MANAGEMENT SCIENCE
Vol. 17, No. 6, February, 1971
Printed in U.S.A.

WORK SAMPLING IN FINANCIAL MANAGEMENT—COST DETERMINATION IN POST OFFICE DEPARTMENT*

B. J. MANDEL†

Post Office Department

Work Sampling has been described by some authorities as the best and most objective method of measuring the time and cost of performing the various functions of complex organizations for use in management, budgeting and setting performance standards. Yet its use for these purposes has been quite limited. This paper is designed to encourage further use of Work Sampling. It describes what is probably the largest continuing application of Work Sampling in the world, namely, that made by the Post Office Department. About 4.5 million observations are made during a year at randomly selected instants of time to estimate the cost of the in-office activities and over 200,000 employees participate each year.

The paper describes the basic objectives served by the application, the size and composition of the universe of interest, the methods of sampling, the sample size used, the estimation procedure, the methods of measuring sampling variation and controlling the quality of the information. In addition, a few areas of research to improve the sample design are described.

The Post Office Department needs to know the cost it incurs to handle and deliver the mail in each class (e.g., first class and air mail) in order to compare it with the corresponding postage or other revenue earned by class. This information is needed primarily to recommend postal rate changes to Congress. While the total cost of postal activities currently amounts to about $7 billion a year, a first step taken in estimating the cost of each class by probability sampling deals with the in-office costs. This cost currently amounts to about $4 billion a year. Later probability sampling applications will be extended to cover out-of-office and other postal costs.

A sophisticated three-stage sample design was adopted. First, a stratified random sample of 500 post offices was selected. Stratification was based on size of office. Then, a digital one percent sample of employees in these offices was selected to participate each week. Finally, the work which the selected employees are performing is observed at four randomly chosen instants of time during their 8-hour work day. Thus, each employee is generally work-sampled at 20 instances a week and a total of about 4.5 million readings are made annually.

A quality control type of edit and control system has been established to detect and correct field reporting errors, thus helping to obtain high quality of results.

Purpose of Paper

Work Sampling has been described by some authorities[1] as the best and most objective method of measuring the time and cost of performing the various functions of complex organizations for use in budgeting and setting performance standards. Yet its use for these purposes is quite limited. To encourage further use of this method in management problems, this paper describes what is probably the largest application of Work Sampling in the world. The application is designed primarily to estimate the in-office cost of performing the large variety of mailhandling functions of the Post Office Department, utilizing a random sample of post offices throughout the country. It describes the basic objectives served by the Work Sampling application in the

* Received October 1969; revised May 1970.

† The opinions expressed here are those of the author and not necessarily those of the Department.

[1] "Management Science Applications in the Bell System," by W. W. Hardgrave and W. H. MacWilliams in *Management Science*, Vol. 15, No. 8, (April 1969) pp. B-387-396.

Department as well as the methodology involved in: Sampling the universes of post offices, employees, and worktime; determining sample size; making estimates from the sample data; measuring sampling error of the estimates and controlling clerical errors at the source. The paper also lists the advantages of using Work Sampling for cost and time estimation and suggests some further research to improve the technique as it is applied to postal functions.

Objectives

In compliance with an act of Congress (February 28, 1925), the Post Office Department determines regularly the cost incurred and the revenue earned in providing postal services for each class of mail (e.g., air mail, first-class mail, etc.) and type of special service (e.g., insured mail, registered mail, etc.). The resulting cost-revenue relationships for each class of mail and service are analyzed each year and used in making recommendations to Congress and the Interstate Commerce Commission for changes in postal rates, where appropriate. Since cost determination is an integral part of the statutory requirement of the Department, the major objective of the Work Sampling application therefore is to inform management of the cost of handling the different classes of mail and of providing special postal services.

While the major reason for the cost determination is to help management meet the requirement of the law, other purposes may also be served by the results. For example, by using the estimates of time required to perform various mail processing functions in conjunction with independently derived data on the number of work units completed, it may be possible to develop performance norms. Also, by knowing the estimated amount of time being devoted to the various projects or activities, management may, through analysis, plan budget allocations or act to reduce costs on excessive time-consumers. Thus, a question such as: How much time is being devoted to the handling of flat mail (or parcel post) in each of the major post offices can be answered accurately by Work Sampling. Large differences in the unit time among the organizational units can point out areas for management evaluation.

