

# Calculating Wage Percentiles in the National Compensation Survey

by Hilery Simpson and James Smith

*Originally Posted: April 28, 2004*

*For its estimates of hourly wage percentiles produced for the National Compensation Survey, BLS recently returned to a method based upon individual wage rates within sampled establishment occupations, following a period in which average occupational wage rates were used. The current percentile estimates allow for a more precise measure of wage dispersion and increase the utility of such estimates for data users.*

Wage percentiles describe the distribution of earnings within published occupations. These estimates, along with establishment and worker characteristics, help to explain the wage variation within occupations. This article explains the methods used in recent years to calculate percentiles in the BLS [National Compensation Survey \(NCS\)](#) locality wage publications and how these methods relate to different data users. The NCS publishes hourly wage data for specific occupations in more than 80 metropolitan areas across the United States.

NCS locality publications present hourly wage rate estimates for the 10th, 25th, 50th, 75th, and 90th earnings percentiles. The percentiles designate position in the earnings distribution within each published occupation, weighted by the scheduled hours of work. At the 50th percentile, the median, half of the hours are paid the same as or more than the rate shown, and half are paid the same as or less than the rate shown. At the 25th percentile, one-fourth of the hours are paid the same as or less than the rate shown. At the 75th percentile, one-fourth are paid the same as or more than the rate shown. The 10th and 90th percentiles follow the same logic.

The usefulness of percentiles is illustrated in the following example. In the 2003 Honolulu, Hawaii, NCS publication,<sup>1</sup> the occupation titled "librarians" had a mean wage of \$27.79 per hour. Using only this information, one would not expect to find librarians in the Honolulu market with wages as low as \$15.00 per hour or as high as \$50.00 per hour. However, when percentile estimates are examined, this range does not appear to be out of line. Librarians in the Honolulu market at the 10th percentile earned \$16.49 per hour, while those at the 90th percentile earned \$49.06 per hour. The size of the wage dispersion in this occupation shows the value of percentile estimates.

## Implementation Of The NCS

Data from the NCS locality wage program were first published in 1997. The purpose of the program is to measure occupational wages and work levels<sup>2</sup> in numerous metropolitan and nonmetropolitan areas. The NCS program uses a probability sample of occupations to represent all occupations in the non-Federal, nonfarm economy.<sup>3</sup> In each sampled establishment, as many as eight occupations are selected for study, based on the establishment's employment size. The greater the number of people working in an establishment's occupation, the greater chance that occupation will be selected. Wage and related information are then collected for all employees in those occupations. During final processing, employees in each selected occupation in the survey are appropriately weighted to represent all employers and employees within the survey area.

When an establishment is first surveyed, a specially trained field economist interviews an official from the company or government agency to gain cooperation, sample establishment occupations, and collect wages and related information. Data from each establishment are updated annually for a period of approximately 5 years, usually via telephone or mail collection.<sup>4</sup>

During NCS's first round of surveys in 1997, individual wages were collected for each incumbent in the sampled occupations. Published hourly wage percentiles were based on those wage rates.

## The First NCS Update

During the first update of these data in 1998, average wage rates for sampled establishment occupations were allowed to be collected. The change from collecting individual wage rates for all workers within a sampled occupation to collecting an

average rate for the sampled occupation was made to balance the unusually heavy survey workload necessary to complete the development of the NCS program. This was one of several steps BLS took to expedite NCS data collection during this transition period.

The introduction into the NCS of average wage rates for sampled occupations complicated the process used to produce the wage percentiles. Wage rates for incumbents could no longer be easily distributed from lowest to highest and published as they fell in the distribution. A new method was needed to take into account reported averages for occupations for which the internal wage dispersion was unknown. BLS decided to use both the individual rates collected during 1997 and the average rates collected during the 1998 update.

