THE REINTERVIEW PROGRAM FOR THE BLS COMPENSATION SURVEYS

Steven P. Paben, Bureau of Labor Statistics
BLS, 2 Massachusetts Ave., N.E., Room 3160, Washington, DC 20212

INTRODUCTION

In order to ensure the quality of its data, the Bureau of Labor Statistics (BLS) conducts a reinterview program for various compensation surveys, including the National Compensation Survey (NCS) and the Employment Cost Index (ECI). This paper discusses the evolution, current sample design, and some of the significant findings of the reinterview program.

The reinterview program, entitled Technical Reinterview Program (TRP), is a quality assessment and improvement process. The program includes, along with a reinterview of the respondent, immediate feedback and discussion with the field economist who did the original collection leading to a continuous quality improvement cycle. Then, by analyzing the reinterview program data, we can assess the quality of the survey data by determining non-sampling error rates and identifying areas of data collection in need of additional study.

BACKGROUND

The National Compensation Survey (NCS) is a business establishment survey for occupational wages. (Eventually, benefit information will also be collected for a portion of the NCS.) This survey includes data on broad occupational classifications such as white-collar workers, major occupational groups such as sales workers, and individual occupations such as cashiers. The job level of an occupational series is derived from generic standards that apply to all occupations and occupational groups. The predecessor to NCS, the Occupational Compensation Survey (OCS), had job levels that were based on narrowly defined descriptions that were not comparable across specific occupations. Since data collectors (or field economists) were matching sampled jobs to preset defined descriptions, OCS had a reinterview program called the Job Match Validation (JMV) program.

The JMV process consisted of reviewers looking at entire schedules on a flow basis during survey collection. A schedule is a report form for all collected data for a business establishment. The reviewers then reinterviewed respondents to verify data deemed to be "suspect" of error. One hypothesis of this process was that review was too subjective to lead to valid conclusions. It was suggested that items chosen to be reviewed depended as much on the particular reviewer as the data being "suspect". In addition, the reviewers were staff whom were also responsible for survey production and the JMV process was often left shortchanged in order for survey production deadlines to be met.

In order to test the hypothesis that review was too subjective, a study of "suspect" data and "non-suspect" data was conducted. For this study, a probability subsample of the non-suspect job matching decisions were reinterviewed in addition to the reinterviewing of all suspect job matching decisions. The probability subsample of non-suspect data required the introduction of differential weights into the review process. The weights for non-suspect data were the inverse of the probability of selection. Since data deemed to be suspect is the result of 100% review, the weights for suspect data were one (or certainty). Table 1 shows the weighted percentage of incorrect job match decisions for the suspect and non-suspect groups.

Table 1. Weighted Percentage of Incorrect Job Match Decisions for Suspect and Non-Suspect Data

	Suspect	Non-Suspect
Incorrect Job Matches	387	468
Number of Job Matches	1796	14,693
Weighted Percentage	21.5%	3.2%

The weighted error rate for suspect job decisions was 21.5% and the weighted error rate for non-suspect job decisions was 3.2%. Testing for a difference between two means, it was found that the weighted error rate for the suspect group was significantly greater than the weighted error rate for non-suspect group at the .01 significance level. One would expect a higher percentage of errors for suspect job match decisions than for non-suspect job match decisions. So, this was an expected result.

However, it was clear that not all errors were captured by examining suspect decisions. In fact, the total weighted number of incorrect decisions for the non-suspect group was 468 while the total number of incorrect decisions for the suspect group was 387. This implied that more than half of the errors would have been missed by examining only suspect data. Table 2 shows the percentage of incorrect job match decisions without including errors from the non-suspect group and with errors from the non-suspect group. Since previously to this study all non-suspect job match decisions were considered correct, the percentage of errors without including non-suspect errors is calculated

as the number of incorrect suspect job match decisions out of the total number of job match decisions.

Table 2. Weighted Percentage of Incorrect Job Matches Without and With Non-Suspect Errors

	Without Non-Suspect	With Non-Suspect
Incorrect Job Matches	387	855
Number of Job Matches	16,489	16,489
Weighted Percentage	2.3%	5.2%

The table above shows the impact of including non-suspect errors into the overall error rate. The percentage of incorrect decisions without including non-suspect errors was 2.3%. It increased significantly with non-suspect errors included to 5.2% at the .01 significance level.

This study along with several other studies showed the JMV process missed a significant number of errors even though every job match was at least reviewed if not reinterviewed. In addition, the discontinuation of OCS brought along an opportunity to improve quality review by making it a more objective and statistically based process.

