All surveys contain incomplete data, even if measures are implemented in order to avoid them before and during the data collection. The problems with nonresponse are that estimators are less accurate (the sample size is lower) and that it creates bias (generally, respondents behave differently than nonrespondents). Different types of nonresponse may occur:

- unit nonresponse refers to the failure of a unit in the sample frame to participate in the survey (noncoverage, not at home, unable to answer, refusal...),
- item nonresponse occurs when most of the questions for a unit are answered, but no answer is given for some questions, or the answer is clearly wrong and must be deleted; or due to under-reporting (questions that are sensitive or for which the respondent does not have the information, boring document to fill out...).

As it appears in the definition, the difference between unit and item nonresponse is that we have less information about a unit nonrespondent than for an item nonrespondent. Indeed, for a selected unit, the only information that we can get without error is geographical, and even if the sample is picked from the census, the figures are not necessarily up to date. We know much more about the units that do not reply to some questions (from the responses given to the other questions). This paper will focus on the correction of item nonresponse. First, we will give a general presentation of the theory on this subject, then we will talk about the existing solution and illustrate our presentation with examples drawn from the last two National Personal Transportation Surveys conducted in France.

Even if nonresponse is inevitable, we should take measures to avoid it. For example, in a mobility survey, is it necessary to ask for trip distance that people do not always know or poorly estimate? Or is it possible to calculate it from origin and destination, on which there are almost no missing data (Flavigny and Madre, 1994)? Missing trips are often due to memory effects or to a too long period under review to be reported in boring documents. Memory joggers seem to be good instruments to avoid memory effects. Anyway, it seems that most people remember well their daily mobility only for the day before the interview, and their long-distance trips for about one month (Armoogum and Madre, 1996).
RE-WEIGHTING OR IMPUTING DATA?

In practice, it is easier to compute results from a survey on a “clean data matrix” (i.e., without missing data). Indeed, you will obtain the same total for any variable, whatever other variable you wish to cross with it. It is also much easier to have only one weight set for each level (household, person, trip, etc.), which can be used for any variable describing this unit.

There are two main methods to correct nonresponse: one involves re-weighting the respondent sample, and the second consists of implementing imputation procedures. Let us suppose that we are measuring two interest variables, Y1 and Y2. If the unit k does not respond to Y1 and Y2 simultaneously, it is common to give the unit k the weight 0, and to increase the weight of the respondents (by determining response mechanism, which is a long procedure). If the unit k responds to Y1 and not to Y2, it is usual to replace the missing value by one (or many) plausible value(s). This process is commonly called “imputation”. We will consider when re-weighting is a reasonable solution, and when it is better to implement imputation procedures.

If the set of missing data for Y1 and Y2 is not the same as for Y1 and Y3, you will have to run two different re-weighting procedures, and their result will probably not be identical for the total of Y1. Thus, re-weighting has to be implemented only when a unit is totally missing (for instance, in case of missing trips), or when almost all important variables describing a unit are missing. For instance, in our NPTS, we skipped all car diaries for which surveyors suspected missing trips (about 5%), or in which there was at least one trip with not enough information to implement the imputation procedure (less than 1% of diaries). Thus, we obtained a data file containing more than 94% of the diaries collected, with a single weight set and with no missing data left for the main variables (trip distance, etc.).

Re-weighting Procedures for an Omitted Part

In the last French NPTS, a questionnaire was handed to the B individual, who was asked to describe his trips of more than 100 km, and to send the questionnaire back three months after the interviewer’s visit. But only 60% of the questionnaires were returned. The problem is that we do not know why some of them were not returned. Is this because the B individual was too busy to fill out the form (for example, too much travel), or because he did not travel at all and assumed the questionnaire would be of no importance? It’s usual to classify this type of nonresponse as non-ignorable nonresponse, since the probability of responding depends on the response itself.

We are interested by the qualitative variable Y, with I modes. The sample is divided into H groups h=1, H (very separative, if possible, regarding the distribution in I categories). Besides, we suppose H to be greater than or equal to I. For example, Y could be the number of trips (0, 1, 2, 3 or more trips), and the groups are regions or zones of residence.

Let us postulate a response model such that only i influences the response, generating a response with a $P_i$ probability and a nonresponse with a $P = 1-P_i$ probability. We intend to estimate $P_i$. 
Thus, Y is a qualitative variable with I modes $y_k = 1, \ldots, i, \ldots, I$.

$R_k = 1$ if the individual $k$ responds, otherwise, $R_k = 0$.

