

# **Update on the Evaluation of Sample Design Issues in the National Compensation Survey**

Gwyn R. Ferguson<sup>1</sup>, Joan L. Coleman<sup>2</sup>, and Chester H. Ponikowski<sup>3</sup>

<sup>1</sup>U.S. Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 3160,  
Washington, DC 20212

<sup>2</sup>U.S. Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 3160,  
Washington, DC 20212

<sup>3</sup>U.S. Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 3160,  
Washington, DC 20212

## **Abstract**

The National Compensation Survey is conducted by the Bureau of Labor Statistics to compute measures of the pay and benefits for America's workers. The current survey uses a three-stage sample design to select samples of areas, establishments, and jobs for which wage and benefit data are collected periodically over a five-year rotation. In recent years, several potential changes to this design have been explored to increase survey efficiency, adjust to budget changes, reduce respondent burden, and reduce design complexity. Design areas that have been studied include sample rotation, allocation, sample frame preparation, establishment selection, and sample initiation scheduling. This paper will update the discussion of these issues, describe the alternative approaches that have been explored, present results from the recent design research, and present the recommended changes to the general survey design. The work in this paper updates and significantly expands upon the work presented in 2010 JSM Paper "Evaluating Sample Design Issues in the National Compensation Survey".

**Key Words:** survey design, sample allocation, dependent sampling, respondent burden, sample rotation

## **1. Introduction**

The National Compensation Survey (NCS) is an establishment-based survey conducted by the U.S. Bureau of Labor Statistics (BLS). Over the last several decades, the NCS has undergone many changes leading up to the survey design currently in operation which has been used by NCS since the mid-1990's. In recent years, several potential changes to this design have been explored due to budget cuts, known issues with the current design, and an on-going effort to make the survey more efficient. In February 2011, the Bureau of Labor Statistics began implementing a change to the Locality Pay Survey (LPS) component of NCS which is used to produce annual occupational earnings data for the nation, each Census Division, and selected geographic areas. When this change is implemented, these occupational earnings estimates will be produced using a modeling technique that combines the national data from the NCS with the locality data from the Occupational Employment Statistics (OES) survey (BLS Handbook of Methods Chapter 3). All other data estimates computed using NCS data, including the Employment Cost Index, Employer Costs for Employee Compensation, and various measures of access and participation in employer provided benefits will continue to be produced. With the

elimination of a need to produce locality estimates directly from the NCS sample, the remaining NCS outputs can follow a more efficient national based sample design. At the same time many of the sample design issues associated with the current design that have been explored will be addressed by the redesign of the NCS current sample design.

The planned changes to the NCS sample design are considered major revisions to a Federal Government survey and are subject to the revision guidelines in the Office of Management and Budget Standards and Guidelines (OMB Standards 2006) and further described in the OMB Memorandum on "Guidance on Agency Survey and Statistical Information Collections" (OMB January 2006). Section 1 of these standards is applicable to the work on a new survey design for NCS. With this redesign and the research supporting the recommended changes, NCS has attempted to address all the concerns in these standards.

The current sample design issues and our plans for evaluating them were presented in the 2010 JSM paper "Evaluating Sample Design Issues in the National Compensation Survey" (Ferguson et al, JSM, 2010). This paper provides an update on the previous work and includes several recommended changes to the current sample design. Section 2 provides an overview of the current NCS sample design. Section 3 provides an overview of the proposed NCS sample redesign. Section 4 describes our proposed recommendations for the new sample design, describes our plans for transitioning from the current design to the new design, summarizes the research and analysis efforts that were done to develop the recommendations, and highlights issues that still need to be resolved. Section 5 provides a summary and conclusion of the redesign efforts.

## **2. Overview of the NCS Current Sample Design**

The NCS studies workers in private industry establishments, and in State and local government, in the 50 States and the District of Columbia. Establishments with one or more workers are included in the survey. The BLS Quarterly Census of Employment and Wages (QCEW) serves as the sampling frame for the NCS survey. The QCEW is created from State Unemployment Insurance (UI) files of establishments, which are obtained through the cooperation of the individual state agencies (BLS Handbook of Methods, Chapter 5).

The NCS sample consists of five rotating replacement sample panels for private industry establishments, an additional sample panel for State and local government entities, and an additional panel for private industry firms in the aircraft manufacturing industry. Each of the sample panels is in the sample for at least five years before it is replaced by a new sample panel selected annually from the most current frame.

The NCS sample is selected using a three stage stratified design with probability proportionate to employment size (PPS) 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 jobs within sampled areas and establishments.

The first stage of the NCS sample occurs at the national level across geographic areas. These Primary Sampling Units (PSUs) are based on the 2003 Office of Management and Budget (OMB) area definitions. Under the OMB definitions there are three types of

statistical areas. These area types are defined as Metropolitan, Micropolitan, and Combined Statistical Areas. Combined Statistical Areas (CSAs) are defined as a combination of adjacent Metropolitan and Micropolitan areas that meet certain conditions set by OMB. A number of counties exist outside of these areas and are referred to as Outside Core Based Statistical Areas (CBSA). For selection purposes, PSUs in these outside CBSA's consist of one or more adjacent counties. Where possible the counties were organized into clusters to create heterogeneous primary sampling units.