The Technical Reinterview Program (TRP) is designed to correct the deficiencies of JMV. The TRP is conducted by a unit separate from survey production to remove production pressures and potential conflicts of interest of the program staff from effecting quality review. The schedules are selected and assigned randomly to reviewers, and the items to be reviewed and reinterviewed are also selected randomly. This more objective process, also helps reduce any perception of reviewer bias against individual field economists. In addition, the TRP incorporated the Employment Cost Index (ECI), which is one of the major leading economic indicators.

It should be noted that the reinterview program is not the sole quality review program for the compensation surveys. Another quality assurance program, entitled Targeted Data Analysis (TDA) is conducted by the regional collection offices of BLS. It serves as more of a performance check on field economists than TRP since "pass/fail" criteria are applied to a schedule as the result of the quality review. There is no subsampling of collected data on selected schedules or recontacting of the respondent in TDA.

SAMPLE DESIGN

NCS uses a rotating panel design with three stages of selection used in selecting each panel: geographic area PSUs; establishments selected from industry strata; and occupations selected separately from each sampled establishment. The establishments are selected with probability proportional to size (pps), with total employment the measure of size. Occupations are selected pps within the establishment, with the number of employees in each occupation the measure of size. The number of occupational selections for each establishment is dependent upon the size of the establishment (see Table 3). Wage data is obtained for all employees having the same detailed job description as a selected occupation. The sampling frame from which the establishments are selected is constructed primarily from the unemployment insurance universe.

For the NCS TRP, a multi-stage subsampling of schedules, occupational selections, and generic standards (or factors) used in determining the level of a job is conducted. Approximately 10% of eligible completed schedules are selected for reinterview about once a week. Completed schedules may be selected only for that week, since it is preferable to reinterview a respondent as soon as possible after the original interview.

Occupational selections are then randomly subsampled for each establishment selected for reinterview depending on the PSO employment of the establishment (see Table 3). The PSO employment is the number of in-scope employees who receive cash payments for services performed. Out-of-scope employees include workers such as contractors or workers who set their own pay.

Generic leveling factors for each occupational selection are also randomly selected. There are 10 generic factors used to determine the level of a job, four of which are considered to be major factors since they have more predictive power in determining the level of a job than the other (minor) six factors. The number of major and minor factors selected for each subsampled occupation is based on the following schedule:

Table 3. Number of Occupations and Factors Selected for the NCS TRP

PSO Employment	# of Occ.	# of Occ. Sub-Sampled	# of Factors Selected for Each Occ.
1 – 49	4	1	All Factors
50 – 99	8	2	3 Major; 3 Minor
100 – 249	10	2	3 Major; 3 Minor
250 – 999	12	3	2 Major; 2 Minor
1000 – 2499	16	3	2 Major; 2 Minor
2500+	20	4	2 Major; 1 Minor

The reinterview program also includes the ECI. The ECI is a fixed-employment-weighted index which tracks quarterly changes in labor costs (including wages, salaries, and employer costs for employee benefits). The sample design of the ECI is somewhat similar to NCS without the first stage of selection (geographic areas). Also updating of wage and benefit information for establishments already initiated into ECI is obviously quarterly, while wage data for establishments already initiated into NCS is updated annually.

The ECI TRP is limited to schedules being updated from the previous quarter. The sample design is similar to the NCS TRP for the final two stages of sampling. The schedules are randomly selected and assigned, then the occupational selections are randomly selected based on the PSO employment of the establishment, although the percentage of schedules reinterviewed and the number of occupational selections is smaller in ECI. The third stage of sampling for the ECI TRP is a subsampling of certain specified benefits rather than factors. In ECI there is no leveling of a job, but benefit data is collected in addition to wage data. The benefits are selected as an unequal weighted sample with the same benefits reinterviewed for every subsampled occupation in a schedule. An unequal weighted sample is used, since there is an interest in studying the more complex benefits. For example, health insurance is sampled for 50% of the schedules to be reinterviewed, while state unemployment insurance which is a legally required benefit is reinterviewed for about 5% of the schedules.

WEIGHTING AND VARIANCE ESTIMATION

The NCS and ECI samples are the population for the TRP. Thus, the weights in the TRP generally pertain to the establishments and occupations in the particular survey. The weights are the inverse of the probability of selection. For example in the NCS TRP, the weight for a particular generic leveling factor, f_{ijk}, would be:

$$f_{ijk} = e_i \cdot o_{ij} \cdot m_{ijk},$$

where e_i = the reinterview weight of the i^{th} establishment, o_{ij} = the reinterview weight of the j^{th} occupational selection in the i^{th} establishment, and m_{ijk} = the weight of the k^{th} major/minor factor for the j^{th} occupational selection in the i^{th} establishment.

