

## BLS Resumes Estimation of Sample Errors for Benefits Measures

by [Omolola E. Ojo](#) and [Jonathan J. Lisic](#)

Bureau of Labor Statistics

*Originally Posted: May 22, 2008*

*Standard errors for the estimates in the National Compensation Survey (NCS) benefits publications have not been available to data users since the integration of the NCS sample. To provide a reliability measure for data users, the BLS is resuming production of standard errors for benefits estimates using Fay's method of Balanced Repeated Replication (BRR).*

Employee benefits measures are one of the four key products derived from the integrated [National Compensation Survey \(NCS\)](#) sample. These measures cover the incidence and detailed provisions of selected employee benefit plans. Incidence data are presented as the percentage of employees who have access to, or participate in, a broad selection of benefits. Provisions data are available for certain benefits, such as paid vacations and holidays, disability insurance (short-term and long-term), life and health insurance, and retirement plans.<sup>1</sup> This article briefly describes the integrated NCS sample; it also discusses the motivation for estimating sample errors and reviews the methodology used to produce sample errors for the NCS benefits data.

The integrated NCS sample provides data for the [Employment Cost Index \(ECI\)](#), the [Employer Costs for Employee Compensation \(ECEC\)](#) program, the estimates of [wages by area and occupation](#), and the [NCS benefits](#) publications. The sample covers civilian workers in private industry establishments and in State and local governments across all 50 states and the District of Columbia. Data are collected from a multistage probability sample consisting of the following three stages: 1) a probability sample of geographic areas, 2) a probability sample of establishments within sampled areas, and 3) a probability sample of occupations within sampled establishments.<sup>2</sup>

Because the benefits measures are derived from a probability sample, they are subject to sampling errors. Sampling errors are the differences between results computed from a sample and those computed from all units within a given population. The statistical value calculated to measure sampling errors is called the *standard error*<sup>3</sup>. Until recently, of the four previously mentioned products that use the NCS integrated sample, the NCS benefits program was the only one for which BLS had not produced standard error estimates. Starting in May 2008, BLS resumed producing standard errors for its benefits publications.

Standard errors for the estimates in the NCS benefits publications have not been available to data users since the integration of the NCS sample. Prior to integration, standard errors for benefit measures were computed from a representative portion of the survey estimates and illustrated as a curve fitted to the standard errors using regression techniques. Chart 1 shows the generalized standard errors for the 1995 estimates of benefits in medium and large private establishments.<sup>4</sup> For example, if a 1995 estimate was 55 percent, chart 1 shows that the standard error for the estimate is 2.2 percent.

With the standard error known, one can then compute a *confidence interval* around an estimate. A confidence interval estimates a range of values that are likely to include the true population value. This likelihood is given as a percentage and generally is referred to as the "confidence level." The NCS, for example, uses a 90-percent confidence level. Using the earlier example, the 90-percent confidence interval for a 55-percent estimate with a standard error of 2.2 percent would range from 51.38 percent to 58.62 percent.<sup>5</sup> This means that if all possible samples were selected to estimate the population value, the interval from each sample would include the true population value approximately 90 percent of the time.

Due to considerable changes in sampling methodology, the prior method and program used to calculate standard errors for benefit products is no longer applicable. Calculating standard errors for BLS benefit measures is difficult due to the multistage design of the sample and the multiple levels of data (establishment, plan type, and occupation). All standard error estimation systems currently used in the NCS use Fay's method of Balanced Repeated Replication (BRR). Fay's method is desirable because of its computational efficiency and ease of application to the complex NCS sample design. In general, the BRR method for standard error calculations is to calculate the estimate of interest from the full sample as well as from a number of

subsamples. The standard error for the full sample is then calculated using the variation among estimates of the subsamples. With many replication methods, a part of the sample is lost due to a weight of zero being applied during the process, which produces a biased, but consistent, estimator. Fay’s method of Balanced Repeated Replication makes less extreme modifications to the weights for the constructed replicates.<sup>6</sup>

