Evaluating Sample Design Issues in the National Compensation Survey October 2010

Gwyn R. Ferguson[1], Chester Ponikowski[2], and Joan Coleman[3]
[1]U.S. Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 3160, Washington, DC 20212
[2]U.S. Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 3160, Washington, DC 20212
[3]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 measure employment cost levels and trends, incidence of employer-provided benefits, benefit plan provisions, occupational earnings by geographic area, and occupational pay comparisons between areas. The current survey design 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 due to budget cuts, known issues with the current design, and an on-going effort to make the survey more efficient. This paper discusses the issues and alternative approaches to the current design being explored and presents some recommended changes to the general survey design.

Key Words: survey design, 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. This paper presents an overview of the NCS that includes its scope and major product lines in Section 2. Section 3 provides an overview of the current sample design and rotation strategy. Section 4 outlines several issues with the design while Section 5 describes the objectives for any design changes. Some challenges and constraints on design changes and research are described in Section 6. Section 7 describes the efforts currently underway to evaluate potential changes to the sample design and rotation strategy and future plans for additional research to evaluate the issues. The paper concludes with a summary of the status of the analysis in Section 8.

2. Survey Background

The NCS provides comprehensive measures of occupational earnings, employer costs of employee compensation, compensation trends, wages in one geographic area relative to other geographic areas, and the incidence and provisions of employer-provided benefits (BLS Handbook of Methods, Chapter 8). The Employment Cost Index (ECI)—a Principal Federal Economic Indicator—is estimated from data collected by the NCS.

The NCS produces several types of data with varying degrees of frequency as summarized below.

- Employment Cost Index (ECI) data are released quarterly
- Employer Costs for Employee Compensation (ECEC) data are released quarterly
- Incidence and Provisions of Employer Provided Benefits data are released annually
- Detailed Provisions for employer provided health insurance, defined benefit retirement plans, and defined contribution retirement plans are released once a year with a focus on one of these benefit areas each year
- Occupational earnings data for the nation, each Census Division, and selected geographic areas are released once a year
- Occupational pay comparisons for approximately 80 geographic areas are released once a year

The NCS covers workers in private industry establishments, and in State and local government, in the 50 States and the District of Columbia. For the NCS, the term *"civilian workers"* denotes workers in private industry and workers in State and local government. Establishments with one or more workers are included in the survey. Excluded from the survey are workers in the Federal Government and quasi-Federal agencies, military personnel, agricultural workers, workers in private households, the self-employed, volunteers, unpaid workers, individuals receiving long-term disability compensation, individuals working overseas, individuals who set their own pay (for example, proprietors, owners, major stockholders, and partners in unincorporated firms), and those paid token wages.

3. NCS Current Survey Design

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 occupations 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 definition 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. Outside of these areas exists a number of counties. These counties 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 are organized into clusters to create heterogeneous primary sampling units.

In 2004, a new area sample was selected for the NCS. This sample contains 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. All establishment samples selected since 2005 have been selected from the 2004 area sample while establishment samples selected in 2004 and earlier were chosen from the previous sample of areas. Until all establishment samples selected before 2005 rotate out of the survey in 2012, data collected from samples selected from the previous area sample and from the 2004 area sample will be included in the various NCS published estimates.

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 occupation, 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 occupations 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. The chart below shows the current rotation strategy. Each row in this chart is an independent establishment sample. The yellow rows indicate the introduction of a new area sample while the vertical red lines indicate the transition period when estimates include establishments selected from two area samples.

