

## SAMPLE ALLOCATION AND SELECTION FOR THE NATIONAL COMPENSATION SURVEY

Lawrence R. Ernst, Christopher J. Guciardo, Chester H. Ponikowski, and Jason Tehonica

Ernst\_L@bls.gov, Guciardo\_C@bls.gov, Ponikowski\_C@bls.gov, Tehonica\_J@bls.gov

Bureau of Labor Statistics, 2 Massachusetts Ave. N.E., Room 3160, Washington, DC 20212-0001

**KEY WORDS:** Employment Cost Index, Employee Benefits Survey, Integrated sample

### 1. Introduction

Three of the Bureau of Labor Statistics compensation survey programs, the Employment Cost Index (ECI), the Employee Benefits Survey (EBS), and locality wage surveys, were integrated, creating one comprehensive National Compensation Survey (NCS) program. The ECI publishes national indexes which track quarterly and annual changes in employers' labor costs and also cost level information annually on the cost per hour worked of each component of compensation. The EBS publishes annually incidence and detailed provisions of selected employee benefit plans. The locality wage surveys program publishes locality and national occupational wage data. These surveys were integrated to expand the data products of the existing compensation programs, eliminate duplicate data collection and processing requirements, lessen the respondent burden, and to maximize the use of limited resources.

Prior to integration of these three surveys, the ECI/EBS and locality wage surveys used independent samples and data were collected separately by regional field staff. The ECI/EBS used a common sample for some time prior to integration, which was selected using a two-stage stratified design with probability proportionate to employment sampling at each stage and we will refer to this sample simply as the ECI sample. The first stage of sample selection was a sample of establishments selected with probability proportionate to establishment employment and the second stage was a sample of occupations selected with probability proportionate to occupation employment within each sample establishment. The locality wage surveys sample was also selected using two-stage stratified design with probability proportionate to employment sampling at each stage, but the first stage of sample selection was a sample of areas (PSUs) selected with probability proportionate to area employment. The areas with employment of 560,000 were selected with certainty. Initially, the first-stage sample consisted of 151 areas of which 33 were certainty areas and 118 were noncertainty areas. Three areas, Richmond VA, Dayton-Springfield, OH, and Huntsville, AL, with large federal work force in each of these areas were added to the list of certainty areas at the request of the President's Pay Agent. The second stage of sample selection was a stratified sample of establishments selected with probability proportionate to establishment employment within each sample area. Within each sample establishment data were collected for a fixed set of occupations.

After the integration of the three surveys, the NCS sample is selected using a three-stage stratified design with

probability proportionate to employment sampling at each stage. The first stage of sample selection is a probability sample of areas; the second stage is a probability sample of establishments within sampled areas; and the third stage is a probability sample of occupations within sampled areas and establishments.

In NCS Wage surveys, the successor to the locality wage surveys, most critical (certainty) areas, particularly smaller ones, are oversampled to insure adequate sample for an area publication. The area oversampling is not done for ECI because most estimates are national estimates. The ECI sample is a subsample of the larger NCS Wage sample, except for certain industries that were oversampled in ECI to reduce the difference in ECI industry variances; hence the two samples may be identical for some areas. The NCS Wage sample of 42,000 establishments is to be selected from 154 areas. The ECI sample of 18,000 establishments is to be selected from the 151 areas, that is, Raleigh-Durham, NC, Kalamazoo, MI, and Huntsville, AL were excluded. These three areas were excluded because they were not part of the original wage sample. The first two areas replaced two others that were originally sampled noncertainty areas that became certainty areas, while Huntsville was added to the list of certainty areas. The original 151-area design may be slightly more efficient for national estimates. Both samples, particularly the ECI sample, are larger than previously.