Using Fay’s Method of Balanced Repeated Replication involves four steps. In the first step, the sample is partitioned into  $S$  variance strata. Benefit products use the same variance strata as other national NCS products. The second step is to divide the sample units in each stratum into two Primary Sample Units (PSUs). The third step is to use the PSUs to construct  $R$  replicate weights, where the number of replicates is greater than or equal to the number of strata  $S$  defined earlier. These replicate weights are constructed by choosing either PSU1 or PSU2 and increasing the chosen PSU’s weight by a proportion  $k$ , while decreasing the other PSU’s weight by the same proportion. Let  $k = 0.5$ . Sample observations within the chosen PSU are weighted by 1.5, and units in the other PSU are weighted by 0.5. Finally, the fourth step is to generate a full sample estimate using full sample weights, and to generate replicate estimates using replicate weights. The sum of the squared differences between the estimate from the full sample  $\hat{Y}$ , and the  $R$  replicates  $\hat{Y}_r$ , can then be formulated to calculate the standard error as follows:<sup>7</sup>

$$SE(\hat{Y}) = \sqrt{\frac{1}{R(1-k)^2} \sum_{r=1}^R (\hat{Y}_r - \hat{Y})^2}$$

For the analysis presented in this article, we applied this method to previously released data from the NCS benefits program and assessed the quality of the results.

The initial investigation focused on percentage estimates published in the March 2006 summary of employee benefits in private industry.<sup>8</sup> Because the size of the standard error depends upon the size of the estimate, nonpercentage estimates will have greater variation. In an attempt to include as many estimates as possible while maintaining consistency, this analysis includes only percentage estimates. Nonetheless, standard errors for nonpercentage estimates are calculated and will be available for this and future employee benefits products. Fortunately, nonpercentage estimates made up a relatively small amount of the total number of estimates in the March 2006 summary. The present analysis utilized 3,657 estimates from the summary. The remaining estimates were either nonpercentage estimates or percentage estimates with a standard error of zero (for example, estimates of 100 percent have a standard error of zero).<sup>9</sup>

Chart 2 illustrates the distribution of the standard errors in the present analysis. The first four columns show that about 80 percent of the estimates have a standard error that is greater than zero and less than 2 percent. For example, the NCS estimate for all workers with access to medical care benefits in the March 2006 summary was 70.56 percent. The calculated standard error for this estimate is 0.75 percent. Thus, the 90-percent confidence interval ranges from 69.33 to 71.79. Furthermore, about 3 percent of the estimates have a standard error that is greater than or equal to 5. These observations provide insight into the quality of the estimates of the NCS benefit products. With such a large percent of total estimates having small standard errors, it is unlikely that there would be a large difference between the estimated values and the actual population being represented.

Further investigation showed that most of the larger standard errors--those greater than or equal to 5--are for estimates with a lower number of contributing observations. The standard errors for estimates of all employees are generally small compared with those for subdomains, especially census divisions. Estimates for census divisions accounted for approximately 90 percent of the estimates with standard errors that were greater than or equal to 5. For example, the standard error on the percent of all employees in the Nation with access to medical care benefits is 0.75 percent, compared with 5.76 percent for employees in the East South Central census division. There was no statistical difference for employees with access to medical benefits among the census divisions.

The quality and utility of an estimate is directly dependent upon the measure of its standard error. Without this measure, there is no way to gauge the validity of any conclusions drawn from the data. This investigation into the standard errors of the

NCS benefit incidence and provisions products appears to support the soundness of the estimates. The National Compensation Survey program intends to apply these standard error calculation methods to future NCS benefit products, beginning with the publication of *National Compensation Survey: Retirement Benefits in State and Local Governments in the United States, 2007*, available on the Internet at <http://www.bls.gov/ncs/ebs/sp/ebsm0008.pdf>.<sup>10</sup> The standard errors will provide users of BLS benefit measures with a sound measure of reliability to use when they employ the data in their individual practices.

Omolola E. Ojo

Mathematical Statistician, NCS Estimation and Review Branch, Statistical Methods Group, Bureau of Labor Statistics.

Telephone: (202) 691-6113; E-mail: [Ojo.L@bls.gov](mailto:Ojo.L@bls.gov)

Jonathan J. Lisic

Mathematical Statistician, NCS Estimation and Review Branch, Statistical Methods Group, Bureau of Labor Statistics.

Telephone: (202) 691-6115; E-mail: [Lisic.Jonathan@bls.gov](mailto:Lisic.Jonathan@bls.gov)

## Notes

<sup>1</sup> For more technical information on sampling and estimation for the *National Compensation Survey*, see "National Compensation Measures," in *BLS Handbook of Methods* (online version); available on the Internet at <http://www.bls.gov/pub/hom/pdf/homch8.pdf>.

