

Reducing Sample Sizes in the National Compensation Survey in Response to Budget Cuts

Christopher J. Guciaro, Lawrence R. Ernst, Gwyn R. Ferguson, Yoel M. Izsak, and Erin E. McNulty
 Guciaro.Chris@bls.gov, Ernst.Lawrence@bls.gov, Ferguson.Gwyn@bls.gov, Izsak.Yoel@bls.gov, McNulty.Erin@bls.gov
 Bureau of Labor Statistics, Room 3160, 2 Massachusetts Ave N.E. Washington DC 20212-0001

Keywords: sample cut, budget cut, sample size, sample reduction, sample supplement

1. Introduction

In 2004, the National Compensation Survey (NCS) faced a reduction in its operating budget. To help compensate, the number of business establishments in the NCS sample was reduced, which allowed a reduction in the number of field economists required to collect NCS data. This report describes the reduction scenarios that were considered, and the degree to which they reduce the sample sizes, the data collection workload, and the number of field economist positions.

2. Background

2.1 Purpose and History of the NCS

The NCS is administered by the U.S. Bureau of Labor Statistics (BLS), to support the following publications:

- 1) Occupational Wages – mean wages, by year, including estimates for individual localities
- 2) Employment Cost Index (ECI) – quarterly changes in hourly wages and benefit costs. The ECI is a principal federal economic indicator.
- 3) Employer Costs for Employee Compensation (ECEC) – mean hourly wages and benefit costs, by quarter
- 4) Incidence and Provisions – percent of employees with access to benefit plans, percent of those with access who participate, and data on key provisions of select benefit plans, by year.

There are no locality estimates for 2, 3, and 4.

The NCS microdata used by the Occupational Wage publications are also used to produce special estimates that assist the “President’s Pay Agent” (consisting of the Secretary of Labor, the Director of the Office of Management and Budget, and the Director of the Office of Personnel Management) in making recommendations on federal locality pay. Annual increases in pay to most federal employees come from two sources: a general increase that is the same for all employees, and a locality pay increase that depends on work location. To determine locality pay, the U.S. is divided into three parts: (1) a set of “locality pay areas”, (2) Alaska and Hawaii, and (3) a set of the remaining localities in the contiguous U.S., called “Rest of U.S.” (RUS). Locality pay increases are computed for the locality pay areas and for RUS (yet not

for Alaska and Hawaii), proportional to the pay disparity between federal and non-federal workers. The NCS microdata is used by the BLS to produce estimates of non-federal pay for the Pay Agent.

2.2 NCS Establishment Sample Design

The microdata used for the Occupational Wage publications and used for Pay Agent estimates comes from one large sample of business establishments, called the NCS sample. The microdata used for the ECI, ECEC, and Incidence and Provisions publications, however, come from a subsample of the NCS sample, called the ECI sample. Wages are collected from all NCS sample units, yet information on benefits (costs, and incidence and provisions data) is collected only from units in the ECI sample. Those units in the NCS sample but not in the ECI sample are called Wage-Only units. Because the ECI is a principal federal economic indicator, all current ECI units were retained during the 2004 budget reduction, and future ECI sample sizes will remain as planned, with no reductions in ECI sample sizes.

The NCS sample and the ECI subsamples each contain three parts: a sample of state and local government establishments, a sample of aircraft manufacturing establishments, and a sample of all other private industry establishments. It was decided neither to reduce the government nor the aircraft manufacturing samples (unless, of course, an entire locality was dropped). Therefore, the following discussion will only focus on the sample design for the private, non-aircraft manufacturing sample.

The current NCS sample design has three phases. First, a sample of 154 localities are selected. Some of these localities were selected with certainty, either because they were locality pay areas, or had extreme employment. The remainder comes from a stratified sample, with one locality selected from each primary sampling stratum. Localities were grouped into primary sampling strata based on Census Division, metropolitan-area status (met./non-met.), and mean annual earnings. Second, a sample of establishments is selected from these 154 NCS localities. Third, a sample of jobs is selected from the NCS sample of establishments. The ECI sample also has three phases, but is integrated into the three phases of the NCS sample. The ECI uses only 151 of the 154 NCS localities. Originally, the NCS and ECI samples were selected from the same sample of 151 areas. Next, the Pay Agent added three more localities to their list of locality pay areas. One locality was new to the sample,

yet two were already sampled noncertainty localities. When these latter two localities were made certainty localities, then they were removed from their primary sampling strata, and two new localities were chosen as replacements to represent these two strata. These changes caused some modifications in PSU-weights (inverse probability of selection of a locality PSU). The ECI sample, however, did not require these changes, so it retained the 151-locality original design, and the original PSU-weights. The ECI sample of establishments is a subsample of the NCS sample of establishments. The ECI sample uses the same jobs used in the NCS sample.