Currently, no non-response adjustments are made for the reinterview program, since the refusal rate for the reinterview program is very low.

Since the TRP has a multi-stage selection scheme, the variance is calculated using a sampling of clusters with replacement formula. For example, the variance of the weighted percentage of occupations in error, $V(\hat{R})$, would be the following:

$$V(\hat{R}) = \frac{1}{\hat{X}^2} \cdot \frac{1}{n(n-1)} \cdot \sum_i (\hat{y}_i - \hat{R}\hat{x}_i)^2,$$

where $\hat{X}$ = the total weighted number of occupational selections, n = the number of establishments, $\hat{y}_i$ = the total weighted number of occupational selections in error for the i^{th} establishment, $\hat{x}_i$ = the total weighted number of occupational selections for the i^{th} establishment, and $\hat{R} = \sum_i \hat{y}_i \Big/ \sum_i \hat{x}_i$.

Since we are usually interested in the weighted percentage of errors, this ratio estimation form is the most frequently used form in the reinterview program analyses.

RESULTS

Changes to survey data occur as a result of the reinterview program only when the reinterviewer and the field economist who collected the original data agree a change should be made. This is because the field economists are considered the owner of a schedule and all changes must be made through him/her. From the field economist perspective, this prevents overestimating the error rates since it gives the field economist an opportunity to reconcile his/her collection with the reinterview results. There may be a valid reason for a difference between the original collection and reinterview. For example, questions that arise from the reinterview are often a matter of lack of documentation on the original decision. However, from a statistical perspective the true error rate may be underestimated, since the original field economist may decide not to make changes to the data because he/she strongly believes his/her original decision was correct. For this reason, we examine both the weighted percent questioned and the weighted percent changed for a particular data element knowing the true error rate lies probably somewhere in between. However, the primary focus is on the weighted percent changed, since this represents the percentage of a particular data element in the survey that would be corrected or changed upon quality review and reinterview.

When analyzing reinterview data, we examine the weighted percent changed and weighted percent questioned for almost every element on the schedule. We then make comparisons of each data element by regional collection office, by field economist, and by reinterviewer to check for any differences among each group. Often, the numbers are too thin to make these comparisons for every data element. In which case, we

aggregate the data elements into four separate sections, establishment background, employment, worker characteristics, and generic leveling factors. Establishment background includes such information as establishment address, contact's name, and phone numbers. Although this is important for professional contact with the respondent, the establishment background section has no bearing on the quality of the compensation data. The employment section contains all elements that pertain to the various types of employment for an establishment. The worker characteristics section contains all job classification information except the generic leveling criteria.

Changes to Levels

In NCS, generic factors are used to determine the overall level of a job. Each of the 10 generic factors is assigned a level which has a corresponding point total. The point totals are then summed to determine the overall level of the job. This generic leveling process is essentially mapping private, state, and local government jobs to federal government GS-levels. Since this was the greatest departure from OCS, the quality review of the generic factors is the primary interest in NCS. The following results are from NCS TRP data collected from mid-March 1997 to mid-March 1998.

One of the more significant findings we uncovered while examining reinterview data came from the analysis of the magnitude and direction (+ or -) of change of the generic factor levels. Using the unweighted sampled factors, Table 4 presents the difference between factor levels before and after the TRP (diff. = level after TRP – level before TRP). The factors in bold type indicate if there were significantly more decreases in level than increases in level for that factor. This test was done using a nonparametric sign test with the null hypothesis being that an increase in level is as likely as a decrease in level. Each test was conducted at the .05 significance level.

TABLE 4. Difference in Factor Levels after TRP

Factor	<=-2	-1	No Diff.	+1	>=+2
Knowledge	4	38	953	16	2
Guidelines	3	23	961	13	.
Complexity	1	32	956	14	.
Scope/Effect	2	22	956	12	1
Sup Received	.	13	700	10	.
Pers Contacts	1	16	695	9	.
Purp of Cont	.	15	701	3	1
Phy Demands	.	6	713	9	.
Work Envir	.	13	696	9	.
Sup Span	1	6	705	9	.
Total	12	184	8,036	104	4

Approximately 96.4% of the time the factor did not change as a result of the TRP. Of the 304 factors that did change level, 288 (or 94.7%) of them were within one factor level of the original field economist decision. However, 8 of the 10 factors had more decreases than increases, and four of them (knowledge, guidelines, complexity, and purpose of contacts) had significantly more decreases than increases. This is indication that when one of these four factors is misleveled, it is more likely the factor was leveled too high. Overall, there were significantly more factors decreased than increased as a result of the TRP.

