

BEFORE THE
POSTAL RATE COMMISSION
WASHINGTON, D.C. 20268B0001

EXPERIMENTAL PRIORITY MAIL FLAT-RATE
Box, 2004

Docket No. MC2004-2

RESPONSE OF UNITED STATES POSTAL SERVICE
WITNESS LOETSCHER TO INTERROGATORIES OF
OFFICE OF THE CONSUMER ADVOCATE
(OCA/USPS-T3-1-6)
(July 8, 2004)

The United States Postal Service hereby files witness Loetscher's responses to interrogatories of the Office of the Consumer Advocate: OCA/USPS-T3-1-6, filed on June 24, 2004.

Each interrogatory is stated verbatim and is followed by the response.

Respectfully submitted,

UNITED STATES POSTAL SERVICE

By its attorneys:

Daniel J. Foucheaux, Jr.
Chief Counsel, Ratemaking

Kenneth N. Hollies

475 L'Enfant Plaza West, S.W.
Washington, D.C. 20260-1137
(202) 268-3083; Fax -3084

OCA/USPS-T3-1. Library Reference USPS-LR-2 indicates at page 1 that the sample design for the data collection was based on the selection of ten sample sites selected at random, with probability proportional to total destinating Priority Mail volume in FY2002.

(a) What was the statistical basis for determining the number of sites to be selected, in this case ten? Please provide references to the sampling literature and/or textbooks.

(b) The study indicated that potential stratification variables were not *a priori* thought to be correlated with the density characteristics. Was there any test of this assumption or, alternatively, was there any previously obtained information to substantiate this assumption?

(c) The Library References USPS-LR-2 states at page 3 that 5,368 sample pieces were obtained as a result of the sampling effort. Please provide the statistical analysis that determined the desired sample size as well as associated levels of confidence and/or any other available relevant statistical information.

(d) Please discuss whether and/or how the distribution of the 5,368 sample pieces among the 10 collection sites was related to the statistical accuracy and precision of the sampling effort.

RESPONSE:

(a-d)

The number of sites selected was determined by the resources available to us for the study. Any statistical analysis conducted to determine the number of sites needed to achieve a desired precision level would have required information on the size distribution of Priority Mail parcels at each site¹. To my knowledge no such data existed. The only site-specific Priority Mail data we had available at the time of sample design was the ODIS originating and destinating volumes for each site.

The decision to sample destination Priority Mail was made because we strongly believed that the distribution of destination volumes at each site was more representative of the population than the distribution of originating volumes at each site. Originating mail volumes at a site may be dominated by one Priority Mail customer such as a fulfillment house. The sizes of Priority Mail parcels from the dominant customer

¹ Cochran, William G. (1977) *Sampling Techniques*. John Wiley & Sons, New York. Chapter 10 pages (280-283)

are likely to be less variable than the population. Conversely by the time Priority Mail reaches its destination, the mail from all customers large and small would be commingled and therefore most representative of the population.

The site-specific data available at the time were measures such as site location, size, facility type, destinating volumes of other classes. These measures are unlikely to be correlated with the size distributions of Priority Mail parcels. These assumptions could not be tested *ex ante* as we lacked the necessary information on the size distribution of Priority Mail parcels at each site.

The sample of 5,368 parcels was achieved by intense sampling at each site for two days. Again the lack of data on the size distribution of Priority Mail parcels precluded us from calculating the number of sample pieces needed to achieve a desired precision level. Because sampling occurred at destination the distribution of sample pieces across sites, is unlikely to have a significant effect on the precision or accuracy of the estimates.

OCA/USPS-T3-2. The Library Reference USPS-LR-2 presents the special study that was initiated to gather the data on the characteristics of Priority Mail. Please provide copies of training manuals, procedural instructions, and other relevant material distributed to data collection personnel.

(a) Please enumerate the training procedures and information provided to the data collection personnel for the data collection.

(b) Please delineate the quality control procedures.

RESPONSE:

(a-b)

A copy of the detailed sampling instructions given to data collectors has been supplied as an attachment to my response to DBP/USPS-T3-3. Prior to sampling all personnel involved in the survey met and reviewed the sampling protocol outlined the training document, discussed possible problems, and solutions to those problems. Each data collection team was staffed with a team leader who had experience with Postal Service operations and sampling techniques. Teams arrived and inspected the sample site the day before data collection began to tour the site and established the site-specific sampling protocol. If the team leader had questions or concerns regarding a site they were discussed with the survey leaders and resolved before data collection began. Throughout data collection the survey leaders communicated frequently with the team leaders to discuss any areas of concern.