Therefore, we can classify the observations into $H$ groups, and thus:

$$ n_h = \sum_{i=1}^{I} n_{h,i} + n_{h,i} \text{, within each group,} $$

where:

- $y_k = 1$ for $n_{h,1}$ units
- $y_k = i$ for $n_{h,i}$ units
- $y_k = I$ for $n_{h,I}$ units
- $y_k = ?$ for $n_{h,?}$ units

Let us pose the following model:

- Probability ($R_k = 1 \mid y_k = i$) = $P_i$
- Probability ($y_k = i \mid k \in (h)$) = $m_{h,i}$

Conditionally, to the sample that gives the size of the known $n_h$, the exact number $n_{h,i}^*$ ($i = 1$ to $I$ and $i = ?$) is one of the model’s parameters. But we will prefer to take $m_{h,i} = \frac{n_{h,i}^*}{n_h}$ as parameters. Note that we have the relation $\sum_{i=1}^{I} m_{h,i} = 1$ in each group. The $n_{hi}$ follows an L multinomial distribution ($n_h ; m_{hi} ; \sum_{i=1}^{I} m_{h,i} \bar{P}_i$), and we can write the likelihood logarithm as:

$$ \log(L) = \sum_{h=1}^{H} \left( \sum_{i=1}^{I} n_{h,i} \log(m_{h,i} \bar{P}_i) + n_{h,?} \log \left( \sum_{i=1}^{I} m_{h,i} \bar{P}_i \right) \right) $$

With H Lagrange multiplier $L_h$ associated with the H constraints ($\sum_{i=1}^{I} m_{h,i} = 1$), the maximum likelihood equations are as follows:

For $h = 1$ to $H$, $i = 1$ to $I$,

$$ \frac{\partial \log(L)}{\partial m_{h,i}} = 0 \iff \frac{n_{h,i}}{m_{h,i}} + n_{h,?} \frac{\bar{P}_i}{\sum_{i=1}^{I} m_{h,i} \bar{P}_i} + \lambda_h = 0 \tag{1} $$
For \( i = 1 \) to \( I \),

\[
\frac{\partial \log(L)}{\partial P_i} = 0 \iff \sum_{h=1}^{H} n_{h,i} \frac{P_i}{P_i} - \sum_{l=1}^{H} n_{h,l} \frac{m_{h,l}}{\sum_{i=l}^{I} m_{h,i} P_i} = 0
\]  

(2)

We can write the first equations at fixed \( h \):

\[
n_{h,i} + n_{h,l} \frac{m_{h,i} P_i}{\sum_{i=l}^{I} m_{h,i} P_i} + \lambda_h m_{h,i} = 0
\]

Then, if we add for \( i = 1 \) to \( I \), we find that \( h = -nh \)

The equations become:

For \( h = 1 \) to \( H \), \( i = 1 \) to \( I \),

\[
m_{h,i} = \frac{n_{h,i}}{n_h - \frac{n_{h,l} P_i}{\sum_{i=l}^{I} m_{h,i} P_i}}
\]  

(3)

For \( i = 1 \) to \( I \);

\[
P_i = \frac{\sum_{h=1}^{H} n_{h,i}}{\sum_{h=1}^{H} \sum_{l=1}^{I} m_{h,l} P_i}
\]  

(4)

We estimate the \( P_i \) with an algorithm. In the first stage of this algorithm, we consider that response is missing only at random, so we initiate the processes by

\[
m_{h,i} = \frac{n_{h,i}}{n_h - n_{h,l}}
\]
Then we calculate the \( P_i = \frac{\sum_{h=1}^{H} n_{h,i}}{\sum_{h=1}^{H} n_h m_{h,i}} \), and estimate \( Q_h = \frac{n_{h,i}}{\sum_{i=1}^{l} (m_{h,i} \bar{p}_i)} \).

The other stages of the processes consist in estimating the \( m_{h,i} = \frac{n_{h,i}}{n_h - Q_h \bar{p}_i} \), making the \( m_{h,i} \) keep the relation for all \( h: \sum_{i=1}^{l} m_{h,i} = 1 \). We then calculate the \( P_i = \frac{\sum_{h=1}^{H} n_{h,i}}{\sum_{h=1}^{H} Q_h m_{h,i}} \), estimate the quantity \( Q_h = \frac{n_{h,i}}{\sum_{i=1}^{l} (m_{h,i} \bar{p}_i)} \), and finally make \( Q_h = \frac{Q_{h,i}}{\sum_{h=1}^{H} Q_h} \).