<sup>2</sup> For further details on the National Compensation Survey sample, see "Sample Allocation and Selection for the National Compensation Survey," on the Internet at <http://www.bls.gov/ore/pdf/st020150.pdf>.

<sup>3</sup> See "National Compensation Measures," *BLS Handbook of Methods*.

<sup>4</sup> *Employee Benefits in Medium and Large Private Establishments, 1995*, Bulletin 2496 (Bureau of Labor Statistics, April 1998); available on the Internet at <http://www.bls.gov/ncs/ebs/sp/ebbl0015.pdf>; for information on the reliability of the estimates, see appendix A, pp. 162-64.

<sup>5</sup> In other words, for this example, in a normal distribution, the 90-percent confidence interval would be computed as  $55 \pm (1.645 \times 2.2) = 51.38$  to 58.62. For a detailed explanation of confidence interval formulas, see Robert V. Hogg and Allen T. Craig, "Confidence Intervals for Means," in *Introduction to Mathematical Statistics* (Prentice Hall, 1995), pp. 268-76.

<sup>6</sup> For a detailed technical discussion on Fay's method of Balanced Repeated Replication (BRR), see "Estimating Variance in the National Compensation Survey, Using Balanced Repeated Replication," on the Internet at <http://www.bls.gov/ore/pdf/st010110.pdf>; also, "Comparison of Variance Estimation Methods for the National Compensation Survey," on the Internet at <http://www.bls.gov/ore/pdf/st990290.pdf>.

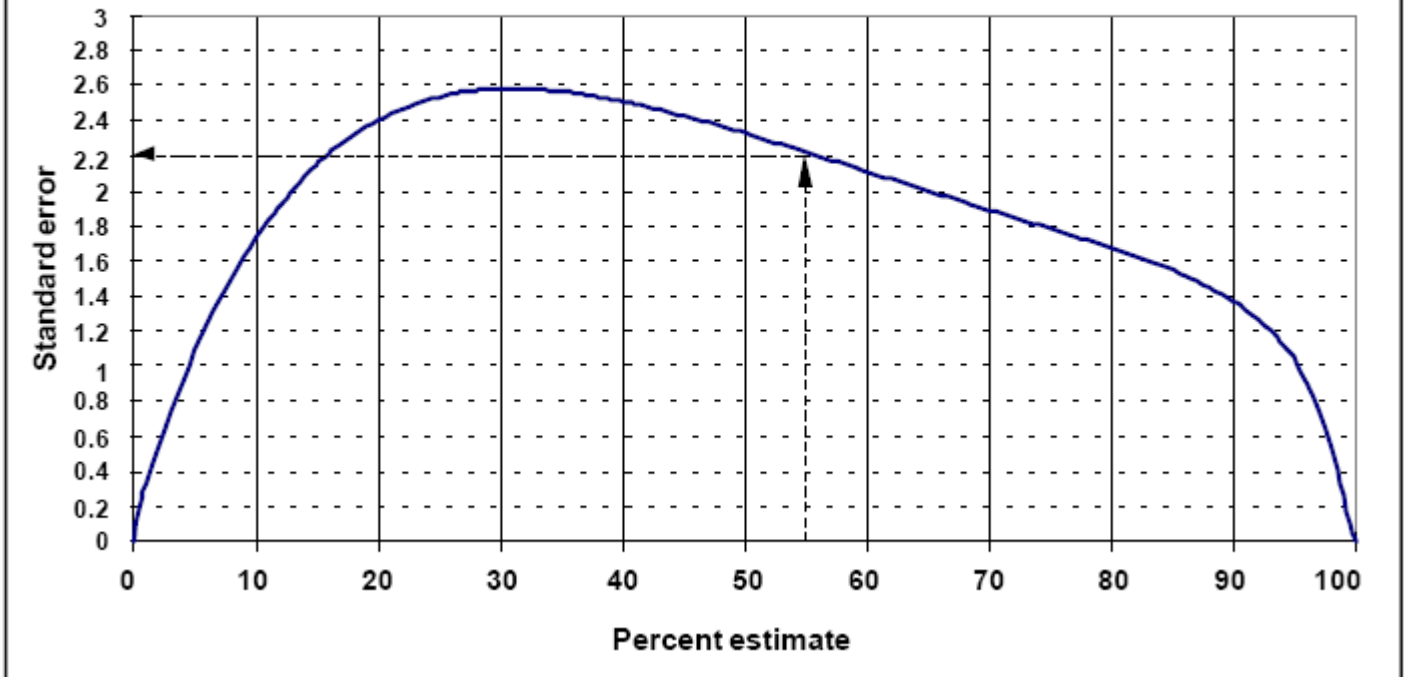
<sup>7</sup> *Ibid.*

<sup>8</sup> "National Compensation Survey: Employee Benefits in Private Industry in the United States, March 2006," Summary 06-05 (Bureau of Labor Statistics, August 2006); available on the Internet at <http://www.bls.gov/ncs/ebs/sp/ebsm0004.pdf>.