The integrated NCS sample consists of five rotating replacement sample panels, designated as NCS 101-105. Each of the five sample panels will be in sample for five years and then it will be replaced. Each year there will be a new panel selected from the most current frame. Units selected with certainty will be in all five replacement sample panels, with the selection using sampling intervals determined as if the NCS 101-105 Wage and ECI sample were each selected as a single large sample instead of being divided into five sample panels. (Note that throughout the paper when we speak of certainty units it is understood that this is conditional on the set of sample PSUs.) Since the NCS Wage sample size in a given sampling cell is usually larger than the ECI sample size, the number of units selected with certainty in the wage sample is larger. The five-panel sample and the first of the five single-panel samples were selected recently. In the process of selecting these samples we came across several interesting issues that may be of interest to survey practitioners who may have need to select an integrated sample that meets sample requirements for two or more surveys.

This paper describes the sample allocation and selection processes for the integrated sample. The allocation and selection processes were done in two passes. Section 2 covers the first pass of the allocation process, which determined the sample sizes for each sampling cell

for the entire NCS 101-105 Wage and ECI samples. This process involved allocating the ECI sample first to set a minimum sample within a sampling cell and then allocating the NCS Wage sample taking into account the ECI minimums, thus insuring that the ECI allocation requirements were met while allowing the ECI sample to be selected as a subsample of the NCS Wage sample. The integrated allocation process was complicated by the fact that the focus in NCS Wage is on allocating first to areas and then to industry strata, while the reverse is true for ECI. Section 3 covers the first pass of sample selection. This pass identifies the NCS Wage and ECI establishments that became certainty units in all five sample panels based on the first pass allocations. Section 4 covers the second pass of the allocation process. The allocation of five-panel ECI noncertainty units takes place during this pass. There are no such units for NCS Wage. In addition, for both NCS Wage and ECI, the allocation of the single-panel sample took place during the second pass. For each sampling cell this allocation prior to rounding is simply one-fifth the difference between the total cell allocation and the five-panel allocation and thus the main part of the work in this allocation consisted of a complex rounding process. Section 5 covers the second pass of sample selection using sample numbers from the second sample allocation pass. This pass selected NCS Wage and ECI 101 single-panel samples from the first pass NCS Wage sample units that are not certainty units for NCS Wage. Also this pass selected ECI five-panel noncertainty units from the set of units that are NCS Wage certainties. These ECI units were assigned to all five panels due to respondent burden concerns, that is, we did not want to have a given NCS Wage certainty unit collected as a wage only unit one year and then as wage and benefit unit the next year. Also, in this section, the plans for selecting the NCS Wage and ECI 102-105 single-panel samples are outlined.

## 2. First Pass Allocations

In this section we present the steps in the first pass allocations. This pass produced allocations for industry strata  $\times$  area cluster sampling cells for both NCS Wage and ECI that would be appropriate if the NCS 101-105 Wage and ECI were each selected as a single large sample instead of being divided into five sample panels. As described in the next section, these NCS Wage allocations were used to select an NCS Wage sample and the certainty units of this sample became certainty units in all five NCS 101-105 Wage sample panels. A subsample of this NCS Wage sample was selected using the ECI allocations, and its certainty units became ECI certainty units for all five panels.

The steps in obtaining the first pass allocations are as follows.

### 1. Inputs into the steps that follow.

The inputs into the steps of the allocation procedure that follow this step include the total sample sizes for 37 area clusters for the NCS Wage sample and 24 industry strata for the ECI sample. The differences in the types of input are due to the fact that in the NCS Wage products the

focus is on locality estimates, while for ECI the focus is on national estimates, including estimates for industry and occupational groups.

To obtain these inputs we began by reducing the total nationwide sample size for these surveys to account for the fact that we are only selecting private sector units in NCS 101-105 and are not selecting a new sample for the Aircraft Manufacturing industry since an entirely new sample had been selected for this industry relatively recently. This reduced the total sample to 37,284 establishments for NCS Wage and 15,980 for ECI.

