Survey of Income and Program Participation

Working Paper Series

METHODOLOGIES FOR IMPUTING
LONGITUDINAL SURVEY ITEMS

No. 8506 7

by

Vicki J. Huggins, Michael E. Samuhel
and Lynn Weidman

December 1985

Table 9
SELECTED INCOMPLETE AND DONOR PATTERNS FOR WANT JOB
January 1979 - December 1979
(0 = yes, 1 = no, 2 = missing item)

Rotation	Incomplete Pattern	Frequency	Donor Patterns for Missing Months	Frequency (%)	Percent Incorrect Imputations
A	000 000 022 211	2	111 000	1 (33.33) 2 (66.67)	(.4444)
B	002 220 001 111	2	111 000	8 (66.67) 4 (33.33)	(.4444)
A	011 111 111 122	2	11 00	56 (94.92) 3 (5.08)	(.0964)
A	100 022 211 111	3	111 000	24 (80.00) 6 (20.00)	(.3200)
C	111 000 222 111	2	111 000	24 (77.42) 7 (22.58)	(.3496)
A	111 111 100 022	2	11 00	23 (82.14) 5 (17.86)	(.2934)
B	111 111 111 112	8	1 0	1086 (98.91) 12 (1.09)	(.0216)
A	111 111 111 122	12	11 00	1067 (99.26) 8 (0.74)	(.0147)
A	111 122 222 222	4	11 111 111 11 111 100 11 100 011 11 100 000 00 011 111 00 011 100 00 000 011	1067 (94.3) 8 (0.71) 23 (2.04) 5 (0.44) 17 (1.51) 1 (0.09) 3 (0.27)	(.1100)

TABLE OF CONTENTS

I. INTRODUCTION

A. Missing Data in a Longitudinal Survey

For each wave of a longitudinal survey, non-response may occur in one of two basic forms: wave nonresponse where the respondent answers no questions for the entire wave interview and item nonresponse where the respondent answers some questions but not all.

Wave nonresponse for a longitudinal survey is different from unit nonresponse for a cross-sectional survey. For the latter, typically the only variables available are those from sample frame data and interviewer notes. But because a longitudinal survey is conducted over time, the values of some variables, such as demographic variables, are available from earlier waves and are not expected to change much. These presumed known values can be used to contribute to the resolution of missing data values. Even more important, for a missing variable one may have access to values of this variable and other correlated variables as reported in other waves (longitudinal patterns). In some cases, these two factors can contribute to a reasonable resolution of missing values for wave nonresponse, and this sort of option is not usually available for cross-sectional surveys.

The focus of this report is an initial study of the imputation of missing items in a longitudinal survey. For a given variable, a respondent can report its value for every time period over a given wave, for none or some. Some variables are wave variables, some monthly, and some are weekly. Missing data in SIPP and ISDP result from noncoverage, household nonresponse, person nonresponse, record nonresponse and item nonresponse. Noncoverage occurs when some units in the survey population are not included in the sampling frame. Household nonresponse occurs when no data are reported for any household members. Person nonresponse results when data are not collected for one or more persons in an otherwise cooperative household. Item nonresponse occurs when some but not all the data are missing for a person. This may consist of individual questions (items) being unanswered or of whole sets of questions (record types) being unanswered. For item nonresponse within a wave, missing values can be imputed based on other reported values for that wave, auxiliary (often demographic) information, or from responses from other waves. We will attempt to describe a unified item imputation strategy to handle each of these cases. Whether these techniques can be applied fruitfully to wave nonresponse is an open question and will not be addressed in this report.

Over the past several years there has been an increased interest in imputation methodologies. We refer the reader to Kalton (1983) for a broad review of techniques and an extensive bibliography. The imputation methods for cross-sectional surveys do not translate well for longitudinal surveys. They do not make use of reported values in other time periods which can provide valuable information in determining a valid and reasonable response. For this reason, the burden in devising a longitudinal imputation strategy is to simultaneously employ and be constrained by both cross-sectional and longitudinal information.

B. Objective of This Report

The broad objective for work described in this report is to develop imputation methodologies for the Survey of Income and Program Participation (SIPP) longitudinal data file. This report is the first of several on this topic and certainly does not fully resolve the question as to the optimal imputation procedures for SIPP.

The first objective in developing such a procedure is to test and examine techniques for using information from more than one interview (longitudinal information) in an imputation strategy. In this report, we only treat one variable at a time and do not employ cross-sectional information (with one exception described below). Subsequent efforts must integrate longitudinal and cross-sectional information, and work is in progress to integrate the use of both longitudinal and cross-sectional information for item imputation.

There are structural reasons why an item may be blank on a data file. One is that a respondent failed to answer a question that should have been answered. Another is that the respondent was never asked that question, and no answer is required. Because of responses on preceeding questions and the skip pattern underlying a questionnaire, all questions are not asked of all respondents. In lieu of a response to a non-applicable question, one should code the field as, say, "NA", to indicate that no response is required and the item is not applicable for this respondent. The determination that a field should be coded as NA is made by observing responses of other variables in the same time frame, and this is the only systematic use of cross-sectional information in this study.

In this report we present an imputation methodology for categorical variables having the following property: for each variable the distribution of longitudinal patterns for months not reported conditioned on values in reported months is the same in the set of imputed records as in the set of completely reported records. Indications are given

for the incorporation of demographic information, and methods are suggested for the imputation of monthly income amounts.

C. Using Income Survey Development Program (ISDP) Test Data

The imputation methodology described here was developed from data collected on the Income Survey Development Program (ISDP). However, its implementation can occur only after further development, in particular, the incorporation of cross-sectional information into an overall methodology. Here, it is described as a general approach applicable to any longitudinal survey. Data from the ISDP are utilized to explain and test the method, provide examples, and explore its applicability for SIPP. The longitudinal data for our imputation research is an annual file, constructed by merging the five waves of ISDP data from the 1979 Research Panel.

The Income Survey Development Program (ISDP) was initiated to gain experience with the data collection and data analysis requirements of SIPP. The ISDP contained two national longitudinal surveys, the 1978 Panel and the 1979 Research Panel. The sample design is a multi-stage stratified sample of the United States population. Sampling elements are housing units at the time of the first interview, individuals within the unit are identified during this interview. From this point on the individuals in sample are the sample elements. The first sampling stage involves classifying the United States in terms of counties or groups of counties, called primary sampling units (PSU's), which are stratified. At the second stage, a sample of addresses within the PSU's is selected. Interviews are conducted every three months, and each household is assigned to one of three rotation groups (A,B,C). Every three months all the households in a rotation group are interviewed and data on labor force participation, income, and program participation are collected for each of the previous three months. A wave is the time period during which each rotation group is interviewed once. The rotation group pattern for the 1979 Research Panel is shown in Figure 1.

The ISDP file used in this project contained six waves of data and was entered into a Scientific Information Retrieval (SIR) database. The organization of this database allows one to retrieve a variety of cross-indexed data records, and this capability was very useful in this study. In addition to manipulating records for data extraction, SIR allows the user to specify the format of output files, in particular, SIR is able to create SPSS files. Much of the data analysis in this report was performed using the SPSS statistical package on data extract files created within SIR.

II. LONGITUDINAL NONRESPONSE PATTERNS

Define a longitudinal record for a survey unit to be the set of responses recorded over a fixed time period. In the ISDP as well as SIPP, the survey unit is a household, but other examples of survey units include the person, family, and employer. In this report, the survey person record is the unit of analysis. The set of responses on the longitudinal record may be any combination of survey items. For this report, we restrict ourselves to a single item recorded monthly for one year.

In the annual file for the 1979 panel of ISDP, missing data occurs in a wide variety of types and patterns. Household, person, and item nonresponse can all occur in any combination on a longitudinal record. In addition, they can occur in any month or group of months. The following table of work force variables from the ISDP database used in this study indicates the extent and diversity to which missing data can occur. (The total number of longitudinal records for persons 16 or over who were in the survey for the wave 1 interview is 19,114.)

item	% of longitudinal records having item non-response but no wave non-response	% of longitudinal records having both wave and item non-response	% of longitudinal records having wave non-response but no item non-response
WORK STATUS	2.7%	3.1%	19.9%
LOOKING/ NOT WORKING	2.7%	3.1%	19.9%
LOOKING/ WORKING	2.7%	3.1%	19.9%
WANTJOB/ LOOKING	4.3%	3.9%	19.1%
RECEIPT	8.1%	10.4%	45.6% •

• The percentages were obtained by evaluating the person responses for job 1 only.

Consider the item WORKSTAT, which indicates whether an individual had a job or business. In Table 1.A, all the monthly nonresponse patterns in our database are listed for the period January 1979 to December 1979. The values refer to the reference month and not the month of the interview. We include every response pattern in this table in order to indicate the diversity of such patterns. Recall that each rotation group is interviewed every three months with the information being collected for the previous three. Note also that in Table 1.A only response and nonresponse are indicated. A reported value is denoted by $X_t = 0$ ($t = 1,2,...,12$). We do not distinguish between "yes" or "no" responses in this table. Months where the item is missing but is not part of a missing wave are indicated by $X_t = 1$. An item missing because it is part of a missing wave is denoted by $X_t = 2$. In Tables 1.B to 1.D, these nonresponse patterns are broken down into disjoint subsets having only record types missing, only items missing, and both record types and items missing. Tables 2.A to 2.D are similar to Tables 1.A to 1.D for the variable LOOKING/WORKING (people working and looking for a job), Tables 3.A to 3.D for LOOKING/NOTWORKING, Tables 4.A to 4.D for WANTJOB, and Tables 5.A to 5.D for RECEIPT of wages. In Tables 2.A to 4.D, $X_t = 3$ means the question was not applicable for those persons in month t.