In 2004, a new area sample was selected for the NCS. This sample contained 152 areas. In this sample 57 areas were selected with certainty, where certainty areas are defined as having employment greater than 80 percent of the final sampling interval, which is obtained through an iterative process. The remaining areas consisted of 60 non-certainty metropolitan areas, 22 non-certainty micropolitan areas, and 13 non-certainty outside CBSA county clusters.

The second stage of this design occurs at the establishment level within each selected area. Establishments in the sampling frame are stratified by ownership and industry. Industries for the NCS are defined using the North American Industry Classification System (NAICS). Within each of the ownership by industry strata, NCS employs PPS systematic sampling with frame employment as the measure of size (MOS). To ensure that no unit has a probability of selection greater than one, we identify all units that would be selected with certainty before the sampling process, designate them as part of the sample, and set their sampling weights to one. These certainty units with a weight of one are identified once every five years and are included in each yearly sample until we identify a new set of certainty units. These units are referred to as multi-year certainties. By including them in every annual sample, we ensure that each sample represents the target population while making it easier operationally to process data for the various NCS outputs. During the selection process, approximately one-half of the establishments, the index portion, are sub-sampled and flagged to support the ECI, ECEC, and NCS Benefits products as well as the NCS wage products. The remaining establishments, the wage-only portion, are flagged to support the NCS wage products only. After the sample of establishments is selected, it is used for the third stage of the sampling process.

The third stage of this design occurs at the occupational level within each selected establishment. A sample of jobs is drawn from each of these establishments using PPS systematic sampling where the number of employees in the job is the measure of size. To ensure consistency across all establishments, the Standard Occupational Classification (SOC) manual is used to classify the selected jobs into occupations based upon the assigned duties. After this selection and classification we create our smallest aggregate unit known as a quote, which is a distinct combination of time or incentive pay, work level, collective bargaining status, full-time or part-time status, and establishment defined occupation.

Establishments in each sample are initiated over a one-year time period. During the initiation process, respondents are identified, jobs are selected, and respondents provide BLS with initial information about each selected job quote. All establishments are asked to provide BLS with employer provided wages and salaries for all workers in each selected job quote. Establishments in the index portion of the sample are also asked to provide the cost of each employer provided benefit, a description of each benefit offered to the employees in each selected job, and benefit access and provisions data such as the

number of employees who are offered the benefit, the number who partake of the benefit, and detailed descriptions of the benefit.

Respondents are asked to provide periodic updates for the initiated jobs for the next five years. Index respondents are asked to provide quarterly updates while wage respondents are asked to update their data annually. At the end of the five year update period, NCS thanks the respondents for supporting our survey and ceases to ask for updated data unless the respondent has been selected in a subsequent sample.

### **3. Overview of the Proposed New NCS Sample Design**

The redesigned sample will have two stages of sampling instead of current three stages and all sampled establishments will be used to support the production of all NCS product lines. In the first stage a sample of establishments will be selected and in the second stage a sample of jobs will be selected from sampled establishments. The new sample design will follow a three-year rotation for private industry with the selection of an aircraft manufacturing sample once every three years at the start of each private rotation. A sample of state and local governments will be selected approximately once every 10 years using a two stage sample design for which the details are still being finalized. Aircraft manufacturing is selected independently from the other private industries due to the desired level of publication detail in this industry and the number of available establishments in the sample frame. State and local governments are sampled less often than private industry because of the traditional high response rates and slower change in occupational mix for this sector of the economy.

The NCS will continue to use the most recent data available on the BLS QCEW (Quarterly Census of Employment and Wages) database to generate each establishment frame. Since the QCEW does not contain a full frame of railroad data, frame data for this industry will be obtained from an outside source in order to fully represent the industry in the NCS sample. For allocation and selection purposes, the NCS will use a measure of size (MOS) based on a modified employment value from the frame. This will ensure that the NCS will continue to maintain the same proportions of industry sample as the current design. The very large establishments will continue to be selected with certainty. The determination of certainty establishments will be based on the total three-year sample size and these units will be self-representing with an assigned weight of one. One-third of the total non-certainty sample units will be selected independently each year and their weights will be the inverse of the probability of selection.

All sample units will be assigned to one of four collection panels for initiation. Once a sample of establishments is selected and collection panels have been assigned, BLS Regional Office employees will review and refine the sample before collection begins. As part of this refinement process, establishments may be moved from one collection panel to another to coordinate initiation in firms with more than one establishment and/or to reduce travel costs associated with initiation efforts. Establishments will be initiated over a one year period with one collection panel required to be completed every three months. Once initiated, a unit will then be updated quarterly until it rotates out of the design, which will be approximately three years after the entire sample is initiated for non-certainty establishments. No newly initiated establishment will be used in the NCS estimates until the entire sample has been initiated and updated for a common/base quarter. During this base quarter, data for the newly initiated sample as well as all the

older samples will be updated. After the base quarter is completed, the entire sample will be added to the data available for estimation while the oldest sample in estimation is dropped from further updates and inclusion in the estimates.