In Fiscal Year 1969, the total cost of operating the postal establishment amounted to nearly $7 billion. The Work Sampling system described in this paper covers activities performed inside the post offices and these activities involved about 60 percent of all costs; therefore, the system covered in-office wage and salary costs of about $4 billion during 1969, paid to some 900,000 employees who worked for the Department at some time during the year.[2]

The Old Method of Costing

Several methods are available for determining the cost incurred for each type of work. The method used by the Department before Work Sampling was adopted in July 1969 was generally "slip reporting"; it was a combination of time reported on a slip by each employee during one week in each quarter and special "mixed mail tests". By this method, all employees in the participating post offices were required to fill out a slip each day and show the time spent on major classes of mail and services during the entire Cost Ascertainment Week of each quarter. By combining these time-sheet figures with the special mail counts, the total time (later converted to Wage and Salary costs) was distributed to detailed classes of mail and service worked on during the

[2] The average number of employees per pay period was approximately 730,000, but with employee turnover the number employed during the year is substantially larger.

week. The results of the four weeks of slip reporting during the year were then combined and expanded to represent the entire year.

No doubt, this conventional method of cost determination by slip reporting and judgment sampling has served the Department well over a long period of time in meeting its objectives. The slip reporting method is used by many agencies throughout the country to determine costs. Yet, it has long outplayed its usefulness, having been outmoded by the modern technique of random Work Sampling. With Work Sampling, recordings are made at randomly selected moments of time throughout the year (or period of study) of the specific work activities actually being performed by the individual employees. The random sample results are then expanded to represent the total period of study. Work Sampling is bound to yield more objective and accurate estimates because it observes and records what is actually being worked on at the observation moments and does not have to rely on memory, as does slip reporting. When properly applied, it has been proven to be a superior method. That it requires less time and effort to develop has similarly been proven. That it is a more interesting and challenging cost determination approach needs to be proven only by actually undertaking a Work Sampling application and seeing the accuracy of its results and its underlying power or by asking for the reaction of employees who have applied it.

The New Method—Work Sampling

Work Sampling harnesses the power inherent in probability sampling to make estimates of the time spent on, and costs incurred in, performing the various functions of an organization. It does this by making observations of these types of activities at randomly determined points of time and, using appropriate mathematical estimation procedures, expands the number of these sample observations to the universe of total time or total cost.

The Universe of Interest

In the Department's Work Sampling application, the universe of interest consists of all the in-office paid worktime and salary costs of all postal employees throughout the country each year except custodial and motor vehicle employees. Expressed differently, the universe of interest consists of the $4 billion in payroll costs of the Department for in-office activities performed by the 900,000 employees who processed the mail and performed the special postal services in its 33,000 post offices at some time during Fiscal Year 1969 (and, of course, subsequent years). The primary objective is to determine how much of this money is spent for handling each of the 60 to 70 different categories of mail and services provided by the Department each year.

The universe may also be considered to consist of all *post offices* geographically dispersed throughout the United States, of all the specified postal *employees* in these offices and of all the paid *worktime* they spend. The sampling procedures, therefore, have been applied first of all to select a random sample of post offices, then a random sample of the employees in these offices, and finally a random sample of instants of time. During these instants, observations and recordings are made of the type of work being done and class of mail and service being processed by each participating employee.

Methods of Selecting the Sample

Since the universe of interest is subdivided into three components—post offices, employees and worktime—the sampling procedure follows a three-stage design, in that order.

Sampling Post Offices

Because a very large proportion of the employees are employed by only a few post offices and these offices pay a very large proportion of the total salaries (e.g. about 50 % of the employees and payrolls are in 0.2 of 1 % of the 33,000 post offices), use of a stratified random sample of post offices was a must, to assure an efficient sample design. Thus, all post offices with annual postage revenue of $8,200,000 (which is highly correlated with payroll size) or more and declining proportions of the smaller post offices were included in the sample. Altogether, 502 first- and second-class post offices or slightly over four percent of the 12,000 such offices were included in the sample as of July 1969. These offices, however, employed about 60 percent of all postal employees. For practical reasons, third- and fourth-class post offices which employ about two percent of the employees were not covered by the Work Sampling technique, but continue to use the old method.

Sampling Postal Employees

Because each employee generally represents a 40-hour week, i.e., each is generally paid for an equal amount of worktime, employees were generally sampled at the same rates. With minor exceptions, one percent of the employees in the sample of post offices were included in the sample each week. The exceptions to this proportional sampling at the second stage were in the smaller post offices, where employees were sampled at a progressively higher rate as the post office size declined.

Several alternative ways of selecting the sample of employees were considered, such as simple random or systematic random selection. However, because each employee is identified by a social security number (a nine-digit number issued serially over the past 30 to 35 years by the Social Security Administration to each employee whenever he applied for it), this number was used to select the weekly sample of employees. In other words, a digital sampling procedure (equivalent to systematic selection) was used to select the sample of employees. Thus, for example, a one-percent sample was obtained by selecting all employees whose social security account number ended in the digits 01, since these two digits represent a single two-digit combination of the 100 possible two-digit combinations (e.g., 00, 01, 02, . . . 99).

