

| DOCUMENTATION CHANGE VERSION CONTROL TABLE | | | |
|---|---|-------------|---|
| Requirement Name | Author | Date | Comments |
| Calculate CD variance and relative standard error (RSE) | Xingyou Zhang, Dee Zamora, Erin Huband, Michelle Myers, Jun Mize, Anurag Katare, Daniel Suh | 07/13/2023 | Final requirement coming out of BigSOII Phase 3 |
| | | | |
| | | | |
| | | | |

| USER STORY LOG | | | |
|-----------------------|-------------------|------------------|--------------|
| Topic | User Story | Iteration | Notes |
| | | | |
| | | | |
| | | | |
| | | | |

Quality Characteristics (If applicable):

Requirements Acceptance Criteria:

(Guide to writing Requirements: Major distinct Components should include FR, NFR, and BR – discreet requirements... that testable. If the testing seems multipart, then break the Requirement)

1. Starting Point

- 1.1. After variances have been calculated for Annual Summary.
- 1.2. After estimation is completed for C&D.

2. Definitions and assumptions

- 2.1. This process generates sampling errors and variances for all counts (weighted estimates) and incidence rates for state and national levels.
- 2.2. The C&D uses a Taylor series linearization methodology to calculate estimates of standard errors for published core C&D estimates.
- 2.3. These requirements go into a lot of detail on when to roll-up and use a higher aggregate TEI for several calculations. Additional information about how to roll-up is in APPENDIX A at the end of this requirement.
 - 2.3.1. What determines when to roll-up should be the same in Annual Summary as in Case and Demographic variance calculations. The only variables that are different from Annual Summary are those calculated for the z variables that are unique to case and demographic estimates.
- 2.4. When counting establishments in the sampling stratum use the sampled/original size class as the reported size class and count the number of establishments in the sampling stratum based on the sampled/original size class.
- 2.5. The N, n, ow_i , w_i , y parts of the equations below should use the establishment dataset generated from the C&D extract program to do the calculations in the formulas throughout the requirements. The calculations for any z variables should use the FCW, finalcases, and the OIICS precategory datasets to do the calculations in the formulas below.

3. Inputs

- 3.1. SAS Datasets with the last successful Process ID
 - ~~3.1.1. FCW Dataset~~
 - 3.1.2. Finalcases SAS dataset
 - 3.1.3. SAS dataset after estimation is completed
 - 3.1.4. The 5 OIICS precategory datasets
- 3.2. TEI Aggregate Levels
- 3.3. SAS dataset with variance information from Annual Summary
- 3.4. Establishment dataset from the C&D extract process that contains all establishments in the sample including mining, railroad, and dummy/prorated establishments from Annual Summary. The original sample weight and the summary final weight variables will be in this dataset to use for the formula calculations and determining if the roll-ups are needed.

4. Processing Steps

- 4.1. Estimation domains for both national and state estimates are combinations of estimation strata.
 - 4.1.1. For total by total, an estimation stratum is defined by collection year, FIPS state code, ownership code, TEI, reported size class, and case type.

- 4.1.2. For one-way estimates, an estimation stratum is defined by collection year, FIPS state code, ownership code, TEI, reported size class, and case type, characteristic.
- 4.1.3. For two-way estimates, an estimation stratum is defined by collection year, FIPS state code, ownership code, TEI, reported size class, case type, characteristic, and cross characteristic.
- 4.1.4. Estimates are based the reported size class of the establishment (LDB Number). The number of usable establishments in an estimation stratum is based on the reported size class, not the size class used for sampling.
- 4.2. For variance estimates, calculate the estimate at the finest detail: collection year, FIPS state code, ownership code, TEI, reported size class, case type, characteristic, and cross characteristic. The variances of *counts* of broader levels are calculated as the sum of the finer levels that comprise the broad level. For example, the variance of counts of the collection year, FIPS state code, ownership code, TEI (i.e., reported size class 0), case type, characteristic, and cross characteristic is the sum of the variances of the collection year, FIPS state code, ownership code, TEI of reported size classes 1-5, case type, characteristic, and cross characteristic. For biennial estimates, the variance of counts of the FIPS state code, ownership code, TEI (i.e., reported size class 0), case type, characteristic, and cross characteristic is the sum of the variances of collection year 1, FIPS state code, ownership code, TEI of reported size classes 1-5, case type, characteristic, and cross characteristic and collection year 2, FIPS state code, ownership code, TEI of reported size classes 1-5, case type, characteristic, and cross characteristic.
 - 4.2.1. The variances of *rates* of broader levels cannot be calculated as the sum of variances of the finer levels that comprise the broad level. Instead, they are calculated using all the usable establishments in the estimation domain. For example, the variance of the rate of the survey year, FIPS state code, ownership code, TEI (i.e., reported size class 0), case type, characteristic, and cross characteristic is calculated using all the usable establishments from size classes 1-5.
- 4.3. If the characteristic name and cross characteristic name for an estimate does not include TEI or TEI aggregate level, then use all industries for determining the calculations described below. Otherwise, match the usable establishments to the TEI or the TEI aggregate level in the characteristic or cross characteristic of the estimate and roll-up from there, if necessary.
- 4.4. Data for the mining (NAICS 212) and railroad industries (NAICS 482) are obtained from MSHA and FRA, respectively, and are considered a census. Since these data are obtained from outside sources for which there is no ability to assess reliability, set the variance and covariance estimates for these industries to 0 before summarization for any estimation domain of interest. This means MSHA and FRA strata only contribute to point estimates (total and rate).
 - 4.4.1. Data for mining, railroad, and dummy records will not have sample information available for them. Set the original sample weight to 1 for the following LDB numbers:
 - 4.4.1.1. LDB number starting with '995': these are dummy records where there were no usable cases across all five size classes within a TEI and FIPS State Code. Proration happens in the benchmarking process for Annual Summary. The system looks for the closest sibling TEI and essentially imputes a number of summary cases from sibling TEIs. The original sample weight shall be set to 1 to be self-representing for LDB numbers starting with '995'.
 - 4.4.1.2. LDB number starting with '996': these are railroad industries (NAICS 482). FRA provides a data set of all injuries and illnesses that happen in the railroad industry. There is no sampling done for this industry. The original sample weight shall be set to 1 for LDB numbers starting with '996'.