Although NCS is implementing a new sample design, we are not making any changes to the data collection processes. The data elements to be collected, the timing of the collection processes, and the collection methods and modes will remain the same as they are under the current design unless they are changed for other business reasons. Thus, we expect that the respondent burden associated with our survey for any single collection effort will not change. In general, we expect that most individual respondents will now have a lower overall survey burden due to the shorter update period (3 years instead of 5). The overall respondent burden for all companies in the survey will be affected by these changes as documented in the OMB clearance package submitted in December 2010.

#### **4. Proposed NCS Sample Design – Recommendation, Research, Transition, and Remaining Efforts**

This section of the paper will describe the various components of the recommended new sample design in more detail, summarize the approach used to research each component of the design, describe the research findings and analysis, and identify any remaining issues and research needed to complete the design.

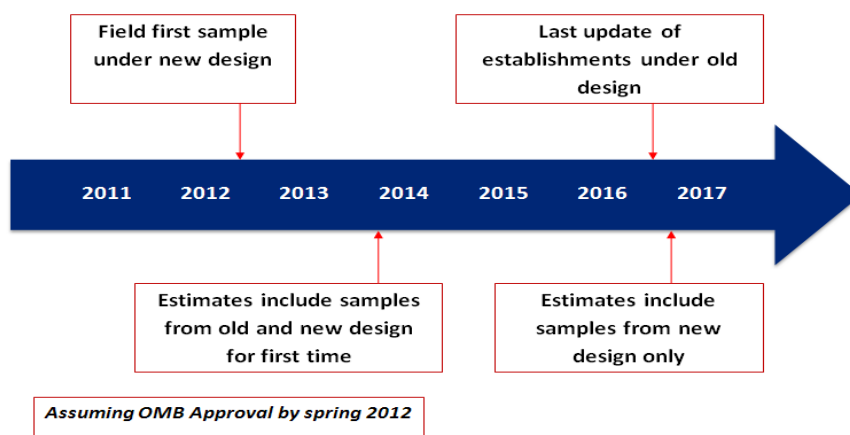
##### **4.1 Design Summary and Transition**

The NCS plans to move forward with this new national based sample design beginning with the production and collection of the next annual sample. Collection of the first sample under the national based design is scheduled to begin in the spring of 2012. After a 15 month initiation period and one quarter for base period wage and benefit cost collection, this sample will be used in NCS estimates for the first time for December 2013 estimates. Implementation of the new design affects all of the major processes starting with sample rotation and frame development through data collection and estimation.

The sample rotation will follow a three year panel design for the private industry sector with a total sample size set for the three year rotation. Large establishments with probabilities of selection greater than their sampling interval based on the three year sample size will be flagged as multi-year certainty establishments. These units will be selected in the first year and will remain fixed for three years. Independent non-certainty private samples will be selected and initiated every year with each sample being 1/3 of the total non-certainty sample size.

Although the details have not yet been completed, the plan for the State and Local Government sectors is to select a new single panel sample once every ten years. The first government sector sample will be selected after the initial rotation of three private samples. During the years of sampling and initiating a new State and Local Government sample, the existing private sample will not be replaced. As shown in Chart 1, transition to the new design will begin in the spring of 2012 with the fielding of the first private industry sample and will continue until late 2016<sup>i</sup> when the state and local government sample enters the estimates. Estimates computed for December 2013 through September 2016<sup>i</sup> will be based on data collected under a mix of the old three-stage sample design and the new two-stage sample design.

**Chart 1: Transition Timeline<sup>i</sup>**



#### **4.2 Private Industry Sample Rotation Options**

To evaluate the process of transitioning from a five year rotation to a three year rotation, research and analysis was conducted to study the effect this change would have on cost, frame creation, respondent burden, and operational complexity. The three year rotation presented several benefits to the NCS program, such as larger usable sample due to lower sample attrition, decrease in respondent burden as a result of shorter time in sample, and quicker implementation of changes such as area definition changes and SOC and NAICS classification system revisions. Of course with any change there will be some disadvantages as well. Some of the negative features of the three year rotation include cooperative responders dropping out sooner, increased initiation costs, and potentially inflated standard errors on year-to-year change estimates.

One of the major resource concerns with NCS is the number of Field Economists available to collect NCS data during the quarterly update period. Due to the complexity of our collection efforts, where up to 70 data elements are collected for up to 8 different jobs in each establishment, update data are obtained through personal contact with respondents during a six week period each quarter. Using average hours per schedule data and expected sample sizes, the number of full-time staff needed to collect the updates each quarter over the next several years was evaluated as well as the number of staff needed to initiate the selected establishments during the period when updates are not occurring.

