

APPENDIX II

DATA SOURCES, SURVEY METHODOLOGY & ERROR ANALYSIS

ACKNOWLEDGMENTS

The project manager for this report is statistician Christine E. Steele (telephone: 202-606-1817 and email owi@opm.gov). The survey and publication were prepared under the direction of Acting Assistant Director Rhonda K. Diaz. We gratefully acknowledge the contributions of Zoraida V. Arledge, Randall T. Matke, and Matthew S. Walters.

INTRODUCTION

Associated with every statistical survey are different types of errors affecting each stage of the survey, from initial data collection to the printing of the publication. This Appendix covers areas of potential errors arising from various survey operations. When possible, we estimated the size of the error. When no information was available about the sources and potential impact of error, we noted it. We did not attempt to determine how the errors interact. The user of this publication should consider the various data limitations discussed below.

DATA SOURCES

Most of the data in this publication came from the Central Personnel Data File; Federal agencies reported the additional data. These two data sources are described briefly below.

Central Personnel Data File. The Central Personnel Data File (CPDF), started in 1972 and maintained by the U.S. Office of Personnel

Management (OPM), is an automated system of individual records for most Federal civilian employees. It is updated on a quarterly basis from data submitted by agency input of employment status files. We edit all input files for validity, and insert functional blanks into data fields with errors. In December 2000, CPDF contained 1,737,221 active records on non-Postal Federal civilian employees. The National Science Foundation had a problem reporting pay status so their 415 inactives were added to the survey.

Non-CPDF Data Collection. Non-CPDF data were collected through three procedures. First, the U.S. Postal Service (USPS) submitted individual employee records on magnetic tape; we converted these data to the CPDF record format. Second, most of the other non-CPDF agencies and the CPDF agencies with foreign nationals employed overseas submitted summarized data on OPM Form 1312 with geographic location, pay system category, work schedule (if located in United States), and U.S. citizenship (if located overseas). Third, we generated survey data for the Congress, Congressional Budget Office, General Accounting Office, U.S. Court of Appeals for Veterans Claims, Office of Compliance, Commission on Security and Cooperation in Europe, Judicial Branch, White House Office, Office of the Vice President, Office of Policy Development, Board of Governors of the Federal Reserve System, and Postal Rate Commission, mostly from their December 2000 Monthly Report of Federal Civilian Employment (SF 113-A).

DATA CORRECTIONS

Automated Edits. Records for all active employees were extracted from the CPDF file excluding seasonal employees (other than the

13,808 teachers and staff with the Department of Defense Education Activity which were included). Special extracts were prepared from CPDF to reflect the seasonal workforces (see pages 12-14).

Special processing was done for the CPDF data by matching CPDF records with unspecified geographic location against the CPDF error files (from CPDF edits) to substitute back in the original geographic location and U.S. citizenship code. These geographic locations were rejected again by the subsequent geographic survey edits but we had data which could be corrected (cannot correct unspecified data).

After the above special processing, CPDF records along with all the other data submitted for the survey were edited by an automated procedure to meet criteria specific to this report. The edits identify and reject all employee records with unspecified or invalid data for geographic location, U.S. citizenship (for overseas locations), work schedule (for USA locations), pay plan, agency, and agency subelement (for Defense Department in the main survey and for other agencies in the special analysis of the seasonal workforce).

Corrections. We analyzed the error listings and made 13,541 corrections to increase the coverage of the report and insure the most accurate representation of the status of agencies and employment.

The unspecified State code 99 was assigned for 10 unspecified or erroneous geographic locations. U.S. Territory codes were corrected for 13 while 628 erroneous foreign country codes in agency reports were corrected. County codes were updated for 10,176 (mostly U.S. Postal Service). City errors were corrected for 175 with name available in agency reports; the other 2,502 city or place code errors (mostly U.S. Postal Service) were changed to unspecified.

No corrections were needed for agency, Defense subelement, and U.S. citizenship codes. The 29 CPDF records with unspecified pay plan were assigned to "other" pay plans. Since most employees work full-time, unspecified work schedule was changed to full-time for 8 employee records.

BENCHMARKS

Table I below shows the distribution of records used for the December 2000 survey (after making all data corrections) compared to benchmarks (BM).