The digital sampling procedure not only provided a quick and easy way of selecting the sample of employees, but it also offered some other advantages. By controlling on the specific two-digit social security number endings previously selected, it was possible to select a different sample of employees each week. Thus, if employees with two-digit endings of 01 were selected the first week, and employees with two-digit endings of 02 were selected the second week, these two groups of employees would not be reselected for another 98 and 99 weeks, respectively. Yet, because of the method of serially issuing social security numbers, each group would represent a random sample of all employees. By use of *different* two-digit endings each week, there was definite assurance of avoiding more than one week of participation a year for each employee. This controlled selection actually gave assurance that all employees (with some exceptions in the case of employee turnover) would be in the sample about once in two years.

There were two other advantages of the digital sampling procedure, aside from the fact that it provided a ready-made random selection procedure. First, it offered a practical and easy way of checking for errors in the sample selection (e.g., when the two-digit endings of 01 were sampled, it was easy to check that no other two-digit endings were erroneously selected in that sample). Secondly, newly recruited employees

were automatically represented by the system, since the social security number issuance system automatically provided for a like proportion of each two-digit ending (as well as other digit endings) because the numbers were issued in sequence by the Social Security Administration.

Sampling Time

An instant (or one second) may be considered as the possible sampling unit for the universe of worktime. The sampling procedure, therefore, was applied to all the seconds within the 24-hour workday (since the post offices generally operate on a three-tour basis, each consisting of eight hours). The time periods for making random observations were selected by a systematic random sampling procedure applicable uniformly to all employees in the sample, to assure that each eight-hour tour is represented. Four instants of observation, two hours apart, during each tour were selected each day by a systematic procedure, with a random start in the first two-hour period. Thus, at a specified time in each two-hour period, a work sampling observation was made on the employees in the sample.

In summary, the entire three-stage sampling method may be classified as a *stratified digital* (systematic) *random-cluster sample*, or in short, a stratified cluster sample.

Size of Sample

The size of sample used attempted to balance the cost of collecting the data with the accuracy sought by the Department in the most critical estimates. Also considered was the concept of a sampling unit, since sample size had to be measured in terms of the number of sampling units. Because there were three stages of sampling, three different sampling units were involved: Post offices, postal employees, and instants of time. These units were described above.

First Stage—Post Offices. The number of post offices to include was determined by a criterion called "span of control". Since the 15 Regional Controllers were made responsible for administering the Work Sampling System in their respective regions and this responsibility included regular visits to and contacts with the participating post offices, the number of post offices to be included in each region had to be reasonably close to the number which could be controlled, i.e., within the span of control. The 502 selected post offices resulted in an average of 33 to 34 offices per region. However, the actual number varied from a low of 14 offices in the Denver Region to a high of 62 in the Philadelphia Region.

The size of sample in the first stage of sampling was also influenced by "administrative convenience"; that is, the 502 offices were the same which were included in another major data-collection sampling system of the Department—the New System for Estimating Revenue by Class of Mail and Service. Thus, regional costs for administering the Work Sampling System were shared with the Revenue Estimation System. (The cost of several other systems is now also being shared by this method.)

Finally, the 502 post offices included in the sample were in a large way selected by the stratified sampling policy, which required the inclusion with certainty of the largest post offices, i.e., those with annual revenue of at least $8.2 million. Under this policy, the 81 post offices with this amount of revenue or more were included with certainty. The remaining 421 offices were selected from the remaining 12,000 offices by size-of-office strata, in which the sampling fraction was made smaller in direct proportion to the amount of postal revenue represented by the stratum. The first-stage sample size

is given in the following table:

Post Office Size[a] (Annual Postal Revenue) (in Millions)	Number in Universe (as of 7/1/69)	Number in Sample	First Stage Sampling Fraction (Percent)
$24.581 or more	29	29	100.0
8.194 to 24.581	52	52	100.0
1.639 to 8.194	312	176	56.4
0.819 to 1.639	360	57	15.8
0.328 to .819	818	66	8.1
0.148 to .328	1,318	45	3.4
0.066 to .148	2,093	38	1.8
0.030 to .066	3,031	25	.8
0.013 to .030	4,307	14	.3
TOTAL	12,320	502	4.1

Upon recapitulation, it was found that these 502 offices included over 60 percent of the total postal revenue as well as over 60 percent of the employees.

Second-Stage—Employees. The decision regarding the number of employees who should participate in the system in the 502 post offices was geared to a "manageable workload" and to the "level of reliability" required of key cost estimates, such as the annual in-office cost of first-class mail. The criteria usually considered in determining the size of sample to use are reliability and confidence in the estimate required by its users, balanced by the reasonableness of the workload and cost of collecting the data. In this case, we considered the latter two criteria first and then evaluated the reliability and confidence to be expected from the manageable size of sample.