For occupations with average wages collected at update, the percent change in the average wage for each sampled occupation from 1997 to 1998 was used to adjust the individual rates collected during 1997. In other words, if the increase in a sampled occupation's average wage between 1997 and 1998 was 3 percent, then the individual wage rates collected in 1997 were increased by 3 percent. These adjusted individual wage rates were used in the calculation of percentile statistics in 1998 publications. The basic assumption of this modeling method was that the wage dispersion within occupations did not change over the year. In 1998, the frequency of the collection of average rates was low.

### **The Second NCS Update**

For the second NCS update in 1999, BLS was no longer able to model the percentiles from a mix of individual and average wages for a number of reasons. First, the assumption that changes in an occupation's individual wages closely emulated that of its mean wage was too simplistic. Numerous examples were cited showing that this assumption was not always valid. For example, when an occupation's mean wage decreased, the modeling occasionally caused rates at the low end of the scale to drop below the legal minimum wage. Also, the modeling did not account for changes in an occupation's wage distribution created by hiring, firing, or promoting workers. Second, the model would need major revisions to account for new data collection scheduling strategies.<sup>5</sup> New software would need to be designed and implemented. Third, the frequency of reporting averages had increased dramatically between 1998 and 1999.

Due to these three factors, the model to calculate percentile estimates for 1998 could not be used for 1999. Two options were available: Create a model based on individual wages that would take into account the issues discussed in the previous paragraph, or develop a simpler percentile estimation process based on average wages. While the first option was preferable from the standpoint of estimate design, such a system would take 2 to 3 years of development and testing before wage percentiles could again be published.

### **An Interim Method**

To prevent an interruption in the publication of percentile estimates, BLS determined that an interim measure needed to be developed that had the following three attributes:

- *Relevance.* Current users of NCS wage information must find the interim measure useful.
- *Feasibility.* The new model must be based on currently collected data and use a method that is relatively easy to implement.
- *Practicality.* Due to the high percentage of averages collected, the system must accept the reporting of average occupational wage rates.

The interim method to calculate wage percentiles used average hourly rates for establishment occupations; thus, the percentiles were calculated from a distribution of occupational averages rather than from individual worker wage rates. Even if individual earnings were collected, the individual wages reported for incumbents in sampled occupations were converted to averages. Like the former method, the interim measure explained some wage variation between occupations--both within establishments and across establishments--but it could not explain wage variation within an establishment occupation. The new measure averaged out within-occupation wage variation, and thus limited the size of the estimate of wage dispersion.

## Research On The Interim Method

BLS conducted a series of research tests to determine the effect of the average rate method.<sup>6</sup> Following are the results of three tests that compared the individual rates method of calculating percentiles with the method using average rates for establishment occupations. All data used in the tests are 1997 individual wage rates.<sup>7</sup>

**First test.** The first test compared percentile estimates calculated using average wages for establishment occupations with those calculated with individual worker data for the same occupations in a large survey area. The purpose of this test was to confirm expectations about how percentile estimates would change using the two methods for some selected occupations.

Four occupations from the Houston survey<sup>8</sup> were selected for testing: Elementary teachers, secondary teachers, cashiers, and waiters and waitresses. Teachers were selected because they tend to have a high degree of wage dispersion, while cashiers and waiters and waitresses were selected because they tend to have less wage dispersion.<sup>9</sup>

The testing confirmed that calculating percentiles with occupational average wage data instead of individual wage rate data reduces, on average, the estimates of wage dispersion for an occupation. Occupations with larger wage dispersions were more affected using this method than those with lower wage dispersions.

The results from this test are shown in table 1. The difference between the 10th and 90th percentile hourly wages was 65 percent for elementary school teachers using individual wage data. When percentiles were calculated using the average wages for establishment occupations, the difference dropped to 13 percent. The percent difference for secondary school teachers showed a similar decline when the average wage method was used.

The percent differences between the two methods were less for cashiers and for waiters and waitresses. These two occupations had low wage dispersions. Cashiers had a difference of 80 percent between the 10th and 90th percentiles using the individual wage method, and a difference of 68 percent using the average wage method. The difference was even less for waiters and waitresses: 177 percent using individual wage data, and 175 percent using average wage data.

