Reconciling Employment Differences between Administrative and Survey Data
October 2008

Margaret Applebaum, Kristin Fairman, Jeffrey Groen, and Polly Phipps
Bureau of Labor Statistics, 2 Massachusetts Avenue NE, Washington, DC 20212

Abstract

This paper investigates differences in employment figures gathered from unemployment-insurance tax filings under the Quarterly Census of Employment and Wages (QCEW) and employment figures reported through the Current Employment Statistics (CES) survey. Since these two Bureau of Labor Statistics (BLS) programs both collect monthly employment from an establishment for the same reference period, the employment figures should generally be identical. However, differences exist at the micro and aggregate levels, both at a point in time and in seasonal patterns. We analyze employment differences for 200 establishments with large differences in 2005-06 and determine the scope of and patterns in the differences. We then report on findings from a response analysis survey (RAS) in which the establishments were asked about reasons for differences in employment reported on the tax and survey forms.

Keywords: measurement error, survey and administrative data, employment, response analysis survey

1. Background

The QCEW is a quarterly census of all U.S. business establishments subject to unemployment-insurance taxes, about nine million establishments, and the CES is a monthly survey of 400,000 business establishments in the U.S. The QCEW is the sampling frame for the CES and all other BLS establishment surveys. Once a year, the QCEW is used to "benchmark" (or realign) sample-based CES estimates to incorporate universe employment counts. Differences in over-the-year employment growth between the two programs can result in benchmark revisions, including a large revision in 2006. Month-to-month employment differences with seasonal movement often contribute to over-the-year differences, and the differences are fairly consistent over time. This seasonal pattern is shown in Figure 1, which plots aggregate differences between QCEW and CES estimates of total U.S. private nonfarm employment for the 13-month period ending in March, from 2004 to 2007 (the 2004 series begins in March 2003 after the benchmark and ends in March 2004 before the benchmark). The difference is zero at the beginning of each period because the annual benchmark forces the CES estimate to align with the QCEW estimate. Several patterns are evident. From July to August, the difference between the two programs increases, as QCEW increases more than CES. From September to October, the difference falls, but from October to December, the difference increases, as both programs increase but QCEW increases more than CES. From December to January, the difference falls, with both programs decreasing in employment and QCEW decreasing more. Finally, from January to March, the difference increases, as both programs increase but QCEW increases by more. In 2006, the two programs fell out of alignment to a large extent due to differences during the fourth quarter of 2005 and the first quarter of 2006; this time period is the focus of this study.

A number of factors have been offered as potential reasons for CES and QCEW employment differences, including non-response, reporting differences, payroll-processing firm and software procedures, imputation formulae, sampling error, and business birth/death modeling; the last two are specific to the CES. Research on the 2006 CES benchmark revision concluded that non-response and payroll-processing firm procedures did not contribute substantially to employment differences, while reporting differences in some industries, birth/death modeling, as well as the effect of Hurricane Katrina due to imputation procedures, did contribute to these differences (Eickman, 2007). For this study, we focus on reporting differences, including the role of payroll-processing firms.

2. Sample Selection and Profile of Employment Differences

Establishments with large differences between the CES and QCEW from October 2005 to March 2006 were identified from ten states with large benchmark revisions in 2006 (Arizona, California, Colorado, Georgia, Nevada, North Carolina, Ohio, Pennsylvania, Texas, and Virginia). For each month, employment growth relative to the prior month was calculated separately for the CES and QCEW. The difference in employment growth between the CES and QCEW was then calculated, and an establishment was chosen for the sample frame if the absolute value of this difference exceeded a threshold for at least one month. The thresholds were based on employment-size class: 10 employees for size class 1-49, 25 employees for size class 50-249, and 50 employees for size class 250+. Only CES establishments with a one-to-one match with the QCEW were included (some companies provide aggregated establishment data to the CES, making it impossible to match to the QCEW at the establishment level), and units reporting via electronic data interchange were excluded, as well as a small proportion of QCEW establishments that had imputed values. Table 1 describes establishments in the sample frame, those that were contacted for the study, and those that responded. Of the 431 establishments included