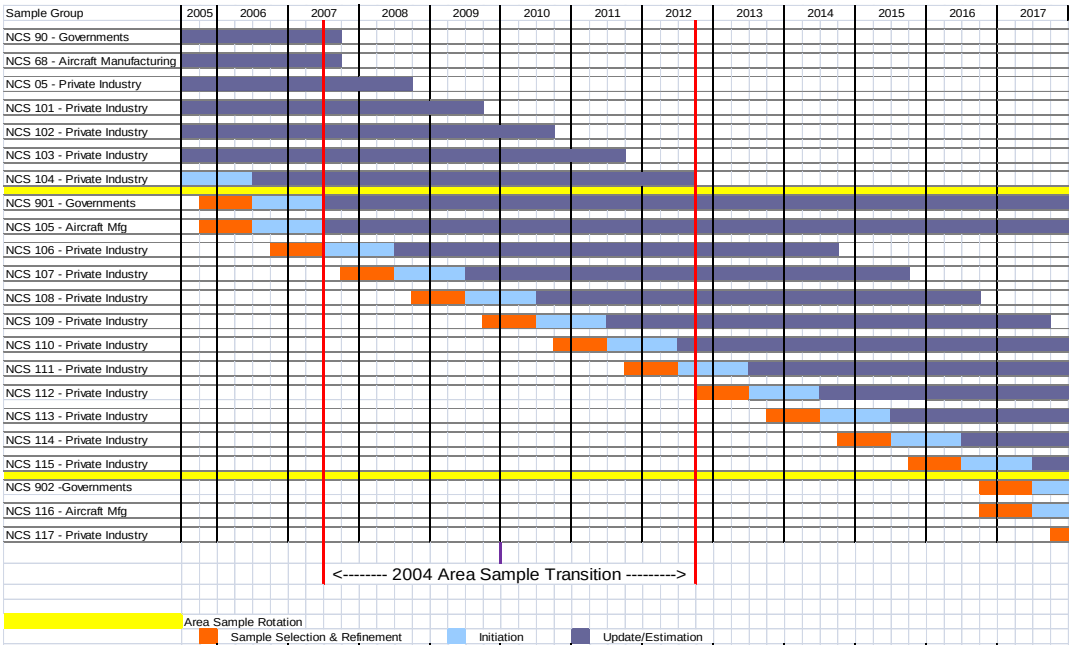

Figure 1: NCS Sample Rotation Chart

4. Issues with the Current Design

Since the introduction of the current sample design, NCS has used the same basic design and rotation strategy. Initially, the design worked well and there were no major problems with its implementation. But as we used the design for a longer period of time, several things changed leading the survey to conclude that we had some potential issues with the design.

In the last several years, NCS has completed several studies of the current sample design.

- In 2009, NCS completed an internal review of the sample design and issues related to it from the NCS employee's perspective. This review identified several concerns with the sample design that prompted us to begin an evaluation of several major components of the design.
- From 2006 to 2009, the BLS compared the National Compensation Survey to the Occupational Employment Statistics (OES) survey in terms of concepts, measurement objectives, and sample design. The OES is an establishment based survey that provides employment counts and earnings by occupation (BLS Handbook of Methods, Chapter 3). Based on this comparison, we identified several areas where the two survey designs were different and could potentially be changed to be more consistent, although we agreed that additional research was needed before making any changes.
- From fiscal year 2004 through fiscal year 2009, the NCS budget was cut four different times, resulting in reduced sample sizes each time. While the budget reductions were implemented and the sample sizes were reduced, the process for implementing the reductions became more difficult due to the design of the sample allocation process which assumes fixed sample sizes for five years at a time. Most recently, the President's proposed fiscal year 2011 budget calls for the elimination of the Locality Pay Survey component of the NCS with a transition to a national design and a reduction in the overall sample size for the remaining portions of the survey. Although no final budget for fiscal year 2011 has been approved as of the writing of this paper, NCS will need to make several changes to the sample design if the President's proposed budget is implemented.

Using the results of each of these studies and budget cuts, NCS staff compiled a comprehensive list of issues with the sample design as described below.