To identify the best transition approach and to finish implementing prior budget cuts, current budget levels, workload levels, and staff availability were evaluated. Several options were evaluated but only one of the options balanced the NCS available staff with work required without requiring major effort and modifications to current systems and sample composition. The transition from the five-year to the three-year rotation will begin with the rotation of the oldest two sample groups before the first sample under the new design is even fielded. Once selection of samples under the new design begins, one sample per year will be rotated out until the third sample under the new nation-wide NCS sample design is complete. At that time, the last two samples selected under the area-based sample design will rotate out and the three-year design for the NCS private industry sample will be completely implemented. This methodology will align sample

sizes with staff levels supported by the NCS budget sooner than previously planned. It will also permit NCS to rotate out the last private industry sample selected under the old area sample one year sooner than currently planned. This methodology will also be implemented without any special sampling or weight adjustments.

#### **4.3 Establishment Sample Frames**

Sample frames will be constructed every year from the BLS QCEW database using the most recent quarter available. The frames (private as well as state and local government when applicable) must include all establishments from all 50 states and the District of Columbia for all in-scope industries. The in-scope industries for NCS are not changing in the new sample design and are listed in Appendix 1 for private industry. State and Local government samples use the same industries with the addition of NAICS 92, except for 928, for Public Administration.

The current NCS sample design does not use all establishments in the QCEW database for each state. In the QCEW, some establishments are not identified as belonging in a specific county within the state. These establishments are assigned a county code between 990 and 999. Since the current NCS sample design included selection of specific geographic areas, establishments in county codes greater than 990 were excluded from the sampling frame. Under the nation-wide design, which includes the entire country in the sampling frame, the establishments in these non-specific county codes could potentially be included in the sample frame. To determine if they should be included, the establishments in these counties were evaluated and their average monthly wages were compared to those of other establishments in the frame. Based on this analysis, NCS determined that we should include all of these establishments in our frame except for those with a county code of 996 which indicates that they are operating in foreign countries. Units with a county code of 996, while comprising of less than 0.1% of the establishments in the frame, operate outside the U.S. and generally pay a wage two to five times greater than the rest of the counties in a given state.

In many states, railroad establishments are not required to report to the State Unemployment Insurance file which is the basis for the BLS QCEW database. Therefore, QCEW does not contain a complete frame for this industry. Research was conducted and although the data are not available below the state level, the NCS will use the supplemental railroad frame constructed by the Occupational Employment Survey which includes all railroads in each state but does not map the railroads to specific cities or counties in the state. This will allow NCS to include railroads in the Census Division, Census Region, and National estimates. However, railroads will not be included in any estimates for the 15 largest metropolitan areas. In the future, the NCS may pursue refining the OES railroad frame to include area specific details for the fifteen largest metropolitan areas so that the railroad data can be used to feed the locality estimates as well.

#### **4.4 Establishment Sample Frame Refinement and Preparation**

For the private non-railroad frames, the NCS does not currently perform any frame refinement activities and does not plan to introduce any new frame refinement activities. To prepare the frame for sampling, several key variables are added to the data extracted from the QCEW database. First, every establishment must be assigned a geographic area code used to stratify the sample for allocation and selection. NCS will use 24 geographic areas, one for each of the 15 largest metropolitan areas by employment and one for the remainder of each Census Division. Second, each establishment is assigned to an

aggregate industry and a detailed industry based on its NAICS code on the QCEW database. For private industry samples other than aircraft manufacturing, there will be five aggregate industry strata. NCS will continue to use 23 detailed industries, each of which is mapped to one and only one of the aggregate industries. See Appendix 1 for a list of these aggregate industries, the detailed industries, and their associated NAICS codes. The aggregate industries being used for allocation and sample selection were created by analyzing the levels of aggregation in the various NCS publication lines.

Each establishment on the frame will be assigned an adjusted employment. For most establishments, the adjusted employment will be set equal to the third month's employment for the most recent quarter of QCEW data. However, if this monthly employment is equal to zero, the adjusted employment will be set equal to 1. In this manner, all in-scope establishments from the QCEW are given a chance of selection even those seasonal establishments with no employees in the designated month. NCS plans to continue research on alternative methods for setting the adjusted employment. Several options have been identified, including using the maximum monthly employment over 12 months, using an average employment over 12 months, changing values of 0 to 4, and changing values of 0 to 8. Another option excludes all establishments with a frame employment of 0 (using each of the options above) from the NCS frame. To date, we have analyzed initiation results for establishments that were assigned an arbitrary employment value of 1 from a sample selected using 2005 QCEW frame data. This analysis showed that more than half of these establishments were actually out of business. Future work will include obtaining a current QCEW frame and assigning each of the alternative adjusted employment values to each establishment. Fifty to one-hundred simulated samples will be selected using each of the adjusted employment values as the basis for the measure of size and mean monthly frame wages will be computed from each sample. Since mean monthly frame wages can be calculated from the entire frame, a comparison will be made between the average mean monthly wages for our simulated samples to those on the frame to help identify the method that generates the best results.

#### **4.5 Establishment Sample Allocation and Multi-Year Certainty Establishment Identification**

For private industry samples, NCS will continue to compute detailed allocations and identify multi-year certainty establishments once every three years under the three-year rotation cycle. If budget or resource levels change significantly, allocation and selection of multi-year certainties will be run prior to the selection of any sample to restart the three-year rotation. This section will first describe the process for allocating sample sizes and identifying the multi-year certainty establishments followed by the proposed methodology for when these processes are not executed for a specified sample.

