on Survey Research Methods, American Statistical Association.

Rao, J.N.K. and Molina, I. (2015). *Small Area Estimation*, 2nd Edition. New York: John Wiley & Sons. DOI:10.1002/9781118735855

Rubin, D. B., and Zanutto, E. (2002). Using matched substitute to adjust for nonignorable non-response through multiple imputation. In *Survey Nonresponse*, R. Groves, R. J. A. Little, and J. Eltinge (Eds.), 389-402. New York: John Wiley.

Schroeder, T. (2019). *Updated 2018 NEISS weights based on the 2017 AHA Health Forum Frame.* Division of Hazard and Injury Data Systems, U.S. Consumer Product Safety Commission.

Valliant, R., Dever, J.A., and Kreuter, F. (2018). *Practical Tools for Designing and Weighting Survey Samples*, 2nd edition, New York: Springer.

Vehovar, V. (1999). Field substitution and unit nonresponse, *Journal of Official Statistics*, *15*(2), 335-350.

Appendix A

Budget Constrained Average Relvariance Approach

The estimated population total for a particular injury *v* in a stratified simple random sample is

$$\hat{t}_v = \sum_{h=1}^{H} N_h \bar{y}_{vh}$$

where *h* is a stratum, *H* is the total number of strata, N_h is the number of hospitals in the population in stratum *h*, and $\bar{y}_{vh}$ is the mean number of injuries of type *v* in the sample for stratum *h*. The relvariance of the estimated total for injury *v* is

$$\text{relvar}\left(\hat{t}_v\right) = \frac{1}{t_v^2} \sum_{h=1}^{H} \frac{N_h^2}{n_h}\left(1 - \frac{n_h}{N_h}\right) S_{vh}^2 = \frac{1}{Y_v^2} \sum_{h=1}^{H} N_h \left(\frac{N_h}{n_h} - 1\right) S_{vh}^2 \qquad (1)$$

where t_v is the population total for injury *v* and S_{vh}^2 is the unit (or population) variance of injury *v* among hospitals in stratum *h*. Since neither of these population values is known, we used estimates in our calculations. The average of the relvariances across $v = 1, \ldots, V$ injury types can be written as

$$\begin{aligned}
\Phi &= \frac{1}{V} \sum_{v=1}^{V} \text{relvar}\left(\hat{t}_v\right) \\
&= \frac{1}{V} \frac{1}{t_v^2} \sum_{v=1}^{V} \sum_{h=1}^{H} N_h \left(\frac{N_h}{n_h} - 1\right) S_{vh}^2 \\
&= \sum_{h=1}^{H} N_h \left(\frac{N_h}{n_h} - 1\right) \left(\frac{1}{V} \sum_{v=1}^{V} \frac{S_{vh}^2}{t_v^2}\right) \\
&\equiv \sum_{h=1}^{H} N_h \left(\frac{N_h}{n_h} - 1\right) \bar{V}_h^2
\end{aligned}$$

where $\bar{V}_h^2 = \frac{1}{V} \sum_{v=1}^{V} \frac{S_{vh}^2}{t_v^2}$ is the average stratum *h* value of S_{vh}^2 / t_v^2 across the *V* injuries. (Note that $\bar{V}_h^2$ is convenient for computations but is not the standard definition of a population relvariance.)

One way of determining an allocation to the strata is to minimize the average relvariance Φ subject to a budget constraint while requiring at least a minimum number of sample hospitals assigned to each stratum. Formally, the problem that we solved was:

Find $\{n_1, n_2, \ldots, n_H\}$ to minimize Φ subject to these constraints

$$\min\{n_h\}_{h=1}^{H} = 8$$

$$C = \sum_{h=1}^{H} n_h c_h = \$2{,}000{,}000$$

This problem was solved for Option 2 in Table 3-1. For Option 1 the additional constraint was added that $\sum_{h=1}^{H} n_h = 100$.