For NCS Wage, 36 of the area clusters consist of a single certainty area, while the remaining cluster consists of the entire set of 118 noncertainty areas. To obtain the allocation among these 37 clusters, we began with the allocation for the prior NCS Wage sample designated as NCS 01-05. This allocation generally oversampled the certainty areas relative to the noncertainty areas in comparison with an allocation that samples from these two universes proportional to their aggregate employment. This is due to the need to produce separate publications for each of the certainty areas. Also, among the certainty areas the sample is greater for the larger areas, but by less than would be the case if the sample had been allocated among these areas proportional to employment. This is a compromise between an allocation appropriate for producing estimates of the same reliability for each of the areas, which would tend to yield a sample of approximately the same size for each area, and an allocation appropriate for producing national estimates, which would be closer to a proportional to employment allocation. For each of the 37 clusters, the NCS Wage 01-05 Wage sample was inflated by a factor to bring the total sample up to the 37,284 total for the NCS Wage 101-105 sample. The factors, which vary by cluster, are proportional to the ratio of the total employment in all establishments in the cluster with at least one employee to the total employment in all establishments in the cluster with at least 50 employees, thereby adjusting for the corresponding change in the universe in the NCS Wage 101-105 panels. That is, the universe for NCS Wage 101-105 consists of all establishments with at least one employee, while for NCS Wage 01-05 it was restricted to establishments with at least 50 employees.

Prior ECI samples were selected from 22 private industry strata that were defined as Standard Industrial Classification (SIC) codes or groups of codes. For ECI certain industries were oversampled in comparison with a proportional to total employment allocation due to the greater data variability in these industries. This is particularly true for industries with a high proportion of employees in commissioned jobs. The NCS 101-105 samples are selected from 24 private industry strata defined according to a different system, the North American Industrial Classification System (NAICS). To convert the ECI allocation from a prior sample, which consisted of 7,106 establishments, to a corresponding NAICS allocation we calculated for each NAICS stratum the sample size in the prior sample for each SIC stratum multiplied by the

proportion of the employment in the SIC stratum that was in the NAICS stratum and summed the results over the 22 SIC strata to obtain an allocation for that NAICS stratum. For each of the 24 NAICS strata, this allocation was then multiplied by 15980/7106 to obtain an ECI 101-105 allocation for the stratum, where this ratio was used to bring the total ECI 101-105 sample size up to the desired total of 15,980.

## 2. Allocation of ECI sample among sampling cells.

Each of the sampling cells for both NCS Wage and ECI is the intersection of one of the 24 industry strata and a cluster of area PSUs. As will be explained, the NCS Wage allocation uses 54 area clusters and the ECI allocation uses 12 area clusters. Thus there are  $24 \times 54 = 1296$  sampling cells for NCS Wage and  $24 \times 12 = 288$  for ECI.

Despite the fact that the ECI sample is a subsample of the NCS Wage sample, the allocation of the ECI 101-105 sample among the ECI sampling cells was determined prior to the corresponding allocation for the NCS Wage sample. This is because even though the desired ECI sample is smaller than the NCS Wage sample, the opposite is the case for certain sampling cells, principally due to the fact that ECI oversamples certain industries while NCS Wage allocates among the industries proportional to frame employment in the industry. An example of such a cell would be Finance in the New York CMSA. By determining the cell allocations for ECI first and then imposing a minimum value constraint for each NCS Wage cell allocation based on the ECI allocations, we insured that the ECI allocations could be obtained while selecting the ECI sample as a subsample of the NCS Wage sample.