**Second test.** In the second test, percentile estimates for broad occupational groups in 15 survey areas were compared using both the average and individual wage methods. The purpose of this test was twofold:

1. To test the consistency of the Houston results for broad occupational groups;
2. To estimate the size of the differences for broad occupational groups.

To test the consistency between the two methods, percentiles were calculated for broad occupational groups using the two methods for each of the 15 areas.<sup>10</sup> For each occupational group, a ratio was calculated that measured the difference in wages, at each percentile, between those calculated with individual wage rates and those calculated with average wages.

The results confirmed that using average wage data instead of individual wage data had the effect of moving percentiles toward the median (50th percentile). At the lower percentiles (10th and 25th), the percentiles that were calculated using averages were higher than those calculated with individual wage rates. At the higher percentiles (75th and 90th), the percentiles calculated using averages were lower than those calculated with individual wage rates. However, an important finding was that differences were almost nonexistent for the medians.

Table 2 presents tabulations of the average ratio difference between the individual rate method and the occupational average rate method of calculating percentiles for all occupations and for three broad occupational groups.<sup>11</sup> Blue-collar occupations had smaller ratio differences than white-collar occupations. The most likely cause of this difference is that individual white-collar occupations have a higher degree of wage dispersion than blue-collar occupations. Thus, the effect, on average, is larger for white-collar occupations. The following conclusions were reached:

- The percentiles calculated using average establishment occupational rates are nearer to the median, irrespective of broad occupational group (white collar, blue collar, or service).

- Differences in average wages for broad occupational groups are small, ranging from 0.1 to 5.3 percent, even though they are larger for individual occupations within the broad groups.
- Irrespective of occupational group, differences are largest for the 10th, then 25th, then 90th percentiles; differences were almost nonexistent for the medians.

The results from the second test demonstrated the broad consistency of the Houston test results--percentiles calculated with average wages had shown less wage dispersion than the other method.

**Third test.** The final test calculated how the ranking of specific occupations by a measure of wage dispersion would change using establishment occupational average rates. For this test, published occupations within each of the 15 areas were ranked by the magnitude of dispersion under each method (individual rate and occupational average) of calculating the percentiles. The number of published occupations within each area varied from a low of 18 to a high of 150.

The correlation of occupational rankings found that the "new" average rate occupational rank order was very consistent with the "old" individual rate occupational rank order. Also, it was found that in the larger areas, those with more published occupations, the new and the old occupational rank orders were more consistent than in the smaller areas, those with fewer published occupations.

**Test summary.** The research was designed to measure the differences between the two constructs, not to determine the benefits of either method. In summary, the amount of wage dispersion estimated with the average wage measure was different from the amount estimated with the individual wage measure. This difference could be greater for individual occupations than for broad groups of occupations. However, both methods order occupations by relative wage dispersion quite similarly.

### Uses And Users For The Two Methods

Percentile estimates produced from the average wage method from 1998 through 2002 differ from those produced using individual wage rates in 1997 and 2003. The average wage method yields a measure of wage distribution that relates to the pay of establishment occupations, while the individual wage measure relates to the pay of individual workers within establishment occupations. Comparing the published estimates from each method shows that the average wage method produces percentiles with less wage dispersion than those produced with the individual rate method.

Those who use percentile estimates can gain value from estimates produced using either method. Data users such as pay administrators who want to place pay points within the distribution estimates can place average pay points for occupations within the published percentiles calculated with averages. General researchers examining pay for specific occupations in the economy can still use the percentiles calculated with the average method. However, researchers examining occupational pay variation within establishment occupations are better served with percentile estimates calculated with the individual wage rate method.

BLS has now returned to the collection of individual wage rates in the NCS program. Beginning with NCS locality wage surveys published in Spring 2003, percentiles are again calculated using individual wage rates. This should allow for a more precise measure of wage dispersion and better serve all users who rely on these estimates.

Hilery Simpson

Economist, Division of Compensation Data Estimation, Bureau of Labor Statistics.

Telephone: (202) 691-5184; E-mail: Simpson.Hilery@bls.gov

James Smith

Economist, Division of Compensation Data Estimation, Bureau of Labor Statistics.