Sample Size Constrained Average Neyman Allocation Approach

The Neyman allocation minimizes the variance of a single estimate based on a fixed overall sample size and does not take into account costs. The Neyman allocation formula we used to calculate the allocated sample size within each stratum for each specific injury was[7]:

$$n_{vh} = n \frac{N_h S_{vh}}{\sum_{h=1}^{H} N_h S_{vh}}$$

where *h* is a stratum, *H* is the total number of strata, N_h is the number of hospitals in the population in stratum *h*, S_{vh} is the population unit standard deviation of injuries of type *v* in the sample for stratum *h*, and n_{vh} is the sample size allocated within stratum *h* for injury *v*.

[7] For Neyman allocation, most textbooks use W_h - the ratio of the population size in stratum h to the total population size. In this instance, W_h and N_h can be used interchangeably, given the constant N. We have used N_h for ease of explanation and understanding.

The average Neyman allocation per stratum across the six key estimates was then calculated as:

$$\bar{n}_h = \frac{\sum_{v=1}^{V} n_{vh}}{\mathrm{V}}$$

The relvariances for Option 3 in Section 3.3.1 were calculated using equation (1) above, with $\bar{n}_h$ replacing n_{vh}.

Appendix B

Discontinuity Analyses

The CPSC provided us with some individual examples of replacement hospitals whose reported firearms and amusement park injuries were substantially higher than for the hospitals that they were replacing. Although a hospital-to-hospital comparison is useful for understanding the problem, a decision on how to handle any discontinuities should also consider the effects of replacement at the full population estimation level. (We were not able to do this in the scope of this contract.) If estimators are not extremely affected by a replacement, then no action may be necessary. If a jump in a time series occurs, then the question becomes whether the series with the original or the series with the replacement produces estimates closer to the truth. Thus, any analyses of discontinuities in a NEISS time series should consider two questions:

1. Is a replacement hospital an unrepresentative outlier compared to other sample hospitals?
2. What is the best estimator to use given the answer to (1)?

If a replacement hospital is not considered an outlier, then an estimated total using the sample with the replacement hospital's reports does not need an adjustment. However, to avoid a difficult-to-explain jump in a series, smoothing out the size of the jump may be desirable. Wedging, described in section B.1, will do this. If a replacement hospital is an outlier, approaches for handing the problem are to modify the number of injury reports (*y*-values) (see section B.3) or reduce the hospital's weight (section B.4). Both will reduce the estimated totals.

B.1 Wedging

If a decision is made that either the new series is at an unacceptable level or using the special procedures described in section B.3 would be too disruptive, then making a gradual transition from the old to the new series can be done. This method, referred to in this report as wedging, is used by the US Bureau of Labor Statistics (BLS) to taper-in a jump in an employment series due to use of administrative data to benchmark survey results (Mance 2016). The BLS publishes many monthly series of employment estimates for different industry sectors based on the Current Employment Statistics (CES) sample of establishments. Each year in March, the series are adjusted (benchmarked) to agree with universe counts from the Quarterly Census of Employment and Wages (QCEW).

Revisions for the previous year's estimates are made to wedge-in the difference between the CES March estimate and the QCEW benchmark value.

To describe a similar procedure for NEISS, we use the following notation:

$\hat{t}_u$	=	estimator of total at time *u* based on the NEISS sample at time *u*, including any substitutes
$\Delta_{u,u-1} = \hat{t}_u - \hat{t}_{u-1}$	=	difference in estimated totals for times *u* and *u*-1

Suppose at time *u* a substitute comes into the sample that reports one or more products at a substantially different level from the hospital it is replacing. Assume that $\Delta_{u,u-1}$ is large and positive (or large and negative) compared to previous changes causing a jump in the series. The general idea below is to phase-in the difference over *p* time periods.