Since the factors are subsampled, the TRP is not designed to determine changes in the overall work level. That is, a change to one or two sampled factors may not change the overall level, but if all factors were reinterviewed then the level may have changed. Despite the intent being to find error rates for the factors, the TRP staff tracked changes to work levels that changed because of factors changed in the TRP process so we are able to obtain a rough estimate of the impact of the TRP on the work levels. Using similar methods as with the factors, we obtain Table 5 for the work levels. (Difference in Level = Level After TRP – Level Before TRP.)

TABLE 5. Difference in Work Level after TRP

Difference in Level	Number of Sampled Quotes	Pct. of Sampled Quotes
-3 or more	4	0.3%
-2	7	0.5%
-1	53	3.7%
No Change	1,341	93.4%
+1	22	1.5%
+2	1	0.1%
+3 or more	2	0.1%
Levelable after TRP	6	0.4%
Total	1,436	100%

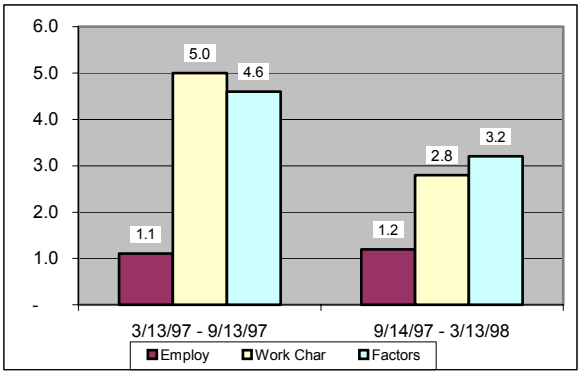
Approximately 6.6% of the occupational selections changed work level as a result of the TRP. There were 6 occupations that were not leveled before the TRP that became leveled after the TRP. Of the 89 occupations that changed work level, 64 (or 71.9%) that changed decreased in level. If we were to conduct a nonparametric sign test with the null hypothesis being that an increase in level is as likely as a decrease in level, then having this many more decreases than increases is significant at the .05 level. This means when a job is misleveled, it is more likely the job was leveled too high. However, 75 of the 89 work levels (or 84.3%) that changed were within one work level of the original field economist level.

Assessing Change Rate over Time

With any quality review program, the goal is to decrease the amount of errors. Since NCS is a relatively new survey and the generic leveling process demands more of the field economists collecting data. It was expected that the initial error rates would be higher than OCS, then as the field economists became more familiar with the process and received feedback on their collection the error rates would decrease.

If we look at the weighted percent change for the first six months versus the weighted percent change for the second six months of the March 1997 to March 1998 study, we obtain the following results (shown in Chart 1) for the employment, workers characteristics, and generic leveling factors section of the TRP.

Chart 1. Weighted Percent Change for Each Section over Time



If we conduct a test for the difference between two means, we find there are no significant differences between the time periods for the employment section. However, both the worker characteristics and factors showed a significant decrease at the .05 significance level from the first to second time period. The challenge for the TRP and other quality review staff will be to find ways to decrease the amount of changes further after the weighted change rates level off.

Impact on ECI Published Estimates

Ideally, we would like to be able to assess the effect of the non-sampling errors discovered during the reinterview process have on published estimates. This is not always a simple task in NCS, since we are dealing with many different locality surveys. However, we have looked at the impact of non-sampling errors in the quarterly update ECI TRP. In June 1996, we conducted a special study known informally as "No Change". The reason for the study was to verify the data was actually being updated each quarter and not simply being carried forward from the previous quarter without being updated.

To carry out this study, a subsample of establishments that reported no change from the previous quarter and a subsample of establishments that reported some change from the previous quarter was taken. The hypothesis was that the wages collected from establishments that reported no change would have a higher error rate than the wages collected from establishments that reported some change. Although "no change" establishments did have a slightly higher error rate, we were unable to confirm this hypothesis since the two types of establishments turned out not to be significantly different. However, one of the interesting results of the study was the comparison of the error rates by whether the wages were reported for each individual in an occupation or collected as an average for an occupation. For this quarter, it was attempted to collect individual wage rates as much as possible for the reinterview. In order to examine which errors would have more impact on published estimates, we used the measure of relative gross effect and the measure of relative net effect.