OCA/USPS-T3-3 The Library Reference USPS-LR-2 mentions on page 3 “data processing and cleaning.” What steps were involved in these efforts?

(a) Please indicate the number of pieces of Priority Mail and the characteristics of the pieces eliminated from the analysis as a result of the implementation of these processes.

(b) Given that a special study was implemented to gather the data, please denote the reasons for needing to implement a cleaning process, if in fact the process involved any major amount of elimination of data from the study.

(c) Please indicate if the cleaning process was perfunctory, involving the elimination of only a small amount of data.

RESPONSE:

(a-c)

The primary purpose of the data processing and cleaning phase of the study was to use the origin ZIP Code, destination ZIP Code, postage affixed, and special services to map each observation to postal zone. To accomplish this we used the zone chart supplied to us by the Postal Service to calculate zone from the origin ZIP Code (taken from the return address) and the destination ZIP Code. We then calculated the postage for each piece based on the measured weight and calculated postal zone. For pieces with postage affixed we compared the calculated postage and the affixed postage. If there were inconsistencies between the calculated postage and the affixed postage that could not be explained by special service fees or if the origin ZIP Code was missing, the observation was discarded. For permit imprint pieces we assumed that the ZIP Code given in the return address was an accurate indicator of origin.

In addition to calculating postal zones, the cleaning process eliminated all pieces from the sample that were flat shaped or not Priority Mail pieces. The systematic sampling technique used would occasionally result in the selection of a flat shaped piece or a piece from another class that had ended up in the Priority Mail parcel stream. During data collection all selected parcels were measured and entered in the database.

For the first sites visited, selected flats were also measured and entered into the database. We realized that entering the data for flats was unnecessary and detracted from the time available to sample parcels. For later sites flats were still selected in the systematic sampling and measured to verify that they met the flat criteria as defined in the *Domestic Mail Manual* but if they were determined to be flats they were not entered into the database.

During collection, information on 5,666 pieces was entered into the database. Of these, 289 pieces were eliminated from the sample for the following reasons:

- 149 pieces were flat shaped pieces,
- 84 pieces were missing origin ZIP Code information,
- 30 pieces were determined to be of foreign origin,
- 16 pieces were determined to be First-Class Mail pieces,
- 8 pieces were determined to be Business Reply Mail pieces,
- 8 pieces had irreconcilable differences between affixed postage and calculated postage,
- 2 pieces were determined to be Parcel Post pieces, and
- 1 piece was determined to be Media Mail.

OCA/USPS-T3-4. On page 2 of the Library Reference USPS-LR-2, is the statement “Sample selection was determined by how best to sample for the selected SCF while not interfering with operations and retaining sufficient time to select and record sample pieces for the sample site.”

(a) Please expand on this statement.

(b) Please indicate whether this procedure biased the sample. If the answer is “No”, please explain why such a procedure did not bias the sample.

RESPONSE:

The incoming secondary operations for Priority Mail are not uniform across sites. The incoming secondary operation for some sites was conducted at a Priority Mail Processing Center (PMPC); for others it was conducted at the P&DC of the selected site. Each location where the Priority Mail incoming secondary operation was being performed had different plant layouts, sortation technologies (manual/SPBS), flows to and from the docks and safety concerns.

Because sampling locations differed so much it was necessary to adapt the sampling protocol to the individual site. At each site the team leader worked with operations personnel to get a complete understanding of the flows of Priority Mail in the plant so that the sampling would include all pieces in the universe. From there a sampling location was chosen that was safe and did not interfere with operations. For some sites this was near the location where containers of Priority Mail were staged after being unloaded from trucks and before being taken to the incoming secondary sortation operation. At other sites sample pieces were drawn from containers as they entered the sortation operation. When the incoming secondary operation for the selected site was conducted at a PMPC and incoming mail for the selected site was commingled with mail for other sites, sample pieces were collected as the mail was dispatched from the operation.

In all cases the systematic skip factor was applied to all relevant containers. At all sites sampling was conducted whenever Priority Mail was in the plant and available for sampling during the two days we were at the plant. For selected containers, the systematic piece skip factors were applied to all pieces in the container until the container was empty to avoid any bias that might be introduced by smaller high-density pieces filtering to the bottom of the container or by any other consequence of mixing. The schedule for sampling at the sites was constructed such that we sampled pieces processed on all days of the week. For these reasons I do not believe that any bias was introduced by the slight variations in the sampling protocols at each site.