The algorithm stops when the differences between 2 successive \( P_i \) are less than

We can approach the solution by two other methods. If we suppose that \( \frac{n_{h,i}}{\sum_{i=1}^{l} m_{h,i} \bar{p}_i} = n_h \), it results that (3) \( \Rightarrow \) (4) and, above all, \( m_{h,i} = \frac{n_{h,i}}{n_h \bar{p}_i} \).

Let \( V_i \) be \( V_i = \frac{n_{h,i}}{n_h} \) so for \( h = 1 \) to \( H \),

\[
m_{h,i} = \frac{V_i}{\bar{p}_i}
\]

(5)

From the equations (5), we propose two other methods to estimate the \( P_i \):

1. Econometric method. We transform the equations (5) in
\[ \sum_{i=1}^{l} \frac{1}{P_i} V_i = 1 + \varepsilon_h \]

because \( \sum_{i=1}^{l} m_{h,i} = 1 \)

where the \( h \) follow a normal distribution with a small variance, and 0 as mean. In the equations of (5), only the \( P_i \) and the \( h \) are unknown, so in doing a regression, we can directly estimate the \( P_i \).

2. b) Data analysis method. First of all, we have to do a principal component analysis, with all the \( V_i \) (\( i=1,...,I \)), \( V_i \) defined as above. Then, we calculate the component of the hyper plan that dimension is equal to \( I-1 \), passing by the central point of gravity of all points. The \( P_i \) are given by the intersection of this hyper plan with the variables \( V_i \).

The assessments of all simulations show that the 3 methods give similar results: a good estimation of the \( P_i \).

**APPLICATION TO THE FRENCH NPTS, 1993-94**

Let us suppose, for the purpose of the study, that we want to know if the \( k \) individual of our sample had traveled or not during the three months following the interviewer’s visit. We have three possible cases for the 14,068 individuals in the sample: no travel, at least one trip, and did not return the questionnaire.

We have some other information concerning these individuals (from the two interviews), in particular, the number of cars in each one’s household (0, 1, 2, or more than 2 cars), and also the number of trips they made during the three months preceding the second interview (0, 1, 2, 3, or more than 3 journeys). By crossing these two variables, we obtain 16 groups. Let us see how people in our sample filled the self-administered questionnaire (Table 1).

Table 1 shows us that the two variables we have used to build the groups are strongly related to the act of sending back the questionnaire; actually, the more a household is equipped with a motor vehicle (likewise, the more trips the individual has made during the three months before the interview), the more the individual is likely to return the duly completed questionnaire. The results of the algorithm are that the probability for a person who had traveled to send back his questionnaire is 0.77, and for those who did not travel at all during this period, 0.47. Thus, if we want to establish an O-D matrix, the loss of information is less important than indicated by nonresponse rates.

We have shown that the results do not vary significantly for the three procedures of estimation tested for the program (algorithmic, main component analysis and econometrics). However, we can wonder to what extent the results depend on the choice made “a priori” (after several attempts) for \( H \) groups. Are there automatic methods (for instance, segmentation) useful to determine more rapidly “optimal” groups? Could we imagine solutions avoiding this “grouping step”, which could seem somewhat arbitrary?
TABLE 1 Number of Self-Administered Questionnaires by Car Ownership and Mobility Groups (Row Percentage)

<table>
<thead>
<tr>
<th>Number of cars in the households</th>
<th>Number of journeys (first interview)</th>
<th>No trip</th>
<th>At least one journey</th>
<th>Questionnaire not returned</th>
<th>Total (Nh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>34.4%</td>
<td>8.3%</td>
<td>57.4%</td>
<td>1733</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>30.3%</td>
<td>20.9%</td>
<td>48.8%</td>
<td>492</td>
</tr>
<tr>
<td>0</td>
<td>2 or 3</td>
<td>14.1%</td>
<td>42.3%</td>
<td>43.6%</td>
<td>227</td>
</tr>
<tr>
<td>0</td>
<td>over 3</td>
<td>5.4%</td>
<td>59.8%</td>
<td>34.8%</td>
<td>112</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>38.1%</td>
<td>19.2%</td>
<td>42.7%</td>
<td>2751</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>29.2%</td>
<td>35.3%</td>
<td>35.5%</td>
<td>1615</td>
</tr>
<tr>
<td>1</td>
<td>2 or 3</td>
<td>17.5%</td>
<td>51.5%</td>
<td>31.1%</td>
<td>1156</td>
</tr>
<tr>
<td>1</td>
<td>over 3</td>
<td>7.7%</td>
<td>56.7%</td>
<td>35.6%</td>
<td>730</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>36.0%</td>
<td>25.2%</td>
<td>38.8%</td>
<td>1408</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>25.9%</td>
<td>39.8%</td>
<td>34.3%</td>
<td>1224</td>
</tr>
<tr>
<td>2</td>
<td>2 or 3</td>
<td>17.4%</td>
<td>49.2%</td>
<td>33.5%</td>
<td>1100</td>
</tr>
<tr>
<td>2</td>
<td>over 3</td>
<td>8.6%</td>
<td>58.8%</td>
<td>32.6%</td>
<td>743</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>33.1%</td>
<td>24.6%</td>
<td>42.4%</td>
<td>236</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>23.8%</td>
<td>39.4%</td>
<td>36.8%</td>
<td>193</td>
</tr>
<tr>
<td>3</td>
<td>2 or 3</td>
<td>13.6%</td>
<td>49.8%</td>
<td>36.7%</td>
<td>199</td>
</tr>
<tr>
<td>3</td>
<td>over 3</td>
<td>8.1%</td>
<td>60.4%</td>
<td>31.5%</td>
<td>149</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>27.0%</td>
<td>33.1%</td>
<td>39.9%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Source: INSEE-INRETS 1993-94 French NPTS.