in the frame, 191 were selected for the study, based on staffing availability at the data-collection centers. Overall, those contacted for the study were quite similar to the frame in terms of industry, size, multi or single unit status, and data collection mode. Compared to the frame, the group contacted had a slightly higher proportion of establishments in the 100-249 size class, multi-establishment units, and fax/mail/web respondents (fax comprised the greatest proportion of this category), and a slightly lower proportion of touchtone data entry (TDE) respondents. Approximately 86% (165 of 191) of those contacted responded to the RAS, and differences between respondents and non-respondents were small.
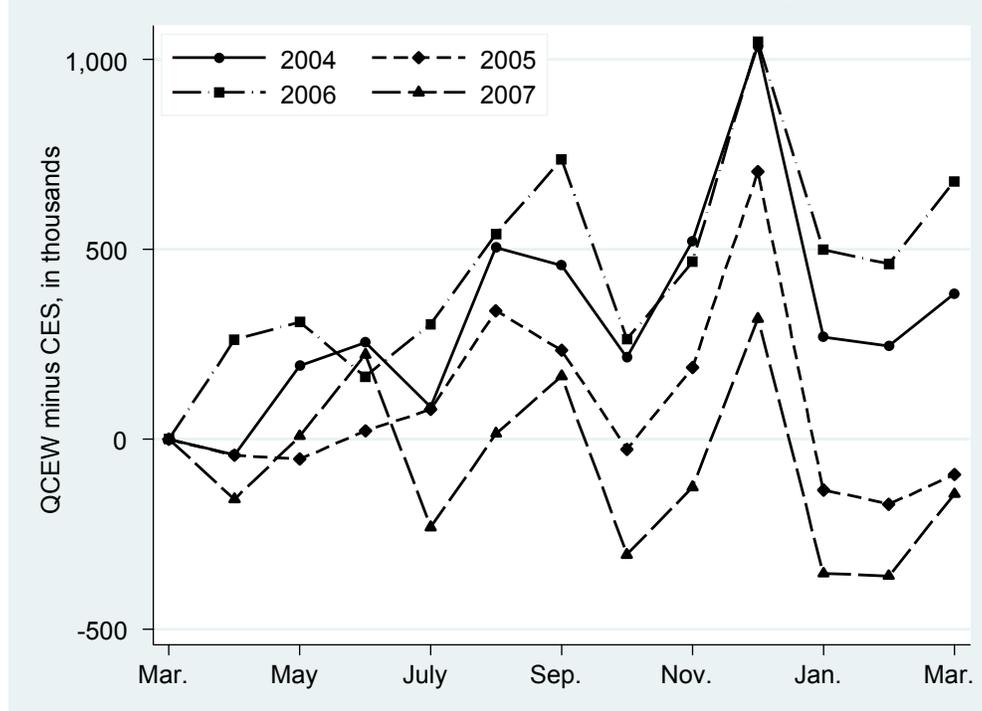

Figure 1. Difference between QCEW and CES Estimates of Private Nonfarm Employment, 2004-2007

Table 1. Characteristics of Frame and Response Rates (in percent)

	Entire Frame	Contacted	Among Contacted: Responded? Yes	No	Response Rate
Industry					
Manufacturing & Mining	11	9	9	8	88
Construction	14	14	16	4	96
Trade, Transportation, Utilities	19	21	20	31	80
Info. & Financial Activities	8	7	7	12	79
Prof. & Business Services	17	21	22	12	93
Educ, Health & Other Services	12	9	9	12	83
Leisure & Hospitality	19	18	17	23	82
Size (employees)					
1 to 19	11	8	8	4	93
20 to 49	26	23	21	35	79
50 to 99	9	8	9	4	94
100 to 249	13	21	22	12	93
250 or more	41	40	39	46	84
Multi or single establishment					
Single	70	61	61	62	86
Multi	30	39	39	38	86
CES collection mode					
TDE	58	40	44	15	95
Fax/Mail/Web	25	38	36	50	82
CATI	17	22	20	35	79