An undetermined portion of the record and household nonresponse in the longitudinal file resulted from difficulties in merging the cross-sectional files. The lack of unique personal identifiers caused some cases to be mismatched or made it infeasible to find a unique match. The extent of the problem is not known, but it does distort the household and person nonresponse described in all the tables of nonresponse patterns. The objective in creating the longitudinal file used in this project was to have a database to develop and test imputation methodologies. The nonresponse rates for the various categories displayed reflect records on this file and should not be used to approximate actual nonresponse rates for ISDP. The tables of nonresponse rates are included in this report to show the diversity of nonresponse patterns and to indicate the variety of nonresponse combinations that must be considered in devising a broad-based procedure to treat nonresponse in a longitudinal survey.

III. IMPUTATION OF MISSING LONGITUDINAL CATEGORICAL SURVEY ITEMS

A. The Basic Procedure

There are two equivalent ways of defining this imputation procedure for a specified

variable. The first uses the distribution of reported longitudinal patterns and the second uses matching. Throughout this report we discuss the distributional approach.

1. Distributional approach

   1a. Take the longitudinal records for the specified variable and one rotation group.

   1b. For a given record that has one or more months missing, take all records from (1a) that have 12 complete months of data and match the given record in all its reported months.

   1c. Estimate the conditional distribution of patterns of reported values for the missing months given the reported values in the responding months.

   1d. Using this distribution, randomly select a pattern of values with which to fill in the missing months on the given record.

2. Matching approach

   2a. Take the longitudinal records for the specified variable and one rotation group.

   2b. For a given record that has one or more months missing, randomly select from the records with 12 complete months one that matches the given record in all its reported months.

   2c. Fill in the missing months on the given record with the corresponding monthly values from the selected record.

The following example illustrates this imputation process using the first procedure and a precise description of the procedure is given below.

Example: Consider the ISDP survey item WORKSTAT indicating whether a person had a job or business during a wave. Further, consider the set of individuals in rotation group B who responded "yes" from January thru November 1979, but did not respond in December,

1979. An example of a longitudinal record for an individual is given by

$$X = (0,0,0,0,0,0,0,0,0,0,0,2),$$

where $X_t = 0$, if the response in the t^{th} month is "yes", $X_t = 1$ if the response is "no", and $X_t = 2$ indicates a missing item for t=1,...,12. Either "0" or "1" is an admissible imputation value for X_{12}. Based on the individuals in rotation group B who reported data in every month from January to December (see Table 6) we estimate

$$\text{Prob}\,(X_{12} = 0 \mid X_1 = 0, X_2 = 0,\ldots,X_{11} = 0)$$

$$= \frac{2313}{2379} = 0.9723$$

$$\text{Prob}\,(X_{12} = 1 \mid X_1 = 0, X_2 = 0,\ldots,X_{11} = 0)$$

$$= 1 - .9723 = 0.0277\,.$$

Generating a random number between zero and one, we impute $X_{12} = 0$ if the random number is less than or equal to 0.9723, otherwise we impute $X_{12} = 1$.

This imputation procedure can be applied to any categorical survey item with any combination of missing months. Consider the item RECEIPT, indicating receipt of wages and salaries, and the following longitudinal record for a person in rotation group A:

$$X = (0,0,0,0,0,0,0,2,2,2,0,0)\,.$$

For persons responding in all twelve months (see Table 10), we estimate:

$$\text{Prob}\,(X_8 = x_8, X_9 = x_9, X_{10} = x_{10} \mid X_1 = 0,\ldots,X_7 = 0, X_{11} = 0, X_{12} = 0) =$$

$\frac{1120}{1140} = 0.9823$ if $x_8 = 0$, $x_9 = 0$, $x_{10} = 0$,

$\frac{10}{1140} = 0.0088$ if $x_8 = 1$, $x_9 = 0$, $x_{10} = 0$,

$\frac{4}{1140} = 0.0035$ if $x_8 = 0$, $x_9 = 0$, $x_{10} = 1$,

$\frac{3}{1140} = 0.0026$ if $x_8 = 1$, $x_9 = 1$, $x_{10} = 0$,

$$\bar{1}\bar{1}\overset{2}{\bar{4}}\bar{0} = 0.0018 \qquad \text{if } x_8 = 0,\ x_9 = 1,\ x_{10} = 0,$$

$$\bar{1}\bar{1}\overset{1}{\bar{4}}\bar{0} = 0.0009 \qquad \text{if } x_8 = 1,\ x_9 = 0,\ x_{10} = 1.$$

Here, we impute the entire subvector (x_8, x_9, x_{10}) based on a random draw from a uniform (0,1) distribution. That is, conditioning on the values of the items reported in other months, missing data patterns are imputed based on the empirical distribution function generated from the completely reporting cases.

Basic Procedure: The imputation process is formalized by letting the random variable X represent the responses (and missing data) on a longitudinal record. The response vector X can be partitioned into subvectors X_m and X_r, representing the missing and recorded monthly values, respectively. On the i^{th} longitudinal record, we impute the missing items X_{m_i} based on the reported values x_{r_i}. The imputed values are a random draw from the conditional distribution $f(X_{m_i} \mid X_{r_i} = x_{r_i})$, empirically estimated from the longitudinal records with values reported in every month.

Let

$$\underline{a} = (a_1, \ldots, a_{12})$$

be a longitudinal record where a_i is a possible response for field X_i, i=1,...,12, let $T'(X = \underline{a})$ be the number of times the vector $\underline{a}$ appears in the data set completed by imputation, and let n be the total number of records in the data set.

When using the procedure outlined above,

$$T'(X = \underline{a})/n$$

in an imputed data set is an unbiased estimator of

$$P(X = \underline{a})$$

if the data are missing at random in the sense that the mechanism giving rise to nonresponse is independent of respondent values, see Kalton (1983) for further discussion. This is what one would expect from this procedure, and we include a proof of this assertion in the Appendix. Of course, data are rarely ever truly missing at random, and in any implementation of this procedure (or any other imputation procedure for that

matter) the non-random aspects of the missing data mechanism must be brought into play to the extent they are known. For this, one must draw upon cross-sectional information and subject-based consideration and blend them with the basic procedure described above and expanded upon in the Appendix.

B. Determining Monthly Applicability

In a longitudinal survey certain survey items may apply to an individual in some months, but not others. For example, certain income questions do not apply to persons who are unemployed or retired, and individuals may retire or become unemployed at any time during the survey. Similarly, survey items concerning social welfare programs only apply to those individuals who participate in these programs. In Tables 2.A through 5.D, we include a sample of nonresponse patterns in which we distinguish between various forms of missing and nonapplicable. Consider the selected nonresponse patterns in Tables 2.A to 2.D and 3.A to 3.D for the item LOOKING, which indicates whether the individual spent any time looking for work. (This is a wave variable which is constant for the three months in a wave.) Tables 2.A to 2.D refer to those individuals who were working, but also looking for other work, and Tables 3.A to 3.D refer to those individuals not working and were looking.

Consider the survey items indicating whether an individual wants a regular full or part time job. On the data file, separate variables distinguish between those individuals who want a job based on whether or not they are looking for a job. For individuals in rotation group A who are looking for a job but not working, consider the longitudinal record for the variable LOOKING/NOT WORKING:

$$x = (1, 2, 2, 2, 4, 4, 4, 4, 4, 4, 4, 4),$$

where $x_t = 0$ ($t = 1,2,\ldots,12$) if the response in t^{th} month is "yes", $x_t = 1$ if the response is "no", $x_t = 2$ indicates missing data and $x_t = 4$ indicates nonapplicability of the item in the t^{th} month (see line 6, Table 8). Individuals with the above longitudinal record were not looking for a job in January, they did not respond to the survey item from February through April, and the item did not apply from May through December.

To impute responses for the vector (x_2, x_3, x_4), we condition on the reported values on the longitudinal record. We want to find, among the completely reporting cases, those individuals for whom this question was not applicable from May through December, but were not looking for a job and not working in January. Accordingly, we

treat nonapplicability as a valid response so that the only longitudinal records that occur in the data set which can be considered for the imputation of subvector (x_2,x_3,x_4) in

$$x = (1,2,2,2,4,4,4,4,4,4,4,4)$$

are

$$(1,0,0,0,4,4,4,4,4,4,4,4) \text{ and } (1,1,1,1,4,4,4,4,4,4,4,4).$$

The concept of nonapplicability allows incorporation of cross-sectional information into the coding of the longitudinal record. With detailed identification of relationships between relevant variables, some of the desired cross-sectional consistency can be induced simultaneously within the proposed longitudinal imputation scheme. Whether through nonapplicability techniques, an editing scheme, decision logic tables, or any combination of the above, it is fundamental that imputed values pass editing criteria and procedures that guarantee this must be implemented at some point in data processing.