In period *u* compute the published estimate as the *t*-1 estimate plus 1/*p* of the change between the *u* and *u*-1 estimates. At time *u*+1 add 2/*p* of $\Delta_{u,u-1}$ to the adjusted time *u* estimate. Keep adding in a larger portion of $\Delta_{u,u-1}$ until, at time *t* + *p*, the estimate based on the sample including the substitute is used without any adjustment. In formulas this procedure is:

$$\hat{t}_u^* = \hat{t}_{u-1} + \frac{1}{p}\Delta_{u,u-1}$$

$$\hat{t}_{u+1}^* = \hat{t}_u^* + \frac{2}{p}\Delta_{u,u-1}$$

$$\ldots$$

$$\hat{t}_{u+p-1}^* = \hat{t}_{u+p-2}^* + \Delta_{u,u-1}$$

$$\hat{t}_{u+p}^* = \hat{t}_{t+p}$$

For example, if *p*=2 and *u*=2, compute the difference $\Delta_{2,1}$. Estimate total injuries at times 2, 3, and 4 as

$$\hat{t}_2^* = \hat{t}_1 + \frac{1}{2}\Delta_{2,1},\ \hat{t}_3^* = \hat{t}_2^* + \Delta_{2,1},\ \text{and}\ \hat{t}_4^* = \hat{t}_4.$$

As described above, the procedure would be applied going forward from some year. However, it could also be applied going backward in time if the CPSC were willing to revise previous estimates. Revising earlier estimates is what the BLS does for the CES series.

B.2 Poststratified and Other Calibrated Estimators

Poststratification may influence the level of an estimator. Suppose that the sample is selected at time 0 and estimates are made for a later year *u*. The poststrata are defined in the same way as the design strata. Assume that a hospital was in design stratum h_0 when the sample was selected at time 0 and is in stratum h_u when estimates are to be made. Suppose that the average weight for hospital *k* in design stratum h_0 across the months when it responds during a year is $\overline{w}_{h_0 k}$. A subscript is used on the weight to emphasize that the hospital was selected as part of stratum h_0.

The poststratified weight for hospital *k* that was in design stratum h_0 and is in stratum h_u when estimates are made is

$$w_{h_u k}^{PS} = \overline{w}_{h_0 k} \frac{N_{h_u}}{\hat{N}_{h_u}}$$

where $\hat{N}_{h_u} = \frac{1}{12} \sum_{m=1}^{12} \hat{N}_{h_u m}$ is an estimator of the number of hospitals in stratum h_0.

The month *m* estimator is $\hat{N}_{h_u m} = \sum_{k \in s_{h_u m}^R} w_{km}$ where w_{km} is the nonresponse-adjusted weight; $s_{h_u m}^R$ is the set of responding sample hospitals that are in stratum *h* at time *u*. Note that this accounts for migrations between strata. Each $\hat{N}_{h_u m}$ estimates the population count of hospitals N_{h_u} in stratum h_u. For these calculations, assume that N_{h_u} is taken to be the end of year count.

Poststratification has the effect of modifying the weight that a hospital was selected with. For example, suppose a substitute hospital replaces one that was selected in the M stratum, but the substitute has migrated to the Large stratum at time *t*. If $\frac{N_{h_u}}{\hat{N}_{h_u}} < 1$ because the sample from stratum

h_u has effectively increased due to the migration of the substitute, then the substitute is down-weighted compared to the original that it replaced. Such down-weighting may decrease the amount of a discontinuity in a time series.

However, there is no guarantee that this will occur. The data will either confirm or refute whether poststratification has this effect. In fact, poststratification reduces variances, as illustrated in section 4.4. Consequently, discontinuities in a series may be more likely to be statistically significant if a poststratified estimator is used. However, if NEISS switches to a more efficient estimator, like poststratification, any method for handling discontinuities should be evaluated using the new estimator.

If poststratification helps reduce discontinuities, the more elaborate calibration included in section 4.4 should also. Poststratification is simpler and should be tried first. If it does not reduce jumps in a time series, then it is unlikely that the more elaborate calibration will either.