1) As a result of independent PPS sampling and identification of multi-year certainty establishments, many units are selected year after year in the NCS. This frequent sampling and the continuous need to obtain current wage and benefits data from employers has the potential of overburdening some respondents by requesting data for multiple samples simultaneously.
2) Under the current five-year rotation, it takes several years to make changes to our scope, definitions, sampled areas, etc. In order to meet publishability criteria, we need to initiate several years of samples under new definitions, areas, and/or processes before fully moving to the new coverage.
3) The five-year rotation scheme is a concern, especially for private industry establishments which remain in the survey for six or more years, up to one year for sample initiation and at least five years for sample updates. This also creates an attrition problem -- the rate of attrition in private industry is approximately 1 percent per quarter; thus, sample groups approaching replacement may be more prone to bias due to larger nonresponse.
4) The current allocation methodology assumes a constant budget and fixed sample size levels over a five year period. Under this methodology, the allocation process should be executed at the beginning of each five year period, multi-year certainty establishments are to be identified and held constant for the five year period, and annual samples should be selected from a new frame that excludes the multi-year certainties using the previously identified sample sizes. The recent budget climate has not supported this continuity. In fact, NCS has seen budget reductions in four of the

last seven budget years, forcing the survey to implement sample size changes in the middle of the five year period. These budget changes have resulted in uneven sample sizes from one year to the next and a set of multi-year certainty establishments that is larger than would be selected given the current sample size.

5) The President's proposed budget for fiscal year 2011 calls for the elimination of the locality pay portion of the NCS and a change to a national sample design instead of the current area-based design. Under this proposal, NCS will cease to produce occupational earnings estimates as direct computations from sample data, eliminating the requirement to sample geographic areas in support of the area-based earnings data. If Congress approves this budget proposal, several changes will need to be made to the sample design including eliminating the first stage of sampling to select areas, preparing a sample frame from establishments across the nation instead of the sampled areas, eliminating the process of identifying establishments that will support only the wage products, and updating the weighting modules to reflect these changes.
6) Including establishments that appear on the frame with zero employment during the frame reference month may be causing NCS to sample establishments that are no longer in business, also leading to sample inefficiencies due to a smaller useable sample. NCS currently looks at a single month of employment data on the frame. In seasonal establishments, this month's employment may be zero even though the establishment has many employees in other months of the year. Thus, NCS decided to include all establishments on the frame and set the employment to one when it is equal to zero on the frame. However, this may lead to inclusion of establishments that have gone out of business during the quarter from which the frame was extracted or giving establishments very small employment values when they typically have much larger employment.
7) Complexity is also an issue with the current sample, but it is inherent. The NCS sample data are used to produce many outputs within a limited budget constraint and different subsets of the NCS data are used for the various outputs. For example, the ECI is produced using all samples that have completed the initiation process and have not yet rotated out of the survey. However the detailed provisions products only include data from the most recently initiated sample. Thus, synchronization of changes to the methods and concepts is difficult especially since the sample affects every aspect of the surveys, from data collection to review, estimation, and publications.
8) Response rates are lower than desired and require nonresponse bias studies. NCS has begun conducting nonresponse bias studies on the wage outputs produced by the survey as documented in Ponikowski et al (2006) and Crockett et al (2008).
9) New establishments on the sample frame, sometimes called birth establishments, are not included in every establishment sample included in the NCS estimates. Since most of the NCS estimates include data from multiple samples, this can lead to underrepresentation of the current frame due to exclusion of these newly formed businesses in samples that have been in the survey for two or more years.

5. Design Objectives

As NCS has begun an evaluation of each of the sample design and rotation issues, we have agreed that all design changes must meet several objectives. First, the new design alternatives must meet the NCS measurement objectives related to the ECI, ECEC, and Benefits products. The critical measurement objectives for the NCS are that the not seasonally adjusted Total Compensation and Wages series of the ECI have a standard

error of change that does not exceed 0.3 for the 3-month changes and 0.5 for 12-month changes at least 75 percent of the time. In addition, the survey should support production of the various outputs for the Civilian, Private, and State and local government sectors of the economy at various specified levels of detail by ownership, industry, and geographic area. Second, the design shall allow quicker implementation of survey changes, especially mandated changes such as area definitions, NAICS, and SOC. Third, any design changes must address respondent burden issues related to frequent sampling of some employers. Fourth, the new design should provide the ability to adjust sample sizes as necessary with fluctuating budgets. Fifth, the new design should address response rate concerns. And, finally, the new design changes should decrease the complexity of the survey design, where possible.

