

Joyce A. Stevens, U.S. Office of Personnel Management

## INTRODUCTION

The Central Personnel Data File (CPDF)/Official Personnel Folder (OPF) Accuracy Survey is an essential component of a comprehensive quality assurance program for the Federal Workforce Information System. The purpose of this program is to determine the accuracy and maintain a desired level of reliability of the data contained within this information system. The CPDF is the major computer file within this system and major information source for policy development and analysis by the Executive Branch and Congress. Persons outside the Federal government also rely upon this data as a source of expertise and authority.

This special study is designed to determine with statistical reliability the accuracy of the data contained in this computer file by comparing CPDF data against source documents from Official Personnel Folders and verifying the information through employee interviews. This paper will review the survey procedures used in this special study.

## TWO-PART SURVEY DESIGN

Part I of this Survey consists of 445 mail-out questionnaires to be completed by personnel specialists in the 36 agencies selected for this study. Information was extracted from the CPDF and preprinted on the mail-out questionnaires. Each personnel specialist was instructed <sup>1/</sup> to compare the preprinted CPDF data against identical information in the sample employee's Official Personnel Folder or its automated equivalent. The mechanics of this verification technique (CPDF Audit), as well as a description of the CPDF and of an OPF will be given later.

Part II of this survey was designed to be the control group for Part I and to pretest two techniques in addition to the verification technique used in Part I. A subsample of 33 Federal employees was chosen for these tasks. Office of Personnel Management evaluators were instructed <sup>2/</sup> to administer the questionnaires. The mechanics of the two additional techniques and how the sample size and subsample size were derived will be explained below.

## CPDF

The Central Personnel Data File is maintained by the Office of Personnel Management. It is an automated data base of personnel information on most of the active Federal civilian work force, approximately 2.8 million persons. The CPDF is used for annual and ad hoc surveys, and for determining important issues such as minority representation in the Federal work force.

The information contained in this file is taken from various input documents, such as, Notification of Personnel Action Forms and Payroll Slip Change Forms, and from minority group designation input, quarterly training input, and handicap self-identification input.

## OPF

An Official Personnel Folder is maintained on each active and inactive Federal employee. Each OPF contains the documentation relative to the individual's employment with the Federal government. This documentation includes forms such as the Personal Qualifications Statement, Notification of Personnel Action, Payroll Slip Changes, certificates for completed training courses, accident reports, and employee evaluations. Minority group designation and handicap designation, however, are kept in separately maintained files in accordance with the Privacy Act.

## WHY AN ACCURACY SURVEY?

Most data on the Federal work force is based on the CPDF: therefore, the accuracy of the data is paramount. There are many areas where errors can be introduced into the data. This study is an attempt to examine the central ones including errors made when the employee first supplies information to the employing agency, when this information is recorded and maintained in the OPF or its automated equivalent, and when the information is entered into the CPDF.

The 30 data elements which this study was designed to determine the accuracy of include the employee's:

1. name
2. social security number
3. birth date
4. sex
5. U.S. citizenship
6. minority group designation
7. handicap designation
8. location of employment
9. grade
10. step
11. salary
12. pay plan
13. pay basis
14. special pay rate
15. work schedule
16. service computation date
17. retirement designation
18. tenure
19. veteran's preference
20. supervisory designation
21. excepted/competitive service designation
22. occupation code

23. scientist and engineer classification
24. special program code
25. education level
26. date of degree
27. academic major
28. Federal Employees' Group Life Insurance Status
29. agency code
30. submitting office number

#### SAMPLE SELECTION

There were several constraints which had to be taken into consideration before deciding upon the best method of sample selection. These can be classified as administrative reasons and as statistical considerations. The administrative reasons include:

- 1) Inadequate time for a pretest.
- 2) Minimizing the cost of the survey.
- 3) Minimizing the time needed to conduct the survey and to get useable results.
- 4) Minimizing the reporting burden on the participating agencies.

The statistical considerations include:

- 1) Minimizing bias that could be introduced when selecting agencies. (A representative distribution of all agencies was required.)
- 2) Keeping the design simple. (Since this was a first look at the error rates by data element it was desirable to keep the design simple. Further studies could then be developed using a stratification technique to probe at the source of the errors.)

Taking these constraints into consideration a systematic random sample was chosen as the method of sample selection.

Since the CPDF is sorted on agency code the grouping of agency records lent itself best to a systematic selection of records. This way the larger agencies were more likely to have more records selected than were the smaller agencies (a self-weighting factor) and each record had an equal chance of being selected. Also, not all big agencies or all small agencies would make up the sample (as could happen with a simple random sample), rather a representative distribution would occur.

A stratified random sample would have resulted in a much larger sample size, and a pretest would have been required for selecting the best stratification criteria.