The measure of relative gross effect gives an estimate of the impact of the wage changes relative to the mean wage after the reinterview. The measure of relative gross effect, RGE, is calculated as follows:

$$RGE = \frac{\sum_j \sum_l W_j \cdot W_j' \cdot |\bar{Y}_{jl} - \bar{Y}_{jl}'|}{\sum_j \sum_l W_j \cdot W_j' \cdot \bar{Y}_{jl}'},$$

where W_j = the individual weight for the *jth* occupational selection in the survey, W_j' = the individual weight for the *jth* occupational selection in the reinterview, $\bar{Y}_{jl}$ = the original average wage for wage record l in the *jth* occupational selection in the survey, and $\bar{Y}_{jl}'$ = the corrected average wage for wage record l in the *jth* occupational selection after the reinterview.

The measure of the relative net effect, *RNE*, is calculated the same as the *RGE* with the exception that no absolute value sign is needed in calculating the difference between the average wages before and after the reinterview.

Table 6 shows the relative gross effect calculated two ways. The first way is calculated with "no change" discrepancies only. The second way includes all discrepancies found during the reinterview. Since during this quarter there was no reconciliation with the field economists regarding differences between the reinterview results and the original data collection, the difference is considered as a "discrepancy" due to the fact it was not confirmed to be an error.

Table 6. Measure of Relative Gross Effect

	"No Change" Discrepancies	All Discrepancies
Overall	.006018	.009597
Update & TRP avg wages	.01572	.01572
Update & TRP ind wages	.00299	.00756
Update avg wages & TRP ind wages	.02108	.02108

In the first part of the above table, the measure of RGE is about 0.6% for "no change" discrepancies and about 1% for all discrepancies. The second part of the table shows the measure of RGE is smallest when both the original collection for the ECI quarterly update and reinterview are collected using individual wages.

The table below displays the measure of relative net effect in the same two ways as before with "no change" discrepancies and then all discrepancies. The measure of relative net effect of the discrepancies gives an estimate of the net impact of the wage changes on the quarterly update relative to the mean wage after the reinterview for "no change" occupations. Since the results are negative, this would imply if the difference were significant that the original collection underestimated the mean wage according to the results of the reinterview.

Table 7. Measure of Relative Net Effect

	"No Change" Discrepancies	All Discrepancies
Overall	- .00564	- .0049
Update & TRP avg wages	- .0222	- .0222
Update & TRP ind wages	- .00139	- .0005
Update avg wages & TRP ind wages	- .0148	- .0148

In the first part of Table 7, it is evident that the measure of RNE is small, about 0.56% for "no change" discrepancies and about 0.49% for all discrepancies. It is interesting to note that the relative net effect is closer to zero for all discrepancies than for "no change" discrepancies. The additional discrepancies have an overall positive contribution. This means the average wage reported for these occupations was higher in the original collection than in the reinterview. The results for the measure of RNE also showed that when the wages for the ECI quarterly update and the reinterview were collected for individuals from an establishment rather than as averages for a job the effect those changes had on the overall published estimate was smallest.

CONCLUSION AND FUTURE

Utilizing the reinterview program, we are able to identify areas of concern in collection, assess whether collection is improving over time, and obtain estimates of the effect collection errors may have on published estimates for BLS compensation surveys. Some of the significant findings of the reinterview program over the last couple years include: the percentage of collection errors in NCS is decreasing over time, when a job is misleveled in NCS it is more likely the field economist leveled the job too high, and collecting individual wage rates is preferable to collecting an average wage rate for an occupation in ECI.

The Technical Reinterview Program has already grown to incorporate establishments being updated for the NCS. In the future, it will grow to incorporate establishments being initiated into the ECI as opposed to only establishments being updated in the ECI. This expansion of the program, along with continuous refinement of the reinterview process, will help BLS to continue to minimize non-sampling error and ensure quality compensation data.

REFERENCES

Cochran, W.G. (1977), *Sampling Techniques* (3rd ed.), New York: John Wiley.

Kydoniefs, L. and Paben, S. (1997), "Analysis of TRP Reason Codes Used in Round 5 NCS Survey Data," Bureau of Labor Statistics report, dated December 11.

Paben, S. (1996), "OCSP JMV (TRP) Round 4 Analysis," Bureau of Labor Statistics report, dated November 26.

Paben, S. (1996), "ECI 'No Change' TRP Analysis," Bureau of Labor Statistics report, dated December 18.

Paben, S. (1998), "Changes to Initiation Data -- March 1997 to March 1998," Bureau of Labor Statistics National Compensation Survey Technical Reinterview Program report, dated April 6.

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