Thus, we first decided that supervisors would hardly feel the impact on their regular duties of the extra workload of making four random observations a day on a sample of their employees. We then decided that the *workload* would be even less telling if the supervisors' readings were to be made on only one out of each 100 employees (one percent). This combination of "four daily observations and one percent of the employees" yielded a total of about four million observations a year in the 502 participating post offices.

Independently, we then determined the reliability which would be associated with the annual in-office cost of first-class mail estimated from a sample of 4 million random observations a year, assuming that management required 99.7 percent confidence. Using a standard formula for computing sampling variation (or sampling error) of a percentage derived from a simple random sample, we determined that the relative error of the estimated annual in-office cost of first-class mail would be well within one percent with 99.7 percent confidence. The following demonstrates how this relative error was determined by use of a simple random sampling formula adjusted for loss of precision due to the use of a three-stage cluster sample: Relative error of a percentage, $E = t(q/np)^{1/2}$

where n = total number of observations

t = confidence coefficient (e.g., $t = 3$ for 99.7 % confidence)

p = estimated percentage of total cost represented by the in-office costs for a given class of mail (e.g., for 1st class, $p = .4$)

[a] These odd breaking points resulted from growth in size as measured by postal revenue, which in the initial stratification period (1965) were in such even units as $15 million or more, $5 million to $14.999 million, etc.

q = 100 percent $-$ p

E = relative sampling error of the estimated percentage (p)

when $t = 3, p = .40, q = .60$, and $n = 4$ million, we obtain $E = \{3[.6/(4\text{mill.})(.4)]\}^{1/2} = .00183$. However, by assuming a three-fold loss of efficiency because we use a multistage cluster sample, we arrive at $E = .0055$ or .55 of 1%. The formula of approximation assumed the use of simple random sampling with a loss of 300 percent efficiency. Efficiency loss is expected from the use of a three-stage cluster sample of post offices, employees and time, rather than a simple-random sample.

To make this total number of observations, it was decided to include a sample of 4,000 employees each week with 20 observations a week on each employee (i.e., four observations each day times five days). Since there were about 400,000 full-time equivalent employees in the 502 participating post offices, this required selecting a one-percent sample each week. By making four random observations each day (or at least 20 a week), the aimed-at sample size of 4,000,000 observations was achieved as follows: 4,000 employees times 52 weeks times 20 observations a week = 4,160,000 observations.

In the second stage of sampling, therefore, we decided to use a sample of about 200,000 employees, but required that each employee participate no more often than one week a year. All of the 400,000 full-time equivalent employees of the participating post offices (with the exception of some employee turnover) were to be included in the system at some time during a two-year period.

Third-Stage—Number of Observations Per Employee. As indicated in the earlier section, the decision was to make four random observations a day on each employee who was selected for the sample. This decision was made to minimize the burden on the local post office supervisors of making the observations and recording the type of work and class of mail being handled at the randomly selected instants of time. For example, it was theoretically possible to make as many as 48 (or even more) readings a day on each employee. To do this would have required burdening the supervisor with about six readings each hour or one every 10 minutes. This frequency of observations would have seriously interfered with the supervisor's performance of his normal duties of supervision and mailhandling. In consultation with a number of supervisors, it was decided that four observations a day (or one each two-hour period) would hardly represent a burden on the supervisor. Therefore, readings were scheduled for each two-hour period during a randomly chosen number of minutes within the two-hour period. The hours and the random minutes within each hour, although the same for all employees in the sample, were different each day, to achieve complete randomness.

In summary, therefore, a sample size was determined for each of the three stages of sampling by a combination of criteria including the following: Span of control, administrative convenience, reliability, and confidence in critical cost estimates and administrative burden on supervisors.

Estimation Procedures

As a rule, Work Sampling techniques yield a count on the number of observations made of each kind of work or activity. Supported by the law of probability, the frequency of observations made for a work-category, taken as a percent of all observations, reflects the proportion of the total amount of time which was devoted to that work-category. An estimate of the number of hours spent on the activity is usually

derived by multiplying the percentage of all observations (or tallies) made on this activity by the total number of hours spent on all activities combined. From the estimate of workhours thus derived, it is then possible to estimate the corresponding cost, i.e., by multiplying the estimated number of workhours by the average cost per hour for each activity.

In the Department's Work Sampling application, it was decided to estimate costs separately for each of the five occupational groups covered[4] in one step directly from the relative frequency of observations (or tallies) applied to total wage and salary costs, rather than by the foregoing two-stage procedure of estimating workhours and then costs. This direct estimation procedure was feasible and later proven to be valid because the average cost per hour for each occupational group of employees was fairly constant from period to period and from one class of mail to another. By using the actual payroll figures for each occupational category (available from the accounting and payroll records), it was possible to derive accurate and unbiased cost estimates in one step, namely, by multiplying the percentage of observations for a given class of mail or service for each occupational group by the total payroll of that group in the period covered. This may be considered to be a post-stratified estimation procedure. The cost of observations not associated with a class of mail or service, such as leave, were later distributed proportionately among the classes.