6. Challenges and Constraints

As with all survey design research, there are several challenges and constraints on this work, some imposed by external forces and some imposed internally. For the NCS, the primary external challenges deal with the Federal budget cycle. If the fiscal year 2011 budget as proposed by the President is approved by Congress, we will need to begin implementing sample design changes quickly. However, the Bureau of Labor Statistics and NCS in particular have not received any extra funds or resources to develop or evaluate changes to our design. In addition, due to the budget cuts in recent years, NCS has been unable to replace staff as they have left the Bureau resulting in fewer available research staff than we had for earlier design efforts. Thus, all research must be conducted using current staff and resources by implementing production efficiencies and delaying other planned improvements to the survey processes. The primary internal constraint facing us as we evaluate potential design changes is the time needed to implement any approved changes. Since NCS selects a single sample each year, design changes must be approved far enough in advance of the selection process to ensure adequate time to develop and test all changes to our sampling systems before the next sample selection process begins. For example, the design changes need to be known and approved in the fall of 2010 for the sample scheduled to begin collection in the spring of 2012.

7. Research and Testing

7.1 Previous Research

Several different research efforts to explore the issues and identify potential alternatives for resolving them have already been completed. As mentioned in Section 4, NCS completed an internal review of the sample design from the NCS employee's perspective in 2009. This review identified several issues with the design but did not propose any specific ways to resolve those issues.

Also in 2009, BLS compared the NCS survey design and measurement objectives to those for the Occupational Employment Statistics survey. This comparison identified several areas where the two BLS surveys were different and made four recommendations for further research. The first recommendation was that the two surveys should adopt common concepts, definitions, and coverage. One of the specific areas mentioned under this recommendation was that the two surveys should look at the definition of an establishment and the establishment's measure of size used for sample selection purposes since their current approaches were different. The second recommendation was that the

BLS should use econometric modeling techniques to present a single set of wage estimates to users. The third recommendation was to streamline and gradually expand coordinated collection and coding of establishments which appear in both of these surveys. The fourth recommendation from this comparison was to implement fully overlapping private industry sampling in 2014 or later. This fourth recommendation included some specific areas for further study between the two surveys to find ways to better coordinate their survey designs. Several of the detailed recommendations included in this report are being researched at this time as described in Section 7.2.

In early 2010, NCS completed a detailed analysis of the effort required to collect NCS data. This analysis included a detailed study of the amount of time needed to initiate an establishment into the survey as well as the amount of time needed to update the data for each establishment on a periodic basis. Using this data, NCS was able to evaluate the staffing levels needed for both a five-year rotation and a three-year rotation. This analysis showed that the NCS did not have sufficient staff levels to implement a three-year rotation under the current design. However, if the proposal to eliminate the locality wage portion of the survey is approved in the fiscal year 2011 budget, the analysis showed that a three-year rotation would be feasible from a staffing viewpoint and would help balance the workload over the course of each year. Since a three-year rotation would permit NCS to implement change faster and helps balance data collection workload, NCS decided to focus our remaining sample design research on a three-year rotation. Where feasible, we will compare the proposed three-year rotation design changes to the results that would be obtained under the current sample design and five-year rotation.

7.2 Current Research

Using the previous research as a base and the design objectives as a goal, NCS identified several areas for further research. By looking at the sample life-cycle, we identified and have begun researching several potential changes related to national design stratification, sample allocation, sample selection of establishments and alternatives for large establishments, the method of assigning an employment measure of size to establishment on the sample frame, and an analysis of the frame coverage.