OCA/USPS-T3-5. The Excel workbook file “Cubic Feet Distribution LR.xls” listed in Library Reference USPS-LR-2, Appendix A, has no headings for other than column A, although one would assume that the headings in “cubic ft dist txt” is related to the data. Since there is not a one to one match between the headings of the two files, please provide the headings for “Cubic Feet Distribution LR.xls”.

RESPONSE:

There are four worksheets in this workbook. The sheet labeled “Distribution” is the output sheet for the macro “distmac” . The outputs are the estimated proportion of pieces in each cubic foot increment by zone and pound increment. The columns in row 10 provide the cubic foot increment 0-0.1 cubic feet, 0.1 to 0.2 cubic feet, 0.2 to 0.3 cubic feet 0.3 to 0.34 cubic feet and 0.34 cubic feet and above. These are repeated for each zone. The zones are labeled in row 9 (centered over the section). The pound increments are presented in column C starting at row 11.

The sheets “1&2 Pound” and “2> Regression” are the standard regression outputs from the regression package available with Excel.

The sheet columns in the sheet “Sample Data” are labeled in row 2. For clarity the column headings are:

- A ObsIID A sequential number of the sample observations.
- B Weight The weight of the sample piece, in pounds, as measured.
- C Length The length of the sample piece, in inches, as measured.
- D Width The width of the sample piece in inches, as measured.
- E Height The height of the sample piece, in inches, as measured.
- F Girth The girth of the sample piece, in inches, as measured.
- G VolumeCubics The calculated cubic volume of the parcel in cubic feet.
- H iZone The calculated zone of the piece.
- I WTI The pound increment of the piece.
- J $\ln(\text{ft}^3)$ The natural log of “VolumeCubics”.
- K nwt The natural log of “Weight”.
- L $\ln\text{wt}^2$ The natural log of “Weight” squared.
- M z4 An indicator for if the piece is zone 4.
- N z5 An indicator for if the piece is zone 5.
- O z6 An indicator for if the piece is zone 6.
- P z7 An indicator for if the piece is zone 7.

- Q z8 An indicator for if the piece is zone 8.
- R 3lbs An indicator for if the piece is over 2 pounds and is subject to the 3 pound rate.

OCA/USPS-T3-6. Please explain your understanding of the Commission's requirements for the development of a statistical sample and how the sample presented in this case fulfills the requirements.

RESPONSE:

It is my understanding that for the development of a statistical study the Commission requires comprehensive descriptions of assumptions made, study plan utilized and the steps undertaken. In my opinion the most significant assumption we made in the study was that the error terms were normally distributed. This assumption is documented explicitly on page 5 of the library reference (USPS-LR-2) and the supporting evidence is discussed on page 4, in footnote 3. The other significant assumption we made, that the volume of pieces originating and destinating in the same 5-digit ZIP Code was negligible, is documented on page 1 of the library reference.

The study plan – to select and sample pieces at those sites, estimate the relationship between cubic volume and weight, and use these estimates to produce the size distributions and density estimates – is described throughout the library reference. The procedures undertaken are also described (e.g., we selected sites proportional to ODIS destinating volume, we sampled pieces using systematic container skip factors, and the regression model employed, etc.).

The Commission requires a description of the survey design, the sampling frame, and units and confidence limits that can be placed on major estimates. The sample design and frame are described in section A beginning on page 3. This section also describes the universe under the study. The major estimates for the study, in my opinion, are the parameter estimates of the two regressions. The confidence limits for these estimates are presented explicitly in the workbook “Cubic Feet Distribution LR.xls”

and the t-statistics from which the confidence limits can be derived are presented on pages 4 and 5 of the library reference.

The method of selecting the sample and the characteristics measured are presented in section B, starting on page 2. In addition the sample data and all programs necessary to replicate the study were provided with the library reference.

CERTIFICATE OF SERVICE

I hereby certify that I have this day served the foregoing document upon all participants of record in this proceeding in accordance with section 12 of the Rules of Practice.

Kenneth N. Hollies

475 L'Enfant Plaza West, S.W.
Washington, D.C. 20260-1137
(202) 268-3083; Fax -3084
July 8, 2004