Essentially, this type of computation was made separately for each post office in the certainty strata (strata 1 and 2), so that the total cost estimate for a given class of mail was obtained by summing across these two strata. In the case of strata with sampling fractions of less than 100 percent, the proportion of readings for each class derived from the sampled offices was applied to the corresponding *total* payroll for each occupational group for all post offices in that stratum.

A slightly different estimation procedure was used for in-office costs of carriers, since total in-office salary costs are not available for them from the accounting records. In this case, the sample determines the proportion of time the carriers spent (on each class of mail) while in the office and on out-of-office time in total. These percentages are then applied to the total salaries paid to the carriers to derive an estimate of in-office costs by class of mail and *total* out-of-office costs (not broken down by mail classifications).

Sampling Variations

The reliability of cost estimates was recently evaluated by use of the following simplified assumptions: That the sample design was a one-stage probability sample of random observations (rather than a three-stage stratified cluster sample) and that a three-fold loss of efficiency resulted from this assumption. On this basis, we used the following standard formula to compute the relative sampling error of an estimated percentage for a given class of mail or service: $E = 3t(q/np)^{1/2}$, where p is the estimated percentage of total tallies for that class of mail or service derived from the sample, q is $100\% - p$, t is the confidence coefficient, and n is the number of tallies. The table below shows the relative standard errors of cost estimates obtained from the first quarterly tabulation of in-office costs under the Work Sampling System for major classes of mail.

[4] That is, mailhandlers, clerks, mail carriers, supervisors and special delivery messengers.

Approximate Maximum Sampling Variation of In-Office Cost Estimates Derived from Work Sampling During First Quarter of Fiscal Year 1969, by Major Class of Mail[5]

Class of Mail	Estimated Cost (000,000)	Maximum Sampling Variation as % of Estimated Cost — Confidence 68%	95%	99.7%
First Class	$474.6	0.3	0.6	0.9
Domestic Airmail	47.4	1.3	2.7	4.0
Second Class	84.8	1.0	1.9	2.9
Third Class	124.6	0.8	1.6	2.4
Fourth Class	98.7	0.9	1.8	2.7
Government	14.2	2.5	5.0	7.5
International	21.9	2.0	4.0	6.0

In the great majority of cases, the standard errors are very small relative to the cost estimates. They will be even smaller (about half of the above) when developed on an annual basis to serve management's need for annual data. These relative standard errors are, of course, computed by reliance on the above simplifying assumptions. A valid assessment of these assumptions and of the accuracy of the foregoing relative errors is being made by use of replication sampling procedures. On this basis, the total sample is split into several equal-sized random subsamples, each of the same three-stage design. For each random subsample, an independent estimate is made of the in-office cost of each class of mail and service, using precisely the same estimating procedures as for the overall sample. The variability among these independently-derived estimates will provide the basis for making a better approximation of the standard error of the estimated costs.

Nonsampling Errors

As indicated earlier, 35,000 to 40,000 group supervisors and/or their assistants are required to implement the Work Sampling procedures in the 502 participating post offices. With such a large number of persons responsible for making observations on the sample of employees at the prespecified random times and for recording the correct activity, there is ample room for incorrect recordings or other deviations from required procedure. (A filled-in copy of the recording form is shown by Appendix 1.) Hence, a systematic program of training and continuous review of completed reports to detect errors and make corrections is necessary. The need to rely on such a large number of persons to make the rounds and record the work being performed makes this Work Sampling system the largest known to date. It also makes for complex problems of controlling the quality of nonsampling errors, requiring special efforts on a continuing basis.

To deal with these problems, a comprehenisve local post office, regional and central system of reviewing the quality of the field reports has been established to detect erroneous tallies and provide training to supervisors who are most in need of it. This system relies on three quality-check and feed-back points; one at the participating post office (whereby trained staff of the Chief Accountant's Office review and edit all the reports originating in their post office and record errors by type); one at the retional level (whereby trained staff under the direction of the Regional Controller review and correct the participating post office reports and record errors and error-

[5] Maximum sampling variation for estimates covering a fiscal year are about one half those shown in the table.

APPENDIX 1—A reporting form completed by a supervisor for a specified employee.

POST OFFICE DEPARTMENT
COST ASCERTAINMENT TALLY SHEET

SEE INSTRUCTIONS ON REVERSE. USE SOFT BLACK PENCIL ONLY.