The current NCS establishment sample design stratifies the frame by geographic area and industry as defined by NAICS. The geographic areas are defined to be the areas which were selected during the area selection phase of the design. If the President's proposed fiscal year 2011 budget is implemented, NCS will move to a national design and cease the selection of geographic areas. In evaluating the stratification needs under a national design, NCS considered the critical estimates desired for the ECI, ECEC, and benefits product lines. The NCS currently produces ECI and ECEC estimates for each of the fifteen largest metropolitan statistical areas and plans to continue these series. In addition, NCS publishes several data products by Census Division and wishes to continue publication of these series. Based on these needs, we identified 24 geographic strata covering the 50 states and the District of Columbia – one for each of the fifteen largest metropolitan areas and one for the remaining states and counties in each of the nine Census Divisions. In addition, NCS publishes many outputs based on industry. Currently the NCS publishes data based on 24 detailed industry groupings for private establishments. Twenty-three of these detailed industry groupings are included in each of the private industry establishment samples while the last detailed private industry

grouping, aircraft manufacturing, is included in a separate sample panel[1]. For stratification purposes, we will continue to use these detailed industry groupings in the sample design. However, all current research efforts will focus on the 23 detailed industries typically included in each private industry establishment sample.

One of the major challenges with the current NCS sample design is that it stratifies the private industry establishment samples by 152 geographic areas and 23 detailed industry groupings, resulting in 3,496 potential sampling cells for private industry establishments. With an overall sample size of approximately 13,400 private industry establishments to be selected over a five year period, there are not enough establishments to fully allocate the sample size to every potential sampling cell. So the NCS implemented a controlled rounding approach to distribute the allocation across the sampling cells and the five years of establishment samples. This process is complicated and not designed for frequent changes in sample size. With the move to a National design, NCS would only have 24 geographic cells. If we keep the 23 detailed private industry cells (excluding aircraft manufacturing) we would have a total of 552 sampling cells. But with a new private industry sample size of slightly less than 10,000 establishments over a full three-year rotation, it would still be difficult to ensure coverage of each sampling cell every year. So, we are researching the impact of moving to a design that has five aggregate industry strata within each geographic area for allocation and creation of independent sampling cells while using the 23 detailed industries for implicit stratification. In this application, implicit stratification is done by sorting the establishments by detailed industry within each aggregate industry sampling cell prior to selecting a systematic sample to ensure that selections are made in each of the detailed industries.

Using the 120 geographic area by aggregate industry cells (i.e. 24 areas by 5 industries), we have begun evaluating the effectiveness of using fewer sampling cells and replacing the controlled rounding approach with a systematic/distributed rounding approach to obtain allocated annual sample sizes. Under the proposed systematic/distributed rounding approach, each area by aggregate industry cell will be assigned a three year sample allocation. This allocation will be divided into three integer allocations by dividing the total allocation by three and then adding any remainder to that number of individual yearly allocations. The individual years to receive the extra allocation will be assigned systematically so that the total sample size within an aggregate industry is held constant each year. The process will also work to ensure that the total sample size across all industries is constant from one year to the next by beginning the systematic assignment of the remainder allocations randomly. The table below shows an example of the resulting yearly rounded allocations with 4 areas (A1, A2, A3, and A4) within one industry stratum where the first random year start is 2.

[1] Aircraft manufacturing is selected separately due to the size of the industry and through a contract with the Aerospace Industries Association under which BLS provides more detailed data for this industry than would otherwise be feasible. A new sample of establishments is selected approximately once every ten years.

Table 1: Example of NCS Systematic/Distributed Allocation Rounding

Area	3-year allocation	Year 1 Rounded Allocation	Year 2 Rounded Allocation	Year 3 Rounded Allocation
A1	10	3	4	3
A2	12	4	4	4
A3	5	2	1	2
A4	7	2	3	2

Our analysis shows that this approach yields approximately the same number of selected establishments in each of the more detailed industries as our current approach that includes controlled rounding. By using a recent frame and the detailed industry sample sizes, we created new sampling cells based solely on the aggregate industry groupings. We then selected 100 simulated samples within each of the aggregate industry sampling cells and computed the mean sample size in each of the geographic area and detailed industry cells across the 100 samples. The mean sample sizes from this simulation were close to the original allocated sizes, showing that we can get comparable sample sizes for the 23 current industries with implicit allocation as we obtain under our current explicit sample allocation process. We also learned that it is acceptable to use the proposed systematic/distributed rounding approach to distribute the sample size across geographic areas within each industry because that is what implicit allocation accomplishes by design.