The NCS private industry sample uses a five-year rotating sample design. Generally, there are five rotation groups in sample at any given time. One group is dropped each year, and replaced with a new group, which is in sample for five years. The private ECI sample also has five rotation groups. Each ECI group is subsampled directly from its associated NCS group. The NCS sample selection process can be broken into two parts. The first part involves the selection of units for the first five rotation groups, 01, 02, 03, 04, and 05. The second part involves the most recent five rotation groups, 101, 102, 103, 104, and 105. The plan was for 01 to drop out when 101 is initiated, for 02 to drop out when 102 is initiated, and so on, until 05 drops out when 105 is initiated.

In 1996, one large NCS sample of establishments was selected. Some of these establishments were sampled with conditional probability less than 1.0 (conditioned on the sample of localities). These establishments were split randomly, yet evenly, into five groups, and each group was assigned to one rotation group in 01-05. These establishments are called “single-group establishments”, because each establishment is contained in a single group. The remaining establishments have a conditional probability of 1.0. Unlike the single-group establishments, each of these remaining establishments were allocated to *all* five rotation groups. Specifically, five “copies” of each establishment are made, one for each group. These establishments are called “five-group establishments”. To compensate for having five copies, the weight applied to each copy is 1/5 the weight of the original sampled unit.

For 101-105, a new sample of establishments was selected. As with 01-05, one copy of each five-group unit is placed in each of 101-105, and the weights divided by 5. Yet the single group establishments used for 101-105 were determined differently than they were for 01-05. For 01-05, the five single groups had come from one large sample selected in 1996. For 101-105, however, a new single group sample is selected each year from the frame (with five-group units excluded). Since an establishment has a chance of being selected in each of the five groups (as long as it stays in the frame), its sampling weight is 1/5 of the

sum, over all groups G in which it was selected, of the inverse probability that it is selected in the single group G .

The ECI samples were chosen similarly, yet rather than being sampled from the NCS sampling frame, the ECI sample groups were subsampled from the corresponding part of the NCS sample. Hence, all ECI five group units are NCS five-group units, and each ECI single-group sample is a subsample of its NCS single group sample.

2.3 Sample Size Allocation Procedure for 101-105

For each rotation group, the sample of establishments is a collection of independent samples selected from a collection of “sampling cells”. Each sampling cell is an “industry sampling stratum” restricted to either a single sampled locality, or to a cluster of sampled localities. There is one set of sampling cells for the NCS sample, and a separate (yet overlapping) set for the ECI sample. Both use the same industries, yet not the same localities or locality-clusters. For NCS, there are two clusters, one contains 99 of the smaller-sized localities, the other contains the 3 non-metropolitan counties sampled from Alaska or Hawaii. The other 52 localities each have their own sets of sampling cells, equal to their industry sampling strata. The 52 localities were chosen either because they were already locality pay areas or there was a chance they would become locality pay areas in the future. For the ECI, there is one cluster of 140 localities, and the other 11 localities each have their own ECI cells. The NCS has more cells because of the need to compute estimates for individual localities.

Separate allocations were run to determine sample sizes for: the NCS five-group units, the ECI five-group units, the NCS 101 single-group samples, and the ECI 101 single-group samples. Single-group samples for 102-105 used the same sizes as 101 (with minor adjustments to account for frame changes over time). The allocation procedures must meet several requirements. First, the total sample sizes for certain collections of cells (see next paragraph) are preset, and must not change. Second, the total number of ECI establishments selected from an NCS cell must not exceed the NCS sample size in the same NCS cell, and this NCS sample size must not exceed the number of units in the sampling frame in the cell. Third, if possible, the sizes must be proportional to “PSU-weighted” frame employment in each cell. The PSU-weighted frame employment is the sum over all establishments in the frame of the establishment employment times the PSU weight (inverse probability of selection of the locality PSU). In most situations, a simple proportional-to-size allocation violates requirement two for some cells. The solution is to first fix the sample sizes for some cells to meet requirement two (using a special methodology), and then allocate the remaining units to the remaining cells, proportional to size. After this a “controlled rounding” is performed, to meet requirement

one. For more details on allocation, see the two Ernst et al. (2002) papers.