C. The Expected Percentage of Incorrect Imputations

The amount of information available longitudinally can be measured by an estimate of the expected percentage of incorrect imputations. Consider the longitudinal record for the monthly receipt of wages and salaries,

$$X = (0,0,0,0,0,0,0,0,0,0,0,2), \tag{1}$$

where $X_t = 0$ indicates receipt and $X_t = 2$ $(t=1,\ldots,12)$ indicates missing data. The probability

$$\text{Prob}(X_{12} = 1 \mid X_1 = 0,\ldots,X_{11} = 0)$$

is estimated from the completely reported cases as 8/1236 = 0.0065. This probability is independent of but equal to the probability of imputing $X_{12} = 1$ when using the basic procedure. Consequently, the probability that $X_{12} = 1$ is imputed and is correct is $(0.0065)^2$ under the assumption that data are missing at random. Similiarly, the probability that $X_{12} = 0$ is imputed and is correct is $(0.9935)^2$. It follows that the estimated probability of an incorrect imputation for the longitudinal record (1) is

$$1 - ((0.0065)^2 + (0.9935)^2) = .013.$$

More generally, if there are n possible imputations, each with probability p_i, $i=1,...,n$ where $\sum_{i=1}^{n} p_i = 1$, then the probability of an incorrect imputation is:

$$H(p_1,...,p_n) = 1 - \sum_{i=1}^{n} p_i^2.$$

Note that $H(p_1,...,p_n)$ obtains its maximum when $p_i = 1/n$ for all $i=1,...,n$, and this is as one would expect. If this value is too large (by some well-defined criteria), alternate procedures may have to be called upon or additional information brought to bear in the choice of an imputation.

Several methods are available for including additional information to use in the selection of the most appropriate donor pattern. One approach is to include related survey items or demographic information as elements in the longitudinal record. For example, to impute the monthly receipt of wages and salaries, we may want to include survey items indicating seasonal or part time workers.

IV. IMPLEMENTATION CONSIDERATIONS FOR CATEGORICAL DATA

A. Selecting Donor Patterns

In order to estimate the feasibility of the methodology presented in this paper, a prototype program to impute categorical data was developed and tested on the ISDP variables WORKSTAT (whether a person worked in the previous three months) and RECEIPT (whether a person received wages and salaries in the previous three months). The coding of a longitudinal record as a twelve character string is not only useful for notational purposes, it is also operationally efficient. The data are viewed in this manner so that a character by character comparison of the longitudinal records can be implemented to identify all incomplete response patterns along with their donor sets and frequency distributions.

A prototype program was written for this project which first identifies and stores the set of incomplete records. For each family of incomplete data patterns, (i.e., some fields missing and identical values on reported fields) the set of potential donors were identified. Frequency of donor patterns (hence probabilities) were computed as described in Section III, and a random draw from the uniform (0,1) distribution was used to select a

donor pattern. The selected donor pattern was imputed for the missing values.

Selected examples of the components of this imputation methodology using the prototype program and data from the ISDP are found in Tables 6 to 10. In each table, selected patterns of response for longitudinal records with missing data, i.e., incomplete response patterns, are given along with their frequencies of occurrence. The donor patterns for the missing monthly values are a subset of data fields on the completely reported cases that agree with the incomplete record on its reported months These are described along with their frequencies of occurrence. The imputation decision is based on the relative frequencies of the donor patterns. This procedure can be extended so that one uses not only completely reported cases for donors, but also cases partially completed at an earlier stage of processing.

B. Rotation Group and Longitudinal Imputation

In order to obtain longitudinal consistency of imputed data for wave variables, it is necessary to condition on the rotation group of the record in addition to conditioning on the values in reported months for wave response variables such as WORKSTAT. To do this, each longitudinal record was coded as a thirteen character string where the first twelve characters represented the responses of the variable in months January to December, and the thirteenth character represented the record's rotation group. If rotation groups A, B and C are denoted by 1, 2 and 3, as the thirteenth record component, then from table 6 the pattern for WORKSTAT:

100 000 000 022 1

is a longitudinal record coming from rotation group A (found on line 7 of Table 6).

WORKSTAT refers to the question whether the individual had a job or business at any time during the three months prior to the interview, so it is a wave variable. Consider the longitudinal record

$$(x_1 = 0, x_2 = 0, \ldots, x_9 = 0, x_{10} = 2, x_{11} = 2, x_{12} = 2),$$

see line 3, Table 6. Because the question refers to the entire three month period, the response must be the same for x_{10}, x_{11}, and x_{12}. Response patterns such as 101 or 010 are not admissible and cannot be imputed.

Referring to Figure 1, we see that this particular pattern can occur only for cases

in rotation group C. By conditioning the imputation on the completely reporting cases in rotation group C, the only patterns considered for imputation are 111 and 000.

Not conditioning on rotation group for wave variables like WORKSTAT readily leads to inconsistent imputations. For example, ignoring the distinction in rotation group could result in the reported months being matched to individuals from rotation group A. In Figure 1, we see that reference months October, November and December are covered by interviews in waves four and five. Consistent responses during the interview could result in the patterns (x_{10} = 1, x_{11} =0, x_{12} =0) and (x_{10} =0, x_{11} =1, x_{12} =1), both of which are inadmissible imputations for an incomplete record in rotation group C.

The categorical imputation program was able to completely impute for the set of incomplete patterns that occured for the variable WORKSTAT, indicating that there was enough variety in the donor set for the imputation scheme to be effective. A donor was matched to an incomplete record if the completely reported months and rotation group of the nonresponse pattern matched the same months and rotation group of the donor pattern.

In order to test if it was practical to include rotation group information on monthly variables, we tested this procedure for the variable RECEIPT. Since RECEIPT is a monthly response variable (whereas WORKSTAT is a wave response variable), many more possible response patterns are generated by the RECEIPT data because of the combination of responses that can occur within waves. By not conditioning on rotation group to identify donors for incomplete records, 96.8% of all incomplete records in the data set were imputed. Conditioning on rotation group resulted in imputation of 96.0% of the incomplete set of records. So only .8% of the possible imputations are lost by conditioning on rotation group.

V. THE IMPUTATION OF CONTINUOUS DATA

Regression and matching are two methods that can be used to capture the longitudinal information for imputing continuous data such as monthly income amounts.

A. Regression Techniques

One regression technique suggested for estimating missing values in multivariate data was proposed by S.F. Buck (1960), and an iterative form of Buck's method is outlined and examined in the paper by E.M.L. Beale and R.J.A. Little (1975). Buck's main purpose was to present a method of estimating the variance-covariance matrix of any k-variate

population. Beale and Little utilized Buck's variance-covariance matrix estimator to iteratively arrive at imputed values and estimators of the survey population where it is not necessary to assume the underlying population is multivariate normal.

An important consequence of the iterated Buck method is that with the assumption of an underlying multivariate normal population, the method is essentially equivalent to an EM-Algorithm which gives maximum likelihood estimates of the population parameters. Without the normal distribution assumption however, the imputed values obtained in the iterative procedure are still regression based estimates and it is not unusual to use regression analyses or non-normal data.

Assume for a set of N observations and n variables that x_{ij} represents the value of the j^{th} variable in the i^{th} observation for j=1,...,n and i=1,...,N. Let m_j denote the sample mean value of the j^{th} variable over all complete observations and u_{jk} denote the sample covariance between variables m_j and m_k over all complete observations. The iterated Buck method uses m_j and u_{jk} to compute:

$$(1)\ x_{ij} = \begin{cases} x_{ij}, \text{ if } x_{ij} \text{ if observed} \\ \text{a linear combination of the set of variables observed in the } i^{th} \text{ observation, if } x_{ij} \text{ is missing} \end{cases}$$

$$(2)\ c_{ijk} = \begin{cases} \text{partial covariance of } m_j \text{ and } m_k, \text{ if } x_{ij} \text{ and } x_{ik} \text{ are both unknown} \\ 0, \text{ otherwise} \end{cases}$$

$$(3)\ \bar{x}_j = \sum_{i=1}^{N} x_{ij}/N$$

$$(4)\ a_{jk} = \sum_{i=1}^{N} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) + c_{ijk}.$$

Then set $m_j = \bar{x}_j$; $u_{jk} = a_{jk}/N-1$ and repeat (1) thru (4) until there are no further changes in m_j and u_{jk}. The term c_{ijk} is a correction term for the bias that would normally occur in the formation of a_{jk}.

In this study, missing values that occured for the continuous variable AMOUNT, the amount of wages and salaries earned in any month by a person, were imputed using the iterated Buck method. Since we were dealing with longitudinal records, the variable AMOUNT was split into 12 subvariables, $AMT_{i,1}$ to $AMT_{i,12}$ where i=1,...,N for the number of longitudinal records that occur and j=1,...,12 represents the months January to

December. The iterated Buck method was then applied as above with $x_{ij} = AMT_{ij}$.

Preliminary results indicate that the method is both efficient and effective. However, subsequent empirical evaluation of this method is underway to better determine the validity of the imputations and the form of the random error that needs to be assigned to these regression estimates.

B. Matching Techniques

For the distance matching methods, let the vector y of monthly income amounts be decomposed as

$$y = (y_r, y_m).$$

The vector y_r represents reported monthly amounts and the vector y_m represents missing monthly amounts. Let the vector x represent monthly income amounts for individuals who report in every month. When matching an incomplete longitudinal record y to the longitudinal records of the completely reporting cases we decompose the vector x into subvectors x_r and x_m, which correspond to the same months as y_r and y_m. The missing monthly amounts for the i^{th} individual y_{m_i} are imputed based on the nearest match of y_{r_i} to some x_r on the set of potential donor records.