For ECI, 11 of the 12 area clusters that define the sampling cells each consist of one of the 11 largest Consolidated Metropolitan Statistical Areas (CMSAs), and the final cluster consists of the remaining 140 ECI sample PSUs. For every industry stratum, each of the 11 largest CMSAs received a separate allocation because in the future we may produce locality ECI indexes for at least some of these areas. These 11 areas are not oversampled. In fact, for each of the 24 industry strata the allocation of the ECI sample, obtained in step 1, among the 12 area clusters is proportional to weighted employment, where the weighted employment was obtained by multiplying the frame employment for each PSU in the cluster by the ECI PSU weight and summing over all the PSUs in the cluster. The ECI PSU weight is the reciprocal of the probability of selecting the PSU in ECI. We also imposed the additional constraint of a minimum sample size of 1 for each cell with nonempty frame employment. By allocating each industry sample separately to each of the 11 largest areas with a minimum allocation of 1 in each of these areas, we insured that each of these areas directly represent themselves. The computer program used to do this allocation insures that the minimum allocation requirements are met, that the allocation does not exceed the frame size for any cell, and that, subject to these constraints, the allocation, ignoring rounding, is as close as possible to being proportional to the cell measure of size (MOS). The program employs a

new algorithm described in Ernst and Guciardo (2002). It is understood that in every step in the allocation process the constraints on the allocation for each industry  $\times$  area cluster cell are always the frame size for the maximum and, unless a different minimum is specified, 1 for the minimum if the frame is nonempty and 0 if it is empty. We refer to this minimum as the default minimum. The cells are either the sampling cells described at the beginning of this step or clusters of sampling cells. The MOS for a cell is its aggregate weighted frame employment, where the weight for each establishment in steps 2 and 3 is the ECI PSU weight, and in steps 4-7 is the NCS Wage PSU weight, which is the reciprocal of the probability of selection of the PSU for the NCS Wage sample.

In general, a proportional to size allocation is not integer-valued. Consequently, for each allocation described in this paper an additional step, known as a controlled rounding, was performed to obtain an integer-valued solution with same total sample size. The controlled rounding rounded up those unrounded sample sizes with the largest fractional remainders and rounded down the rest. The number rounded up is equal to the sum of the fractional remainders.

## 3. Minimum sample sizes for NCS Wage industry $\times$ area clusters based on the ECI allocations.

In order to insure that the NCS Wage allocations in the NCS Wage sampling cells are sufficient to obtain the ECI allocation in each of the ECI sampling cells, these ECI allocations were used to determine constraints on the minimum allocations to industry  $\times$  area cluster cells for the NCS Wage sample.

Separate minimums were set for the 35 NCS Wage certainty areas that are also ECI sample areas. The remaining 116 ECI sample PSUs comprise a single area cluster for this purpose. The three PSUs that are NCS Wage sample PSUs but not ECI sample PSUs obviously do not have any constraints on the minimum cell allocations for NCS Wage determined by the ECI allocations. Thus there are 36 area clusters for this step in the process, a smaller number than for the set of clusters used in defining the NCS Wage sampling cells.

For the 11 largest areas, the NCS Wage sample minimum for each industry  $\times$  area cell is simply the ECI minimum for that cell. For the remaining 25 area clusters, which together comprise the cluster of 140 ECI PSUs, the minimum was determined by allocating the ECI sample size for each industry stratum in the cluster of 140 PSUs among these 25 area clusters proportional to weighted frame employment for that industry, where the weight is the ECI PSU weight. No minimum value constraints were used in allocating the ECI sample size among these 25 area clusters. However, if the allocation for an industry was less than 16 for the cluster of 116 PSUs it was increased to 16 and if it was 0 for any of the other 24 area clusters it was increased to 1. It is explained in step 7 why for each industry stratum the minimum allocation must be at least 16 for the cluster of 116 PSUs.

#### 4. Industry allocations for NCS Wage certainty areas.

For each of the 36 certainty areas, the total sample size for the area, obtained in step 1, was allocated among the 24 industry strata proportional to frame employment. For each of the areas except Huntsville, the minimum for each sampling cell is the value determined in step 3. For Huntsville, the minimum for each cell is the default minimum since it is not an ECI sample area.