The ECI sample requires good industry coverage spanning the nation, so the ECI sample sizes were fixed for each industry, and then allocated to ECI cells. The NCS sample, however, focused more on having good coverage for each locality wage publication, which included the locality pay areas. Therefore, the NCS sample sizes were fixed for 36 localities, which included all 33 locality pay areas, and 3 other certainty localities. The remaining sample was allocated, proportional to size, to Raleigh, Kalamazoo, and one large cluster (called “R116”) of the remaining 116 localities. Raleigh and Kalamazoo required special treatment in the allocation procedure because they are not in the ECI sample.

3. Reduction Scenarios

3.1 Reducing the Sample Sizes

At the time of the sample reduction, the non-aircraft, private industry sample for both NCS and ECI consisted of groups 03, 04, 05, 101, and 102. All these samples were in update collection mode, meaning they had been initiated prior to 2004. Rotation group 103 was already sampled, but not yet initiated for data collection.

It was decided to preserve the ECI sample intact, as well as the collection of NCS five-group establishments. It was also decided, in general, to attempt to reduce the NCS sample sizes for most of the larger localities. The details of what localities were targeted for reduction, and why, is described later. For now, it is enough to know that, for each single-group sample, each of these large localities had originally been assigned a fixed NCS sample size, which then had been allocated to its industry strata. So the goal was to have a reduced *single-group* sample size for *each* rotation group, for *each* industry stratum, for *each* of these localities.

For each locality, one approach would be to apply many small reduction procedures, independently by cell. Yet this produces new problems, mostly involved with meeting the minimum and maximum constraints on sample sizes, and the effects of rounding. For example, suppose a 20% reduction is required for a locality. Then under this faulty method a 20% reduction is made to each cell, and then various adjustments and rounding are performed, all independently by cell. Yet the reduced size for the entire locality may no longer be 20% of the original size.

The solution was to perform the reduction procedure first on the largest possible collection of sample units in the locality, and then apply the original allocation procedure to split this new aggregate size among the cells. The original

allocation procedure automatically handles all the problems described in the previous paragraph.

For each locality, a reduction was made first to the “overall” sample size in the locality (this overall size includes five-group units, and all single-group samples). This process is first done for 01-05, and then independently, for 101-105. The resulting two sample sizes are only *targets*; however, because adjustments have not yet been made to meet minimum and maximum constraints. Second, new targets are determined for the single group sample sizes. The overall target sample size equals the number of five group units plus five times the number of single group units in any *one* rotation group. Hence, the target reduced single group size for 03 equals 1/5 the difference of the original overall size for 01-05 minus the number of five-group units in 01-05. The same target reduced size is used for 04 and 05. The target reduced single group size for 101 equals 1/5 the difference of the original overall size for 101-105 minus the number of five-group units in 101-105. The same target reduced size is used for 102, 103, 104, and 105. Table 3 shows the “target” numbers that were used for Scenario 4 (the scenarios are defined Section 3.3). The third step is to make adjustments to each target, for each rotation group, to accommodate all minimum and maximum constraints. The final step is to reallocate sizes to the industry strata.

Common to all scenarios was the desire to eliminate the “Pay Agent supplement”. The original intent of the NCS was for sample sizes to be allocated to locality pay areas, proportional to PSU-weighted employment. Yet this produced sample sizes for many locality pay areas that were deemed inadequate to provide good estimates for the Pay Agent. Therefore, the sample sizes for these localities were increased, and this increase was called the “Pay Agent Supplement”.

Before the Pay Agent supplement, sample sizes were proportional to employment, yet after the supplement was applied, they were no longer proportional to employment. So the simplest method to eliminate the supplement was to force the overall sample size for a locality to be proportional to its PSU-weighted employment. This can be done by multiplying the PSU-weighted frame employment for a locality by an adjustment factor equal to the ratio of two terms:

- (1) Total sample size before the Pay Agent supplements were added, summed over all localities and all five rotation groups.
- (2) PSU-weighted frame employment at the time units were first sampled, summed over all localities.