Number of establishments	431	191	165	26

To aid in the analysis of the RAS, we define a set of "difference groups" based on differences between CES and QCEW employment. Each group includes establishments whose reported employment over a particular time period exhibits a different pattern in the CES compared to the QCEW. For each time period (October to December, December to January, and January to March), change in employment from the first to the last month of the period is calculated for the CES and QCEW separately, and the absolute value of the difference of the changes is constructed. The difference group is represented by a variable that identifies whether or not the establishment's absolute value exceeds a size-based threshold (for which size is measured as the average of QCEW and CES employment for the first month of the period). The threshold is 3 employees for establishments with 1-9 employees, and the remaining thresholds are 7, 10, 15, and 20, respectively, for establishment sizes 10-49, 50-99, 100-249, and 250+. Table 2 shows the breakdown of the sample frame by the three difference groups (due to missing data on employment, error groups are not defined for 9 cases for October-December and 7 cases for January-March). For each group, about half of the sample meets our definition. An analysis of correlations indicates that the groups are reasonably distinct; the correlation coefficient for each of the pairwise combinations of group indicators is less than .15 in absolute value. However, there is overlap across groups, as 38% of RAS respondents are in one of the three groups, 34% fit into two groups and 17% fit into three groups (11% did not meet the definition of any group). Table 2 also presents employment statistics by the difference groups (for median employment, the data were first averaged over months for each establishment, using only the months that the establishment reported data for both CES and QCEW, then the median across establishments was taken from these averages). Establishments that meet our definition of a difference group are typically much larger than establishments that don't meet the definition. In addition, employment growth over the time periods reveals differences between CES and QCEW that are consistent with the definitions of the difference groups. However, given the selection criteria for our sample, they are not always consistent with the seasonal patterns we discussed earlier. For example, the magnitude of the employment growth/decline in the three groups is usually greater in the CES as compared to the QCEW, which is contrary to the usual seasonal pattern.

Table 2. Breakdown of Sample Frame by Difference Group

	All	Oct.-Dec. Group Yes	Oct.-Dec. Group No	Dec.-Jan. Group Yes	Dec.-Jan. Group No	Jan.-Mar. Group Yes	Jan.-Mar. Group No
N	431	215	207	259	172	207	217
Percent of total	100	50	48	60	40	48	50
Median employment							
QCEW	142.2	291.2	62.7	223.7	64.8	538.7	54.0
CES	151.0	262.5	58.5	212.5	72.8	618.0	52.0
Growth, Oct. to Dec.							
QCEW	16.3	23.5	8.9	29.1	-3.1	40.0	-5.6
CES	22.2	35.5	8.4	40.7	-6.0	57.2	-11.0
Growth, Dec. to Jan.							
QCEW	-25.4	-40.4	-10.1	-35.8	-9.8	-29.9	-22.1
CES	-33.2	-53.7	-10.9	-49.0	-9.3	-47.3	-19.9
Growth, Jan. to Mar.							
QCEW	7.4	4.9	10.2	9.0	5.1	10.4	4.6
CES	8.7	7.0	8.7	9.7	7.3	12.9	4.7

3. Response Analysis Survey

BLS has used the RAS method to investigate establishment survey data quality since the early 1980s (Goldenberg, Phipps, and Butani, 1993). The method involves contacting a respondent soon after survey completion, usually by telephone, and asking a short set of standardized questions on record-keeping practices, records availability and use, understanding of survey instructions and definitions, discrepancies between survey definitions and answers, and other data-quality issues. A RAS allows for direct questions to the respondent on the quality of the data provided, such as reasons for potential discrepancies, and also allows for indirect questions on survey definitions and data sources, both of which are useful in assessing data quality. For example, in this RAS we ask respondents directly about differences in employment figures. In addition, we ask questions about records sources and adherence to definitions such as the reference period and inclusion and exclusion of employee types; these items can be used independent of or in tandem with their answers to employment differences to evaluate data quality and identify problems. Another advantage of the RAS method is a larger sample size than other data-quality evaluation methods, such as cognitive interviews and respondent debriefings (the sample size of BLS RAS studies has ranged from 100 to several thousand respondents). The one major difference between this RAS and

others is a longer recall period for respondents. Since benchmarking production is done on a yearly basis, respondents had to be asked questions about prior-year data.

