Deriving Inputs for the Allocation of State Samples

The SOII Universe Refinement Process Allocation and Sampling System (SURPASS) includes a sample allocation module which applies historical injury and illness data on Total Recordable Cases (TRC) as part of the sample determination. The process described below was developed to produce pseudo-estimates by size classes which aren't directly computed for the SOII.

1. Issue: State total recordable case (TRC) rate estimates by size class that are needed as inputs for the allocation of State samples do not exist. Size class estimates are available only at the sector and supersector levels for State estimates, but not for individual target estimation industries (TEIs). Since TEI size class estimates are needed as inputs for allocation, a proxy must be created where State TEIs are set below the sector or supersector level. Initial proposals for a proxy considered using national TEI size class estimates that would be applied directly to State TEIs. This approach ignores the overall State TEI estimate and does not account for any variation that may exist between State and national estimates for a TEI. A second proposal considered applying across all size classes the overall TRC rate (size 0) for the State TEI. This approach ignores often significant variation between TRC rates across size classes. Also considered was calculating actual State TEI size class estimates. This approach would require the release of such estimates to a Freedom of Information Act (FOIA) request and therefore would require the application of reliability and confidentiality processes for these estimates which could result in the suppression of some TEIs that are currently publishable.

2. Approach: Available State TEI estimates should be used to impute proxy *pseudo* TRC rates by size class for State TEIs. For State TEIs, the overall TRC rate (size 0) will be prorated against national TEI size class data (size 1-5) to create *pseudo* TRC rates for State TEI size classes. In cases where new State TEIs are established, State TEI employment estimates will be used to prorate against national TEI TRC rates by size class in order to calculate *pseudo* State TEI TRC size class estimates. (See New TEIs, below.)

3. Location of required inputs: Weighted case and hour estimates needed for this process can be derived from the estimates table in the OSHSUM database. Employment data, when needed, can be derived from the frame unit and employment counts in the SRS.

4. Required components for calculating rates:

a. Take the components of the State TEI TRC rate:
   i. The weighted count of injuries and illnesses, C_0
   ii. The weighted number of hours, EH_0
b. Take the components of the national TEI TRC size class (0-5) rates:
   i. The weighted counts of injuries and illnesses by size class, $CN_0 to CN_5$
   ii. The weighted number of hours by size class, $EHN_0 to EHN_5$

5. Calculating the rates:
   a. Calculate $\hat{C}_1 = C_0 \left(CN_1 / CN_0 \right)$ and $E\hat{H}_1 = EH_0 \left(EHN_1 / EHN_0 \right)$. Do the same for $\hat{C}_2 to \{ \hat{C}_5 ¿$ and $E\hat{H}_2 to E\{ \hat{H}_5 ¿$.
   b. Calculate a *pseudo* State TEI TRC rate for size class 1 as, $\left(\hat{C}_1 / E\hat{H}_1 \right) \times 200,000$. Do the same for size classes 2 through 5.
   c. Use the pseudo State TEI TRC rates by size class for the allocation.

6. Decision tree for missing values:

> Note: steps a. through e. apply only to private sector industries. Because data collection began in 2008 for government data in all States, input for the 2011 government sample will only include historical data from the 2008 survey.

   a. If the weighted components, i.e. weighted count of TRC and weighted hours, of the national TEI TRC rates by size class are missing, use the components of the State sector size class TRC rates. Output these components to the historical file;
   b. If the weighted components of the State sector size class TRC rates are missing, use the weighted components of the national sector size class TRC rates. Output these components to the historical file;
   c. If the weighted components of the national sector size class TRC rates are missing, use the components of the State supersector size class TRC rates. Output these components to the historical file;
   d. If the weighted components of the State supersector size class TRC rates are missing, use the components of the national supersector size class TRC rates. Output components to the historical file;

e. If a new State TEI has been created such that no estimates exist to be prorated against national size class estimates, use employments of the State TEI and national TEI to prorate. Do not output to historical file; (See New TEI, below.)
f. Where a. – e. do not produce a value for some cells (i.e. this most often occurs in State and local government), use the weighted components of the State TEI size class zero to fill these missing cells, but do not output these components for this cell to the historical file for these cells.
g. Where a. through f. do not produce a value for all cells (i.e. this most often occurs in State and local government), use a default value of 5 to fill the missing cell, but do not output from this cell to the historical file.