The distance matching method is appealing because of its flexibility, generality and ease of implementation. Several distance measures are available for defining the nearest match, including Euclidean and Mahalanobis. The actual amounts imputed may be from the nearest case, or they may be randomly selected from the K nearest. It is expected that the distance matching method can be applied with minor modifications to many income types in SIPP.

VI. SUMMARY

This study examined the problem of imputing missing items in a longitudinal survey. These missing items may occur in single months or in groups of consecutive months. The method discussed takes a longitudinal record with missing items for a specified variable and finds all records complete for this variable that match the designated record on its reported items. Using the complete records, the distribution of responses for the missing months conditioned on the reported months is determined. The missing items are then filled in based on a random draw from the distribution of values in the missing months conditioned on the values in reported months. An alternative

approach to imputation conforming to the overall structure of the basic procedure discussed in this report can be based on a matching process. One selects a complete record whose values agree (or are close to by some specified criterion) with the incomplete record in reported months and imputes the vales found on the reported record into the incomplete record for the missing months. Using either approach it is proved that in the imputed data set the fraction of records having a specified set of values is an unbiased estimate of the probability of that set if the data are missing at random.

This procedure treats only one variable at a time and does not use information from any other variables other than the one being imputed. It is directly applicable to categorical variables and can be adapted for use with continuous variables by selecting a distance function for defining a matchng criterion. Current research is looking into the use of information from other variables in order to improve imputes and maintain cross-sectional consistency of variables upon imputation.

The procedure discussed in this report should be viewed as an underlying methodology which must be expanded upon prior to implementation on any survey. One must draw upon cross-sectional information and subject-matter expertise and blend them with the basic procedure to create a sensitive and practical imputation program.

References

Beale, E.M.L. and Little, R.J.A. (1975). Missing values in multivariate analysis. Journal of the Royal Statistical Society. B37: 129-146.

Buck, S.F. (1960). A method of estimation of missing values in multivariate data suitable for use with an electronic computer. Journal of the Royal Statistical Society. B22: 302-306.

Kalton, G. (1983). Compensating for Missing Survey Data. Survey Research Center, University of Michigan. Ann Arbor.

APPENDIXES

APPENDIX I

We first prove a lemma about conditional expectations of multinomial random variables. This lemma will be used to prove that the basic procedure for imputing longitudinal data introduced in Section III of the text can be used to provide an unbiased estimate of

$$P(X = \underline{a})$$

as asserted earlier.

LEMMA: Let $(X_1,\ldots,X_N)$ be multinomial $(n; P_1,\ldots,P_N)$ random variables, and let α, β, $\gamma \in \{1,\ldots,N\}$ such that α, β and γ are distinct. Then

(1) X_α and X_γ are independent given $X_\alpha + X_\beta = k$, and

$$(2) \quad E\left[\frac{X_\gamma}{X_\alpha + X_\beta}\right] = \frac{nP_\gamma}{P_\alpha + P_\beta}.$$

PROOF: (1) For the sake of simplicity, assume $\alpha = 1$, $\beta = 2$ and $\gamma = 3$.

$$\text{Prob}(X_1 = k_1, X_2 = k_2, X_3 = k_3) =$$

$$\frac{n!}{k_1!\,k_2!\,k_3!\,(n-k_1-k_2-k_3)!}\,P_1^{k_1}P_2^{k_2}P_3^{k_3}(1-P_1-P_2-P_3)^{n-k_1-k_2-k_3}.$$

$$\text{Prob}(X_1 = k_1, X_2 = k_2, X_3 = k_3 \mid X_1 + X_2 = k)$$

$$= \frac{\text{Prob}(X_1 = k_1,\ X_2 = k-k_1,\ X_3 = k_3)}{\text{Prob}(X_1 + X_2 = k)}$$

$$= \frac{\dfrac{n!}{k_1!\,(k-k_1)!\,k_3!\,(n-k-k_3)!}\,P_1^{k_1}P_2^{k-k_1}P_3^{k_3}(1-P_1-P_2-P_3)^{n-k-k_3}}{\dfrac{n!}{k!\,(n-k)!}\,(P_1+P_2)^k\,(1-P_1-P_2)^{n-k}}$$

$$= \left[\frac{k!}{k_1!\,(k-k_1)!} \cdot \frac{(n-k)!}{k_3!\,(n-k-k_3)!} \left(\frac{P_1}{P_1+P_2}\right)^{k_1} \left(\frac{P_2}{P_1+P_2}\right)^{k-k_1} \right] \times$$

$$\left[\left(\frac{P_3}{1-P_1-P_2}\right)^{k_3} \left(\frac{1-P_1-P_2-P_3}{1-P_1-P_2}\right)^{n-k-k_3} \right]$$

$$= \left[\frac{k!}{k_1!\,(k-k_1)!} \left(\frac{P_1}{P_1+P_2}\right)^{k_1} \left(\frac{P_2}{P_1+P_2}\right)^{k-k_1} \right] \times$$

$$\left[\frac{(n-k)!}{k_3!\,(n-k-k_3)!} \left(\frac{P_3}{1-P_1-P_2}\right)^{k_3} \left(\frac{1-P_1-P_2-P_3}{1-P_1-P_2}\right)^{n-k-k_3} \right]$$

Thus, assertion (1) is proved.

(2) It follows from (1) that

$$E_1\left(\frac{X_3}{X_1+X_2} \frac{X_1}{X_2} \mid X_1 + X_2 = k\right) = \frac{(n-k)\,P_1}{(P_1+P_2)} \frac{P_3}{(1-P_1-P_2)}.$$

Taking the expectation with respect to k,

$$E\left(\frac{X_3}{X_1+X_2} \frac{X_1}{X_2}\right) = \frac{(n - n(P_1+P_2))\,P_1P_3}{(P_1+P_2)(1-P_1-P_2)} = \frac{nP_3P_1}{P_1+P_2}.$$

THEOREM: Consider the longitudinal record $(X_1 = a_1, X_2 = a_2,...,X_N = a_N)$, where the a_i are possible valid response values for variable X_i, i=1,...,N. In a simple random sample of size n, completed by imputation, let the multinomial random variable $T'(X_1 = a_1,...,X_N = a_N)$ represent the number of occurrences of the longitudinal record $X = (a_1,...,a_N)$. Assuming the data are missing at random,

$$\frac{1}{n} T'(X_1 = a_1,...,X_N = a_N)$$

is an unbiased estimate of

$$\text{Prob}\,(X_1 = a_1,...,X_N = a_N)$$

PROOF:

For ease of exposition, we assume N=2, a is a value in the range of X_1 and b is a value in the range of X_2. It is not very difficult to see how this argument can extend to an arbitrary integer N. The pattern $(X_1 = a, X_2 = b)$ can arise in the imputed sample in four ways:

1) $(X_1 = a, X_2 = b)$ is reported,
2) $X_1 = a$ is imputed given $X_2 = b$ is reported
3) $X_2 = b$ is imputed given $X_1 = a$ is reported, or
4) $(X_1 = a, X_2 = b)$ is imputed.

Define the multinomial random variable $T(\cdot)$ as the number of occurrences of the event in parentheses. For example, using an asterisk to indicate imputed counts,

$$T^*(X_1 = a \mid X_2 = b)$$

represents the number of times that $X_1 = a$ is imputed given that $X_2 = b$ is reported.

The total number of times the pattern

$$(X_1 = a, X_2 = b)$$

occurs in the sample, completed by imputation, can be decomposed into terms corresponding to the ways the pattern (a, b) arises:

$$T'(X_1 = a, X_2 = b) = T(X_1 = a, X_2 = b) \qquad (A.1)$$

$$+ T^*(X_1 = a \mid X_2 = b) + T^*(X_2 = b \mid X_1 = a) + T^*(X_1 = a, X_2 = b)$$

Let the indicator vector $Y = (Y_1, Y_2)$ represent the reporting status of the elements in the longitudinal record. That is,

$$Y_i = \begin{cases} 1 & \text{if } X_i \text{ is reported } (i=1,2) \\ 0 & \text{otherwise} \end{cases}$$

The expected value of the sum (A.1) with respect to the data reported in the sample is

$$E_1(T'(X_1 = a, X_2 = b) \mid T(X_1 = a, X_2 = b)) = \qquad (A.2)$$

$$T(X_1 = a, X_2 = b, Y_1 = 1, Y_2 = 1) +$$

$$T(X_2 = b, Y_1 = 0, Y_2 = 1) \left(\frac{T(X_1 = a, X_2 = b, Y_1 = 1, Y_2 = 1)}{T(X_2 = b, Y_1 = 1, Y_2 = 1)}\right) +$$

$$T(X_1 = a, Y_1 = 1, Y_2 = 0) \left(\frac{T(X_1 = a, X_2 = b, Y_1 = 1, Y_2 = 1)}{T(X_1 = a, Y_1 = 1, Y_2 = 1)}\right) +$$

$$T(Y_1 = 0, Y_2 = 0) \left(\frac{T(X_1 = a, X_2 = b, Y_1 = 1, Y_2 = 1)}{T(Y_1 = 1, Y_2 = 1)}\right).$$