Unfortunately, data for (1) was incomplete or insufficiently documented, yet a solution was found. One characteristic of a proportional-to-size sample is that the adjustment factor defined above remains the same even if the sums in the

numerator and denominator are made over a smaller set of localities. So attention was reduced to the cluster of 116 localities (R116) defined in Section 2.3. This cluster represented the largest collection of localities for which “proportional-to-size” sample sizes were available. None of these 116 localities had been supplemented, and none had their sample sizes reduced. Two adjustment factors were computed, one for 01-05, and one for 101-105.

Nine separate reduction scenarios were considered, which differed in the set localities reduced and the magnitude of the reductions. For both 01-05 and 101-105, for each locality being cut, we first multiply the PSU-weighted employment for the locality by the adjustment factor that had been computed from R116 data. The difference between the “original overall sample size” (that includes the Pay Agent Supplement) and this new “target, reduced, overall sample size” is referred to as the “full-cut” to the Pay Agent supplement. For a given locality, some scenarios proceed with the reduction process by applying this “full-cut”, yet other scenarios apply only *half* of this “full-cut”. This latter approach is called the “half-cut” to the Pay Agent supplement. Yet regardless of whether or not the full-cut or half-cut is applied, once a reduction is made to the overall sample size, the target single group sample sizes are still found by subtracting off the number of five-group establishments from the target reduced overall size, and then dividing by 5.

In some scenarios, an attempt was made to apply a full-cut to Chicago, Los Angeles, New York, and San Francisco. In each case, however, the target reduced sample size that was obtained (after applying the adjustment factor) amounted to a sample size increase, rather than a decrease. As a result, it was decided to not decrease the sample size for any of these four localities, for any rotation group. This decision was also extended to those scenarios where a half-cut was planned, since 1/2 of a zero cut is a zero cut.

3.2 Sampling the Reduced-size Samples

Before the sample reduction, the NCS and the ECI single-group samples for groups 03, 04, 05, 101, and 102 were already initiated and in update collection mode. One goal was to reduce the NCS sample sizes, yet at the same time keep the ECI sample already being collected intact. Therefore, any size reduction had to come from the Wage-Only sample (NCS units not in ECI sample). The reduced Wage-Only sample size equals the reduced NCS sample size minus the ECI sample size. One possible approach is select a new Wage-Only sample from the same frame (excluding ECI units) used originally to select the NCS sample. Yet this would force a new round of initiations, rather than retaining units already initiated and being collected. Also, the new set of Wage-Only units may not accurately reflect what might have occurred had a new

NCS sample of the reduced size been selected first, then a new ECI sample selected. Finally, any units reselected from an older frame may no longer be in existence or in scope.

Therefore, for each NCS sampling cell in 03-102, subsamples were selected from the set of Wage-Only establishments that were already in the sample before the reduction. The reduced NCS sample is then this reduced Wage-Only sample plus the untouched ECI sample.

Before the sample reduction, the NCS and ECI samples for 103 had been selected, but not yet initiated for data collection. Therefore, there is no danger of altering ECI sample data, so for 103 it was decided to let the reduced NCS sample be subsample of the *entire* original NCS sample. Since the ECI must be subsample of the NCS, we had no choice but to select a *new* ECI sample as a subsample of this new, reduced NCS sample. Yet the ECI sample sizes remained unchanged.

For future groups, since there are no “original” NCS or ECI samples, all new NCS samples are selected directly from the frame (with five-group units excluded), using sample sizes equal to the reduced NCS sample sizes. The ECI sample is a subsample of this new sample, with sizes equal to what had been planned before the NCS sample reductions.

3.3 Reduction Scenarios That Were Considered

Scenario 1. Drop the three localities that were not in the ECI sample (Huntsville, Raleigh, and Kalamazoo).

Scenario 2. Implement Scenario 1, and then apply the full-cut to the Pay Agent supplement to Kansas City, Orlando, and St. Louis. These localities originally were locality pay areas, and had a Pay Agent supplement. Yet they will no longer be locality pay areas starting in 2005 (they will be placed in Rest of U.S.), so that reducing the sample sizes will not effect the individual locality estimates computed for the Pay Agent.

Scenario 3. Implement Scenario 2, and then apply the full-Cut to the Pay Agent supplement to three more localities: Charlotte, Phoenix, and Tampa. These are certainty localities, but are not locality pay areas. It was originally felt that these might become locality pay areas; hence, all had been given a Pay Agent supplement.