The sample allocation process starts with a total budgeted sample size. The NCS will use targeted percentages across industries along with the frame data to determine how to distribute the sample units among the sampling cells. Due to the small sample size of the NCS, for the private non-aircraft manufacturing allocation, NCS will use a five aggregate-industry stratum allocation with a modified measure of size within each of the 23 detailed industries. This adjustment to the MOS will allow fewer strata but still control the number of units needed in the twenty-three detailed industries for which the NCS wants to publish estimates.

The total three-year NCS private non-aircraft manufacturing sample size is first allocated to the five aggregate industry strata. The size of each stratum is calculated so that the



distribution of the new sample mirrors the desired distribution of the sample in order to maximize the ability to meet publication goals. Next, each of the five aggregate stratum allocations is divided among the 24 geographic areas in proportion to the total adjusted employment of the frame units in the areas, resulting in 120 initial area-industry cell allocations. The measure of size (MOS) of a frame unit is the product of the unit's adjusted employment and an adjustment factor that is used to maintain the current distribution of the sample among the 23 detailed NCS industries.

Multi-year certainty units are identified using the initial cell allocations and the adjusted MOS. Each initial area-industry cell allocation is then reduced by the number of certainty units in the cell to create 120 non-certainty area-industry cell allocations. The MOS adjustment factors are recalculated to exclude the certainty units. Finally, the non-certainty allocations are divided between the three years of the sample design by distributing the non-certainty sample sizes across each of the three years. This distribution is done by dividing each size by three and assigning the integer portion of the result to each of the three years. The remainder is assigned to the appropriate number of years, one establishment at a time, in a manner that allows each annual sample to be the same size and the size for each sampling cell to vary by no more than one from year to year.

Under normal processing, sample allocation and multi-year certainty units will be executed once every three years. During years when prior year allocations and multi-year certainties are being used, the most recently identified set of multi-year certainty establishments will be removed from the frame for operational purposes and will be added to the final selected sample. This will ensure that each sample group represents the entire frame. The most recent non-certainty allocations and the sample frame without the multi-year certainties will be used to compute the final measure of size adjustment factor and to set the non-certainty sample size for each of the 120 area-industry sampling cells.

Although this approach has been tested and the NCS is satisfied with the outcome, there are still other approaches that should be researched as they may result in an even more efficient approach. In particular, yearly allocation and selection of certainty units would allow the option to run allocation and select independent single-year samples every year based on a sample size equal to  $\frac{1}{3}$  of the overall NCS sample. This would result in even fewer certainty units and make collection burden a much smaller concern than has been in the past. This would also eliminate the need to identify and remove certainty establishments from the frames in years two and three. The allocation research completed to date is for private industry only. Methodology is still needed for State and Local government samples. NCS currently uses 10 detailed industry sampling cells for State government samples and 10 detailed industry sampling cells for local government samples and selects the public sector samples only once every 10 years. The allocation process used for private industry will not work for these sectors, so a new process needs to be developed and tested for implementing the new allocation approach for State and Local Government samples.

#### **4.6 Establishment Sample Selection**

Under the proposed new design, NCS will select an independent non-certainty sample of private industry establishments (other than aircraft manufacturing) every year within each of the five aggregate industry and 24 geographic area sampling cells. Within each of the sampling cells, units will be sorted by detailed (23) industry, final adjusted MOS, and establishment identification number. The selection process will follow a systematic

Probability Proportionate to Size (PPS) approach where the measure of size includes the adjustment factor as defined in the allocation section. The multi-year certainty units identified in the previous step will be added to each non-certainty sample to form the entire establishment sample each year.

Sample weights will be assigned to each of the selected non-certainty establishments in the sample to represent the non-certainty portion of the frame. Units selected as certainty will be self-representing and will carry a sample weight of one. The sample weight for the non-certainty units will be the inverse of the probability of selection, or total MOS for the sampling cell (area by aggregate industry) divided by the adjusted MOS for the sampled unit.

As a result of following an independent sampling approach, there could be non-certainty establishments that are selected in multiple samples. As in the past, these will be referred to as overlap units and will not be reinitiated with the new sample. Instead, the data collected from the prior sample will be used in estimation with the weights from both the prior sample(s) and the newly selected sample. Although there should be fewer overlap units under the new national based design than exist in the area-based design, they could, and likely will, still exist. NCS is still evaluating other options for reinitiating these overlapping establishments and may decide to allow the use of prior sample data in some situations in the future.

#### **4.7 Initiation Collection Panel Assignment**

In the current NCS design, some geographic areas are initiated in a single quarter or collection panel and the remaining areas are designated for initiation across all four quarters. In areas where initiation occurred across all four quarters, establishments were assigned to a specific quarter or panel for initiation and all initiation work was required to be completed by the close date for that quarter. Since the wage only schedules under the old design were only updated once a year, it was vital that the panel assignments were evenly distributed across the year for areas that were initiated across all four panels. Under the new design, every establishment in the NCS will be updated every quarter. Therefore, all geographic areas will be treated as four-panel initiation areas. All sampled units will be assigned to one of four collection panels for initiation. The units will be assigned such that there is equal distribution across all four collection panels within Regional Offices, geographic areas and industries. During the initiation process, special procedures are followed for establishments with seasonal employment to ensure that the seasonal workers are given a chance of inclusion in the sample and are reflected in the NCS sample, regardless of when the initiation occurs.