The expectation of (A.2) with respect to all possible samples follows from the lemma. E.g., for the second term on the right hand side of (A.2) let

$$Z_\alpha = T(X_1 = a, X_2 = b, Y_1 = 1, Y_2 = 1),$$

$$Z_\beta = T(X_1 \neq a, X_2 = b, Y_1 = 1, Y_2 = 1) \text{ and}$$

$$Z_\gamma = T(X_2 = b, Y_1 = 0, Y_2 = 1)$$

be X_α, X_γ and X_β of the lemma, respectively. In addition, the assumption that the data are missing at random (as defined earlier) asserts the independence of the indicator

random vector Y and the random variables X_i in the longitudinal record.

$$E(T'(X_1 = a, X_2 = b)) =$$

$$E_2\, E_1\, (T'(X_1 = a, X_2 = b) \mid T(X_1 = a, X_2 = b)) =$$

$$n\ \text{Prob}\,(X_1 = a, X_2 = b)\ \text{Prob}\,(Y_1 = 1, Y_2 = 1) +$$

$$\left[n\text{Prob}(X_2 = b)\ \text{Prob}(Y_1 = 0, Y_2 = 1)\right] \times$$

$$\left[\frac{\text{Prob}(X_1 = a,\ X_2 = b)\ \text{Prob}(Y_1 = 1,\ Y_2 = 1)}{(\text{Prob}(X_1 = a,\ X_2 = b) + \text{Prob}(X_1 \neq a,\ X_2 = b))\ \text{Prob}(Y_1 = 1,\ Y_2 = 1)}\right]$$

$$+\left[n\ \text{Prob}\,(X_1 = a)\ \text{Prob}(Y_1 = 1, Y_2 = 0)\right] \times$$

$$\left[\frac{\text{Prob}(X_1 = a,\ X_2 = b)\ \text{Prob}(Y_1 = 1,\ Y_2 = 1)}{(\text{Prob}(X_1 = a,\ X_2 = b) + \text{Prob}(X_1 = a,\ X_2 \neq b))\ \text{Prob}(Y_1 = 1,\ Y_2 = 1)}\right]$$

$$+ n\text{Prob}(Y_1 = 0, Y_2 = 0)\ \frac{\text{Prob}(X_1 = a,\ X_2 = b)\ \text{Prob}(Y_1 = 1,\ Y_2 = 1)}{\text{Prob}(Y_1 = 1,\ Y_2 = 1)}$$

$$= n\text{Prob}(X_1 = a, X_2 = b)\ \text{Prob}(Y_1 = 1, Y_2 = 1) + n\text{Prob}(X_1 = a, X_2 = b)\ \text{Prob}(Y_1 = 0, Y_2 = 1)$$

$$+ n\text{Prob}(X_1 = a, X_2 = b)\ \text{Prob}(Y_1 = 1, Y_2 = 0) + n\text{Prob}(X_1 = a, X_2 = b)\ \text{Prob}(Y_1 = 0, Y_2 = 0)$$

$$= n\text{Prob}(X_1 = a, X_2 = b)$$

APPENDIX II

FIGURE AND TABLES

The response rates shown in these tables should not be used to approximate actual nonresponse rates for ISDP. These tables are based on data extract files taken from the 1979 ISDP Panel used to develop and test methodologies discussed in this report. The tables are included to show the diversity of nonresponse patterns and to indicate the variety of nonresponse combinations that must be considered in developing procedures to adjust for nonresponse in a longitudinal survey.

APPENDIX II

FIGURE AND TABLES

The response rates shown in these tables should not be used to approximate actual nonresponse rates for ISDP. These tables are based on data extract files taken from the 1979 ISDP Panel used to develop and test methodologies discussed in this report. The tables are included to show the diversity of nonresponse patterns and to indicate the variety of nonresponse combinations that must be considered in developing procedures to adjust for nonresponse in a longitudinal survey.

Additional tables available from:

DANIEL KASPRZYK
Special Assistant
Office of the Chief
Population Division
Washington, D.C. ,20233

Table 2.A Nonresponse Patterns for Looking/Working January 1979 to December 1979 (11,520 Complete respondents and not applicables not included below.)

Table 2.B Nonresponse Patterns for Looking/Working January 1979 to December 1979 (Patterns with only missing items are included.)

Table 2.C Nonresponse Patterns for Looking/Working January 1979 to December 1979 (Patterns with only missing records are included.)

Table 2.D Nonresponse Patterns for Looking/Working January 1979 to December 1979 (Patterns with only missing items and missing records are included.)

Table 3.A Nonresponse Patterns for Looking/Not Working January 1979 to December 1979 (11,520 Complete respondents and not applicables not included below.)

Table 3.B Nonresponse Patterns for Looking/Not Working January 1979 to December 1979 (Patterns with only missing items are included.)

Table 3.C Nonresponse Patterns for Looking/Not Working January 1979 to December 1979 (Patterns with only missing records are included.)

Table 3.D Nonresponse Patterns for Looking/Not Working January 1979 to December 1979 (Patterns with only missing items and missing records are included.)

Table 4.A Nonresponse Patterns for Want Job January 1979 to December 1979 (Patterns with only missing records are included.)

Table 4.B Nonresponse Patterns for Want Job January 1979 to December 1979 (11,465 Complete respondents and not applicables not included below.)

Table 4.C Nonresponse Patterns for Want Job January 1979 to December 1979 (Patterns with only missing items are included.)

Table 4.D Nonresponse Patterns for Want Job January 1979 to December 1979 (Patterns with only missing items and missing records are included.)

Table 5.A Nonresponse Patterns for Receipt January 1979 to December 1979 (4,080 Complete respondents not included below.)

Table 5.B Nonresponse Patterns for Receipt January 1979 to December 1979 (Patterns with only missing items are included.)

Table 5.C. Nonresponse Patterns for Receipt January 1979 to December 1979 (Patterns with only missing records are included.)

Table 5.D Nonresponse Patterns for Receipt January 1979 to December 1979 (Patterns with only missing items and missing records are included.)

FIGURE 1

ISDP 1979 PANEL

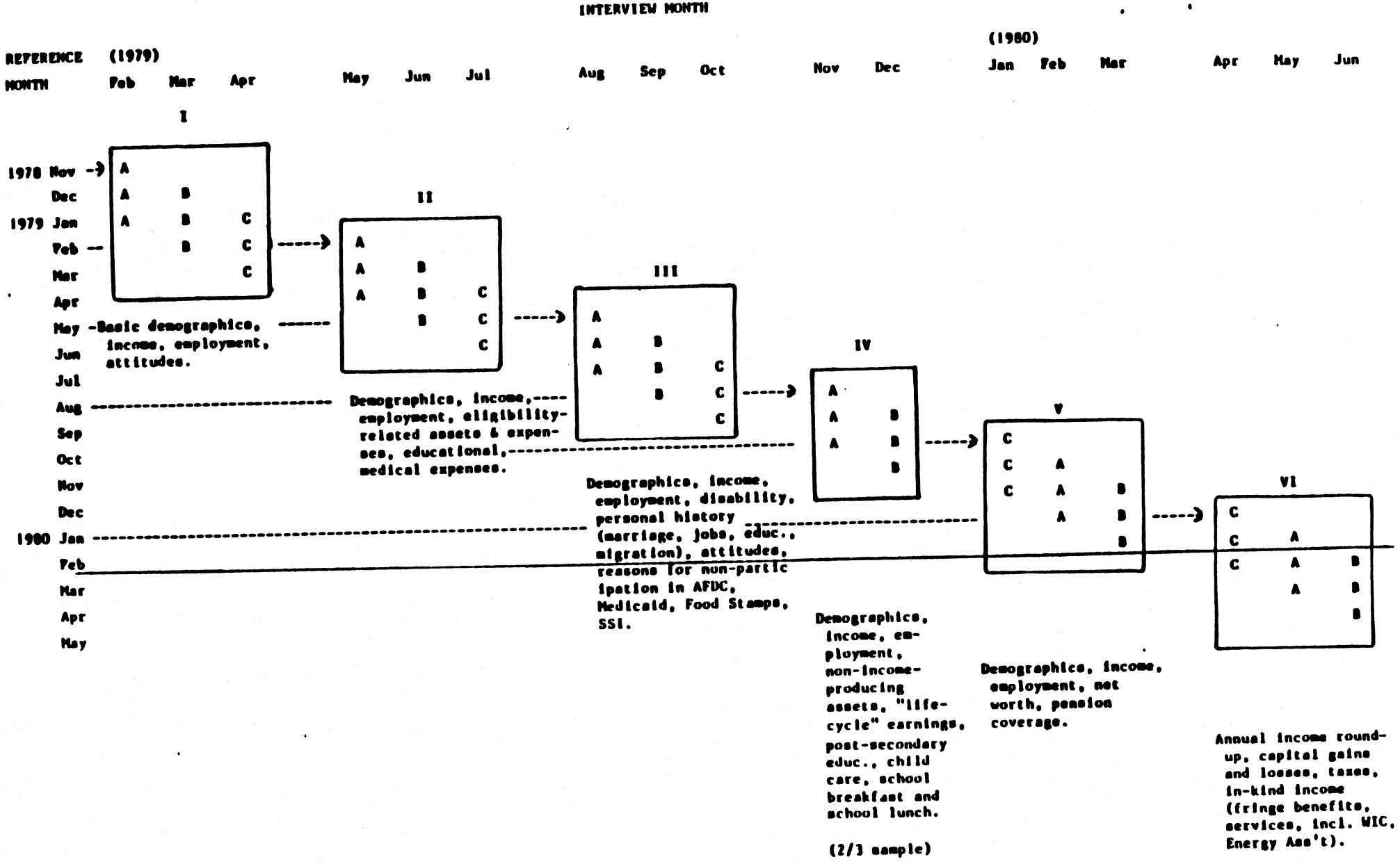

Table 1.A

NONRESPONSE PATTERNS FOR WORK-STAT January 1979 to December 1979
(0=Reported, 1=Missing item, 2=Missing record)
14,181 Complete Respondents Not Included Below