#### SAMPLE SIZE

[Part I]

A sample size of 385 for the first part of the survey was calculated using the following formula:

$$n = p q \left[ \frac{1.96}{E} \right]^2 \quad \frac{3}{/}$$

where: n = sample size  
 p = percent of expected error  
 q = (1-p)  
 E = sample error deviation

Since p was unknown a conservative approach was taken. The value that maximizes pq and thereby maximizes the sample size is p = .5. The sample error deviation was set at ± 5 percent. Fitting that into the formula:

$$n = (.5)(.5) \left[ \frac{1.96}{.05} \right]^2 = 384.16$$

$$n = 385$$

However, it was known ahead of time that the address list that would be used to mail out the questionnaires was not up-to-date and an unknown number of nonmatches could be attributed to it. Also, it was necessary to match two files to the CPDF in order to be able to have access to the employee's name and to the address of the office that submits the personnel data for that employee. An additional unknown number of nonmatches could be anticipated for this. Therefore, 65 records were added to ensure obtaining a sample of 385.

[Part II]

For the second part of the survey 33 records were selected. This size was arrived at by matching the location of employment on each sample record against the locations where Office of Personnel Management on-site evaluations were already taking place. These on-site evaluations were in no way connected with the CPDF/OPF Accuracy Survey. In addition the method by which installations were originally selected for on-site evaluations was totally unrelated to this accuracy survey. Even though the 33 records were not a random selection from the 445 records, this subsample should prove to be useful as a pretest to provide information on problems encountered and on the time required to administer the three techniques.

#### MECHANICS OF SURVEY

[CPDF AUDIT]

The questionnaires<sup>4</sup> used for Part I and Part II of this study were preprinted with actual data from the September 1978 CPDF. (See Attachments

1 and 2.) A personnel specialist (or evaluator) compared the preprinted information to identical data maintained in the employee's OPF or its automated equivalent. The comparison was made using Notification of Personnel Action Forms dated as of or prior to September 30, 1978. However, the minority group designation had to be verified with the Equal Employment Officer, and the handicap data was to be compared with the separately maintained file of Self-Identification of Disability Forms. A single match code was to be selected from eight possible match codes. A match code of 1 represented a match while the remaining codes indicated specific types of nonmatches. This extensive breakdown of non-matches will enable a thorough analysis of the error rate. In addition, an analysis of the nonresponse (to each item and to the survey) will be done.

[OPF Audit]

While the CPDF Audit locates the errors produced when the information is recorded in the CPDF, the OPF Audit measures the error in the information that is maintained in the OPF. This part of the survey verifies the information found on the Notification of Action Forms dated as of or prior to September 30, 1978, against all other supporting documents in the OPF. The minority group designation and the handicap code were not verified for this part of the survey since they are not maintained in the OPF.

[Employee Interview]

The Employee Interview captures the errors made in the information supplied to the employing agency by asking the employee to verify those data elements she/he supplied when first employed. These data elements include 13 of the original 30. (See Attachment 2.)

REASON FOR THE SECOND PART OF THE SURVEY

The second part of the survey was to be used as a control group for the CPDF Audit, and as a pilot test for two additional procedures, the OPF Audit and the Employee Interview.

A control group of OPM evaluators was used because there was the possibility that agency personnel specialists might introduce personal bias into the answers to the questions because of:

1) their employment by the agency, 2) the nature of their duties, and 3) possible knowledge of the employee.

Should the results of the control group compare favorably to the results obtained by the agency personnel specialists, further studies would use the latter group. This would result in a reduction of the survey cost. (Postage for mail-out questionnaires is less expensive than travel expenses incurred by the evaluators.)

SUMMARY

When the results from the first part of the survey are analyzed the Office of Personnel Management will be able to say with 95% confidence that the sample error rate of each data element is within  $\pm 5\%$  of the true error rate in the CPDF on a government - wide basis. OPM will know what the error rate is for each data element for a given point in time. From these findings quality control techniques for monitoring the error levels can be developed.

Other expectations from the first part of the survey include determining:

- (1) if large agencies have errors primarily in the same data elements,
- (2) if error rates by data element differ significantly among large agencies,
- (3) the relationship of error rates between data elements, and
- (4) what type of stratification to use for future surveys.

The second part of the survey will indicate if there is a large difference between using agency personnel specialists and Office of Personnel Management evaluators, thus impacting how much OPM can rely on the results of the first part of the survey. Also, it will describe the time and cost involved and the problems encountered when doing the OPF Audit and Employee Interview. It can also be used as input to a cost/benefit analysis of sending evaluators to the locations selected by the sample rather than just using those who are already at an installation.

FOOTNOTES

1. Part I of the survey was authorized by Federal Personnel Bulletin 298-7, March 8, 1979.
2. Part II of the survey was authorized by Office of Personnel Management Operations Letter 273-892, March 8, 1979.
3. A finite population correction factor was not needed. The population of all CPDF records (less Post Office) was 2.1 million.
4. The questionnaires were computer-generated in order to reduce the cost and time needed to produce them.