4.4.1.3. LDB number starting with '997': these are mining industries (NAICS 212). MSHA provides a data set of all injuries and illnesses that happen in the mining industry. There is no sampling done for this industry. The original sample weight shall be set to 1 for LDB numbers starting with '997'.

- 4.5. Variance Estimation: Calculate the variance for totals of case counts (weighted estimates) for each case type (DART, DAFW, DJTR) using the unbiased formula below for weighted samples.
- 4.5.1. For biennial years, calculate the variance for one collection year at a time and then add the two variances together for the biennial variance.
- 4.5.2. If the number of Summary usable establishments in the estimation stratum and the number of establishments in the sampling stratum equal 1, then set the variance equal to 0.
- 4.5.3. If the number of usable establishments in the estimation stratum is greater than 1 and the number of establishments in the sampling stratum is greater than 1, then the variance is calculated by the following:

$$\text{var}(\hat{z}) = \text{var}(N\bar{z}) = N^2 * \text{var}(\bar{z}) = N^2 \frac{\text{var}(z)}{n}$$

$$\hat{z} N^2 \frac{1}{n}$$

$$\hat{z} \frac{N^2}{n}$$

4.5.3.1. Where:

- 4.5.3.1.1. \hat{z} is the total weighted estimate of the variable (case counts for each collection year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday)) for all usable establishments in the estimation stratum: $\hat{z} = \sum_{i=1}^n w_i z_i$. For biennial years, w_i and z_i will be different based on collection year. See below for more clarification on the individual variables.

4.5.3.1.1.1. If \hat{z} is equal to 0, then the variance, covariance, and RSE for \hat{z} and $\frac{\hat{z}}{\hat{y}}$ are equal to 0.

- 4.5.3.1.2. N is the sum of Summary final weights for usable establishments in the estimation stratum: $N = \sum_{i=1}^n w_i$. For biennial years, N will be the sum of the Summary final weights for usable establishments in the individual collection year. Note: N will be different based on collection year.

- 4.5.3.1.3. n is the number of Summary usable establishments reporting in the estimation stratum. For biennial years, n will be the count of the usable establishments in the individual collection year. Note: n will be different based on collection year.

- 4.5.3.1.4. OW_i is the Summary original sampling weight of establishment i in the estimation stratum. For biennial years, OW_i will be the sum of the Summary

original sampling weights for all establishments in the individual collection year. Note: OW_i will be different based on collection year.

4.5.3.1.5. w_i is the Summary final weight of establishment i in the estimation stratum. For biennial years, w_i will be the Summary final weights for usable establishments in the individual collection year. Note: w_i will be different based on collection year.

4.5.3.1.6. Z_i is the weighted reported case value of the variable (case counts for each collection year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday)) for establishment i in the estimation stratum

4.5.3.1.6.1. For each establishment, count the number of cases within each collection year, FIPS state code, ownership code, reported size, case type, characteristic name, characteristic value, cross characteristic name, and cross characteristic value for the associated estimate.

4.5.3.1.6.1.1. If the count is 0, then set Z_i to 0.

4.5.3.1.6.1.2. Otherwise, if the count is positive, multiply the count from the previous bullet by the CSSF, CNRAF, and CRAF for the case type for establishment i .

4.5.3.1.7. \bar{z} is the weighted average value of the variable (case counts for each collection year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday)) for all usable establishments in the estimation stratum:

$$\bar{z} = \frac{\sum_{i=1}^n w_i Z_i}{\sum_{i=1}^n w_i}$$

4.5.4. If the number of usable establishments in the estimation stratum equals 1 and the number of establishments in the sampling stratum is not equal to 1, the mean of the stratum, \bar{z} , cannot be used because it will yield an incorrect variance of 0. To correct for this, use the mean from the roll-up hierarchy (TEI aggregate levels - TEI6, TEI5, TEI4, TEI3, Sector, Super_Sector, and Domain).

4.5.4.1. The sum of final weights of all usable establishments in the estimation stratum is the frame size for this estimation stratum. Therefore, there would be no missing values for frame size.

4.5.4.2. That is, if there is only one usable establishment in the estimation stratum and the number of establishments in the sampling stratum is not equal to 1, calculate the variance as the variance of the roll-up level where the mean is the mean of the collection year, FIPS state code, ownership code, reported size class, case type, characteristic, cross characteristic, and parent TEI of the usable establishments in this roll-up level. However, if the one usable establishment in the estimation stratum is also the only usable establishment for the roll-up estimation stratum, continue to roll-up by parent TEI level as far as the industry domain level until the number of usable establishments is greater than 1.

4.5.4.3. If the one usable establishment in the estimation stratum is also the only usable establishment for the collection year, FIPS state code, ownership code, reported size class, and industry domain, case type, characteristics, and cross characteristics, calculate the variance as the variance of the roll-up level where the mean is the mean of collection year, FIPS state code, ownership code, reported size class 0 (all sizes) and TEI. Continue to roll-up the parent TEI level as far as all industries until the number of usable establishments is greater than 1.

4.5.5. Therefore, if the number of usable establishments in the estimation stratum equals 1 and the number of establishments in the sampling stratum is not equal to 1, we assign the roll-up level variance for this single usable establishment; in other words, we use the roll-up level variance to approximate the variance for this single establishment stratum. The variance is calculated in the same way as those strata with more than one usable establishment; it just includes all usable establishments at the roll-up level. Specifically, the variance for the stratum with single usable establishment is calculated as the following:

$$\text{var}(\hat{z}) = \text{var}(z_{RU}) = \text{var}(N \bar{z}_{RU}) = N^2 * \text{var}(\bar{z}_{RU}) = N^2 \frac{\text{var}(z)_{RU}}{n_{RU}}$$

$$\hat{z} N^2 \frac{1}{n_{RU}} \hat{z}$$

$$\hat{z} \frac{N^2}{n_{RU}} \hat{z}$$

4.5.5.1. Where:

4.5.5.1.1. \hat{z}_{RU} is the total weighted estimate of the variable (case counts for each collection year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday) for all usable establishments in the roll-up level: $\hat{z}_{RU} = \sum_{i=1}^{n_{RU}} w_i z_i$. For biennial years, w_i and z_i will be different based on collection year. See below for more clarification on the individual variables.