#### 5. NCS Wage total sample sizes for Raleigh, Kalamazoo and the cluster of 116 PSUs.

The NCS Wage allocations were determined for the 36 certainty areas in step 4. The remaining 118 areas were divided into three clusters in the current step, namely Raleigh, Kalamazoo, and the set of all the other 116 PSUs. The allocations for Raleigh and Kalamazoo were determined separately from the cluster consisting of 116 PSUs since Raleigh and Kalamazoo are not ECI sample areas and the minimums determined in step 3 do not apply to them. The total sample size across all industries for each of these three clusters was determined by allocating the total sample size for these 118 areas to these clusters proportional to aggregate weighted frame employment, where the weight for each PSU is the NCS Wage PSU weight. A minimum sample size constraint of 24 was imposed on each of these clusters to insure that there was enough sample in each cluster to allocate at least 1 unit to each industry stratum in the next step.

#### 6. Allocation to industries for 3 clusters of prev. step.

For each of these three clusters the total sample was allocated among the 24 industry strata proportional to the aggregate NCS Wage PSU weighted frame employment for the industry within the cluster. For Raleigh and Kalamazoo the minimum in each sampling cell is the default minimum. For the cluster of 116 PSUs minimums are as determined in step 3.

#### 7. Formation of 16 clusters from the cluster of 116 PSUs and allocations to these clusters.

We have now explained the allocations for 38 area clusters, that is the 36 critical areas, Raleigh, and Kalamazoo. We next explain how the sample for the cluster of 116 PSUs was allocated among 16 area clusters formed from the 116 PSUs cluster, the final 16 of the 54 area clusters that define the NCS Wage sampling cells.

Of these 16 clusters, 14 consist of a single PSU. These PSUs are the largest Metropolitan Statistical Areas (MSAs) that are not currently Pay Agent deliverable areas, that is areas for which separate estimates are provided to the Pay Agent. It is possible that some or all of these areas may become Pay Agent deliverable areas in the future and for that reason we wanted to give them separate allocations, analogous to the separate allocations for the 11 largest CMSAs for ECI. (Raleigh also is in this category, but its allocation was determined in steps 5 and 6.) The 15th cluster consists of the three Alaska-Hawaii noncertainty PSUs and the 16th cluster is the cluster consisting of the 99 remaining PSUs. The reason that the three Alaska-Hawaii noncertainty PSUs form a separate cluster is that Pay Agent estimates are only produced for the original 48 states.

For each of the 24 industries, the sample for the cluster of 116 PSUs determined in step 6 was allocated among

these 16 clusters proportional to the NCS Wage PSU weighted frame employment. The minimum for each NCS Wage sampling cell in these 16 clusters is the default minimum, which explains why a minimum sample size of 16, imposed in step 3, is needed for the entire cluster of 116 PSUs for each industry. Actually, for any industry for which the frame was empty for some of the 16 clusters, the minimum for the cluster of 116 PSUs could have been lowered to the number of nonempty clusters among the 16 clusters.

### 3. **First Pass Sample Selection**

As stated before, the NCS 101 samples were selected in two passes, one pass principally for selecting the five-panel units and one pass principally for selecting the NCS 101 single-panel units. Within each sampling pass, the NCS Wage sample was selected first and the ECI sample second. One of the requirements was that the ECI sample be a subsample of the NCS Wage sample. Therefore it was necessary to select the NCS Wage sample first.

The first sampling pass selected a sample based on the total, five-panel allocation. The sample size for each sampling cell was based on the full survey sample size for both the NCS Wage and ECI surveys. Using this allocation from the NCS Wage survey, a sample was selected for each area cluster. The MOS for each frame unit was calculated by multiplying the employment of the unit by the NCS Wage PSU weight corresponding to the geographical area of the unit. After the sample was selected, any sample unit with a first pass NCS Wage weight (that is, the reciprocal of the conditional probability of selecting the unit given the set of sample PSUs) of 1 was designated as a five-panel unit for the NCS Wage survey. All of the remaining units (with weight > 1) in the first pass NCS Wage sample become the frame for the selection of single-panel NCS 101 Wage units, as described in Section 5.