3.1 Data Collection Procedures

Data collection for the RAS was carried out by two CES data-collection centers and by CES and QCEW program staff in the BLS national office via telephone mode of administration. CES contacts in the sample frame were called in January 2007 to set up an appointment. At that time, they were faxed an advance letter explaining the purpose of the survey and a table with their CES and QCEW employment values and differences from October 2005 to March 2006. The RAS instrument included questions on both CES and QCEW data, so the CES contact was asked if someone else prepared the QCEW in case another person needed to be contacted for QCEW questions. In 36% of the establishments that responded to the RAS, the CES respondent also completed the QCEW, and we collected both question sets from that person. In another 19% of establishments, we collected the CES and QCEW questions from different respondents, while in 1% we collected only from a QCEW respondent and in 3% we collected only minimal data and were unsure whether the respondent prepared data for one or both programs. Finally, in 41% of the establishments we were able to collect only CES data; these were often establishments that used payroll services (payroll processors), and the respondent indicated that he or she was not knowledgeable about the payroll-processor procedures. We reviewed the characteristics of these establishments and found they were fairly evenly distributed across industry and size class. We did not attempt to contact payroll-processing companies for this survey, but have a separate study ongoing to evaluate payroll-processing company procedures and definitions for the QCEW and CES.

3.2 Research Questions and Data Items

Our question of interest is: Why do establishments report different employment numbers? The major data source for this question is the respondent's answer to why the numbers differed, which was recorded verbatim by interviewers and categorized by the authors. Table 3 shows the eight categories we used and the detailed codes that comprise them. The detailed codes are listed in order of highest to lowest frequency. For example, the 'human error' category includes clerical or posting errors by the respondent (the most frequently reported reason in the category), errors attributed to BLS, and also instances where a respondent indicated the difference was due to the count being an estimate, recalled from memory, or involving a manual-counting process.

Table 3. Respondent-Provided Reasons for Differences between Survey and Administrative Records

Category	Detailed Reasons
Human error	Clerical/posting error, BLS error, manual/memory/estimation count
Reference-period problems	Not reporting for pay period of 12th; month, quarter, or cumulative employee count
Automated reporting issues	Error/change in payroll software/processor, switched payroll software/processor
Employee-type reporting issues	Different employee types included or excluded, layoffs/closings
Worksite differences	Data includes more than one location, business structure or UI account changes
Counting checks	Counting of checks rather than employees
Data source or timing difference	Reports compiled at different times or from different data sources
Did not know/not respondent at time	Do not know, not contact person at time, response unclear

Table 4. Other Data Items Used in RAS Analysis

Data Item	Source	Definition
Industry	CES	Manufacturing, construction; trade, transportation, utilities; services
Size class	CES & QCEW	1-49, 50-249, 250+
Multi-or single establishment within state	QCEW	
Data-collection mode	CES	TDE, fax/mail/web, CATI
Pay groups	CES	Single or multiple pay groups
Records source	RAS	Internally developed payroll software, commercial payroll software, other internal data, payroll processor
Same or different record source for CES/QCEW	RAS	
Same or different person prepares CES/QCEW	RAS	
Difficulty reporting for pay period of 12th of month	RAS	
Report could include persons not working	RAS	

Program associated with the reason for difference	RAS	CES, QCEW, both programs, no program identified

For our analyses, we are interested in both the reasons for employment differences and also how they vary by difference group. For example, certain patterns, such as the drop in employment from December to January, might be associated with a specific difference reason, such as the counting of bonus checks in December, which might inflate December figures. Also, we are interested in a number of other potential explanatory variables, such as industry, size, multiple-establishment status, data-collection mode, records sources, and differences in respondents. Table 4 sets out additional data items we use in the analysis by source—the CES program, the QCEW program, or the RAS instrument.

3.3 Data Analysis

Table 5 shows the reasons for employment differences provided by the respondents (respondents sometimes reported reasons in more than one category; this table shows the percent of all establishments that reported at least one reason for the difference in a category). Overall, 37% of all establishments reported problems associated with including or excluding certain groups of workers as a reason for employment differences. This is followed by reference-period problems, human error, and respondents not being able to provide a reason (27%, 24%, and 18% of establishments, respectively). A smaller proportion of establishments reported data sources/timing, check counting, automated reporting, and worksite issues as reasons for the differences. Table 5 also breaks out the reasons by difference groups, but there are only a few differences by group. For example, the December-January difference group is less likely to be associated with human error, and slightly more likely to have employee-type reporting and automated reporting issues. Also, the January-March difference is less likely to be an employee-type problem.