TOUR: 2

TIME (Hour and Minutes): REPORTING 0800 ENDING 1630

POST OFFICE NAME	POST OFFICE NO.	EMPLOYEE NAME	PAY LOCATION	ROSTER DES. NO.	LEVEL STEP	DATE	DAY	SOCIAL SECURITY NO.	DAY CODE
CHICAGO, ILLINOIS	161542	JOHN SMITH	072	11	4	7-1-69	TUES.	100100001	4

SCHEDULED TIME (Hour and Minutes)	A	0019	0219	0419	0619	0819	1019	1219	1419	1619	1819	2019	2219
FOR CHANGE IN ROSTER DESIGNATION	B												
CHECK IF: EARNING NIGHT DIFFERENTIAL	C												
CHECK IF: IN OVERTIME STATUS	D												
CHECK IF: AT AMF	E												
OPERATION OR TYPE OF ROUTE CODE	F					2	1		5	5			
BASIC FUNCTION & ACTIVITY CODE: OUTGOING MAIL	G	1	1	1	1	1 105	1 413	1	1	1	1	1	1
BASIC FUNCTION & ACTIVITY CODE: INCOMING MAIL	H	2	2	2	2	2	2	2	2 221	2 222	2	2	2
BASIC FUNCTION & ACTIVITY CODE: TRANSIT MAIL	I	3	3	3	3	3	3	3	3	3	3	3	3
BASIC FUNCTION & ACTIVITY CODE: LUNCH OR LEAVE	J	4	4	4	4	4	4	4 913	4	4	4	4	4
BASIC FUNCTION & ACTIVITY CODE: OTHER	K	5	5	5	5	5	5	5	5	5	5	5	5
BASIC FUNCTION & ACTIVITY CODE: SPECIAL SERVICE CODES	L	6	6	6	6	6	6 001 004	6	6	6	6	6	6
WHEN REQUIRED, DESCRIPTION OF WORK PERFORMED OR OTHER REMARKS	M												
SUPERVISOR'S NAME (Print)	N					T. Jones	T. Jones	T. Jones	F. Carr	F. Carr			
ACTUAL TIME OF READING	O					0820	1019	1219	1419	1619			

APPENDIX 2—A typical quality report sent to the regions.

POST OFFICE DEPARTMENT
NEW SYSTEM FOR ESTIMATING IN-OFFICE COSTS

ERRORS FOUND BY HEADQUARTERS QUALITY CONTROL OPERATION, BY REGION AND TYPE OF ERROR

PERIOD COVERED
APRIL 5 - JUNE 27
~~POSTAL QUARTER IV FY 1969~~

REGION	ENTRIES MISSING FROM HEADING			WRONG ENTRIES IN HEADING	WHOLE READING MISSED	ENTRIES (when required) MISSING FROM THE FOLLOWING LINES IN READINGS TAKEN										WRONG ACTIVITY CODES	ACTIVITY CODES IN WRONG LINE	WRONG ENTRIES IN LINE F	OTHER WRONG AND MISSING ENTRIES*	READINGS OVER 10 MINS. LATE OR MORE THAN 3 MINS. EARLY	TOTAL NO. OF ERRORS	WEIGHTED TOTAL NO. OF ERRORS	NO. OF TALLY SHEETS	AVERAGE WEIGHTED ERRORS PER TALLY SHEET	QUALITY RANKING OF REGION
	TOUR	REPORT TIME	ENDING TIME			B	C	D	E	F	G-K	L	M	N	O										
ERROR WEIGHTS	1	1	1	1	10	5	5	5	5	5	10	10	10	5	1	10	5	5	5	10					
1. BOSTON						4											16				20	100	2,496	.004	
2. NEW YORK						87						5	14			26	92	9			233	1390	60,046	.023	
3. PHILADELPHIA						4			58	4						58					124	910	31,124	.029	
4. CINCINNATI						4				21						16	4	20			65	405	27,581	.015	
5. WASHINGTON						20				16	4						29				69	365	17,390	.021	
6. ATLANTA						14				18	8		82			38	103	4			267	1975	13,459	.107	
7. CHICAGO										12	4			4		16	12				48	340	37,951	.009	
8. ST. LOUIS																5	13		8		26	155	13,912	.011	
9. MINNEAPOLIS										26						4	23		4		57	305	12,169	.024	
10. WICHITA										4	4					4		3			15	115	7,986	.014	
11. DALLAS										4	8		28			36	74				150	1110	17,457	.063	
12. SAN FRANCISCO						4				8						36	36	4	4		92	640	41,243	.015	
13. MEMPHIS										4						3	8				15	90	7,269	.012	
14. DENVER																16	3				19	175	7,763	.023	
15. SEATTLE						8				4		4				29	12				57	450	9,016	.050	
TOTAL						145			58	121	22	9	124	4		287	425	40	16		1252	8523	333,691	.026	
PERCENT OF TOTAL						11.5			4.6	9.7	2.2	.7	9.9	.3		22.8	33.8	3.2	1.3		100.0				

rates by office); and one at Washington Headquarters (whereby trained staff of the Revenue/Cost Analysis Division make a sample review of the reports from all post offices). Here, too, a record is made of errors by type and erroneous reports are returned for corrective action.