Table I: Geographic Survey Benchmarks

Source	Survey	Benchmark	% of BM
CPDF	1,709,504	1,740,251	98.23
USPS	956,574	873,424	109.52
Agencies	73,920	74,054	99.82
Noncitizens			
Overseas	25,894	26,124	99.12
Total	2,765,892	2,713,853	101.92

This survey achieved 101.92 percent coverage worldwide when benchmarked against December 2000 data collected on the Monthly Report of Federal Civilian Employment (Standard Form 113-A) from all departments and agencies of the Federal Government except the Central Intelligence Agency, National Security Agency, Defense Intelligence Agency, and National Imagery and Mapping Agency. Data in this survey were benchmarked against the December 2000 113-A data adjusted to geographic survey coverage by including all intermittent employees and excluding Commissioned Corps.

Some agency survey totals were more than 5 percent different from their benchmark totals. Most were small agencies with small actual differences which yielded percentage differences greater than 5 percent. Treasury, Agriculture, Interior, Commerce, and Small Business

Administration survey totals were more than 5 percent lower than their bench-mark totals because the survey excludes their seasonal employees. U.S. Postal Service data includes inactives.

After completion of all the editing and data corrections, we merged the accepted CPDF and non-CPDF data and improved place names for 191 records in four locations. The statistical table outputs were then generated. Checks were made for consistency among the reports for the current survey. The 2000 data were also compared with the 1998 survey data. Differences in category totals more than 5 percent above or below the overall change were investigated to determine the cause of 1998 to 2000 changes.

SOURCES OF ERROR

Data Collection. Each quarter, all Federal agencies participating in the CPDF system update their agency files to document their employees' status. Timely file submission by the agencies may be prevented by the loss of automated input in transit to our agency or by system problems within the submitting agency.

To measure the completeness of agency employment reflected by the December 2000 Central Personnel Data File, the number of employee records by agency extracted for the survey was compared with the employment counts by agency (including all intermittents and excluding Commissioned Corps and foreign nationals employed overseas) from an independent reporting system, the Monthly Report of Federal Civilian Employment (SF 113-A system). Comparing the 1,709,504 CPDF records available for the survey against the 1,740,251 adjusted 113-A data for the same agencies yielded a 1.8 percent difference. This difference is due to the 27,716 seasonals not in the main survey, definitional differences (such as effective dates) between the CPDF and 113-A populations, and miscellaneous reporting errors

and processing errors in both the CPDF and 113-A data collection systems.

Survey Processing Errors. All known errors in the edit programs, edit lookup tables, and report programs were corrected.

Excluded Records. The main survey excluded these employee records: 27,716 seasonal employees as of December 31, 2000 (other than the Department of Defense Education Activity).

Data Quality. A September 1994 CPDF Accuracy Survey estimated the percentage of valid (passes CPDF system edits) but inaccurate (differs from the Official Personnel Folder record) data contained in the CPDF. The following Table II shows Governmentwide estimates for the key data elements included in the Geographic Survey.

Table II: Data Element Error Rates
(September 1994 CPDF Accuracy Survey)

Data Element	Percentage Error
Agency	0.7
Pay Plan	0.7
Work Schedule	0.7
Location	0.0
Pay Status	0.0
U.S. Citizenship	0.0

The results of the most recent CPDF accuracy evaluation show the percentage of valid but inaccurate data is about 1 percent for agency (all errors in agency subelement), pay plan (used for assigning pay system category), and work schedule and zero percent for geographic location, pay status (used for extracting CPDF employees in pay status), and U.S. citizenship.

The results of this evaluation cannot be directly applied to this Geographic Survey data because of timeframe and definitional differences in the survey populations; however, these presented results give an overall indication of the quality of the data in this publication. The quality of the non-CPDF data has not been measured.

Data Definition Changes. Appendix I covers the changes in definitions and coverage of agency, geographic location, and pay system since the previous survey; these changes can affect 1998 to 2000 data comparisons.

Redefinitions of a data element are a source of error that can be difficult to measure. New work schedule codes for seasonal employees were effective October 1981 and new work schedule codes for on-call employees were effective November 1985. Agencies had to identify these employees and submit work schedule code corrections for them. Furthermore, the agencies have to process personnel actions at the beginning and end of each work season to activate and de-activate seasonal employees; failure to do so is another potential source of error.