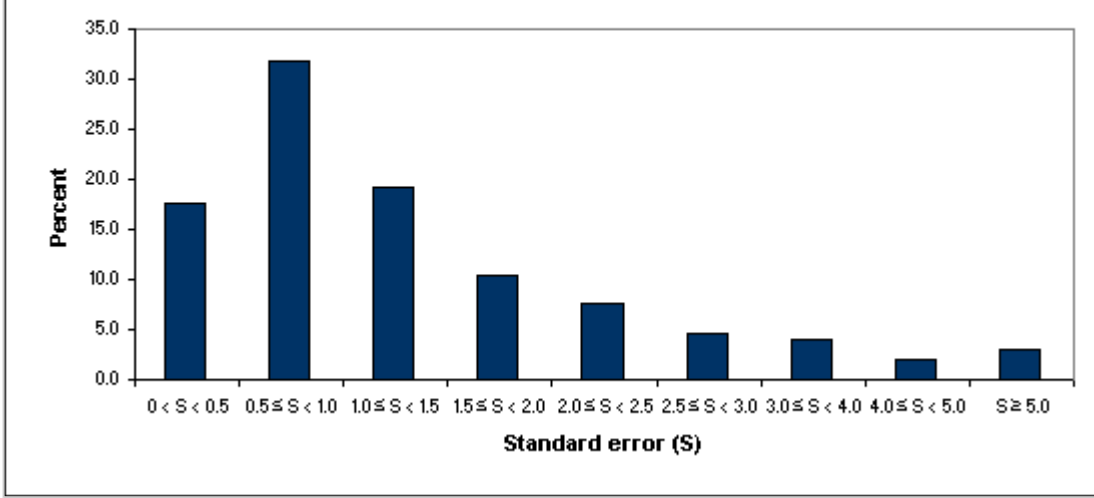
<sup>9</sup> Standard errors for the estimates contained in National Compensation Survey: Employee Benefits in Private Industry in the United States, 2006, are available by calling the BLS Office of Compensation and Working Conditions at 202-691-6199 (E-mail: [NCSInfo@bls.gov](mailto:NCSInfo@bls.gov)).

<sup>10</sup> *National Compensation Survey: Retirement Benefits in State and Local Governments in the United States, 2007*, Summary 08-03 (Bureau of Labor Statistics, May 2008); available on the Internet at <http://www.bls.gov/ncs/ebs/sp/ebsm0008.pdf>.

**Chart 1. Generalized standard errors, medium and large private establishments, National Compensation Survey, Benefits, 1995**



**Chart 2. Standard error distribution, benefits incidence and provisions summary, March 2006**



**Data for Chart 2. Standard error distributions, benefits incidence and provisions summary, March 2006**

| Standard error | Percent          |
|----------------|------------------|
| 0 < S < 0.5    | 17.5006836204539 |
| 0.5 ≤ S < 1.0  | 31.7746786983867 |

| Standard error          | Percent          |
|-------------------------|------------------|
| <b>1.0 d S &lt; 1.5</b> | 19.1960623461854 |
| <b>1.5 d S &lt; 2.0</b> | 10.4183757178015 |
| <b>2.0 d S &lt; 2.5</b> | 7.68389390210555 |
| <b>2.5 d S &lt; 3.0</b> | 4.59392945036916 |
| <b>3.0 d S &lt; 4.0</b> | 3.99234345091605 |
| <b>4.0 d S &lt; 5.0</b> | 1.91413727098715 |
| <b>S e 5.0</b>          | 2.92589554279464 |

---

U.S. Bureau of Labor Statistics | Division of Information and Marketing Services, PSB Suite 2850, 2 Massachusetts Avenue, NE Washington, DC 20212-0001 | [www.bls.gov/OPUB](http://www.bls.gov/OPUB) | Telephone: 1-202-691-5200 | [Contact Us](#)