JFMAMJJASOND	Relation	Frequency	% Frequency	Cumulative Frequency
000222222222	3	367	7.44	7.44
022222222222	1	351	7.12	14.56
002222222222	2	311	6.30	20.86
000022222222	1	272	5.51	26.37
000002222222	3	270	5.47	31.85
000000000222	3	228	4.62	36.47
000002222222	2	201	4.07	40.54
000000000002	2	188	3.81	44.35
000000000022	1	186	3.77	48.12
000000022222	1	166	3.37	51.49
000022200000	1	144	2.92	54.41
000000002222	2	136	2.76	57.17
000000222000	3	124	2.51	59.68
002220000000	2	123	2.49	62.17
000222000000	3	115	2.33	64.50
000002220000	2	97	1.97	66.47
000000022200	1	96	1.95	68.42
000000002220	2	82	1.66	70.08
022200000000	1	66	1.34	71.42
000002222220	2	51	1.03	72.45
000000000111	3	46	.93	73.38
011122222222	1	41	.83	74.21
011100000000	1	38	.77	74.98
000111222222	3	32	.65	75.63
001110000000	2	32	.65	76.28
000000000011	1	31	.63	76.91
000022222200	1	30	.61	77.52
111222222222	3	30	.61	78.13
000001110000	2	29	.59	78.71
000000000001	2	28	.57	79.28
001112222222	2	28	.57	79.85
000000111000	3	26	.53	80.38
122222222222	1	26	.53	80.90
000011100000	1	24	.49	81.39
000111000000	3	23	.47	81.86
000222222000	3	23	.47	82.32
112222222222	2	23	.47	82.79
000000001110	2	22	.45	83.24
000022200022	1	22	.45	83.68
022222200000	1	22	.45	84.13
000000111111	3	19	.39	84.51
000000111222	3	19	.39	84.90
000000222111	3	19	.39	85.28
000002220002	2	18	.36	85.65
000011122222	1	18	.36	86.01
000111111111	3	18	.36	86.38
110000000000	2	[illegible]	.36	86.74

0000011112222	2	17	.34	87.09
0002222000222	3	17	.34	87.43
0022222220000	2	17	.34	87.78
1000000000000	1	17	.34	88.12
1110000000000	3	17	.34	88.47
0111111111111	1	16	.32	88.79
0000011111111	2	15	.30	89.09
0000000111100	1	14	.28	89.38
0000000111122	1	13	.26	89.64
0022220002220	2	12	.24	89.88
0022222220002	2	12	.24	90.13
0222222222200	1	12	.24	90.37
0000011111112	2	11	.22	90.59
0001111111222	3	11	.22	90.82
0001111222000	3	11	.22	91.04
0011111111111	2	10	.20	91.24
1111111111111	2	10	.20	91.45
0022220000002	2	9	.18	91.63
0000000011111	2	8	.16	91.79
0000000011112	2	8	.16	91.95
0000111111111	1	8	.16	92.11
0222000222200	1	8	.16	92.28
0000000111111	1	7	.14	92.42
0000111222200	1	7	.14	92.56
0011111112222	2	7	.14	92.70
0022222222220	2	7	.14	92.84
0111000222222	1	7	.14	92.99
0222000222222	1	7	.14	93.13
1000000000022	1	7	.14	93.27
1111111111111	1	7	.14	93.41
1111111222222	3	7	.14	93.55
1111122222222	1	7	.14	93.70
0002220001111	3	6	.12	93.82
0002221112222	3	6	.12	93.94
0011122211111	2	6	.12	94.06
0022220002222	2	6	.12	94.18
0111111222222	1	6	.12	94.30
0111122211111	1	6	.12	94.43
0000011100002	2	5	.10	94.53
0111222222211	1	5	.10	94.63
0222000000022	1	5	.10	94.73
1112220002222	3	5	.10	94.83
0000011122220	2	4	.08	94.91
0000011122221	2	4	.08	94.99
0000111000022	1	4	.08	95.07
0001111000222	3	4	.08	95.16
0011110002222	2	4	.08	95.24
0011111110000	2	4	.08	95.32
0022220001110	2	4	.08	95.40
0022220001112	2	4	.08	95.48
0022211110000	2	4	.08	95.56
0111111000000	1	4	.08	95.64
0111222222200	1	4	.08	95.72
0222000222211	1	4	.08	95.80
0222222000022	1	4	.08	95.88
1111100002221	2	4	.08	95.97
1111111111222	3	4	.08	96.05
1111111122222	1	4	.08	96.13
0000000222211	1	3	.06	96.19
0000011111110	2	3	.06	96.25
0000022221111	2	3	.06	96.31
0000222222211	1	3	.06	96.37
0001111111000	3	3	.06	96.43
0001111222111	3	3	.06	96.49
0002222111000	3	3	.06	96.55
0002222222111	3	3	.06	96.61
0011110002220	2	3	.06	96.68
0011110002221	2	3	.06	96.74
0011111110002	2	3	.06	96.80
0011111111112	2	3	.06	96.86

001111112221	2	3	.06	96.92
001112220000	2	3	.06	96.98
002221112222	2	3	.06	97.04
022211122200	1	3	.06	97.10
100000022200	1	3	.06	97.16
110002220001	2	3	.06	97.22
110002222222	2	3	.06	97.28
111000222222	3	3	.06	97.34
111111111111	3	3	.06	97.41
112220000000	2	3	.06	97.47
000001110001	2	2	.04	97.51
000002220001	2	2	.04	97.55
000011100011	1	2	.04	97.59
000011111122	1	2	.04	97.63
000022211100	1	2	.04	97.67
000022211122	1	2	.04	97.71
001110000001	2	2	.04	97.75
001110000002	2	2	.04	97.79
001110001112	2	2	.04	97.83
001111110001	2	2	.04	97.87
001112220001	2	2	.04	97.91
001112222220	2	2	.04	97.95
002220002221	2	2	.04	97.99
011100000022	1	2	.04	98.03
011111122200	1	2	.04	98.07
022211111111	1	2	.04	98.11
022211122222	1	2	.04	98.16
022222211122	1	2	.04	98.20
100000011100	1	2	.04	98.24
100011122211	1	2	.04	98.28
110002222220	2	2	.04	98.32
111000000222	3	2	.04	98.36
111000111000	3	2	.04	98.40
111100000000	1	2	.04	98.44
111111112222	2	2	.04	98.48
111111222111	3	2	.04	98.52
111222222111	3	2	.04	98.56
122200000022	1	2	.04	98.60
122211111111	1	2	.04	98.64
122222211122	1	2	.04	98.68
000000002221	2	1	.02	98.70
000002221110	2	1	.02	98.72
000002221112	2	1	.02	98.74
000002222221	2	1	.02	98.76
000022200011	1	1	.02	98.78
000022211111	1	1	.02	98.80
000222111111	3	1	.02	98.82
001110001110	2	1	.02	98.84
001110001111	2	1	.02	98.86
001111111110	2	1	.02	98.88
001111112220	2	1	.02	98.91
001112220002	2	1	.02	98.93
002220001111	2	1	.02	98.95
002221110001	2	1	.02	98.97
002221111111	2	1	.02	98.99
002221111112	2	1	.02	99.01
002222221111	2	1	.02	99.03
002222222221	2	1	.02	99.05
011100022200	1	1	.02	99.07
011100022211	1	1	.02	99.09
011111111100	1	1	.02	99.11
011111111122	1	1	.02	99.13
011122200000	1	1	.02	99.15
011122200022	1	1	.02	99.17
022200000011	1	1	.02	99.19
022200011100	1	1	.02	99.21
022211100000	1	1	.02	99.23
022211111122	1	1	.02	99.25

100000011111	1	1	.02	99.31
100000011122	1	1	.02	99.33
100000022222	1	1	.02	99.35
100011100000	1	1	.02	99.37
100011100011	1	1	.02	99.39
100011122200	1	1	.02	99.41
100011122222	1	1	.02	99.43
100022222222	1	1	.02	99.45
110000001112	2	1	.02	99.47
110001111111	2	1	.02	99.49
110001112222	2	1	.02	99.51
110002220000	2	1	.02	99.53
111000222000	3	1	.02	99.55
111110000000	2	1	.02	99.57
111110002222	2	1	.02	99.59
111111100000	1	1	.02	99.61
111111100022	1	1	.02	99.64
111111110002	2	1	.02	99.66
111111111112	2	1	.02	99.68
111111112221	2	1	.02	99.70
111111122211	1	1	.02	99.72
111111222000	3	1	.02	99.74
111112221110	2	1	.02	99.76
111112221112	2	1	.02	99.78
111112222222	2	1	.02	99.80
111122222200	1	1	.02	99.82
111222000000	3	1	.02	99.84
111222111111	3	1	.02	99.86
111222222000	3	1	.02	99.88
112220000002	2	1	.02	99.90
112222221112	2	1	.02	99.92
122200000000	1	1	.02	99.94
122211122222	1	1	.02	99.96
122222200022	1	1	.02	99.98
122222222211	1	1	.02	100.00
TOTALS=		4933	100.00	

Members of the original sample sixteen and older.