4.5.5.1.2. N is the sum of Summary final weights for usable establishments in the estimation stratum: $N = \sum_{i=1}^n w_i$, in this case it is the single usable establishment estimation stratum, so $N = w_i$. For biennial years, N will be different based on collection year.

4.5.5.1.3. N_{RU} is the sum of Summary final weights for usable establishments in the roll-up level: $N_{RU} = \sum_{i=1}^{n_{RU}} w_i$. For biennial years, N_{RU} will be different based on collection year.

4.5.5.1.4. n_{RU} is the number of usable establishments reporting in the roll-up level. For biennial years, n_{RU} will be the count of the usable establishments in the individual collection year. Note: n_{RU} will be different based on collection year. For example, one collection year could have more than one usable establishment and the other collection year have only one usable

establishment meeting this roll-up condition within the same FIPS state code, ownership code, reported size, case type, characteristic, and cross characteristic.

- 4.5.5.1.5. OW_i is the Summary original sampling weight of establishment i . For biennial years, OW_i will be the Summary original sampling weights for all establishments in the individual collection year. Note: OW_i will be different based on collection year. In addition, the OW_i is the original weight of establishment i at time of sampling, so the TEI used at time of sampling may not be the same TEI used in the roll-up.
- 4.5.5.1.6. W_i is the Summary final weight of establishment i in the roll-up level. For biennial years, w_i will be the Summary final weights for usable establishments in the individual collection year. Note: w_i will be different based on collection year.
- 4.5.5.1.7. Z_i is the weighted estimate value of the variable (case counts for each collection year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday)) for establishment i in the roll-up level
- 4.5.5.1.7.1. For each establishment, count the number of cases within each collection year, FIPS state code, ownership code, case type, reported size, characteristic name, characteristic value, cross characteristic name, and cross characteristic value for the associated estimate.
- 4.5.5.1.7.1.1. If the count is 0, then set z_i to 0.
- 4.5.5.1.7.1.2. Otherwise, if the count is positive, $Z_i =$ the count from the previous bullet * CSSF * CNRAF * CRAF for the case type for establishment i
- 4.5.5.1.8. \bar{Z}_{RU} is the weighted average value of the variable (case counts for each collection year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday)) for all usable establishments in the roll-up level:

$$\bar{Z}_{RU} = \frac{\sum_{i=1}^{n_{RU}} w_i Z_i}{\sum_{i=1}^{n_{RU}} w_i}$$

4.6. Variance Estimation: Calculate the variance for hours using the unbiased formula below for weighted samples. Note: These requirements for the variance of weighted hours are the same as Annual Summary. To save processing time, obtain the information on the weighted hours (W) and the hours variance (W_Variance) from Annual Summary where reported size is 1-5. This should be brought into the program prior to summing within an estimation cell or across estimation cells for roll-ups.

4.6.1. For biennial years, calculate the variance for one collection year at a time and then add the two variances together for the biennial variance.

4.6.2. If the number of Summary usable establishments in the estimation stratum and the number of establishments in the sampling stratum equal 1, then set the variance equal to 0.

4.6.3. If the number of usable establishments in the estimation stratum is greater than 1 and the number of establishments in the sampling stratum is greater than 1, then the variance is calculated by the following:

$$\text{var}(\hat{y}) = \text{var}(N \bar{y}) = N^2 * \text{var}(\bar{y}) = N^2 \frac{\text{var}(y)}{n}$$

$$\hat{y} N^2 \frac{1}{n} \hat{y}$$

$$\hat{y} \frac{N^2}{n} \hat{y}$$

4.6.3.1. Where:

4.6.3.1.1. \hat{y} is the total weighted hours worked of the variable (hours for each collection year, FIPS state code, ownership code, reported size, case type, LDB number) for all usable establishments in the estimation stratum: $\hat{y} = \sum_{i=1}^n wtd_{i,t}$.

For biennial years, $wtd_hrs_worked_i$ will be different based on collection year. See below for more clarification on the individual variable.

4.6.3.1.1.1. If there are multiple cases reported for a single LDB number within a given estimation stratum, then only include the weighted hours worked once for the establishment and not for each individual case.

4.6.3.1.1.2. If \hat{y} is equal to 0, then the variance and covariance for \hat{y} are equal to 0.

4.6.3.1.1.3. $wtd_hrs_worked_i$ is the Summary final weight * total_hrs_worked of establishment i in the estimation stratum.

4.6.3.1.2. N is the sum of Summary final weights for usable establishments in the estimation stratum: $N = \sum_{i=1}^n w_i$. For biennial years, N will be the sum of the Summary final weights for usable establishments in the individual collection year. Note: N will be different based on collection year.

4.6.3.1.3. n is the number of Summary usable establishments reporting in the estimation stratum. For biennial years, n will be the count of the usable establishments in the individual collection year. Note: n will be different based on collection year. For example, one collection year could have more than one usable establishment and the other collection year have only one usable establishment meeting this roll-up condition within the same FIPS state code, ownership code, reported size, case type, characteristic, cross characteristic.

4.6.3.1.4. ow_i is the Summary original sampling weight of establishment i in the estimation. For biennial years, ow_i will be the sum of the Summary original sampling weights for all establishments in the individual collection year. Note: ow_i will be different based on collection year.

4.6.3.1.5. w_i is the Summary final weight of establishment i in the estimation stratum. For biennial years, w_i will be the Summary final weights for usable establishments in the individual collection year. Note: w_i will be different based on collection year.

4.6.3.1.6. y_i is the reported value of the variable (total hours worked for each collection year, FIPS state code, ownership code, reported size, case type, LDB number) for establishment i in the estimation stratum. For biennial years, y_i will be different based on collection year.