Table 5. Differences between Survey and Administrative Records by Category and Seasonality-Pattern Groups

Reason	All Respondents (%)	In Any Group (%)	Oct.-Dec. Group (%)	Dec.-Jan. Group (%)	Jan.-Mar. Group (%)
Human error	24	20	19	13	19
Reference-period problems	27	28	28	30	30
Automated reporting issues	5	5	6	8	6
Employee-type reporting issues	37	35	35	38	30
Counting of checks	6	7	6	9	10
Worksite differences	8	9	5	7	7
Data source or timing difference	9	9	11	9	12
Do not know/Not respondent at time	18	19	23	18	22
Number of establishments	165	147	79	99	81

In Table 6, we investigate the categories that account for the largest proportions of reasons—human error, reference period, employee type and don't know—cross tabulating them with other possible explanatory data items. As with Table 5, respondents can be in more than one category and thus the percents can sum to more than 100 in a given row. Here we find stronger patterns than we did in the difference-group comparisons. Human error is more likely in manufacturing and construction (compared to other industries) and in smaller, single establishments. Human error is also greater in TDE, a self-reporting data collection method, and in establishments with the same record source and respondent. Reference-period differences, reporting employment counts for a period other than the pay period of the 12th of the month, are more likely in small establishments and among those using internal payroll software and other internal data as the records source. Employee-type problems, the inclusion or exclusion of certain types of employees in the employment counts, accounting for the largest proportion of reasons for employment differences, are more likely to be in trade, transportation and utilities and services, in the medium and large size classes, and in establishments with one payroll. These problems are also more likely when a payroll processor is used as the records source and different persons prepare the reports. This indicates there is not always a clear understanding of what is included in the payroll reports and therefore what is included in the employee counts submitted to CES and QCEW when different people prepare each report and the payroll is outsourced. We included a separate RAS question about employee types—asking if there was any possibility that the employment numbers could include persons not working during the reference period. Respondents saying yes to this question for either the CES or QCEW are more likely to be in the employee-type problem category. Further analysis of this question indicates that these establishments are in the 250+ employee size class and are more likely to be in the service sector. Finally, the most complicated establishments are more likely to be represented among those who do not know the reason for the difference: establishments in a large size class with multi-establishment status, multiple payrolls, and with a different respondent preparing the CES and QCEW reports.

Table 6. Reasons for Differences between Survey and Administrative Records by Category and Establishment Characteristics

	N	Human Error (%)	Reference Period (%)	Employee Type (%)	Don't Know (%)
Industry					
Manufacturing, construction	41	32	27	27	20
Trade, transportation, utilities	33	21	27	39	27
Services	91	21	27	40	13
Size					
1 to 49	48	35	33	27	15
50-249	52	37	25	37	13
250 or more	65	5	25	43	23
Multi or single establishment					
Single	101	32	29	35	15
Multi	64	11	25	39	22
Pay groups					
Multiple pay groups	29	23	30	29	20
Single pay group	125	28	24	72	10
Missing	11	18	9	27	9
CES collection mode					
TDE	79	37	25	34	16
Fax/mail/web	59	12	31	36	20
CATI	33	15	27	42	15
CES records source					
Internally developed PR software	16	19	44	25	38
Commercial PR software	70	24	29	34	19
Other internal data	23	13	35	43	13
Payroll processor	34	26	21	41	6
Missing	22	32	14	36	23
QCEW records source					
Internally developed PR software	11	27	55	18	27
Commercial PR software	48	27	33	35	19
Other internal data	7	14	43	71	0
Payroll processor	40	35	15	53	18
Missing	59	14	24	25	17
CES/QCEW records source					
Internally developed PR software	11	27	55	18	27
Commercial PR software	44	27	32	34	20
Other internal data	1	100	0	0	0
Payroll processor	23	39	13	52	4
Mixed records sources	51	14	25	25	22
One source missing	15	27	13	33	13
Both sources missing	20	15	35	65	15
Same/different records source					
Same	79	32	29	37	16
Different	71	14	28	37	20
Unable to determine/missing	15	27	13	33	13
Same/different person prepares reports					
Same	63	29	34	28	11
Different	90	20	23	42	23
Unable to determine/missing	12	17	25	42	17

In Table 7, we break down the reasons for the difference by whether the respondent associated the CES or QCEW with the reasons for the difference (here we tabulated all the detailed responses and used the totals for each category, rather than whether the establishment had at least one detailed response in the category, as we did in Table 5). For the CES program,