Table 1.C

NONRESPONSE PATTERNS FOR WORKSTAT January 1979 to December 1979
(0=Reported, 1=Missing item, 2=Missing record)
Patterns With Only Missing Items Are Included

JFMAMJJASOND	Rotation	Frequency	% Frequency	Cumulative Frequency
000000000111	3	46	8.78	8.78
011100000000	1	38	7.25	16.03
001110000000	2	32	6.11	22.14
000000000011	1	31	5.92	28.05
000001110000	2	29	5.53	33.59
000000000001	2	28	5.34	38.93
000000111000	3	26	4.96	43.89
000011100000	1	24	4.58	48.47
000111000000	1	23	4.39	52.86
000000001110	2	22	4.20	57.06
000000111111	3	19	3.63	60.69
000111111111	3	18	3.44	64.12
110000000000	2	18	3.44	67.56
100000000000	1	17	3.24	70.80
111000000000	3	17	3.24	74.05
011111111111	1	16	3.05	77.10
000001111111	2	15	2.86	79.96
000000011100	1	14	2.67	82.63
001111111111	2	10	1.91	84.54
111111111111	2	10	1.91	86.45
000000001111	2	8	1.53	87.98
000011111111	1	8	1.53	89.50
000000011111	1	7	1.34	90.84
111111111111	1	7	1.34	92.18
001111110000	2	4	.76	92.94
011111100000	1	4	.76	93.70
000001111110	2	3	.57	94.27
000111111000	3	3	.57	94.85
111111111111	3	3	.57	95.42
000001110001	2	2	.38	95.80
000011100011	1	2	.38	96.18
001110000001	2	2	.38	96.56
001111110001	2	2	.38	96.95
100000011100	1	2	.38	97.33
111000111000	3	2	.38	97.71
111100000000	1	2	.38	98.09
001110001110	2	1	.19	98.28
001110001111	2	1	.19	98.47
001111111110	2	1	.19	98.66
011111111100	1	1	.19	98.85
100000011111	1	1	.19	99.05
100011100000	1	1	.19	99.24
100011100011	1	1	.19	99.43
110001111111	2	1	.19	99.62
111110000000	1	1	.19	99.81
111111100000	1	1	.19	100.00
TOTALS=		524	100.00	

Members of the original sample sixteen and older.

Table 1.B

NONRESPONSE PATTERNS FOR WORKSTAT January 1979 to December 1979
(0=Reported, 1=Missing item, 2=Missing record)
Patterns With Only Missing Records Are Included

JFMAMJJASOND	Rotation	Frequency	% Frequency	Cumulative Frequency
000222222222	3	367	9.65	9.65
022222222222	1	351	9.22	18.87
002222222222	2	311	8.17	27.04
000022222222	1	272	7.15	34.19
000002222222	3	270	7.10	41.29
000000000222	3	228	5.99	47.28
000002222222	2	201	5.28	52.56
000000000002	2	188	4.94	57.50
000000000022	1	186	4.89	62.39
000000022222	1	166	4.36	66.75
000022200000	1	144	3.78	70.54
000000002222	2	136	3.57	74.11
000000222000	3	124	3.26	77.37
002220000000	2	123	3.23	80.60
000222000000	3	115	3.02	83.63
000002220000	2	97	2.55	86.18
000000022200	1	96	2.52	88.70
000000002220	2	82	2.16	90.85
022200000000	1	66	1.73	92.59
000002222220	2	51	1.34	93.93
000022222200	1	30	.79	94.72
000222222000	3	23	.60	95.32
000022200022	1	22	.58	95.90
022222200000	1	22	.58	96.48
000002220002	2	18	.47	96.95
000222000222	3	17	.45	97.40
002222220000	2	17	.45	97.84
002220002220	2	12	.32	98.16
002222220002	2	12	.32	98.48
022222222200	1	12	.32	98.79
002220000002	2	9	.24	99.03
022200022200	1	8	.21	99.24
002222222220	2	7	.18	99.42
022200022222	1	7	.18	99.61
002220002222	2	6	.16	99.76
022200000022	1	5	.13	99.89
022222200022	1	4	.11	100.00
TOTALS*		3805	100.00	

*Members of the original sample sixteen and older.

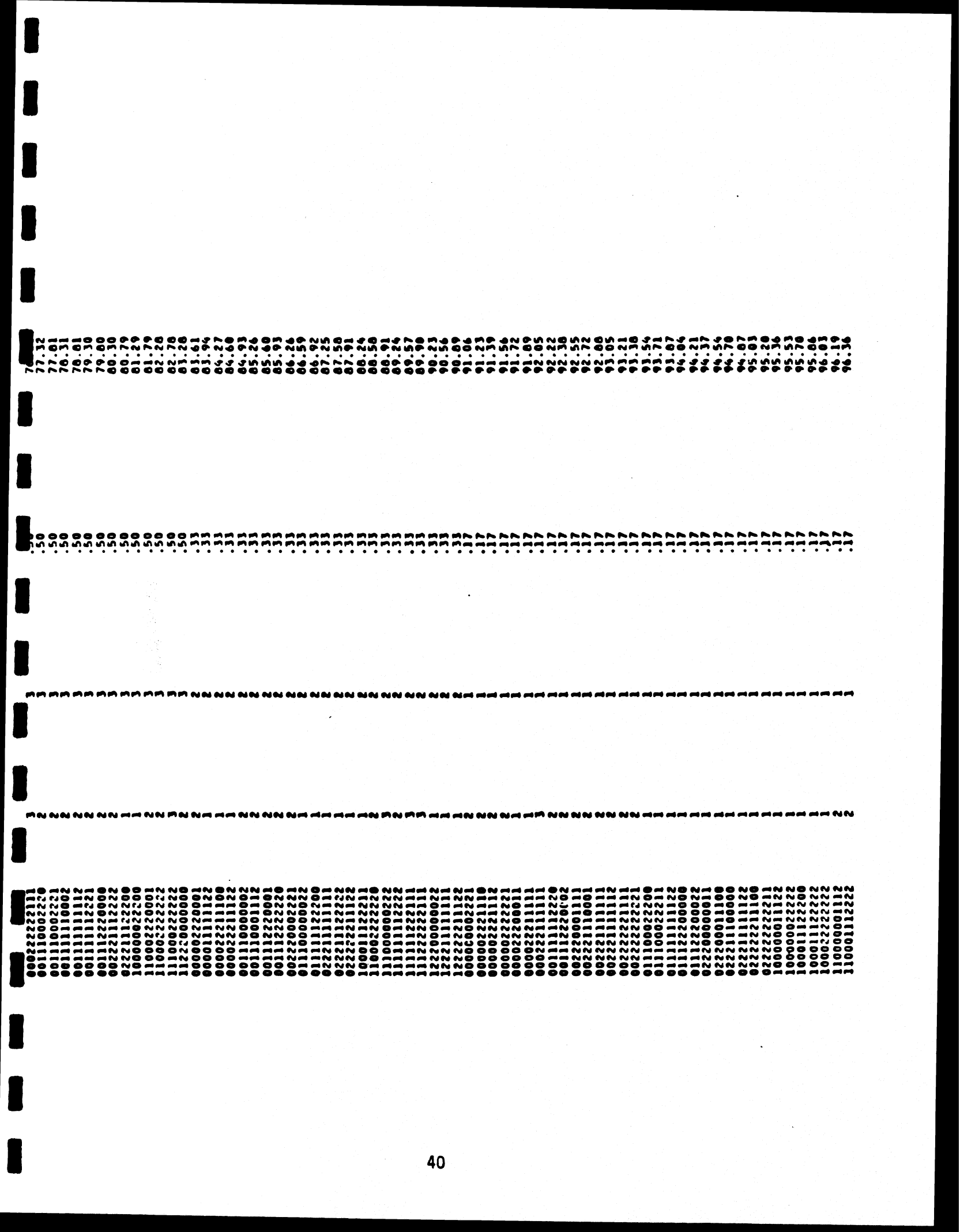

Table 1.D

NONRESPONSE PATTERNS FOR WORKSTAT January 1979 to December 1979
(0=Reported, 1=Missing item, 2=Missing record)
Patterns With Only Missing Items And Missing Records Are Included

JFMAMJJASOND	Rotation	Frequency	X Frequency	Cumulative Frequency
011122222222	1	41	6.79	6.79
000111222222	3	32	5.30	12.09
111222222222	3	30	4.97	17.05
001112222222	2	28	4.64	21.69
122222222222	1	26	4.30	25.99
112222222222	2	23	3.81	29.80
000000111222	3	19	3.15	32.95
000000222111	3	19	3.15	36.09
000011122222	1	18	2.98	39.07
000001112222	2	17	2.81	41.89
000000011122	1	13	2.15	44.04
000001111112	2	11	1.82	45.86
000111111222	3	11	1.82	47.68
000111222000	3	11	1.82	49.50
000000001112	2	8	1.32	50.83
000011122200	1	7	1.16	51.99
001111112222	2	7	1.16	53.15
011100022222	1	7	1.16	54.30
100000000022	1	7	1.16	55.46
111111222222	3	7	1.16	56.62
111122222222	1	7	1.16	57.78
000222000111	3	6	.99	58.77
000222111222	3	6	.99	59.77
001112221111	2	6	.99	60.76
011111122222	1	6	.99	61.75
011122211111	1	6	.99	62.75
000001110002	2	5	.83	63.58
011122222211	1	5	.83	64.40
111222000222	3	5	.83	65.23
000001112220	2	4	.66	65.89
000001112221	2	4	.66	66.56
000011100022	1	4	.66	67.22
000111000222	3	4	.66	67.88
001110002222	2	4	.66	68.54
002220001110	2	4	.66	69.21
002220001112	2	4	.66	69.87
002221110000	2	4	.66	70.53
011122222200	1	4	.66	71.19
022200022211	1	4	.66	71.85
111110002221	2	4	.66	72.52
111111111222	3	4	.66	73.18
111111122222	1	4	.66	73.84
000000022211	1	3	.50	74.34
000002221111	2	3	.50	74.83
000022222211	1	3	.50	75.33
000111222111	3	3	.50	75.83
000222111000	3	3	.50	76.32