4.6.3.1.7. \bar{y} is the weighted average value of the variable (total hours worked for each collection year, FIPS state code, ownership code, reported size, case type, LDB number) for all usable establishments in the estimation stratum:

$$\bar{y} = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}$$

4.6.4. If the number of usable establishments in the estimation stratum equals 1 and the number of establishments in the sampling stratum is not equal to 1, the mean of the stratum, \bar{y} , cannot be used because it will yield an incorrect variance of 0. To correct for this, use the mean from the roll-up hierarchy.

4.6.4.1. That is, if there is only one usable establishment in the estimation stratum and the number of establishments in the sampling stratum is not equal to 1, calculate the variance as the variance of the roll-up level where the mean is the mean of the collection year, FIPS state code, ownership code, reported size class, case type, characteristic, cross characteristic, and parent TEI of the usable establishments in this roll-up level. However, if the one usable establishment in the estimation stratum is also the only usable establishment for the roll-up estimation stratum, continue to roll-up by parent TEI level as far as the industry domain level until the number of usable establishments is greater than 1.

4.6.4.2. If the one usable establishment in the estimation stratum is also the only usable establishment for the collection year, FIPS state code, ownership code, reported size class, and industry domain, case type, characteristics, and cross characteristics, calculate the variance as the variance of the roll-up level where the mean is the mean of collection year, FIPS state code, ownership code, reported size class 0 (all sizes) and TEI. Continue to roll-up the parent TEI level as far as all industries until the number of usable establishments is greater than 1.

4.6.4.2.1. Therefore, if the number of usable establishments in the estimation stratum equals 1 and the number of establishments in the sampling stratum is not equal to 1, we assign the roll-up level variance for this single usable establishment; in other words, we use the roll-up level variance to approximate the variance for this single establishment stratum. The variance is calculated in the same way as those strata with more than one usable establishment; it just includes all usable establishments at the roll-up level. Specifically, the variance for the stratum with single usable establishment is calculated as the following:

$$\text{var}(\hat{y}) = \text{var}(y_{RU}) = \text{var}(N \bar{y}_{RU}) = N^2 * \text{var}(\bar{y}_{RU}) = N^2 \frac{\text{var}(y)_{RU}}{n_{RU}}$$

$$N^2 \frac{1}{n_{RU}}$$

$$i \frac{N^2}{n_{RU}} i$$

4.6.4.3. Where:

4.6.4.3.1. \hat{y}_{RU} is the total weighted estimate of the variable (hours for each collection year, FIPS state code, ownership code, reported size, case type, LDB number) for all usable establishments in the roll-up level: $\hat{y}_{RU} = \sum_{i=1}^{n_{RU}} wtd_{i}$. For biennial years, $wtd_hrs_worked_i$ will be different based on collection year. See below for more clarification on the individual variable.

4.6.4.3.1.1. $wtd_hrs_worked_i$ is the Summary final weight * total_hrs_worked of establishment i in the estimation stratum.

4.6.4.3.2. N is the sum of Summary final weights for usable establishments in the estimation stratum: $N = \sum_{i=1}^n w_i$, in this case it is the single usable establishment estimation stratum, so $N = w_i$. For biennial years, N will be different based on collection year.

4.6.4.3.3. N_{RU} is the sum of Summary final weights for usable establishments in the roll-up level: $N_{RU} = \sum_{i=1}^{n_{RU}} w_i$. For biennial years, N_{RU} will be different based on collection year.

4.6.4.3.4. n_{RU} is the number of usable establishments reporting in the roll-up level. For biennial years, n_{RU} will be the count of the usable establishments in the individual collection year. Note: n_{RU} will be different based on collection year. For example, one collection year could have more than one usable establishment and the other collection year have only one usable establishment meeting this roll-up condition within the same FIPS state code, ownership code, reported size, case type, characteristic, cross characteristic.

4.6.4.3.5. ow_i is the Summary original sampling weight of establishment i . For biennial years, ow_i will be the Summary original sampling weights for all establishments in the individual collection year. Note: ow_i will be different based on collection year. In addition, the ow_i is the original weight of establishment i at time of sampling, so the TEI used at time of sampling may not be the same TEI used in the roll-up.

4.6.4.3.6. w_i is the Summary final weight of establishment i in the roll-up level. For biennial years, w_i will be the Summary final weights for usable establishments in the individual collection year. Note: w_i will be different based on collection year.

4.6.4.3.7. y_i is the reported value of the variable (total hours worked for each collection year, FIPS state code, ownership code, reported size, LDB number)) for establishment i in the roll-up level. For biennial years, y_i will be different based on collection year.

4.6.4.3.8. \bar{y}_{RU} is the weighted average value of the variable (total hours worked for each collection year, FIPS state code, ownership code, reported size, LDB_number)) for all usable establishments in the roll-up level:

$$\bar{y}_{RU} = \frac{\sum_{i=1}^{n_{RU}} w_i y_i}{\sum_{i=1}^{n_{RU}} w_i}$$

For biennial years, w_i and y_i will be different based on collection year. See above for more clarification on the individual variables.

4.7. Covariance Estimation: Calculate the covariance for total case counts for each collection year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2 - female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday) with hours.

4.7.1. If the number of usable establishments in the estimation stratum and the number of establishments in the sampling stratum equal 1, set the covariance equal to 0.

4.7.2. If the number of usable establishments in the estimation stratum is greater than 1 and the number of establishments in the sampling stratum is greater than 1, the covariance is calculated by the following:

$$\text{cov}(\hat{z}, \hat{y}) = N^2 \frac{1}{n} \hat{c}$$

$$\hat{c} \frac{N^2}{n} \hat{c}$$

4.7.2.1. Where:

4.7.2.1.1. \hat{z} is the total weighted estimate of the variable (case counts for each collection year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2 - female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday)) for all usable establishments in the estimation stratum: $\hat{z} = \sum_{i=1}^n w_i z_i$.

For biennial years, w_i and z_i will be different based on collection year. See below for more clarification on the individual variables.