After the first pass sample was selected for the NCS Wage survey, the first pass sample was selected for the ECI survey. The frame units for the ECI selection were the NCS Wage first pass sample units. This insured that the ECI first pass sample would be a subset of the NCS Wage sample. As with the NCS Wage selection, the sample size for each sampling cell was based on the full survey allocation. The MOS used to select the ECI first pass sample was calculated by multiplying the employment of the unit, the ECI PSU weight for the unit, and the first pass NCS Wage weight. After the sample was selected, any ECI first pass sample unit with both an NCS Wage first pass weight of 1 and an ECI first pass weight of 1 (the reciprocal of the conditional probability of selecting the unit in the first pass ECI sample given the first pass NCS Wage sample) was designated as a five-panel certainty unit for the ECI survey. The remaining units in the ECI first pass sample are not necessarily in the final ECI sample. In fact, the remaining ECI sample units are selected during the second pass from two frames, one consisting of the NCS Wage second pass sample, the other the NCS Wage five panel units that are not one of these ECI five-panel certainty units from the first pass.

#### 4. Second Pass Allocations

The previous two sections described the first passes of sample allocation and selection, where the sample sizes for the NCS Wage and ECI 101-105 samples were determined and then allocated among sampling cells, and where the five-panel certainty units were determined for NCS Wage and ECI 101-105. This section describes the second pass of sample allocation, where the allocation among the cells of the single-panel sample sizes for both NCS Wage and ECI were determined, along with the allocation of the five-panel ECI noncertainty units.

For NCS Wage, all five-panel units, are certainty units. For ECI, because of the smaller ECI sample size some of these NCS Wage five-panel units would be noncertainty ECI sample units. If we assigned these ECI units to a single panel then there would be added respondent burden among those units in this group that were first selected in the ECI sample in any of the samples 102-105, since a wage initiation would be required for these units for the 101 sample followed by a later benefits initiation when selected for ECI. Consequently, these units if selected for ECI would be selected for all five panels. This required that the second pass sampling frame be split into two pieces for the ECI allocation and selection as described below.

##### 1. ECI 101-105 five-panel noncertainty sample sizes.

The frame for selecting the ECI 101 second pass consisted of two pieces: the NCS Wage five-panel units that are not ECI certainty units, from which the ECI five-panel noncertainty sample was selected, and the NCS Wage first pass sample units that are not NCS Wage five-panel units, from which the ECI 101 single panel sample was selected.

For the ECI five-panel noncertainty sample, the final sample size  $A'$  for a cell was found directly from the first pass allocation, rather than, as was done in the first pass, using an allocation proportional to some measure of size. First, for each cell, we computed a "target" sample size  $A = B * C$  for ECI five-panel noncertainty units, where  $B$  was the fraction of weighted employment in the ECI noncertainty frame (in the cell) that came from ECI five-panel units, and  $C$  was the ECI second pass target sample size, which was the total ECI sample size for the cell from the first pass minus the number of ECI five-panel certainty units. The weight in this calculation was the product of the ECI PSU weight and the NCS Wage first pass weight. The value  $A$  was then rounded to a value  $A'$ , using ordinary rounding with some exceptions. If  $C > 1$  and  $0 < A < C$ , and ordinary rounding of  $A$  was either 0 or  $C$ , then we set  $A'$  to 1 or  $C - 1$ , respectively. If  $C = 1$  and  $A > 0$ , then we set  $A' = 1$  and incremented  $C$  to 2, which increased the total sample, but we anticipated that this would occur so rarely that it would not be a practical issue. These exceptions were designed to ensure  $A' > 0$  whenever the piece of the frame used to select the ECI five-panel noncertainty sample units was nonempty, and  $C - A' > 0$  whenever the remainder of the ECI second pass frame was nonempty.