From an analysis of errors detected at Washington Headquarters, a report showing the frequency of various types of errors and the rate of error, regionally and nationally, is released each quarter to the regions. (See Appendix 2.) These competitive evaluation reports serve as a challenge to the high-error ranking regions to strive for improved accuracy. They also indicate the major causes of detectable errors, so that remedial action can be taken on the most troublesome problem areas. Similar error summaries are sent to the post offices by the Regional Controllers.

What is not readily possible with this quality control system is the identification of certain hidden errors or errors which cannot be detected at the edit and review points. For example, when a reasonable looking but wrong class of mail code is recorded by a supervisor, there is no way of detecting this erroneous entry by the edit and review teams at any of the three levels of quality control. To minimize such errors, a system of monitoring was established whereby the higher level supervisors take periodic readings on the recording supervisor to assure that the latter knows the codes for the classes of mail, services, and operations. Special refresher training programs and independent on-site audits[6] have also been initiated recently to minimize this problem.

All in all, significant attention is being given to the problem of nonsampling errors which besets this (as it does any) data-collection system. Without this special attention, serious questions could be raised regarding the validity of the results and the confidence in the system, even though Work Sampling per se is an objective data-collection system in principle. In the Department's Work Sampling System, continuous quality control maintenance programs represent the backbone of the system, since sampling errors which depend on the sample design and sample size are being kept well within the precision requirements of the users of the cost estimates and nonsampling errors represent the major problem.

Advantages of Work Sampling

In the short period of operation of the Department's Work Sampling System, sufficient evidence has been gathered to pinpoint the advantages of this system over the old Cost Ascertainment System and, in general, over the conventional slip-reporting or time and motion systems. Among these advantages can be included the following:

1. Objectivity and accuracy of the results,
2. Measurable reliability of estimates,
3. Less costly than other methods for the same detailed information,
4. Substantially more detailed and useful management information,
5. Enthusiastic acceptance by the participating employees.

These advantages are detailed in the following sections:

Accuracy

A comparison was made of the cost estimates for the first quarter of Fiscal Year 1969, by class of mail and service, obtained by Work Sampling and the old Cost Ascertainment System. Some large differences, some as large as 15 to 25 percent, were obtained. There were also some small differences.

[6] An audit was conducted during the period November 8–21, 1969 in the 15 largest post offices and the findings were useful in adjusting some of the in-office cost estimates.

The question arises: Which of these two methods provided more accurate estimates? Since none of these differences could be explained by sampling variation (see section on Sampling Variations), the differences are primarily a result of the accuracy of the cost determinations made under the old and new systems.

Under the old system, cost data were collected during one week each quarter (i.e., the Cost Ascertainment Week). Under the New System, cost data are collected each day of the quarter. For this reason alone, the method used by the New System is more likely to yield more accurate results, since it gives representation to cost variations throughout the quarter and does not assume that the variations during the quarter are the same as those shown during the Cost Ascertainment Week. Furthermore, the New System includes a probability sample of post offices whereas the old system did not. Thus, estimates in the New System are more likely to be objective and representative of the universe of in-office costs in all post offices.

A further reason why the New System is more likely to yield more accurate results is the minimal amount of reliance on judgment and memory in the New System as compared with the old system. In the latter system, estimates of the number of hours spent on each class of mail and service were made subjectively by each participating employee during the entire week (i.e., by slip reporting and mixed mail tests). It is a well-known fact that memory can and often does fail and that subjective reporting of time is prone to biases. The New System, on the other hand, uses random-time sampling, whereby at random instants of time each day employees in the sample are observed by someone else (their supervisor) who records the class of mail or service actually being worked on at that time. Thus, there is no reliance on judgment or memory to determine which activity is being worked on. This built-in objectivity of the work sampling recordings is more likely to yield accurate results in general. In fact, objectivity is a well-known feature of random work sampling techniques.

Another fact generally favoring the New System is the use of supervisors rather than the clerical staff to determine the specific type of mail and service being worked on at the random observation moments. Their greater length of postal experience and knowledge of mail classes and other activities is more likely to yield more accurate coding of the actual mail handled or service given by the employee. Furthermore, a system of monitoring a sample of the supervisors' readings has been established which helps detect a need for training to improve the accuracy of the readings.

It is true that supervisors and postmasters who are included in the sample report their activities themselves, i.e., by self-reporting. However, we have not found any particular problem with self-reporting.

Measurable Sampling Variability of Estimates

Systems of data-collection based on probability sampling provide estimates whose reliability, or sampling variation, is measurable. In the section on Sampling Variations, approximate measures of reliability were provided, indicating that highly reliable estimates are being derived from the Work Sampling System. While more refined ways of measuring sampling variation are yet to be developed, no measures of reliability could possibly have been made for the old system of cost estimation, since it was based on judgment samples and subjective determinations.