4.7.2.1.1.1. If \hat{z} is equal to 0, then the variance, and covariance and RSE for \hat{z} and $\frac{\hat{z}}{\hat{y}}$ are equal to 0.

4.7.2.1.2. \hat{y} is the total weighted estimate of the variable (total hours worked) for all usable establishments in the estimation stratum: $\hat{y} = \sum_{i=1}^n w_i y_i$. For biennial years, w_i and y_i will be different based on collection year. See below for more clarification on the individual variables.

4.7.2.1.3. N is the sum of Summary final weights for usable establishments in the estimation stratum: $N = \sum_{i=1}^n w_i$. For biennial years, N will be the sum of the

Summary final weights for usable establishments in the individual collection year. Note: N will be different based on collection year.

4.7.2.1.4. n is the number of usable establishments reporting in the estimation stratum. For biennial years, n will be the count of the usable establishments in the individual collection year. Note: n will be different based on collection year.

4.7.2.1.5. OW_i is the Summary original sampling weight of establishment i in the estimation stratum. For biennial years, OW_i will be the sum of the Summary original sampling weights for all establishments in the individual collection year. Note: OW_i will be different based on collection year.

4.7.2.1.6. w_i is the Summary final weight of establishment i in the estimation stratum. For biennial years, w_i will be the Summary final weights for usable establishments in the individual collection year. Note: w_i will be different based on collection year.

4.7.2.1.7. Z_i is the weighted reported value of the variable (case counts for each collection year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday)) for establishment i in the estimation stratum.

4.7.2.1.7.1. For each establishment, count the number of cases within each collection year, FIPS state code, ownership code, case type, reported size, characteristic name, characteristic value, cross characteristic name, and cross characteristic value for the associated estimate.

4.7.2.1.7.1.1. If the count is 0, then set Z_i to 0.

4.7.2.1.7.1.2. Otherwise, if the count is positive, $Z_i =$ the count from the previous bullet * CSSF * CNRAF * CRAF for the case type for establishment i

4.7.2.1.8. \bar{z} is the weighted average value of the variable (case counts for each collection year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday)) for all usable establishments in the estimation stratum:

$$\bar{z} = \frac{\sum_{i=1}^n w_i Z_i}{\sum_{i=1}^n w_i}$$

4.7.2.1.9. y_i is the reported value of the variable (total hours worked for each collection year, FIPS state code, ownership code, reported size, LDB number) for establishment i in the estimation stratum. For biennial years, y_i will be different based on collection year.

4.7.2.1.10. \bar{y} is the weighted average value of the variable (total hours worked for each collection year, FIPS state code, ownership code, reported size, LDB

number) for all usable establishments in the estimation stratum: $\bar{y} = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}$.

For biennial years, w_i and y_i will be different based on collection year. See above for more clarification on the individual variables.

4.7.3. If the number of usable establishments equals 1 and the number of stratum sample establishments is not equal to 1, the means, \bar{z} or \bar{y} , cannot be used because they will yield an incorrect covariance of 0. To correct for this, we need to use the means from the roll-up hierarchy:

4.7.3.1. If there is only one usable establishment in the estimation stratum and the number of establishments in the sampling stratum is not equal to 1, calculate the variance as the variance of the roll-up level where the mean is the mean of the collection year, FIPS state code, ownership code, reported size class, case type, characteristic, cross characteristic, and parent TEI of the usable establishments in this roll-up level. However, if the one usable establishment in the estimation stratum is also the only usable establishment for the roll-up estimation stratum, continue to roll-up by parent TEI level as far as the industry domain level until the number of usable establishments is greater than 1.

4.7.3.1.1. If the one usable establishment in the estimation stratum is also the only usable establishment for the collection year, FIPS state code, ownership code, reported size class, and industry domain, case type, characteristics, and cross characteristics, calculate the variance as the variance of the roll-up level where the mean is the mean of collection year, FIPS state code, ownership code, size class 0 (all sizes) and TEI. Continue to roll-up the parent TEI level as far as all industries until the number of usable establishments is greater than 1.

4.7.3.1.1.1. Therefore, where the number of usable establishments in the estimation stratum equals 1 and the number of establishments in the sampling stratum is not equal to 1, similarly we use the roll-up level covariance as the covariance for the single usable establishment stratum. The approximate covariance for the stratum with single usable establishment is calculated by the following:

$$\text{cov}(\hat{z}, \hat{y}) = \hat{c} \text{cov}(\hat{z}_{RU}, \hat{y}_{RU}) = N^2 \frac{1}{n_{RU}} \hat{c}$$

$$\hat{c} \frac{N^2}{n_{RU}} \hat{c}$$

4.7.3.1.1.1.1. Where:

4.7.3.1.1.1.1.1. \hat{z}_{RU} is the total weighted estimate of the variable (case counts for each collection year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic

value (e.g., 3 - Tuesday)) for all usable establishments in the

roll-up level: $\hat{z}_{RU} = \sum_{i=1}^{n_{RU}} w_i z_i$. For biennial years, w_i and z_i will

be different based on collection year. See below for more clarification on the individual variables.

4.7.3.1.1.1.1.2. \hat{y}_{RU} is the total weighted estimate of the variable (total hours worked) for all usable establishments in the roll-up

level: $\hat{y}_{RU} = \sum_{i=1}^{n_{RU}} w_i y_i$. For biennial years, w_i and y_i will be

different based on collection year. See below for more clarification on the individual variables.

4.7.3.1.1.1.1.3. N is the sum of Summary final weights for usable

establishments in the estimation stratum: $N = \sum_{i=1}^n w_i$, in this

case it is the single usable establishment estimation stratum, so $N = w_i$. For biennial years, N will be different based on collection year.

4.7.3.1.1.1.1.4. N_{RU} is the sum of Summary final weights for usable

establishments in the roll-up level: $N_{RU} = \sum_{i=1}^{n_{RU}} w_i$. For biennial

years, N_{RU} will be different based on collection year.

4.7.3.1.1.1.1.5. n_{RU} is the number of usable establishments reporting in the roll-up level. For biennial years, n_{RU} will be the count of the usable establishments in the individual collection year.