Before the sample files are finalized, Regional Office staff review the initial panel assignments and have the opportunity to move establishments from one panel to another to better coordinate the collection efforts. This process is important especially when multiple locations of a firm have been sampled or when the sample only contains a few units in a remote geographic area. However, the regions will be instructed to move units in a way that results in a relatively equal distribution of establishments across each of the four initiation panels. National Office staff review the panel change requests to ensure that the sample distributions are still relatively even.

Since we will continue to use our current processes, no detailed analysis or research was conducted on how to assign initiation panels. But extensive analysis was done to determine how many field staff will be needed to complete the initiation efforts each

quarter. While conducting the allocation and multi-year certainty identification processes, we also researched the selection of the non-certainty establishments. In addition to the analysis that was done for allocation validation, we computed average sample sizes by BLS regional office areas for the year 1, the year 2, and the year 3 sample allocations across all the simulated samples. These average sample sizes were used to analyze BLS data collection resource needs by regional office once the new design is fully implemented. Staffing levels have been set for each region to ensure that sufficient staff levels are available to complete the initiation and update processes. Processes are also being implemented to evaluate staffing needs on a regular basis and to transfer work from one region to another as needed to balance workload across the nation.

Regional collection staff want to have more flexibility in the initiation process. They have asked for the ability to designate some establishments for collection after the assigned collection deadline for an initiation panel has passed. They have also asked for more flexibility in determining which establishments get collected at each point in the collection process. However, the current NCS data collection and processing systems will not allow for either of these options without major modifications. Until those modifications can be made and the potential effect on the quality of the survey data as a result of these possible changes can be evaluated, these processes will remain unchanged.

#### **4.8 Initiation, Data Updates, Response Rates, Estimation, and Publication**

Under the current NCS design, some establishments are flagged for wage only collection, while others are flagged for wage and benefit collection. Under the proposed new design, wage and benefit data will be collected from the entire sample. NCS will continue all collection activities under existing on-going processes and procedures. For all NCS outputs except the occupational earnings data, NCS will use the current estimation processes and procedures and will attempt to publish all levels of detail currently included in our publications. As described at the beginning of this paper, occupational earnings data by worker characteristics will be estimated using a modeling technique that combines earnings data from the OES with worker characteristics from the NCS. The specific modeling technique is still being finalized and will be described in a future paper. We will also continue to calculate variances for each of our estimates using a modified Faye's approach to balanced repeated replication.

As the new design rotates into estimation, NCS will be generating estimates and computing variances using a mix of samples selected from the old area-based design and the new national design. This will add some complexity to the estimation processes, and may cause some short-term loss of publishability due to the mix of sample types.

To determine the effect of the new design on the expected response rates, we computed response rates for the three most recent NCS samples and applied those rates to the simulated samples obtained during the allocation research. We used response rates computed by establishment size class and detailed industry grouping for the initial analysis. We then did an analysis by looking at response rates by metropolitan versus non-metropolitan areas. By applying our historical response rates to the average sample sizes, we were able to determine that the overall response rates for our samples should not change with the new design. Table 1 below shows this comparison for the Manufacturing Industry by Establishment Size Class and for the entire non-certainty sample. In this Table, the expected sample sizes are rounded estimates for a single sample and may not add to the total due to rounding. The sample sizes for recent samples cover all samples used in estimation before recent sample size reductions. While the viable

rates are the same within each industry and employment size category, the composition of the sampled establishments under the proposed design is different than under the current design. This different sample composition causes the aggregate viable rates for aggregate industries and for the entire sample to be slightly different between the current and proposed designs. Similar tables were evaluated for all industries, and various geographic groupings.