Table 7 shows that employee-type reporting issues account for the largest percent of the reported differences. This can be attributed to misunderstanding of definitions about who to include or exclude, and also to a disconnect in those establishments where reports are prepared by different persons. Even though employee-type issues are a common reason for the differences, it is often difficult to interpret respondent answers to identify whether the program is incorrectly reporting. The second largest reason for differences for CES is human error, partly attributed to one of the self-reporting methods, TDE. Also, since the CES is a voluntary monthly survey, there may be a higher propensity to manually count or have less-involved procedures to calculate employee counts, leading to a higher occurrence of human error. The third largest reason for differences in the CES program is reference-period problems, which are likely attributed to a misunderstanding of how to report for a certain week, or to the lag time between the availability of in-house payroll system reports for the requested reference period and the CES data collection period. Table 7 also presents the reasons for differences associated with the QCEW program, the most common of which is reference-period problems. Since the QCEW is submitted quarterly it appears to be difficult for respondents to understand that the reference period is the same as for the CES. A common problem among QCEW respondents is the reporting of employment for the entire month rather than the pay period of the 12th. The second most common reason for respondent reported differences associated with the QCEW is employee-type reporting issues. Since the QCEW is more often prepared offsite than is the CES (either by payroll processors or accountants), there is a disconnect in those establishments where reports are prepared by different persons (as noted with CES).

Table 7. Reasons for Differences Associated with CES and QCEW Programs

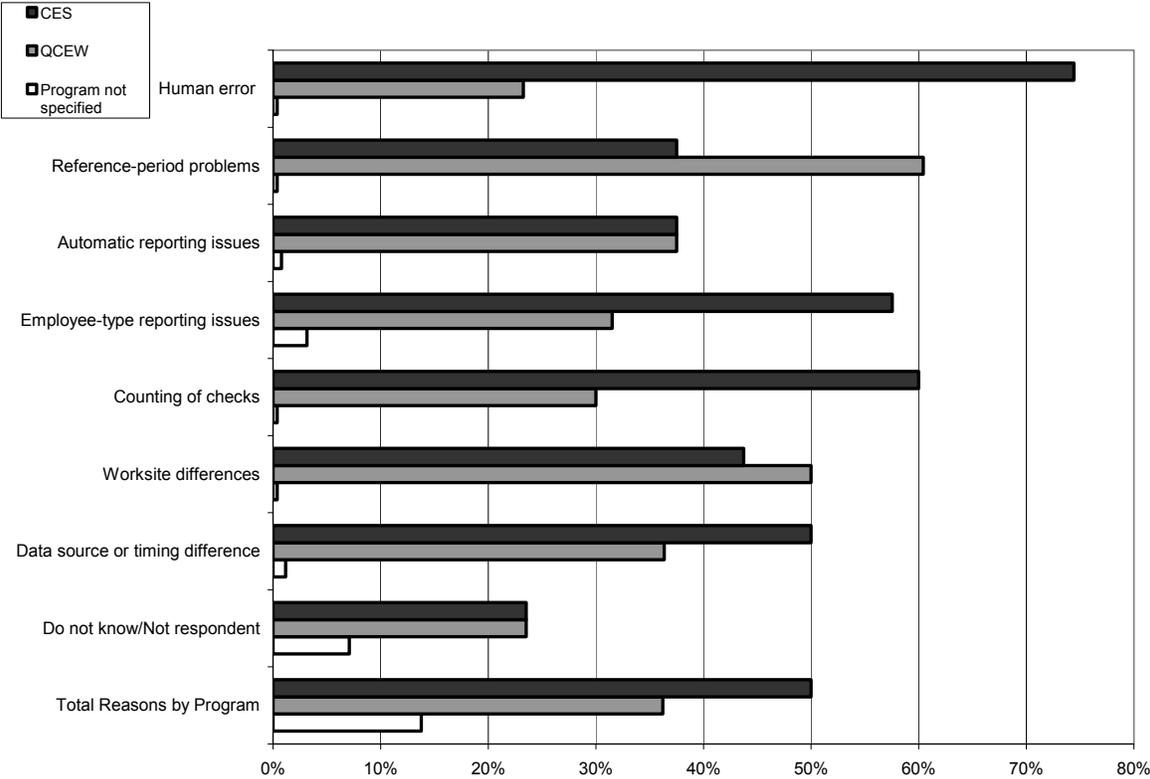| Reason	CES	QCEW
Number of responses	127	92
Human error	25%	11%
Reference-period problems	14%	32%
Automatic reporting issues	2%	3%
Employee-type reporting issues	33%	25%
Counting of checks	5%	3%
Worksite differences	6%	9%
Data source or timing difference	9%	9%
Do not know/Not respondent at frame	6%	9%