Note: n_{RU} will be different based on collection year. For example, one collection year could have more than one usable establishment and the other collection year have only one usable establishment meeting this roll-up condition within the same FIPS state code, ownership code, reported size, case type, characteristic, cross characteristic.

4.7.3.1.1.1.1.6. ow_i is the Summary original sampling weight of establishment i . For biennial years, ow_i will be the Summary original sampling weights for all establishments in the individual collection year. Note: ow_i will be different based on collection year. In addition, the ow_i is the original weight of establishment i at time of sampling, so the TEI used at time of sampling may not be the same TEI used in the roll-up.

4.7.3.1.1.1.1.7. w_i is the Summary final weight of establishment i in the roll-up level. For biennial years, w_i will be the Summary final weights for usable establishments in the individual collection year. Note: w_i will be different based on collection year.

4.7.3.1.1.1.1.8. z_i is the reported value of the variable (case counts for each collection year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross

characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday)) for establishment i in the roll-up level.

4.7.3.1.1.1.1.8.1. For each establishment, count the number of cases within each collection year, FIPS state code, ownership code, case type, reported size, characteristic name, characteristic value, cross characteristic name, and cross characteristic value for the associated estimate.

4.7.3.1.1.1.1.8.2. If the count is 0, then set Z_i to 0.

4.7.3.1.1.1.1.8.3. Otherwise, if the count is positive, Z_i = the count from the previous bullet * CSSF * CNRAF * CRAF for the case type for establishment i

4.7.3.1.1.1.1.9. \bar{Z}_{RU} is the weighted average value of the variable (case counts for each collection year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday)) for all usable establishments in the roll-up level:

$$\bar{Z}_{RU} = \frac{\sum_{i=1}^{n_{RU}} w_i Z_i}{\sum_{i=1}^{n_{RU}} w_i}$$

4.7.3.1.1.1.1.10. y_i is the reported value of the variable (total hours worked for each collection year, FIPS state code, ownership code, reported size, LDB number) for establishment i in the roll-up level. For biennial years, y_i will be different based on collection year.

4.7.3.1.1.1.1.11. \bar{y}_{RU} is the weighted average value of the variable (total hours worked for each collection year, FIPS state code, ownership code, reported size, LDB number) for all usable establishments in the roll-up level. For biennial years, w_i and y_i will be different based on collection year. See above for more clarification on the individual variables.

$$\bar{y}_{RU} = \frac{\sum_{i=1}^{n_{RU}} w_i y_i}{\sum_{i=1}^{n_{RU}} w_i}$$

4.8. Variance of Ratios: Calculate the variance of ratios for each incidence rate estimate.

4.8.1. If the number of usable establishments in the estimation stratum and the number of establishments in the sampling stratum equal 1, set the variance equal to 0.

4.8.2. If the number of usable establishments in the estimation stratum is greater than 1 and the number of establishments in the sampling stratum is greater than 1, calculate the variance of the ratio of total case counts for each collection year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday) (z) and hours (y) ($Rate = \frac{\hat{z}}{\hat{y}}$) using the following unbiased formula for weighted samples

$$Var\left(\frac{\hat{z}}{\hat{y}}\right) = \left(\frac{\hat{z}}{\hat{y}}\right)^2 * \left(\frac{var(\hat{z})}{\hat{z}^2} + \frac{var(\hat{y})}{\hat{y}^2} - \frac{2 * cov(\hat{z}, \hat{y})}{\hat{z} * \hat{y}}\right)$$

4.8.2.1. Where:

4.8.2.1.1. \hat{z} is the total weighted estimate of the variable (case counts for each collection year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday)) for all usable establishments in the estimation stratum: $\hat{z} = \sum_{i=1}^n w_i z_i$. For biennial years, w_i and z_i will be different based on collection year.

4.8.2.1.1.1. If \hat{z} is equal to 0, then the variance, and covariance and RSE for \hat{z} and $\frac{\hat{z}}{\hat{y}}$ are equal to 0.

4.8.2.1.2. w_i is the Summary final weight of establishment i in the roll-up level. For biennial years, w_i will be the Summary final weights for usable establishments in the individual collection year. Note: w_i will be different based on collection year.

4.8.2.1.3. z_i is the reported value of the variable (case counts for each collection year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday)) for establishment i in the roll-up level.

4.8.2.1.3.1. For each establishment, count the number of cases within each collection year, FIPS state code, ownership code, case type, reported size, characteristic name, characteristic value, cross characteristic name, and cross characteristic value for the associated estimate.

4.8.2.1.3.1.1. If the count is 0, then set z_i to 0.

4.8.2.1.3.1.2. Otherwise, if the count is positive, $z_i =$ the count from the previous bullet * CSSF * CNRAF * CRAF for the case type for establishment i

4.8.2.1.4. \hat{y} is the total weighted estimate of the variable (total hours worked) for all usable establishments in the estimation stratum: $\hat{y} = \sum_{i=1}^n w_i y_i$. For biennial years, w_i and y_i will be different based on collection year.

4.8.2.1.5. y_i is the reported value of the variable (total hours worked for each collection year, FIPS state code, ownership code, reported size, LDB number)

for establishment i in the estimation stratum. For biennial years, y_i will be different based on collection year.