We then continued an analysis of a new allocation process for a three-year rotation and a national sample design. To conduct this analysis, we obtained a full frame of establishment data from the Bureau's QCEW database and assigned a geographic area, detailed industry grouping, and aggregate industry grouping to every establishment on the frame. Under this approach, we assigned a fixed proportion of the total size to each industry using prior response and variance information and distributed the industry size across each of the 24 national geographic areas in proportion to the total employment in the area. As with the current design, we identified multi-year certainties in each geographic area by aggregate industry cell and set the probability of selection for these units equal to one. The remaining non-certainty sample size for each sampling cell was distributed across the three years using the systematic/distributed rounding approach described above. We then selected fifty simulated samples from the full frame for each of the yearly allocations, i.e. Year 1, Year 2, and Year 3 allocations. We computed the mean sample size across all 150 simulated samples by aggregate industry, detailed industry, geographic area, and BLS collection region. Along with each mean sample size, we also computed the standard error of the mean sample size as well as the minimum, maximum, mode, and median sample sizes across the samples. We are currently using this data to help us better understand the impact of using this design and allocation approach.

We are also evaluating some sample design alternatives to deal with establishments that appear in multiple samples. The main alternative being investigated at the current time is dependent sampling. Under the current sample design, all establishments on the sample frame (except for the multi-year certainties) are eligible for inclusion in the sample. Under dependent sampling, we would exclude all establishments in the two most recent

samples from the frame used to select the current sample. This research effort has been conducted using the current area-based sample and five-year rotation design and is described in the paper by Ojo and Ponikowski (2010). We plan to repeat this research using the three-year rotation and national sample design before making any changes from our current independent sample design with multi-year certainties.

The NCS sample frame is drawn from the BLS QCEW database. This database contains one record for each quarter that an establishment reports to its state unemployment insurance file. Each record contains the establishment's monthly employment for each month in the quarter and the total wages paid to employees for the entire quarter. On the BLS QCEW database, it is possible for an establishment to exist on the database with an employment value of zero for one or more months of the quarter. This can occur for many reasons including newly forming establishments (sometimes called births), seasonal employment, and establishments that are in the process of going out of business. When the NCS encounters establishments on the frame with an employment value of zero, we currently set their employment to one so that the establishment has a chance of being selected in the sample. But this assumes that all of these establishments are seasonal or birth establishments and ignores the potential that some of them will not be in business by the time the sample is selected and data collection begins. So we are currently evaluating the nature of these establishments on the frame by tallying their frequency and employment trends over a multi-year time period. We are also evaluating several different approaches to decide when to use these establishments in the frame and how to set their frame employment values. For each option, we are setting the assigned employment value, selecting 100 simulated samples, computing the mean monthly wages across the simulated samples, and comparing the mean monthly wages from the samples to those computed using the full frame. Until this analysis is completed, we will not change this portion of the sample design.

NCS is also evaluating the frame for a National design. On the QCEW database, each establishment is assigned numeric codes representing the state and county in which the company conducts business. Under the current area based design, only establishments with a state and county code located in one of the sampled areas are included in the NCS sample frame. Under the proposed National design, all establishments in each county will be included on the frame as long as they belong to the in scope industries. But some of these establishments do not have a distinct county code on the QCEW as they represent firms that conduct business state-wide, outside the given state, or in foreign locations. NCS is currently evaluating the magnitude and employment of firms with these county designations and gathering information about how other BLS surveys treat these establishments in order to decide whether to include some or all of these unusual reporters in the NCS sample frame.

7.3 Future Research Plans

As we continue to evaluate the above mentioned research topics, we have also identified several other questions for future research. We expect to begin research on these new topics in the near future and plan to report on the results of all research efforts through published papers.