##### 2. NCS Wage and ECI 101 single-panel allocation

The NCS Wage and ECI single-panel sample sizes were determined together using the same allocation algorithm used in the first pass, except that we used a new set of values that served as inputs to the process. We needed new total NCS Wage sample sizes for each of the 37 NCS Wage area clusters described in step 1 of section 2, and new total ECI sample sizes for each of the 24 NAICS industry strata. We also needed new frame sizes, measures of size, and minimum sample sizes for each NCS Wage and each ECI sampling cell.

The minimum sample sizes for both NCS Wage and ECI were defined or computed the same way as in the first pass allocation, except where NCS Wage single-panel minimums were defined based on the ECI single-panel allocation. As part of this process, described in step 3 of Section 2, the ECI sample for the cluster of 140 PSUs was allocated among 25 NCS Wage area clusters proportional to the weighted frame employment. For the five-panel case, this frame contained all NCS Wage frame units and the weight was the ECI PSU weight. For the single-panel case, the frame was the NCS Wage first pass sample units that are not NCS Wage five-panel units and the weight was the product of the ECI PSU weight and the first pass NCS Wage weight.

For each cell the frame size for the NCS 101 single-panel allocations is the number of units in the NCS Wage first pass sample minus the number of NCS Wage five-panel certainty units.

The paragraphs below describe the remaining inputs for the NCS 101 single-panel allocations, which are cell measures of size, total NCS Wage sample sizes for each of the 37 NCS Wage area clusters, and total ECI sample sizes for each of the 24 NAICS industry strata. There is no plan to redo the entire single-panel allocation process for 102-105. But in these subsequent years, if the frame size drops below the 101 sample size for a sampling cell, then the new sample size must be reduced to the new frame size. The units that are freed up may be reallocated to the other cells, or dropped entirely. This has yet to be determined, depending on how many units are affected.

##### 3. NCS Wage 101 single-panel allocation inputs.

For each NCS Wage cell, the MOS for the single-panel allocation was found by taking the NCS Wage total sample size for the cell (from the first pass), subtracting the number of NCS Wage five-panel certainty units, and then dividing the difference by 5. This MOS replaces NCS Wage PSU weighted frame employment in the allocations in steps 4-7 of Section 2. The MOS for each cell for each annual single-panel sample is then roughly 1/5 the single-panel sample size for 101-105 for the cell. However, the MOS is generally not an integer and the allocation algorithm basically rounds the measures of size.

The total sample sizes for the 37 NCS Wage area clusters were found in the following way. First these cell measures of size were summed over all NCS Wage sampling cells and then the result, if not an integer, was rounded up to obtain the total single-panel sample size. Next, this rounded value was allocated across the 37 NCS Wage area clusters proportional to the MOS, followed by a

controlled rounding. The resulting total sample size for each area cluster is close, if not equal, to its total MOS, but all are integer values.

#### 4. ECI 101 single-panel allocation inputs.

For each cell, the ECI single panel MOS was computed by subtracting the ECI five-panel noncertainty sample size  $A'$  from the ECI target second pass sample size  $C$ , and then dividing the difference by five. This MOS replaces ECI Wage PSU weighted frame employment in the allocations in step 2 of Section 2.

The total sample size for each of the 24 NAICS industries was found by summing these cell measures of size over all ECI sampling cells, rounding the result upwards if not an integer to obtain the total single-panel ECI sample size, and then allocating it across the 24 industries proportional to the MOS, followed by a controlled rounding.