Table 1: Unweighted Response Rates – Current & Projected

|                       |           | Expected Sample Sizes under Proposed Design |        | Sample Sizes and Response Rates for Recent Samples Under Current Design |        |               |             |               |             | Projected Results under Proposed Design |             |               |             |
|-----------------------|-----------|---------------------------------------------|--------|-------------------------------------------------------------------------|--------|---------------|-------------|---------------|-------------|-----------------------------------------|-------------|---------------|-------------|
|                       |           | Sample Size                                 | % dist | Estab Count                                                             | % dist | Viable Estabs | Viable Rate | Usable Estabs | Usable Rate | Viable Estabs                           | Viable Rate | Usable Estabs | Usable Rate |
| Industry              | Estab Emp |                                             |        |                                                                         |        |               |             |               |             |                                         |             |               |             |
| Mfg excl aircraft mfg | <50       | 85                                          | 2.58%  | 235                                                                     | 1.95%  | 191           | 81%         | 140           | 73%         | 69                                      | 81%         | 51            | 74%         |
|                       | 50-99     | 45                                          | 1.37%  | 152                                                                     | 1.26%  | 125           | 82%         | 83            | 66%         | 37                                      | 82%         | 25            | 68%         |
|                       | 100-249   | 71                                          | 2.15%  | 295                                                                     | 2.45%  | 256           | 87%         | 163           | 64%         | 62                                      | 87%         | 39            | 63%         |
|                       | 250-499   | 50                                          | 1.50%  | 251                                                                     | 2.08%  | 216           | 86%         | 148           | 69%         | 43                                      | 86%         | 29            | 67%         |
|                       | 500+      | 89                                          | 2.71%  | 703                                                                     | 5.84%  | 651           | 93%         | 441           | 68%         | 83                                      | 93%         | 56            | 67%         |
| Industry total        |           | 341                                         | 10.31% | 1636                                                                    | 13.58% | 1439          | 88%         | 975           | 68%         | 294                                     | 86%         | 200           | 68%         |
|                       |           |                                             |        |                                                                         |        |               |             |               |             |                                         |             |               |             |
| Total                 |           | 3,302                                       | 100%   | 12,044                                                                  | 100%   | 10659         | 89%         | 7329          | 69%         | 2865                                    | 87%         | 1942          | 68%         |

To confirm that the sample design will support our current levels of publication, we compiled a master list of all the levels of detail currently being released for the ECI, ECEC, and the Benefits Incidence and Key Provisions products using private industry samples. This list of estimates included various industry break-outs, size of establishment categories, the 15 largest metropolitan areas, Census Divisions, and Census Regions. We then used the results from the simulated samples obtained during the simulation research process to determine if we expect to have enough data to continue releasing each of our current estimates. Based on this analysis, there were no published estimates lost as a result of published estimates having 50 or fewer sampled establishments. If a publication line had at least 200 sampled establishments, NCS assumes that we will be able to continue to publish the estimates. If a publication line had 50 or more but fewer than 200 sampled establishments, NCS expects that we will likely still be able to release the data but will monitor the situation carefully to ensure that the estimates meet our reliability and confidentiality publication criteria. There are only eight estimates that are currently published that fall into this area of concern, including two detailed industries and six of the 15 largest metropolitan areas. As the new design is implemented, NCS will need to pay careful attention to these estimates to ensure that they continue to meet all publication standards. Should any of them fail to meet publication standards, they may need to be dropped from one or more of our publication lines.

Data used to generate the detailed provisions estimates come from a single sample using only data collected during the initiation process. Since each initiation sample will be larger under the new design than under the old design, we do not expect any adverse effects on our ability to publish detailed accurate data for this product line. However, we have not yet done any research to validate this assumption or determine if we will be able to release additional detailed provisions estimates.

Some further work is needed to complete the estimation methodology for the new design and during the transition period. We still need to determine exactly how to create the variance strata used for variance estimation that will be used during the transition. We also need to determine how to create the variance strata to be used once the transition is complete. In all cases, we plan to continue computing variances using the current modified Faye's methodology described in the BLS Handbook of Methods, Chapter 8. The rest of the estimation processes will use the current methods and formulae. NCS also plans to continue publishing all current outputs for the ECI, ECEC, and NCS Benefits products in the same format and releases as are currently being used.

## **5. Summary and Conclusion**

The NCS has conducted several different research and analysis efforts to identify options and recommended approaches for a new sample rotation strategy and establishment selection design. Barring any unforeseen issues, we plan to implement this design starting with the next sample to be selected. That sample will be drawn in the fall of 2011 and will be fielded for initiation in the spring of 2012. After a 15 month initiation period, the sample will rotate into estimation with estimates generated for the December 2013 reference period. It will take two more years before all private industry samples selected under the old area-based design rotate out of the NCS estimates and an additional year before the State and Local government sample selected under the old area-based design rotates out of the NCS estimates. Thus, the first set of estimates computed from private industry samples selected only under the new national design with a 3-year rotation will be generated for the December 2015<sup>ii</sup> reference period. The first set of estimates computed for all workers, private and government, based on samples selected only under the new design will be generated a year later in December 2016<sup>i</sup>. During this transition period, most NCS estimates will be generated using a combination of establishments selected under the two designs. While we have a design that has been carefully reviewed and tested using sample simulations, there are still many implementation issues that need to be resolved and other design options that can be explored. We will work on these issues as resources permit.

## **References**

- Cochran, W. G. (1963), *Sampling Techniques*, New York: John Wiley & Sons, Inc.
- Ernst, L.R., Guciardo, C., Ponikowski, C.H., and Tehonica, J. (2002), "Sample Allocation and Selection for the National Compensation Survey," 2002 *Proceedings of the Section on Survey Research Methods*, Alexandria, VA: American Statistical Association.
- Ferguson, Gwyn R., Ponikowski, Chester, and Coleman, Joan (2010), "Evaluating Sample Design Issues in the National Compensation Survey", 2010 *Proceedings of the Section on Survey Research Methods*, Alexandria, VA: American Statistical Association.