Cost Comparisons

Generally, Work Sampling has been found to be less expensive for collection of the required data and development of cost estimates by function. Direct cost comparisons

with the old system are not possible because the Work Sampling System at the Department provides a far greater amount and variety of data than the old system. Recent cost studies have indicated that the total annual cost of making and monitoring the readings under the New System, even with its extensively greater amount of detailed data, is about $200,000 less that that of the old system. Thus, over a period of years, the total savings are likely to run into the millions of dollars.

Detailed Data

The New System provides more detailed information than the old system. Data are now available for more than 500 combinations of shapes and classes of mail, as well as basic operating functions, services and activities. This compares with but a small fraction of this number of informational items under the old system. Were the cost comparisons of the two systems made on the basis of cost per item of data, the cost differentials between the two systems would have overwhelmingly favored the New System.

Greater Acceptance of Work Sampling

Interviews with numerous data-collection employees and a questionnaire study have indicated very substantial support for the Work Sampling System. Over 90 percent of the supervisors preferred the New System because of its nondisruptive effect on ongoing work and particularly its objectivity. The latter has given confidence to the participants that accurate results are being derived from the New System.

Some Remaining Research

It is a fact of life that mankind has never remained satisfied with what it has achieved and therefore is constantly striving to improve on its achievements. The challenge of discovering new frontiers and of doing things better and more efficiently is inherent in human beings. So it is with the already successful adaptation of Work Sampling as a tool of cost estimation at the Post Office Department. The existence of the system and availability of data from it will make it possible to conduct research designed to improve the design. Several areas of refinement and further development are readily apparent from a review of what has already been accomplished and what yet remains to be done.

These are:

Optimum Sample Size—As indicated, the size of sample was decided on the basis of a number of administrative as well as technical considerations. The question still remains whether the present size of four million random observations is fully acceptable. In view of the very small relative standard errors which appear to result for many classes of mail and services, one alternative would indicate a smaller sample size. On the other hand, the large relative standard error for the least costly activities would indicate that a larger sample size may be needed.

Optimum Choice Between Sample of Employees and Number of Random Observations Per Employee—The current sample is composed of a 1 % sample of employees (in most instances) and four random observations per day. In combination, they yield the total sample of four million observations a year. Is this the combination of employees and number of daily observations which yields the greatest reliability? In this connection, the question also arises whether more, or fewer, post offices should be included in the first stage of sampling.

Standard Errors—At present, standard errors are computed by reliance on simpli-

fying assumptions of simple random sampling with an allowance of a three-fold loss of efficiency because of the use of a three-stage sample and a cluster sample of an employee-week of random time observations. How closely does this approximation procedure compare with the correct procedure for computing standard errors? This research could also answer the question regarding optimum combination of employees and number of observations per employee each day.

Optimum Sampling Unit—The present second-stage sampling unit is an employee-week. The question exists whether an employee-day, or employee-two weeks, or employee-month, etc., would make a more efficient sampling unit.

Audit-Control Sample—At present, some of the most important errors are not detectable. How many such undetected errors are there? What is their effect on the various cost estimates? These questions remain to be answered, although a recent limited audit of the system has provided some initial information on this question. One way of measuring the magnitude and direction of undetected errors, on a comprehensive basis, is through a careful audit of a random subsample of the total sample. In this way, errors which would normally not be detected by the usual review and edit process would be detected by specialists in data collection. Their findings could then be generalized to the universe of interest and thus provide a basis for statistical adjustment of the original estimates. The planning, development and implementation of this type of audit-control subsample requires research of a challenging type.

More Efficient Quality Control—At present, *all* reports completed in local post offices are reviewed at that post office as well as at the region. It would appear inefficient to perform two such 100% reviews of the recordings. It is also inefficient to review the work of those data collectors who hardly make any errors. If, and when, data collectors can be classified into error-proneness groups, such as low, medium, and high, 100% review can be concentrated on the high-error producers and sample review can be applied to the better performers.

Summary

This paper describes a large-scale application of Work Sampling, using a three-stage sample design. The application covers the in-office costs incurred by the Post Office Department to handle mail prior to delivery, classified by class of mail and service. The work Sampling System covers 900,000 postal employees at some time during a year and about 270,000 employees and 35,000 supervisors participate in the system annually.

The paper describes the basic objectives served by the Work Sampling application, the size of the universe of interest, sample size, sampling methods, estimation procedures, methods of measuring sampling variability and controlling the quality of the basic information. In addition, some of the more essential research studies needed to improve and refine the system are outlined. Above all, this paper tried to illustrate the advantages of using random time sampling for cost determination. While this illustration covered an area of great importance to postal management, the technique is sufficiently adaptable to serve many other management purposes.