#### 5. Second Pass Sample Selection

The second sampling pass selected NCS Wage and ECI samples mainly based on the single-panel allocations. Using this allocation for NCS Wage, an NCS 101 Wage single-panel sample was selected for each of the 54 area clusters in the NCS Wage survey. The frame for the sample consisted of the first pass NCS Wage sample without the previously designated five-panel NCS Wage units. The MOS used to select the NCS Wage second pass sample was calculated by multiplying the employment of the unit, the NCS Wage PSU weight for the unit, and the first pass NCS Wage weight. The selected sample units were designated as NCS 101 Wage single-panel sample units. The final weight of an NCS 101 Wage single panel unit was calculated by multiplying its NCS Wage PSU weight, its first pass NCS Wage weight and its second pass NCS Wage weight (that is, the reciprocal of the conditional probability of selecting the unit in this sample given the first pass NCS Wage sample). The second pass sample units were added to the list of NCS Wage five-panel sample units to form the final NCS 101 Wage sample.

After the second pass sample was selected for the NCS Wage survey, the second pass sample was selected for the ECI survey. Two separate selections were done to create the ECI second pass sample. First, the ECI 101 single-panel sample was selected from the NCS 101 Wage single-panel sample. The MOS was obtained by multiplying the employment of the unit, the ECI PSU weight for the unit, and the product of the first and the second pass NCS Wage weights. The selected sample units were designated as ECI single-panel sample units. The final weight of an ECI 101 single-panel unit was calculated by multiplying all the weights used in calculating the MOS for the unit in this selection and the second pass ECI sample weight (that is, the reciprocal of the conditional probability of selecting the unit in the ECI 101 single-panel sample, given the second pass NCS Wage sample).

The other portion of the ECI second pass sample was selected from the five-panel NCS Wage certainty units that were not selected as five-panel ECI certainty units in the first pass. Unlike the other portions of the second pass

sample, this piece was selected based on a five-panel allocation. The MOS is the product of the employment of the unit and the ECI PSU weight for the unit. The final weight of a selected unit was calculated as the product of the ECI PSU weight and the reciprocal of the conditional probability of selecting the unit given the set of ECI sample PSUs. Even though the second term in this product is generally greater than 1, the units selected for this piece of the ECI sample were designated as five-panel units since these units had already been designated as five-panel units for the NCS Wage survey. Units from this frame not chosen for this sample were designated to be five-panel NCS Wage only units, that is they are not candidates for selection as ECI sample units in any of the five single-panel samples, which explains why this piece was selected based on a five-panel allocation. As explained previously, this was done to alleviate respondent burden. The final ECI 101 sample file was formed by combining the original ECI five-panel units, the ECI 101 single panel units, and the ECI noncertainty five-panel units.

The process for selecting samples 102 through 105, panels 2-5, is slightly different than the process for the 101 sample. Since the 101 sample selection process included the selection of the five-panel units, only the single panel sample units need to be selected for the remaining sample panels. The frames from which both the Wage and ECI sample units will be selected is not a first pass sample, but the original entire frame minus the previously selected NCS Wage five-panel units. Similarly, the sample weights of the 102-105 Wage and ECI units will have a single weight component replacing the first and second pass NCS Wage weights, reflecting the selection of the NCS Wage units in a single stage from the frame. As in the 101 samples, the ECI 102-105 samples will be selected from the NCS Wage 102-105 samples, respectively.

The final NCS Wage and ECI samples fulfill all of the requirements of the NCS sample design. The ECI sample is a complete subsample of the NCS Wage units. For each survey the total single panel sample size for each of the five panels will be approximately the same. Each survey has a number of designated five-panel units that when added to the five single panel samples will result in the full survey sample size.

#### 6. References

- Cochran, W. G. (1977). *Sampling Techniques*, 3rd ed. New York: John Wiley.
- Ernst, L. R. and Guciardo, C. J. (2002). *Allocating Sample to Strata Proportional to Aggregate Measure of Size with Both Upper and Lower Bounds on the Number of Units in Each Stratum*. 2002 Proceedings of the American Statistical Association, Section on Survey Research Methods, [CD-ROM], Alexandria, VA: American Statistical Association.

*Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics.*