B.3 Winsorized Estimator

Discontinuities in time series of injury estimates can be caused when a replacement reports injuries at a much different level than the original hospital. Section 4.2 gave two examples of this. A replacement may be considered an outlier if its reports for one or more injuries are much different than those for other sample hospitals. Beaumont & Rivest (2009) review much of the literature on outlier treatment. A Winsorized estimator of an injury total is one example and is constructed by truncating unusually large injury reports. First, define a residual as

$$y_k - \hat{y}_k$$

with $\hat{y}_k$ being a predicted value for hospital *k*. The stratum that *k* is in will be implied by the set of sample hospitals used in summations below. How the residuals are computed will depend on the form of estimated totals:

- If the expansion estimator that uses inverse probability weights is used, $e_k = y_k - \hat{\bar{y}}_{h_0}$ with y_k being the total for a year for hospital *k* and $\hat{\bar{y}}_{h_0} = \sum_{k \in s_{h_0}} \bar{w}_k y_k \Big/ \sum_{k \in s_{h_0}} \bar{w}_k$ where s_{h_0} is the set of sample hospitals assigned to stratum h_0 at the time of sample selection (time 0) and $\bar{w}_k$ is the average weight for hospital *k* in the months that it responded during a year.

- If a poststratified estimator is used, the implied model is that all hospitals in a poststratum have the same mean. A poststratum is the stratum that a hospital is a member of at the time *u* of estimation. In this case, $e_k = y_k - \hat{\bar{y}}_{h_u}$ with $\hat{\bar{y}}_{h_u} = \sum_{k \in s_{h_u}} w_k^{PS} y_k \Big/ \sum_{k \in s_{h_u}} w_k^{PS}$ where s_{h_u} is the set of sample hospitals classified in a stratum at time *u* and w_k^{PS} is the poststratified weight defined in section B.2.

The idea behind Winsorization is to truncate *y* values if the residual associated with the hospital is unusually large (or small). That is, a hospital with an abnormally large (or small) value compared to other sample hospitals is deemed an outlier and has its *y* value adjusted. To account for spread among *y* values, each residual is standardized as shown below. The estimator of total at time *u* if the NEISS nonresponse-adjusted, expansion estimator is used is

$$\hat{t}_{\psi u} = \sum_h \sum_{k \in s_h} \bar{w}_k y_{ku}^*$$

$$\text{where } y_{ku}^* = \begin{cases} y_{ku} & k \in s_{h_0}^{no.out} \\ \hat{\bar{y}}_{h_0} + c\sigma_k \operatorname{sgn}(e_k) & k \in s_{h_0}^{out} \end{cases}$$

c	=	3 is a standard choice, although *c* is adjustable for different amounts of trimming;
$s_{h_0}^{out}$	=	set of sample hospitals that are outliers in stratum h_0; i.e., ones for which $\lvert e_k \rvert / \sigma_k \geq c$
$s_{h_0}^{no.out}$	=	set of sample hospitals that are not outliers in stratum h_0; i.e., ones for which $\lvert e_k \rvert / \sigma_k < c$

Thus, a standardized residual is considered large if it is bigger than 3 in absolute value. If number of injuries reported by hospitals were normally distributed, then 99% of the e_k / σ_k would be less than 3 in absolute value. Although the distribution of injuries is probably skewed and non-normal, this procedure should still truncate a small percentage of reports.

Assuming that only large positive injury reports are truncated, $y_{ku}^{*} = \hat{\bar{y}}_{h_0} + c\sigma_k$ for $k \in s_{h_0}^{out}$. The estimator of σ_k with the expansion estimator is

$$\hat{\sigma}_k = \sqrt{\sum_{k \in s_{h_0}} \bar{w}_k \left(y_k - \hat{\bar{y}}_{h_0} \right)^2 \Big/ \sum_{k \in s_{h_0}} \bar{w}_k}\ ,\ k \in s_{h_0}$$

If the poststratified estimator is used, the formulas above are the same but substitute h_u for h_0 and w_k^{PS} for w_k. Extensions of this idea are possible if the more elaborate GREG estimators in section 4.4 are used.

One of the practical advantages of a Winsorized estimator is that it selectively adjusts the injury reports. For example, if firearms injuries are extremely large in a hospital but power saw injuries are not, reports for the former may be truncated but report for the latter will not be. Editing software can be written to apply the rule above, and a public use file released that contains the Winsorized values. Weight calculation can proceed independent of this editing process.