7. Extreme values for computed pseudo rates: Rates over 100 are not acceptable because they create a negative measure of size value. The measure of size is based on the square root of p(1-p) times the employment of the strata. Thus the maximum value for p(1-p) creates the largest possible measure of size. In this case the max p value would be p = 0.5 or a rate of 50 per 100 employees. In cases where the pseudo TRC rate exceeds 50, a maximum value of 50 will be applied to these cells. Output the actual weighted components from these cells to the historical table.

8. Residual TEIs: For residual State TEI TRC rates by size class follow the above procedure using State sector TRC rates by size class, employing the decision tree, steps b. through e., when needed. Since the composition of residuals can change from year to year based on a State's TEI selections, weighted estimates of counts and hours for residual TEIs will not be carried over from year to year. Residual size class *pseudo* rates will be calculated newly each year as discussed above.

9. Unpublishable estimates: If a State (or national) TEI TRC rate is unpublishable, use it anyway.

10. Null TRC rates: If the State TEI TRC rate is null, use the national TEI TRC rate by size class, then the State sector TRC rates by size class, then the national sector TRC rates by size class, then the State supersector TRC rates by size class, and then the national supersector TRC rates by size class. If national estimates (TEI, sector, or supersector) are used to prorate, then the estimate should be weighted using the employment ratio described below for new TEIs.

11. New TEIs: If a State selects a new TEI, no State TEI estimates will exist that can be prorated against national TEI size class estimates. In such cases, the components of national TEI size class estimates (size 0-5) for that corresponding State TEI will be used and weighted against employment for the State TEI. To do so:
    a. Determine the ratio of employment by size class (size 0 to 5) in the new State TEI as a proportion of employment in the national TEI, $\left(EMP_{State_{TEI}} / EMP_{National_{TEI}}\right)$ to derive $\hat{E}_0 to \{\hat{E}_5 ¿$.
    b. Take the components of the national TEI TRC size class (0-5) rates:
        i. The weighted counts of injuries and illnesses by size class, $CN_0 toCN_5$
        ii. The weighted number of hours by size class, $EHN_0 toEHN_5$.
    c. Use the components of the national TEI size class rates for this allocation.
    d. Multiply the employment ratio, $\hat{E}_0 to \{\hat{E}_5 ¿$, against the weighted counts and weighted hours, $CN_0 toCN_5$ and $EHN_0 toEHN_5$ to yield $\hat{C}_0 to \{\hat{C}_5 ¿$ and $E\hat{H}_0 toE\{\hat{H}_5 ¿$.
    e. Calculate a *pseudo* State TEI TRC rate for size class 1 as, $\left(\hat{C}_0 / E\hat{H}_0\right) \times 200,000$. Do the same for size classes 1 through 5.
    f. Use the pseudo State TEI TRC rates by size class for the allocation.

12. Building an historical file: Each year, save the weighted components of the State TEI size class values for $\hat{C}_1 to \{\hat{C}_5 ¿$ and $E\hat{H}_1 toE\{\hat{H}_5 ¿$, as described in 6a. – 6e above. This file will be the input into the next year's cycle as we use multiple years' worth of data. For example, for year 2 the formulas for $\hat{C}_1$ would be $\hat{C}_1^{CurrentYear-1}$ from the saved file plus the $\hat{C}_1^{CurrentYear}$ from the current year's data; same for EH_1. Also, if a State changes TEIs, the new TEI starts anew (like this year); however, we would continue to keep the old TEI values in the dataset from year to year in case the State decides to change the TEI back in a future year.