The first new area we would like to study is what impact the proposed sample design changes would have on our survey response rates. Since we are not changing the data being collected or the collection methods or forms, we expect to be able to achieve our current response rates for establishments of similar size and in the same industries. So we

will compute our historical response rates by establishment size class and detailed industry across all samples currently supporting the ECI and ECEC product lines. We will then compute the mean number of establishments in the same establishment size classes and detailed industry groupings from our 150 simulated samples. By applying the historical response rates to the simulated sample results, we will compute an aggregate expected response rate across all size classes and industries.

Next, we want to evaluate the impact of the new design on our published outputs. To do this, we are identifying all critical and high priority estimate publication categories. We will then tabulate the number of establishments in the sample for each of these categories and compare that to the minimums needed to support the NCS published outputs.

There are several other issues that we believe need to be addressed as we transition from our current design to a new design. For example, we would like to conduct additional research to evaluate the potential impact on the survey accuracy due to the design changes by evaluating the theoretical aspects of the survey design as well as results of related types of changes to the design in the past. We will also need to evaluate the current strategy of selecting establishment samples for the State and local government establishments and the aircraft manufacturing industry approximately once every ten years to determine if we should continue with this approach of separate samples or if we should include these establishments in every sample. In addition, we need to determine if we will have sufficient data collection staff to continue to update all the establishments in the current private industry samples until we finish transitioning to the new design. If not, we will need to figure how to reduce the size of the samples in update mode[2], figure out how to update them more efficiently, and/or identify additional short-term resources to conduct the update process.

Each of the samples in update mode reflects the establishments in the frame at the time the sample was drawn. The sample weights are updated to reflect losses due to establishments that have gone out of business since the sample was drawn. But these samples (up to 80% of the overall private industry sample) do not include any establishments that have begun business since the frame for that sample was prepared. These are called the birth establishments. We expect to conduct future research to determine if we should supplement the older samples with birth establishments.

8. Conclusion

As described in section 7.3, NCS is currently conducting several pieces of research on potential changes to the sample design. We plan to implement survey design changes that are based on research results with positive outcome, that is, ones that lead to survey design improvements and that are not very costly to implement starting with the selection of our sample in late summer 2011. The transition from the current to a new design will need to be completed one establishment sample at a time over a multi-year time frame – three years of initiation efforts for private industry followed at some point by a new State

[2] Since it could be several years before we reselect the State and local government sample, NCS plans to reduce the size of the current sample for these establishments by subsampling the existing sample and continuing to update the data only for those establishments selected during this process. However, no decision has been made about whether or not to reduce the size of the private industry samples at this time.

and local government sample and a new private aircraft manufacturing sample. Outputs during the transition will be computed using a combination of data selected under the old and new designs.

Although there is still a lot of work to do before a new sample design is completed for the NCS, there are many positive results that we have learned from both prior knowledge and experience as well as from current research project results. The research will continue and even after the new sample design has been implemented, the NCS staff will continue to look for ways to improve our sample design to provide our users with the best possible outputs.

References

Cochran, W. G. (1963), Sampling Techniques, New York: John Wiley & Sons, Inc.

Crockett, J., Ponikowski, C.H., and McNulty, E.E. (2008), "Update on Use of Administrative Data to Explore Effect of Establishment Nonresponse Adjustment on the National Compensation Survey Estimates," 2008 *Proceedings of the Section on Survey Research Methods,* Alexandria, VA: American Statistical Association.

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.

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

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.

Ponikowski, C. H., and McNulty, E. E. (2006), "Use of Administrative Data to Explore Effect of Establishment Nonresponse Adjustment on the National Compensation Survey Estimates," 2006 *Proceedings of the Section on Survey Research Methods,* Alexandria, VA: American Statistical Association.

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/homch5.htm

U.S. Bureau of Labor Statistics (2008) *BLS Handbook of Methods*, National Compensation Measures, Chapter 8. http://www.bls.gov/opub/hom/homch8_a.htm

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

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