B.4 Weight Trimmed Estimator

Weight trimming would be effective only for certain types of injuries for which some hospitals have outlying values. For injury estimates that are not affected by outliers, the full weight should be used in order to produce unbiased estimates. That is, a given hospital would have its weight trimmed for some estimates but not for others, leading to the necessity of having multiple weights for some hospitals. This seems too impractical to be used in NEISS, especially considering its policy of releasing public use files.

Appendix C

NEISS Variance Estimation

The NEISS is a stratified, equal probability sample. Product-related and other injuries are collected monthly from the sample hospitals. Because the set of responding hospitals can vary somewhat from month-to-month, nonresponse adjustments to weights have been made separately by month. An additional ERV ratio adjustment is made which does not vary by month. For the monthly nonresponse adjustments, the original design strata are used. For the ERV adjustment, three groups of sample design strata are formed to make the adjustments: Small + Medium, Large + Very Large, and Children's (Schroeder 2019). An annual estimate of a total number of injuries is the aggregation of the nonresponse adjusted, ERV ratio adjusted, weighted, monthly estimates. An annual estimated injury total is computed as

$$\hat{t} = \sum_{h_0} \sum_{k \in s_{h_0 m}^R} \sum_{m=1}^{12} w_{km} y_{km} = \sum_{h_0} \sum_{k \in s_{h_0 m}^R} \sum_{m=1}^{12} w_k a_{km} r_k y_{km}$$

where

h_0	=	stratum identifier at the time the sample was designed;
$s_{h_0 m}^R$	=	set of responding sample hospitals in design stratum h_0 in month *m*;
w_k	=	N_{h_0}/n_{h_0} is the base weight for hospital *k* in stratum h_0 with N_{h_0} being the frame count in h_0 when the sample was initially selected; n_{h_0} is the number of initial sample selections from h_0;
w_{km}	=	$w_k a_{km} r_k$ is adjusted weight for hospital *k* in month *m*;
a_{km}	=	$\frac{n_{h_0 m}}{n_{h_0 m}^R}$ nonresponse adjustment for responding hospital *k* in stratum h_0 in month *m*; if the hospital does not respond, then $a_{km} = 0$.
$n_{h_0 m}$	=	number of in-scope sample hospitals in month *m* initially selected in stratum h_0;

$n_{h_0 m}^R$	=	number of responding sample hospitals in month *m* in stratum h_0 ;
r_k	=	$\frac{V_{h_0^*}}{\hat{V}_{h_0^*}}$ is an annual adjustment to the updated frame total of ERVs;
		h_0^* is a group of the design strata;
		$V_{h_0^*}$ is the frame ERV total for in-scope hospitals in the current year in h_0^* and $\hat{V}_{h_0^*}$ is the estimate of that total based on hospitals that responded at least once during the year. Base weights are used for this.
y_{km}	=	reported injuries for month *m* by hospital *k*.

The fact that a_{km} can vary from month to month complicates variance estimation if an estimator is desired that accounts for that variation. We first provide a variance estimator in section C.1 that would be easy to implement for users of NEISS data but does not estimate correct standard errors (SEs). Section C.2 discusses why it is difficult to reflect all sources of variance if the current, monthly nonresponse adjustments are used. In section C.3, an annual nonresponse adjustment is presented that will simplify variance estimation. We also sketch the steps that would be needed for a variance estimator that captures the nonresponse adjustment and ERV ratio adjustment in section C.3.