- Izsak, Y., Ernst, L. R., Paben, S. P., Ponikowski, C.H. and Tehonica, J. (2003). "Redesign of the National Compensation Survey." *2003 Proceedings of the Section on Survey Research Methods*, [CD-ROM], Alexandria, VA: American Statistical Association.
- Izsak Y., Ernst, L. R., McNulty E., Paben, S. P., Ponikowski, C. H., Springer G., and Tehonica, J. (2005). "Update on the Redesign of the National Compensation Survey." *2005 Proceedings of the Section on Survey Research Methods*, [CD-ROM], Alexandria, VA: American Statistical Association
- Office of Management and Budget, "Standards and Guidelines for Statistical Surveys", September 2006, [http://www.whitehouse.gov/omb/assets/omb/inforeg/statpolicy/standards\\_stat\\_surveys.pdf](http://www.whitehouse.gov/omb/assets/omb/inforeg/statpolicy/standards_stat_surveys.pdf).
- Office of Management and Budget, "Guidance on Agency Survey and Statistical Information Collections" Memorandum, January 20, 2006, [http://www.whitehouse.gov/omb/assets/omb/inforeg/pmc\\_survey\\_guidance\\_2006.pdf](http://www.whitehouse.gov/omb/assets/omb/inforeg/pmc_survey_guidance_2006.pdf)
- Ojo, O. E. and Ponikowski, C. H. (2010), "Evaluating the Effect of Dependent Sampling on the National Compensation Survey Earnings Estimates", *2010 Proceedings of the Section on Survey Research Methods*, [CD-ROM], Alexandria, VA: American Statistical Association.
- Schumann, Richard E., "Occupational Selection and Leveling in the National Compensation Survey", U.S. Bureau of Labor Statistics, Compensation and Working Conditions Online, Originally Posted on August 31, 2011, <http://www.bls.gov/opub/cwc/cm20110829ar01p1.htm>
- U.S. Bureau of Labor Statistics (1997) *BLS Handbook of Methods*, Employment and Wages Covered by Unemployment Insurance, Chapter 5. <http://www.bls.gov/opub/hom/pdf/homch5.pdf>
- U.S. Bureau of Labor Statistics (2008) *BLS Handbook of Methods*, National Compensation Measures, Chapter 8. <http://www.bls.gov/opub/hom/pdf/homch8.pdf>
- U.S. Bureau of Labor Statistics (2008) *BLS Handbook of Methods*, Occupational Employment Statistics, Chapter 3. <http://www.bls.gov/opub/hom/pdf/homch3.pdf>

***Note: Any opinions expressed in this paper are those of the author(s) and do not constitute policy of the Bureau of Labor Statistics.***

Appendix 1: NCS Industry Cells for Sample Selection and Allocation

| <b>Aggregate Industry</b>                         | <b>Detailed Industry</b>                         | <b>Included NAICS Codes</b> |
|---------------------------------------------------|--------------------------------------------------|-----------------------------|
| Education                                         | Educational Services (Rest of)                   | 61 (excl 6111-6113)         |
| Education                                         | Elementary and Secondary Schools                 | 6111                        |
| Education                                         | Junior Colleges, Colleges and Universities       | 6112, 6113                  |
| Finance, Insurance and Real Estate                | Finance (Rest of)                                | 52 (excl 524)               |
| Finance, Insurance and Real Estate                | Insurance                                        | 524                         |
| Finance, Insurance and Real Estate                | Real Estate, Renting, Leasing                    | 53                          |
| Goods Producing                                   | Mining                                           | 21                          |
| Goods Producing                                   | Construction                                     | 23                          |
| Goods Producing                                   | Manufacturing (excluding aircraft manufacturing) | 31-33 (excl 336411)         |
| Health Care, including Hospitals and Nursing Care | Healthcare, Social Assistance (Rest of)          | 62 (excl 622, 623)          |
| Health Care, including Hospitals and Nursing Care | Hospitals                                        | 622                         |
| Health Care, including Hospitals and Nursing Care | Nursing and Residential Care Facilities          | 623                         |
| Service Providing                                 | Utilities                                        | 22                          |
| Service Providing                                 | Wholesale Trade                                  | 42                          |
| Service Providing                                 | Retail Trade                                     | 44-45                       |
| Service Providing                                 | Transportation and Warehousing                   | 48-49                       |
| Service Providing                                 | Information                                      | 51                          |
| Service Providing                                 | Professional, Scientific, Technical              | 54                          |
| Service Providing                                 | Management of Companies and Enterprises          | 55                          |
| Service Providing                                 | Admin., Support, Waste Management                | 56                          |
| Service Providing                                 | Arts, Entertainment, Recreation                  | 71                          |
| Service Providing                                 | Accommodation and Food Services                  | 72                          |
| Service Providing                                 | Other Services (excl Public Administration)      | 81 (excl 814)               |

---

<sup>i</sup> Note that the version of this paper published through the American Statistical Association has an error. The date has been corrected in this version to be 2016, not 2017.

<sup>ii</sup> Note that the version of this paper published through the American Statistical Association has an error. The date has been corrected in this version to be 2015, not 2016.