Scenario 4. This was the method chosen for the actual NCS sample cuts; it is also called the “final reduction scenario”. First, implement Scenario 2, yet *retain* the entire sample in Huntsville (see the beginning of Section 3.4 for a rationale). Then, *attempt* to apply the half-cut to the Pay Agent supplement to a special collection of 30 localities, which consists of all certainty localities in the NCS sample, except for the following eight certainty localities: Anchorage,

Honolulu, Birmingham, Huntsville, Raleigh, Kansas City, Orlando, and St. Louis. Anchorage and Honolulu are excluded because Alaska and Hawaii are not part of the Locality Pay Program. Birmingham is excluded because it has never been a possible candidate to become a locality pay area, and therefore did not receive a Pay Agent supplement. The remaining localities are excluded because they are already handled: Huntsville is uncut, and Raleigh, Kansas City, Orlando, and St. Louis are already handled by Scenario 2. This collection of 30 localities receiving the half-cut *includes* Charlotte, Phoenix, and Tampa.

Note the use of the word “attempt” here. In practice (for reasons described in Section 3.1), no reductions could be made to Chicago, Los Angeles, New York, or San Francisco.

Scenario 5. Implement Scenario 4, yet *drop* Huntsville (as Scenario 2 requires).

Scenario 6. Implement Scenario 5, except apply the full-cut, rather than the half-cut, to Charlotte, Phoenix, and Tampa. Note that Scenario 6 attempts to apply the half-cut to 27 localities.

Scenario 7. Implement Scenario 6, then drop the entire Wage-Only sample from all non-metropolitan localities.

Scenario 8. Implement Scenario 6, except apply the full-cut, rather than the half-cut, to the 27 localities mentioned in the description of Scenario 6.

Scenario 9. Implement Scenario 8, then drop the entire Wage-Only sample from all non-metropolitan localities.

Table 1 summarizes the nine scenarios, and details what types of localities were cut, and the degree of the cut. The “33 localities” referred to in Table 1 include Chicago, Los Angeles, New York, San Francisco, and the 29 localities listed in Table 3.

Table 1. Types of Cuts, by Scenario and Locality

Group of Localities		S.1	S.2	S.3	S.4 Final	S.5	S.6	S.7	S.8	S.9
33 Localities	Kansas City, Orlando, St. Louis	N	F	F	F	F	F	F	F	F
	Charlotte, Phoenix, Tampa	N	N	F	H	H	F	F	F	F
	27 Localities	N	N	N	H	H	H	H	F	F
Raleigh, Kalamazoo		D	D	D	D	D	D	D	D	D
Huntsville		D	D	D	N	D	D	D	D	D
Alaska-Hawaii Sample		N	N	N	N	N	N	N	N	N
Remaining Metropolitan		N	N	N	N	N	N	N	N	N
Rem. Non-Metropolitan		N	N	N	N	N	N	D	N	D

N = No Cut, H = Half-Cut, F = Full-Cut, D = Dropped

We were not able to get reduced sample sizes for each cell for each scenario (except for the final scenario) because of the complexity and time constraints. However, projections were made on the percent of sample that would be cut. Table 2 displays two things: (1) the sample sizes for single-group samples for 03-102 that existed before reduction (summed over all 154 localities), and (2) the percent that these sizes would be reduced, for each of nine reduction scenarios.

Table 2. Percent Reduction in Sample Sizes, by Group and Scenario, for All 154 Areas

Single Group Sample	Sample Size, with No Cut	Percent Reduction								
		S.1	S.2	S.3	S.4 Final	S.5	S.6	S.7	S.8	S.9
03	5,600	1	2	4	8	9	10	16	16	29
04	5,584	2	2	4	9	9	10	15	16	29
05	5,601	2	3	4	9	9	10	16	16	29
101	6,935	2	4	6	13	14	15	20	23	28
102	6,936	2	4	6	13	14	15	20	24	29
103	6,938	2	4	6	14	14	15	20	24	28

3.4 Final Scenario Implemented

Scenario 4 was chosen for implementation not because of the size of the sample cut, but because of the size of the staff cut. See Section 4 for details on staff size reductions.