- 4.8.2.1.6. $var(\hat{z})$ is the variance estimate of the variable in the estimation stratum (total case counts for each biennial/survey year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday) (\hat{z}))
- 4.8.2.1.7. $var(\hat{y})$ is the variance estimate of the total hours worked in the estimation stratum total hours worked for each biennial/survey year, FIPS state code, ownership code, reported size, LDB number) (\hat{y})
- 4.8.2.1.8. $cov(\hat{z}, \hat{y})$ is the covariance estimate between two variables in the estimation stratum (case counts of each case type and hours)

If the incidence rate per 10,000 employees (equivalent 20,000,000 employee hours per year) is reported \hat{i} , then the **variance of incidence rate** for case types DART, DAFW, DJTR is

$var \hat{i}$

- 4.8.3. If the number of usable establishments in the estimation stratum equals 1 and the number of establishments in the sampling stratum is not equal to 1, calculate the variance of the ratio as the variance of total case counts for each case type (\hat{z}) and hours (\hat{y}) of the roll-up level used above, using the following unbiased formula for weighted samples

$$Var\left(\frac{\hat{z}}{\hat{y}}\right) = (20,000,000)^2 \left(\frac{\hat{z}}{\hat{y}}\right)^2 * \left(\frac{var(\hat{z}_{RU})}{\hat{z}^2} + \frac{var(\hat{y}_{RU})}{\hat{y}^2} - \frac{2 * cov(\hat{z}_{RU}, \hat{y}_{RU})}{\hat{z} * \hat{y}}\right)$$

- 4.8.3.1. Where:

- 4.8.3.1.1. \hat{z} is the total weighted estimate of the variable (case counts for each collection year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday)) for all usable establishments in the estimation stratum: $\hat{z} = \sum_{i=1}^n w_i z_i$
 . For biennial years, w_i and z_i will be different based on collection year.

- 4.8.3.1.1.1. If \hat{z} is equal to 0, then the variance, and covariance and RSE for \hat{z} and $\frac{\hat{z}}{\hat{y}}$ are equal to 0.

- 4.8.3.1.2. w_i is the Summary final weight of establishment i in the roll-up level. For biennial years, w_i will be the Summary final weights for usable establishments in the individual collection year. Note: w_i will be different based on collection year.

- 4.8.3.1.3. z_i is the reported value of the variable (case counts for each collection year, FIPS state code, ownership code, reported size, case type, characteristic

name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday)) for establishment i in the roll-up level.

4.8.3.1.3.1. For each establishment, count the number of cases within each collection year, FIPS state code, ownership code, case type, reported size, characteristic name, characteristic value, cross characteristic name, and cross characteristic value for the associated estimate.

4.8.3.1.3.1.1. If the count is 0, then set Z_i to 0.

4.8.3.1.3.1.2. Otherwise, if the count is positive, Z_i = the count from the previous bullet * CSSF * CNRAF * CRAF for the case type for establishment i

4.8.3.1.4. \hat{y} is the total weighted estimate of the variable (total hours worked) for all usable establishments in the estimation stratum: $\hat{y} = \sum_{i=1}^n w_i y_i$. For biennial years, w_i and y_i will be different based on collection year.

4.8.3.1.5. y_i is the reported value of the variable (total hours worked) for establishment i in the estimation stratum. For biennial years, y_i will be different based on collection year.

4.8.3.1.6. \hat{z}_{RU} is the total weighted estimate of the variable (case counts for each collection year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday)) for all usable establishments in the roll-up level: $\hat{z}_{RU} = \sum_{i=1}^{n_{RU}} w_i Z_i$

4.8.3.1.7. \hat{y}_{RU} is the total weighted estimate of the variable (total hours worked) for all usable establishments in the roll-up level: $\hat{y}_{RU} = \sum_{i=1}^{n_{RU}} w_i y_i$. For biennial years, w_i and y_i will be different based on collection year. See below for more clarification on the individual variables.

4.8.3.1.8. $var(\hat{z}_{RU})$ is the variance estimate of the variable (total case counts for each biennial/survey year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday) (\widehat{z}_{RU})) for the roll-up level.

4.8.3.1.9. $var(\hat{y}_{RU})$ is the variance estimate of the total hours worked in the estimation stratum (total hours for each biennial/survey year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday) (\widehat{y}_{RU})) for the roll-up level.

4.8.3.1.10. $COV(\hat{z}_{RU}, \hat{y}_{RU})$ is the covariance estimate between two variables (case counts of each case type and hours) for the roll-up level.

If the incidence rate per 10,000 employees (equivalent 20,000,000 employee hours per year) is reported ($r_{ysogab} = \frac{\hat{z}}{\hat{y}} * 20,000,000$), then the **variance of incidence rate** for the roll-up level for case types DART, DAFW, DJTR is

$$\text{var}(r_{ysogab}) = (20,000,000)^2 \left(\frac{\hat{z}}{\hat{y}} \right)^2 * \left(\frac{\text{var}(\hat{z}_{RU})}{\hat{z}^2} + \frac{\text{var}(\hat{y}_{RU})}{\hat{y}^2} - \frac{2 * \text{COV}(\hat{z}_{RU}, \hat{y}_{RU})}{\hat{z} * \hat{y}} \right)$$

4.8.4. Very small variances should be set to zero. Set variances less than the absolute value of 0.000005 to zero.

4.9. Percent Relative Standard Error (RSE) Estimation: Calculate the percent relative standard error using the formulas below.

4.9.1. Exclude calculating percent relative standard errors for total hours worked (\hat{y}).

4.9.2. Calculate percent RSE for counts (weighted estimates):

$$\%RSE(\hat{z})_{Count} = \frac{100 * \sqrt{\text{Var}(\hat{z})}}{\hat{z}} \quad \text{for totals}$$

4.9.2.1. Where:

4.9.2.1.1. \hat{z} is the total weighted estimate of the variable (case counts for each biennial/survey year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday)) for all usable establishments in the stratum. Note that the estimate may be derived from a single usable establishment or multiple usable establishments.

4.9.2.1.1.1. If \hat{z} is equal to 0, then the variance, and covariance and RSE for \hat{z} and $\frac{\hat{z}}{\hat{y}}$ are equal to 0.

4.9.2.1.2. $\text{var}(\hat{z})$ is the variance estimate of the variable (case counts for each biennial/survey year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday)). Note where the number of usable establishments in the estimation stratum equals 1 and the number of establishments in the sampling stratum is not equal to 1, the roll-up level variance was used to approximate the variance for the single establishment stratum.