HOW TO ESTIMATE UNDER-REPORTING DUE TO MEMORY EFFECT

Correction of Memory Effects with an Underreport Estimate Function

Whenever we want to measure a rare event (long-distance trips, for example) with precision in a retrospective survey, an easy method consists of enlarging the period under review. But in doing so, we may face memory effects. For example, in the last French NPTS, persons were asked to list all the journeys they had made during the three past months, but in order to avoid nonresponse high-mobility persons (who had made more than 6 trips during the last month), they had to describe in detail only the trips they had made during the last month). Since seasonal effects have been avoided by conducting the survey throughout the year, the average distribution of journey has to be uniform during the 13 weeks (3 months) of the survey period. Figure 1 shows that it is true for those high-mobility persons. But for those with a lower mobility, three phenomena interfere:

- There are few trips made during the week just before the interview because of the constraint of presence at home (the long-distance interview takes place during the second visit of the interviewer); for the same reason, few journeys reported in the self-administered questionnaire filled after the interview end during its first weeks.
- The omission of trips due to memory effects is more and more obvious as we refer to earlier weeks.
• For the first week in the scope, there is an “edge effect”: people hesitating about the exact date of a journey situate it at the beginning of the period under review.

Between the two visits of the interviewer, the interviewee had to fill a memory jogger to help the recall of trips. The experience of our colleagues from Norway tends to show that it has been useful, since we have collected 29% journeys during the first month, 34% during the second month, and 37% during the third, while those figures are 15%, 35% and 50% for the survey conducted in Norway (MEST, 1996a).

N.B.: Week i means the i-th week before the interview.

FIGURE 1  Distribution of journeys during the period under review for high- and low-mobility persons.

In order to correct the memory effects for low-mobility persons, we have to avoid the problem of presence at home, so we took into account only the persons who did not change their appointment with the interviewer to build the memory effects model. As we can see in Figure 2, the distribution of trips for those people decreases from the first week. We classify the trips into 3 homogeneous categories, shown on Figure 2: private trips of less than 500 km, private trips of more than 500 km and business trips. It seems that people forget their business trips more often than their short private trips and people remember more their long private trips than their short private trips. In each category, we estimate the number of trips as a function of the number of weeks between those trips and the interview. Empirically, an exponential function gives a more stable result; that is why we choose this function to adjust the data.

Besides this test of different functions to adjust memory effects, we could also wonder if the three categories of trips (according to their frequency, their purpose and their length) that we have considered are the best ones in order to estimate memory effects models. It is also interesting to take into account personal characteristics or means of transport (for instance, are car trips better reported when they are made as driver than versus passenger?). Are there methods to identify trips that are well-reported and those subject to memory effects?
Correction of Memory Effects with an Underreport Ratio

In the last French NPTS, a person in the household had to describe the trips he (or she) made during the day before the interview and during the last weekend. As last Saturday can be one week before, we suspect memory effects. The car diary gives a more homogeneous image throughout the week. Table 2 compares the results from these two survey instruments.

For weekdays, the two survey instruments give the same image of car trips. It can be noted that car diaries cannot be filled out by persons too long away from home; thus, more long-distance trips are described by interview (for instance, the trip back from holidays). So, if we limit the scope to trips within 80 km crow-flight distance from the residence of the household, total traffic (in vehicles-kilometers) is almost the same. There are 2% less trips collected in the car diary, but their average length is a little higher (9.9 km in the car diary, 9.7 km for car drivers in daily trips), denoting a slightly different understanding of the notion of trip when the driver fills out his diary alone, without the assistance of the interviewer (short trips are probably omitted).