C.1 Variance Estimates that Ignore Weight Adjustments

This section presents a variance estimator that does not account for the nonresponse and ERV adjustments but would be easy for data users to implement. Define

$$\delta_{km} = \begin{cases} 1 & \text{if hospital } k \text{ reponds in month } m \\ 0 & \text{if not} \end{cases}$$

δ_{km} is a random variable, assuming that response is treated as a random process. In particular, suppose R denotes the response distribution and that the response indicators are independent and Bernoulli with

$$E_R\left(\delta_{km}\right)=\rho_k \,,\quad V_R\left(\delta_{km}\right)=\rho_k\left(1-\rho_k\right).$$

The estimated total can then be written as a sum over the full sample s_{h_0} in each stratum:

$$\hat{t}=\sum_{h_0}\sum_{k\in s_{h_0}} w_k \underbrace{\sum_{m=1}^{12}\delta_{km}a_{km}r_k y_{km}}_{\tilde{y}_k} . \tag{C.1}$$

The term $\tilde{y}_k$ is a nonresponse- and ERV-adjusted, annual number of injuries for hospital *k* in the sample from stratum h_0.

If we ignore the fact that $\tilde{y}_k$ contains the random quantities δ_{km} and r_k, and treat w_k as if it is a base weight, the variance in a stratified simple random sample (*stsrs*) selected without replacement (conditional on the sample hospitals that respond in each of the 12 months) is

$$\begin{aligned} V_\pi\left(\hat{t}\,\middle|\, s_m^R, m=1,\ldots,12\right) &= \sum_{h_0}\frac{N_{h_0}^2}{n_{h_0}}\left(1-\frac{n_{h_0}}{N_{h_0}}\right)S_{h_0}^2 \\ &= \sum_{h_0} n_{h_0} w_{h_0}\left(w_{h_0}-1\right)S_{h_0}^2 \end{aligned} \tag{C.2}$$

where $w_k = w_{h_0} = N_{h_0}/n_{h_0}$ is the base weight for all $k \in s_{h_0}$, s_m^R is the set of all sample hospitals across all strata in month *m*, and $S_{h_0}^2=\frac{1}{N_{h_0}-1}\sum_{k=1}^{N_{h_0}}\left(\tilde{y}_k-\frac{1}{N_{h_0}}\sum_{k'=1}^{N_{h_0}}\tilde{y}_{k'}\right)^2$. The subscript π in (C.2) means that the variance is taken with respect to the *stsrs* distribution. Thus, the variance ignoring the effect of the nonresponse and ERV adjustment is obtained via the standard *stsrs* formula.

If the CPSC wanted to make it easy for users to calculate this standard error in software like SAS, Stata, and the R survey package based on (C.2), it could provide a public use file with the base weight w_k for each hospital, the design-stratum identifier, and the adjusted, annual injury totals, $\tilde{y}_k$, for each hospital/product. Users would then be advised to specify the sample design as *stsrs*. The CPSC would also instruct users that the sampling fractions, n_{h_0}/N_{h_0}, are equal to the inverse of the base weights. for each design stratum. The stratum variances in (C.2) would be estimated by using

$$\hat{S}^2_{h_0} = \frac{1}{n^R_{h_0} - 1} \sum_{k=1}^{n^R_{h_0}} \left(\tilde{y}_k - \frac{1}{n^R_{h_0}} \sum_{k'=1}^{n^R_{h_0}} \tilde{y}_k \right)^2$$

where $n^R_{h_0}$ is the number of hospitals that responded in at least one month during the year, and, consequently, have a record in the file of reported annual injury totals.

C.2 Problems in Reflecting All Sources of Variance with the Current Monthly Nonresponse Adjustment

If the goal is to reflect the randomness of monthly response and the ERV adjustment, then a more elaborate variance formula must be derived. The variance of the estimated total in (C.1), accounting for both the variation in random selection of the sample, random response, and the ERV ratio adjustment is

$$V(\hat{t}) = E_R V_\pi \left(\hat{t} \middle| s^R_m, m = 1,\ldots,12 \right) + V_R E_\pi \left(\hat{t} \middle| s^R_m, m = 1,\ldots,12 \right) \quad \text{(C.3)}$$

The first term is the average over the response mechanism of the design-variance of $\hat{t}$, which must incorporate the variance due to the ERV ratio adjustment. The second term adds a positive contribution due to the random response. To completely specify (C.3), E_R and V_R would have to be evaluated with respect to the Bernoulli distribution.