1100022220000	2	1	.17	96.52
1110002220000	3	1	.17	96.69
1111110002222	2	1	.17	96.85
1111111100022	1	1	.17	97.02
1111111110002	2	1	.17	97.19
1111111111112	2	1	.17	97.35
1111111112221	2	1	.17	97.52
1111111122211	1	1	.17	97.68
1111111222000	3	1	.17	97.85
1111112221110	2	1	.17	98.01
1111112221112	2	1	.17	98.18
1111112222222	2	1	.17	98.34
1111122222200	1	1	.17	98.51
1111222000000	3	1	.17	98.68
1111222111111	3	1	.17	98.84
1111222222000	3	1	.17	99.01
1112220000002	2	1	.17	99.17
1112222221112	2	1	.17	99.34
1122200000000	1	1	.17	99.50
1122211122222	1	1	.17	99.67
1122222200022	1	1	.17	99.83
1122222222211	1	1	.17	100.00
TOTALS=		604	100.00	

Members of the original sample sixteen and older.

Table 6

SELECTED INCOMPLETE AND DONOR PATTERNS FOR WORK-STAT
January 1979 - December 1980
(0 = yes, 1 = no, 2 = missing item)

Rotation	Incomplete Pattern	Frequency	Donor Patterns for Missing Months	Frequency (%)	Expected Numbe of Incorrect Imputations
B	000 000 000 002	15	0 1	2313 (97.23) 66 (2.77)	(.0539)
A	000 000 000 022	9	11 00	105 (4.43) 2264 (95.57)	(.0847)
C	000 000 000 222	23	111 000	121 (4.83) 2386 (95.17)	(.0920)
B	000 000 002 220	8	111 000	38 (1.62) 23.13(96.38)	(.0319)
B	000 000 002 221	2	111 000	83 (55.7) 66 (44.3)	(.4935)
B	000 000 002 222	5	1111 1110 0001 0000	83 (3.32) 38 (1.52) 66 (2.64) 2313 (92.52)	(.1420)
A	100 000 000 022	3	11 00	12 (12.12) 87 (87.88)	(.2130)
A	100 022 200 011	1	111 000	4 (25.0) 87 (75.0)	(.3750)
C	111 000 000 222	4	111 000	63 (31.82) 135 (68.18)	(.4339)
B	111 111 111 112	8	1 0	1494 (97.39) 40 (2.61)	(.0508)
B	111 112 221 111	12	111 000	1494 (96.08) 61 (3.92)	(.0753)

Table 7
SELECTED INCOMPLETE AND DONOR PATTERNS FOR LOOKING/WORKING
January 1979 - December 1980
(0 = yes, 1 - no, 2 = missing item, 4 = not applicable)

Rotation	Incomplete Pattern	Frequency	Donor Patterns for Missing Months	Frequency (%)	Percent Incorrect Imputations
B	000 000 001 112	1	1	4 (80.0)	(.3200)
			0	1 (20.0)	
C	000 111 111 222	3	111	84 (92.31)	(.1420)
			000	7 (7.69)	
C	000 222 222 222	4	111 111 111	84 (60.43)	(.6088)
			111 111 000	7 (5.04)	
			111 000 111	6 (4.32)	
			111 000 000	4 (2.88)	
			000 111 111	13 (9.35)	
			000 111 000	8 (5.76)	
			000 000 111	12 (8.63)	
			000 000 000	5 (3.60)	
A	022 211 111 111	3	111	58 (71.60)	(.4067)
			000	23 (28.40)	
B	110 000 002 222	2	1111	10 (83.33)	(.2778)
			1110	2 (16.67)	
B	111 111 111 112	9	1	2031 (98.12)	(.1369)
			0	39 (1.88)	
A	111 111 111 122	8	11	1982 (98.66)	(.0264)
			00	27 (1.34)	
A	111 111 122 233	2	111	63 (100.0)	(0.0)
B	111 112 222 222	9	111 111	2031 (96.16)	(.0748)
			111 111 0	39 (1.85)	
			111 000 1	26 (1.23)	
			111 000 0	8 (0.38)	
			000 111 0	1 (0.05)	
			000 000 1	6 (0.28)	
			000 000 0	1 (0.05)	

Table 8
SELECTED INCOMPLETE AND DONOR PATTERNS FOR LOOKING/NOT WORKING
January 1979 - December 1979
(0 = yes, 1 = no, 2 = missing item, 4 = not applicable)

Rotation	Incomplete Pattern	Frequency	Donor Patterns for Missing Months	Frequency (%)	Percent Incorrect Imputations
B	110 002 221 111	2	111 000	12 (63.16) 7 (36.84)	(.4654)
C	111 111 222 111	5	111 000	1449 (99.18) 12 (0.82)	(.0163)
B	111 112 222 222	5	1 111 111 1 111 110 1 110 001 1 110 000 0 001 111 0 001 110 0 000 001 0 000 000	1318 (95.58) 17 (1.23) 5 (0.36) 3 (0.22) 18 (1.31) 4 (0.29) 5 (0.36) 9 (0.65)	(.0860)
A	111 122 211 111	5	000 111	14 (1.05) 1318 (98.95)	(.0208)
B	112 221 111 111	11	111 000	1318 (99.1) 12 (0.90)	(.0096)
A	122 244 444 444	2	111 000	45 (73.77) 16 (26.23)	(.3870)
B	221 111 111 111	7	11 00	1318 (97.99) 27 (2.01)	(.0594)
C	222 444 444 444	11	111 000	93 (68.89) 42 (31.11)	(.4285)

Table 9
SELECTED INCOMPLETE AND DONOR PATTERNS FOR WANT JOB
January 1979 - December 1979
(0 = yes, 1 = no, 2 = missing item)

Rotation	Incomplete Pattern	Frequency	Donor Patterns for Missing Months	Frequency (%)	Percent Incorrect Imputations
A	000 000 022 211	2	111 000	1 (33.33) 2 (66.67)	(.4444)
B	002 220 001 111	2	111 000	8 (66.67) 4 (33.33)	(.4444)
A	011 111 111 122	2	11 00	56 (94.92) 3 (5.08)	(.0964)
A	100 022 211 111	3	111 000	24 (80.00) 6 (20.00)	(.3200)
C	111 000 222 111	2	111 000	24 (77.42) 7 (22.58)	(.3496)
A	111 111 100 022	2	11 00	23 (82.14) 5 (17.86)	(.2934)
B	111 111 111 112	8	1 0	1086 (98.91) 12 (1.09)	(.0216)
A	111 111 111 122	12	11 00	1067 (99.26) 8 (0.74)	(.0147)
A	111 122 222 222	4	11 111 111 11 111 100 11 100 011 11 100 000 00 011 111 00 011 100 00 000 011	1067 (94.3) 8 (0.71) 23 (2.04) 5 (0.44) 17 (1.51) 1 (0.09) 3 (0.27)	(.1100)

Table 10
SELECTED INCOMPLETE AND DONOR PATTERNS FOR
RECEIPT OF WAGES AND SALARIES
January 1979 - December 1980
(0 = yes, 1 = no, 2 = missing item)

Rotation	Incomplete Pattern	Frequency	Donor Patterns for Missing Months	Frequency (%)	Incorrect Imputations
B	000 000 000 002	17	1	8 (0.65)	(.0129)
			0	1228 (99.35)	
A	000 000 000 022	26	11	3 (0.26)	(.0275)
			01	18 (1.58)	
			00	1120 (98.6)	
B	000 000 000 102	1	0	2 (100.0)	(0.0)
A	000 000 000 122	1	11	1 (100.0)	(0.0)
C	000 000 000 222	30	111	5 (0.39)	(.0930)
			110	2 (0.16)	
			101	1 (0.08)	
			100	4 (0.31)	
			011	21 (1.63)	
			010	4 (0.31)	
			001	26 (2.02)	
			000	1225 (95.2)	
C	000 000 001 222	1	101	1 (9.09)	(.5454)
			100	2 (18.18)	
			001	1 (9.09)	
			000	8 (72.73)	
B	000 000 002 220	29	111	2 (0.16)	(.0272)
			100	2 (0.16)	
			011	6 (0.48)	
			010	2 (0.16)	
			001	6 (0.48)	
			000	1228 (98.63)	
A	000 000 022 200	38	110	3 (0.26)	(.0330)
			101	1 (0.09)	
			100	10 (0.88)	
			010	2 (0.18)	
			001	4 (0.35)	
			000	1120 (98.33)	