Telephone: (202) 691-6207; E-mail: Smith.James@bls.gov

### Notes

<sup>1</sup> See *Honolulu, HI, National Compensation Survey, February 2003*, Bulletin 3115-78 (Bureau of Labor Statistics, July 2003).

- 2 Each occupation is evaluated based on nine factors, such as knowledge, complexity, and work environment. Points are assigned based on the occupation's rank within each factor. The points are summed to determine the overall level of the occupation.
- 3 The predecessor to NCS, the Occupational Compensation Survey, surveyed a fixed list of about 38 occupations.
- 4 Certain establishments who participate in the locality wage portion of the National Compensation Survey also participate in the Employment Cost Index. For these establishments, wages are updated quarterly, not annually.
- 5 Data collection has changed to a panel-based design, where each establishment is assigned to one of four collection periods, or panels during a year. Each panel runs for 4 months; and the start and end dates overlap so that four panels are collected within 13 months. The total sample for most locality publications is spread over four panels. Thus, the collection period for most locality publications is 13 months.
- 6 Test tabulations were prepared by Brooks Pierce, in the BLS Office of Compensation and Working Conditions, Compensation Research and Program Development Group.
- 7 Data from the 1997 surveys were used because they did not include any reported average wages. Standard errors have not been calculated for published percentile estimates. Consequently, none of the statistical inferences made in this report could be verified by a statistical test.
- 8 See *Houston-Galveston-Brazoria, TX, National Compensation Survey, October 1997*, Bulletin 3090-40 (Bureau of Labor Statistics, revised March 1999).
- 9 Tip information is not included in NCS wage data.
- 10 The 15 areas were selected to represent a variety of geographic regions, industry mixes, and sample sizes.
- 11 The individual occupations used in NCS are grouped into the broad occupational groups titled white collar, blue collar, or service. White collar includes professional, technical, executive, sales, and clerical occupations. Blue collar includes production, machine operators, transportation, and handlers and helpers. Service includes protective service, food service, health service, cleaning and building service, and personal service occupations.

**Table 1. Hourly wage percentiles calculated with individual worker data and with occupational averages data: Selected occupations from Houston-Galveston-Brazoria, TX, National Compensation Survey, October 1997.**

Occupation	Hourly wage at percentile					90th-10th percentile difference (in percent)
	10th	25th	50th	75th	90th	
Elementary school teachers						
<b>Individual worker data</b>	17.92	19.73	22.87	26.53	29.56	65.0
<b>Occupational averages data</b>	21.98	22.04	23.32	24.66	24.88	13.2
Secondary school teachers						
<b>Individual worker data</b>	17.92	19.66	23.05	26.97	29.58	65.1
<b>Occupational averages data</b>	22.58	23.23	23.23	23.60	25.77	14.1
Cashiers						
<b>Individual worker data</b>	5.00	5.25	6.00	7.50	9.00	80.0
<b>Occupational averages data</b>	5.13	5.42	6.10	7.75	8.62	68.0
Waiters and waitresses						
<b>Individual worker data</b>	2.13	2.13	2.13	2.38	5.90	177.0
<b>Occupational averages data</b>	2.13	2.13	2.13	2.38	5.86	175.1

**Table 2. Difference ratio(1) between hourly wage percentiles calculated with individual worker data and with occupational averages data: 15 areas from National Compensation Survey, 1997**

Occupational Group	Percentile				
	10th	25th	50th	75th	90th
<b>All</b>	0.045	0.030	0.004	-0.011	-0.023
<b>Blue Collar</b>	0.032	0.024	-0.003	-0.007	-0.019
<b>White Collar</b>	0.053	0.035	0.004	-0.014	-0.026
<b>Service</b>	0.042	0.019	0.006	-0.001	-0.020

Footnotes:

(1) Weighted average over occupations of  $(w_a - w_m) / w_m$ , where  $w_a$  = the wage rate at the given percentile using occupational average data and  $w_m$  = the wage rate at the given percentile using individual worker microdata.