Special Treatment for Huntsville in Final Scenario:

Originally, there were only eight scenarios (1-3, 5-9), and no Scenario 4. Initially, Scenario 5 was chosen as the “final scenario”; hence, the initial plan was to drop Huntsville from the sample. This plan was changed, however, because the Pay Agent still needed locality estimates for Huntsville. Therefore, a new “final scenario” was defined, called Scenario 4, to be equal to Scenario 5, yet with no sample cut from within Huntsville.

Before the adoption of Scenario 4, all other eight scenarios dropped Huntsville, Raleigh, and Kalamazoo; that is, they only retained the 151 localities in the ECI sample. So before Scenario 4 was created, the plan was to replace the NCS PSU weights with their corresponding ECI PSU weights. When Scenario 4 was adopted, there were two options for computing *national* (such as All U.S.) estimates: (1) include Huntsville in the calculation, which would have forced some localities to use their NCS PSU weights, and some to use their ECI PSU weights; or (2), drop Huntsville and only use ECI PSU weights. Option 2 was chosen because it was both easier to explain and would lower the “between-locality” component of the variance.

Table 3 lists the 29 localities whose sizes were reduced under the final reduction scenario (Scenario 4). Listed are all the pre- and post-reduction sample sizes, for all groups 03-103, along with some of the numbers important to the reduction process. Table 3 excludes Raleigh and Kalamazoo, which were dropped entirely from the sample. For Kansas City, Orlando, and St. Louis, an attempt was

made to apply the full-cut. This was not always possible, because the ECI sample size exceed the NCS “target reduced sample size” for some NCS cells, resulting in the reduced sample size being set to the ECI sample size. The half-cut was applied to the remaining 26 localities listed in Table 3.

Table 3. Sample Sizes Used in Final Reduction Process

Locality	For Groups 01-05										For Groups 101-105							
	Original					Reduced					Original			Reduced				
	"Overall"	5-Group*	Group 03	Group 04	Group 05	Target		Actual			"Overall"	5-Group*	Single-Group 101, 102, or 103	Target		Actual		
						"Overall"	Any Single Group	Group 03	Group 04	Group 05				"Overall"	Any Single Group	Group 101	Group 102	Group 103
Kansas City	450	68	80	76	73	171	21	29	31	33	634	33	120	226	39	43	48	39
Orlando	262	18	46	40	51	144	25	25	26	26	358	22	67	210	38	42	38	38
St. Louis	395	44	70	73	79	249	41	41	41	41	584	35	110	314	56	59	56	56
Detroit	513	54	98	94	73	520	93	93	93	73	706	49	131	664	123	123	123	123
Washington	624	27	118	126	112	614	117	117	117	112	937	27	182	899	174	174	174	174
Atlanta	633	72	102	116	108	493	84	84	84	84	843	31	162	682	130	130	130	130
Boston	659	55	123	114	122	618	113	113	113	113	895	44	170	819	155	155	155	155
Charlotte	300	24	63	47	56	224	40	40	40	40	441	21	84	321	60	60	60	60
Cincinnati	439	57	63	81	77	315	52	52	52	52	564	37	105	403	73	73	73	73
Cleveland	520	40	93	98	98	406	73	73	73	73	761	59	140	560	100	100	100	100
Columbus	319	38	57	60	55	239	40	40	40	40	446	28	84	323	59	59	59	59
Dallas	568	37	97	108	116	517	96	96	96	96	804	23	156	731	142	142	142	142
Dayton	263	26	54	44	43	214	38	38	38	38	385	37	70	292	51	51	51	51
Denver	469	40	80	86	85	342	60	60	60	60	740	35	141	540	101	101	101	101
Hartford	280	28	49	55	49	198	34	34	34	34	404	30	75	273	49	49	49	49
Houston	626	73	115	112	107	504	86	86	86	86	811	27	157	665	128	128	128	128
Indianapolis	400	43	62	67	83	280	47	47	47	47	583	51	106	397	69	69	69	69
Miami	582	50	115	112	103	431	76	76	76	76	927	45	177	653	122	122	122	122
Milwaukee	386	57	74	61	67	288	46	46	46	46	487	32	91	358	65	65	65	65
Minneapolis	462	32	85	87	80	389	71	71	71	71	661	34	125	541	101	101	101	101
Philadelphia	623	55	112	117	118	589	107	107	107	107	897	50	169	792	148	148	148	148
Phoenix	437	57	81	75	78	346	58	58	58	58	543	29	103	460	86	86	86	86
Pittsburgh	446	71	79	72	72	330	52	52	52	52	726	65	132	497	86	86	86	86
Portland	388	51	65	68	79	275	45	45	45	45	562	23	108	409	77	77	77	77
Richmond	344	67	57	60	57	268	40	40	40	40	462	35	85	359	65	65	65	65
Sacramento	299	34	60	50	51	195	32	32	32	32	495	24	94	328	61	61	61	61
San Diego	481	70	78	75	91	328	52	52	52	52	686	33	131	479	89	89	89	89
Seattle	515	63	85	94	85	395	66	66	66	66	712	31	136	550	104	104	104	104
Tampa	429	48	72	78	83	313	53	53	53	53	582	52	106	427	75	75	75	75