4.9.3. Calculate percent RSE for incidence rates.

4.9.3.1. Only calculate percent relative standard errors for non-demographic incidence rates. Exclude demographic rates on age group, gender, SOC, and SOC groups.

$$\%RSE\left(\frac{\hat{z}}{\hat{y}}\right)Rate = 100 * \frac{\sqrt{\text{var}(r_{ysogab})}}{r_{ysogab}} \quad \text{for rates}$$

4.9.3.1.1. Where:

4.9.3.1.1.1. \hat{z} is the total estimate of the variable (total case counts for each biennial/survey year, FIPS state code, ownership code, reported size, case type, characteristic name (e.g., gender_code), characteristic value (e.g., 2-female), cross characteristic name (e.g., day_of_week), cross characteristic value (e.g., 3 - Tuesday)) for all usable establishments in the estimation stratum

4.9.3.1.1.2. \hat{y} is the total estimate of the variable (total hours worked) for all usable establishments in the estimation stratum

4.9.3.1.1.3. $\text{var}(r_{ysogab})$ is the variance estimate of the incidence rate. Note where the number of usable establishments in the estimation stratum equals 1 and the number of establishments in the sampling stratum is not equal to 1, the roll-up level variance of the ratio was used to approximate the variance of the ratio for the single establishment stratum.

4.9.3.1.1.4. r_{ysogab} is the incidence rate of case counts to hours depending on which case type is being used.

5. Outputs

5.1. SAS dataset

5.1.1. All SAS datasets shall be saved for all successful process IDs

5.2. Estimates shall be stored at their full accuracy in a database table.

5.2.1. CD_estimates table – store the following variables associated with their weighted estimate and/or incidence rate.

5.2.1.1. Variance_counts

5.2.1.2. Variance_hours

5.2.1.3. Variance_rates

5.2.1.4. Covariance_counts_hours

5.2.1.5. RSE_counts

5.2.1.6. RSE_rates

5.2.2. Only the most recent successful run of Estimates shall be stored to the database table.

6. Error checking

6.1. Check all counts (weighted estimates) have a variance_counts and RSE_counts. That is variance_counts and RSE_counts are non-missing.

6.2. RSE_rates should not be calculated for demographic incidence rates.

APPENDIX A: SOII variance estimation roll-up algorithm

When a detailed SOII stratum, defined by collection year, FIPS state code, ownership code, TEI, and reported employment size, case type, characteristic, and cross characteristic has a single usable establishment, a rolling-up procedure is needed to form a more aggregated stratum with at least two usable establishments to calculate the variance and covariance estimates and assign them for this single-unit stratum. This approach uses the variance estimates for the aggregated stratum to approximate the true variances for this single-unit detailed stratum.

The specific conditions for a stratum rolling-up would be: a stratum has a single establishment with a sampling frame size not equal to one (defined by original weight or final weight) and from industries other than MSHA (mining – NAICS 212) and FRA (railroads – NAICS 482).

The rolling-up procedure involves two main stratum combination or merging to form an aggregated stratum for estimating variance and covariance for SOII outcomes:

- 1) Rolling-up within TEI with strata from the same reported employment size group
- 2) Rolling-up within TEI with strata from all five reported employment size groups

With the guidance from subject experts, the rolling-up should start with within TEI with strata of the same reported employment size and, if this does not work, then consider within TEI with strata from all reported employment size groups. The rest of this document will first explain the SOII stratum TEI hierarchy and then provide the details on the rolling-up procedure for single-unit stratum variance and covariance estimation.

SOII stratum TEI hierarchy

The SOII stratum TEI hierarchy has seven level structure, from the most detailed level (TEI6: 6-digit NAICS) to the most aggregated level (Domain: equivalent to the combinations of 2-digit NAICS)

1. TEI6 (6-digit NAICS)
2. TEI5 (5-digit NAICS)
3. TEI4 (4-digit NAICS)
4. TEI3 (3-digit NAICS),
5. Sector (two-digit NAICS or combinations of two-digit NAICS),
6. Super-Sector (more aggregated combinations of two-digit NAICS)
7. Domain (two types of combinations of two-digit NAICS: GP1AAA and SP1AAA)

A TEI for a SOII stratum could be of any level above except Domain that has only two categories: GP1AAA and SP1AAA (see examples in Table 1).

Table 1. Examples of SOII stratum TEI

| SOII stratum TEI | TEI Hierarchy Level |
|------------------|---------------------|
| 111411 | TEI6 |
| 111410 | TEI5 |
| 111400 | TEI4 |
| 111000 | TEI3 |
| GP2AFH | Sector |
| GP1NRM | Super Sector |
| GP1AAA | Domain |

I. Rolling-up within TEI with strata from the same employment size group

- Step one: a single establishment stratum with TEI will merge with other detailed strata along the TEI hierarchy structure (from **TEI5 to Domain**) to form a more aggregated stratum until this aggregated stratum has two or more usable establishments. All these detailed strata for aggregation have the same collection year, FIPS state code, and ownership code and belong to the same reported employment size group, case type, characteristic, cross characteristic. The starting aggregation level of TEI depends on the original stratum TEI. If the stratum TEI is TEI6, then the starting aggregation level of TEI would be TEI5; if the stratum TEI is a super-sector, then the starting aggregation level of TEI would be Domain.
- Step two: variance and covariance estimates will be calculated for the aggregated stratum and then assigned to the original single establishment stratum with TEI.

If the rolling-up within TEI with strata from the same reported employment size group could not form an aggregated stratum with two or more usable establishments, the rolling-up within TEI with strata from all five reported employment size groups will be applied.

II. Rolling-up within TEI with strata from all five reported employment size groups

- Step one: a single establishment stratum with TEI will merge with all other strata having the same TEI from different reported employment size groups to form a more aggregated stratum until this aggregated stratum has two or more usable establishments. All these detailed strata for aggregation have the same collection year, FIPS state code, and ownership code.
- Step two: if step one could not form an aggregated strata with two or more usable establishments, then repeat step one but at a more aggregated hierarchy TEI level (from **TEI5 to Domain**) until the aggregated stratum has at least two usable establishments.
- Step three: variance and covariance estimates will be calculated for the aggregated stratum and then assigned to the original single establishment stratum with TEI.

It is rare but it is possible that the above two main rolling-ups could not produce an aggregated stratum with two or more usable establishments. If this is the case, then all the strata that have the same collection year, FIPS state code, and ownership code as the single-unit detailed stratum will be used to form a most aggregated strata to calculate the variance and covariance estimates.