Alternatives that will approximately reflect the variation due to adjusting for monthly nonresponse would be the jackknife or bootstrap. However, when monthly nonresponse adjustments are used, a replication variance estimator would probably be impractical. To implement the full jackknife, for example, one hospital from the full sample would be dropped from the sample to form a replicate. The composition of the retained hospitals in the replicate affects the monthly nonresponse adjustments, a_{km}. A replicate value of $\tilde{y}_k = \sum_{m=1}^{12} \delta_{km} a_{km} r_k y_{km}$ could be recomputed using the monthly nonresponse adjustments and the annual ERV ratio adjustment based on the respondents in the replicate. The impracticality of this approach is that a different replicate value of $\tilde{y}_k$ would be needed for each product. Alternatively, monthly replicate values of the monthly weights, $w_k a_{km} r_k$, could be put on the file. However, this is only a small simplification compared to replicating the $\tilde{y}_k$'s because a separate replicate weight would be required for each hospital/month combination due to making nonresponse adjustments separately for every month.

C.3 A Simpler Annual Nonresponse Adjustment and Improved Variance Estimator

If a single annual nonresponse adjustment were used, variance estimation would be simplified. For example, an adjustment in the spirit of the one now used would be

$$a_k = \frac{12 n_{h_0}}{\sum_{m=1}^{12} n_{h_0 m}^R} \equiv \frac{n_{h_0}}{\overline{n}_{h_0}^R}$$

where the average number of respondents per month is $\overline{n}_{h_0}^R = \sum_{m=1}^{12} n_{h_0 m}^R \Big/ 12$. The term a_k is the inverse of a type of average response rate. With complete response every month, $a_k = 1$. The nonresponse- and ERV-adjusted weight for hospital k would be $\tilde{w}_k = w_k a_k r_k$, and an estimated, annual total is $\hat{t} = \sum_{h_0} \sum_{k \in s_{h_0}^R} \tilde{w}_k y_k$ with $s_{h_0}^R$ being the set of sample hospitals that respond sometime during the year. The data file for users would have one record for each hospital that responded in at least one month during the year along with the $\tilde{w}_k$ weight and the total number of injuries for each product, y_k.

Variances that do not account for variation due to the nonresponse adjustment could be estimated with standard survey software packages using the default with-replacement variance estimator, with an *ad hoc* finite population correction factor, as

$$v_\pi\left(\hat{t} \,\middle|\, s_m^R, m = 1, \ldots, 12\right) = \sum_{h_0} \frac{\overline{n}_{h_0}^R}{\overline{n}_{h_0}^R - 1}\left(1 - \frac{\overline{n}_{h_0}^R}{N_{h_0}}\right) \sum_{k \in s_{h_0}^R} \left(\tilde{w}_k y_k - \frac{1}{\overline{n}_{h_0}^R} \sum_{k' \in s_{h_0}^R} \tilde{w}_{k'} y_{k'}\right)^2 \qquad \text{(C.4)}$$

with $\overline{n}_{h_0}^R$ being the mean number of respondents per month in design stratum h_0.

Because this estimator ignores the adjustments, estimated SEs based on (C.4) will be incorrect. In fact, the ERV ratio adjustment reduces standard errors substantially as illustrated in section 4.4. As a result, analysts who use the default variance estimator in survey software are likely to get SE estimates that are overestimates by as much as 30% (see Table 4-4.)

Replication is one option for producing correct SEs that users can easily implement using available survey software. This will require that the CPSC provide replicate weights on its public use files. To use a replication variance estimator like the jackknife, the nonresponse adjustment a_k and the ERV ratio adjustment r_k would be recomputed for each replicate and a series of nonresponse-adjusted replicate weights, $\tilde{w}_k$, would be included on each hospital record along with the full-sample, nonresponse- and ERV-adjusted weight. If the more elaborate calibration methods suggested in section 4.4 were used, the calibration would be replicated in the same way as the ERV ratio adjustment. In their software, users would then specify the appropriate replication variance estimation method that was used to create the weights.