* Number of "5-group units" in a *single* rotation group

4. Workload Reduction

An analysis was made of how each of the sample reduction scenarios would reduce the workload of the field offices, including projections of the number positions that could be cut. The goal was to cut the budget by a certain dollar amount through staff reductions. This dollar amount was translated into a target range for the number of positions to be cut. The final scenario (Scenario 4) was chosen with this target range in mind.

4.1 Estimates of Workload Reduction

Table 4 shows reduced sample sizes for each “collection-panel year” (four-quarter period starting with the second quarter of each year) and reduction scenario. It also has information on the initiations and updates that would be retained and dropped each year, as well as approximations of the percent of the positions to be eliminated as a result of the reduction. All numbers in Table 4 assume that the response rate at initiation is 70%, and the response rate during all update cycles is 100%; that is, if 100 units are initiated, then 70 units will respond initially, and these same 70 units will respond for each update cycle. The number of positions saved was computed by dividing the number of establishments that are reduced by the average number of Wage Only establishments collected per position per year. Initiations take about twice as long to process as update units, so that one calculation was made for initiations, one for updates, and then the results were summed.

Table 4. Percent Reduction in Sample Sizes and Staff by Year and Scenario, for All 154 Areas

Item	Coll Year*	Number With No Cut	Percent Reduction									
			S.1	S.2	S.3	S.4 Final	S.5	S.6	S.7	S.8	S.9	
Establishment	Initiations	04	6,938	2	4	6	14	14	15	20	24	28
		05	6,938	2	4	6	14	14	15	20	24	28
		06	6,938	2	4	6	14	14	15	20	24	28
		07	6,938	2	4	6	14	14	15	20	24	28
	Updates	04	25,829	2	3	4	9	10	11	15	16	24
		05	26,765	2	3	5	10	11	12	16	18	24
		06	27,713	2	3	5	11	12	12	17	19	24
		07	26,330	2	4	6	13	13	14	19	22	26
Staff Positions	04	-	2	3	5	11	11	12	17	19	26	
	05	-	2	3	5	11	12	13	17	20	26	
	06	-	2	4	5	12	13	13	18	21	26	
	07	-	2	4	6	13	14	15	19	23	27	

* Begins in 2nd quarter of given year, ends in 1st quarter of following year

4.2 Projected Workload Reductions for Final Scenario

Once it was decided to use the final scenario (Scenario 4), an attempt was made to get a more refined estimate as to the number of positions that would be saved. This analysis looked at data on a cycle by cycle basis, rather than just an annual basis. Also, rather than first focusing on positions, it focuses first on the number of worker-days (workers times days worked, based on an 8 hour day) required for all types of work. Later, these worker-days are converted into estimates of numbers of workers.

Table 5 shows the number of initiation and update units and the number of worker-days, by cycle, for the final reduction scenario. Note that worker-days for update establishments depends not only on the number of update establishments, but also the proportions that are Index and Wage-Only, which changes over time. The column “other tasks” covers time spent analyzing “summary plan descriptions” (brochures describing benefit plans) and time spent in on-the-job training. Table 5 also shows the surplus in available worker-days that would result if the final reduction scenario were implemented, yet no staff positions are cut. Surplus worker-days is computed by subtracting the projected worker days for the final reduction scenario from the maximum available worker-days that exists before any staff reductions are made.