Figure 2. Major Reasons for Differences by Reason and Program

Figure 2 looks at the reason for the differences by category and must be viewed with some caution, as there are more CES than QCEW respondents, and thus more reasons for differences are associated with the CES. With that in mind, the results indicate that the CES accounts for 74% of the human-error reasons, while the QCEW is associated with 60% of the reference-period problems. Employee-type reporting issues are larger in the CES, as is the counting of checks, although the latter has a very small number of responses (10).

4. Summary and Future Work

There are a number of important findings from the RAS that will help build a foundation for future research and program improvements. First, the analysis of the difference-group patterns (reflecting timing) we set out indicates that the reasons for differences between CES and QCEW seem to cross the various groups rather than reside in one or another. However, since this RAS is on the low end of the usual RAS sample sizes, we think further research on patterns is likely to be fruitful with a larger sample (e.g., the December-January pattern is less associated with human error, indicating more systemic reasons for differences that can be explored further). Second, there are some fairly strong associations between characteristics of establishments and certain error types. Human errors are more likely to be found in smaller, single-unit establishments, those with a single respondent for both programs, and in the self-reporting mode of TDE. Reference-period problems are more frequent in smaller establishments using an internally developed payroll system. Employee-type problems are associated with medium and large size classes and in establishments that outsource their payroll. Looking at individual programs, the CES is more likely to have problems with employee-type reporting, human error, and reference-period problems (in that order). In contrast, QCEW has the largest problem with reference periods and (secondly) employee-type reporting.

There is likely much more to be found in these data by looking at more detailed cross tabulations, and perhaps through use of multivariate analysis, although we plan to do the latter with a larger sample size in the future. In addition, we may want to look at the reasons for the differences by the sheer number of difference groups an establishment fell into. Establishments in two or three of the difference groups may be different than other establishments by virtue of simply having more problematic differences. Finally, the magnitude of the differences may be important and should be evaluated, but again, we plan to do this in the future with a larger sample size.

This RAS has been a "pilot" for a larger RAS focusing on seasonal differences between the CES and QCEW. Our work on the current RAS was used to inform the sample design of the larger RAS and its instrument and data-collection procedures. In June 2008, we completed data collection for a sample of 3,000 establishments. In addition to a larger sample size, the new study covers CES and QCEW data from January 2006 to March 2007, providing us with a longer time period over which to identify seasonal patterns and establishments with differences that most strongly fit the patterns. We studied the difference patterns extensively over this time frame, and establishments were selected into the sample frame based on a number of patterns: (1) a different over-the-month change from December 2006 to January 2007 in the CES and QCEW; (2) a different over-the-quarter change for the fourth quarter in QCEW relative to CES; (3) a larger over-the-year growth in QCEW relative to CES; or a larger buildup in QCEW over the final three quarters of the year, followed by a larger drop in QCEW in the first quarter of the following year, all relative to CES; and (4) constant employment within or across quarters, in only one of the QCEW and CES, or a "stair-step" phenomenon shown in the QCEW. In addition, a control group with none of the patterns was selected for comparison purposes. The RAS instrument was extensively revised to provide more of the detailed data we need to understand the differences. For example, there is now a battery of questions asked of both the CES and QCEW respondents on exact employee types that they include and exclude, so we can pinpoint employee-reporting problems. Also, we asked respondents if they could provide corrected figures, so we might be able to better evaluate measurement error. Data collection is difficult for this type of study, due to the long recall period and multiple-establishment respondents. To improve data collection, we centralized data collection in one center, provided extensive training, monitored interviews, and worked with the interviewers. We are optimistic that the new study will provide useful information for both the CES and QCEW programs about reporting differences and possible solutions to reduce them.

Acknowledgments

References

Eickman, Michelle (2007). "March 2007 Benchmarks for the Nonfarm Payroll Survey." http://www.bls.gov/ces/cesbmart06.pdf
Goldenberg, Karen, Shail Butani, and Polly Phipps (1993). "Response Analysis Surveys for Assessing Response Errors in Establishment Surveys." http://www.bls.gov/ore/abstract/st/st930240.htm