The number of sample units in Table 5 represents the number of initiations and the number of respondent updates. It was based on the same type of information in Table 4, yet was enhanced since it also took into account what would happen each quarter. For example, the number of units from each rotation group that would be rotated out of sample, had there been no sample reduction, was already known for each cycle. This data was used to approximate the number of units that would be rotated out under the final reduction scenario. The projected numbers of initiations per quarter for 2004 and beyond, were computed by taking the total from 2003, subtracting off the annual reduction from Table 4 (for 2nd qtr 04 to 1st qtr 05), and dividing by 4. The projected number of updates was a more complex computation. For a given cycle *t*, it was set equal to the number from quarter (*t*-4), minus the units from this set that are dropped over the last four quarters preceding *t*, plus the number of respondents from initiations in quarter (*t*-4), assuming a 70% response rate.

The lack of initiations in the first half of 2004 is caused by a delay in the initiation of group 103. This delay gave us extra time to determine how the sample sizes should be cut, and also allowed cuts to be applied to group 103 before data was collected for the first time, rather than after the first round of collection (without this delay, some units might be collected only once, then dropped the following cycle because of the sample reduction).

Table 5. Projections for Final Reduction Scenario

Year	Quarter	Establishments Collected Under Final Scenario*		Worker-Days Required Under Final Scenario				Surplus Worker-Days **
		Initiations	Updates	Initiations	Updates	Other Tasks	Total	
2004	1	0	15,267	0	6,851	1,125	7,976	2,474
	2	0	13,822	0	6,237	1,125	7,362	3,088
	3	1,441	13,714	2,190	6,169	1,125	9,484	966
	4	1,441	13,627	2,190	5,993	1,125	9,308	1,142
2005	1	1,441	14,503	2,190	6,419	1,125	9,734	716
	2	1,441	14,427	2,190	6,250	1,125	9,565	885
	3	1,441	15,260	2,190	6,649	1,125	9,964	486
	4	1,441	14,537	2,190	6,382	1,125	9,697	753
2006	1	1,441	15,363	2,190	6,777	1,125	10,092	358
	2	1,441	15,304	2,190	6,618	1,125	9,933	517
	3	1,441	16,166	2,190	7,035	1,125	10,350	100
	4	1,441	15,456	2,190	6,777	1,125	10,092	358
2007	1	1,441	16,228	2,190	7,137	1,125	10,452	-2
	2	1,441	15,898	2,190	6,809	1,125	10,124	326

* The number of units collected is less than the sample sizes in Table 4, because wage-only units are collected once every four quarters, rather than every quarter.

** Surplus = Worker-days available in 2003, minus total worker-days required under the "final reduction scenario"

The surplus workload computation in Table 5 gives a good picture of how the quarterly workload might change from the pre-reduction level. Over time, however, this surplus tends to shrink in size. This is because the sample sizes under the final reduction scenario start out below 2003 levels, but then increase over time, approaching the 2003 levels. The increase is simply because the rotation groups 103-105 that would be cycled into the sample have larger sample sizes than the groups 03-05 that would be cycled out.

A similar, and in fact larger, rise in sample size would occur if no reductions are made at any time, for these same reasons. Therefore, although the workload under the final reduction scenario will approach the 2003 levels, it will tend *not* to approach the rising workload levels that would have existed had no reductions been made. In fact, based on the annual projections in Table 4, the workload levels for these two scenarios would likely diverge over time. For example, in Table 4 the final scenario starts out with a workload 11% below the "no reduction" scenario, and ends with a workload 13% below the "no reduction" scenario. In addition, the "no reduction" sample size rises between 2004 and 2007.

5. Conclusion

The work was challenging because the problem was not anticipated, a solution needed to be found quickly, and the research involved a complex sample design and allocation procedure. The number of scenarios considered and localities reduced was large, and all scenarios had to preserve the reliability of our key estimates, such as the ECI and the locality wage estimates. The processes of reallocation and resampling required a great deal of specialized effort by the team. At the same time, the normal production activities had to be maintained. Yet the team persevered and was successful in meeting these challenges.

References

Ernst, L. R., Guciardo, C. J., Ponikowski, C. H., and Tehonica, J. (2002). "Sample Allocation and Selection for the National Compensation Survey," *2002 Proceedings of the American Statistical Association*, Section on Survey Research Methods, [CD-ROM], Alexandria, VA: American Statistical Association, 920-925.

Ernst, L. R. and Guciardo, C. J. (2002). "Allocating Sample to Strata Proportional to Aggregate Measure of Size with Both Upper and Lower Bounds on the Number of Units in Each Stratum," *2002 Proceedings of the American Statistical Association*, Section on Survey Research Methods, [CD-ROM], Alexandria, VA: American Statistical Association, 914-919.

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