The last weekend is too far away to avoid memory effects. Underestimation is important (about 30%) for very short trips (under 2 km). Sunday trips are a little less underreported than Saturday trips, probably because they are more recent. Thus, we propose to use the figures in the two right columns of Table 2 as correction coefficients for all motorized weekend trips.

TABLE 2  Total Number of Car Driver Trips: Comparison Car Diary/Daily Trips

<table>
<thead>
<tr>
<th>Trip distance</th>
<th>Weekday</th>
<th>Saturday</th>
<th>Sunday</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;2 km</td>
<td>1.01</td>
<td>1.29</td>
<td>1.32</td>
</tr>
<tr>
<td>2-11 km</td>
<td>0.97</td>
<td>1.21</td>
<td>1.16</td>
</tr>
<tr>
<td>12-44 km</td>
<td>0.98</td>
<td>1.19</td>
<td>1.12</td>
</tr>
<tr>
<td>&gt;44 km</td>
<td>0.91</td>
<td>1.06</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Source: INSEE-INRETS 1993-94 NPTS.
THE STANDARD METHODS OF IMPUTATION

Imputation procedures are common methods of adjusting data sets for missing values. “Clean data” (complete rectangular data matrix) avoid problems raised by estimations from response sets of various sizes. Indeed, most computer packages are built for rectangular data files. Thus, the wide availability of statistical packages leads the statistician to use imputation in order to obtain a “clean data matrix”. The main imputation methods are the following ones:

a) Deductive imputation: We use this method for those cases where a missing value can be filled with a perfect prediction, attained by a logical conclusion. The deduction may be based on response given to other items on the questionnaire. This case is not so rare in travel diaries: for instance, travel distance can be checked and calculated from origin and destination places. For instance, is it necessary to answer for trip distance, which people often poorly estimate? Or is it possible to calculate it with a GIS and to ask only origin, destination and other simple questions on the itinerary (for instance, route choice between motorway and road)?

b) Overall mean imputation: We replace all missing values for a given item by the respondent mean for that item. It is a dangerous method, unless the nonresponse is negligible. This procedure may lead to severely understated variance estimates and to invalid confidence intervals.

c) Class mean imputation: This method consists of partitioning the unit response set into imputation classes such that elements in the same class are considered similar. Auxiliary variables are used for this classification. There will be some distortion of the “natural” distribution of values, but the bias is less severe than with the overall mean imputation.

d) Hot-deck and cold-deck imputation: In hot-deck imputation, missing responses are replaced by values selected from respondents in the current survey, while cold-deck procedures use sources other than the current survey. A number of hot-deck procedures have been proposed, including random overall imputation, random imputation within classes, sequential hot-deck imputation, and hierarchical hot-deck imputation.

e) Regression Imputation: We use respondent data to fit a regression of a variable for which one or more imputations are needed on other available variables.

f) Multiple imputation: For each missing value, we impute m (m > 2) responses. Thus, these m imputations create m complete data sets, and they are analyzed as if they were m “clean data matrices.”

Imputed values have to be signaled in the data file. Generally, a character variable tells how imputed values have been obtained for each important variable in the data set. Anyway, it is preferable not to change a missing value into some value that could have been a valid one. For instance, some small value can be added to a date value: say, 0.1 when the month has been supplied, and 0.4 when the day has been supplied (15.4 is closer to the middle of the month, anyway). When SAS uses formats to print out the values, rounding takes place and you won’t notice anything. But when you want to make a frequency table of the day in month, you still can filter the days number 15 you supplied from the ones in the original data.
VALIDATION AND CORRECTION OF DAILY TRIPS DATA IN THE FRENCH NATIONAL PERSONAL TRANSPORT SURVEY

Trips were described in a weekly stage diary for 1981-82, by interviews on the previous day and the last weekend for 1993-94, and in a weekly car diary for both surveys. The main characteristics of the trips are

- Their origin and destination (coded in terms of municipality in France and in terms of NUTS3 regions for neighboring countries in the last survey),
- Their length (estimated by interviewed persons, calculated in car diaries as the difference on the odometer between the origin and destination),
- Their duration (computed as the difference between the arrival and departure times),
- Transport mode (up to four different modes, in the case of a multi-modal trip),
- Trip purpose.

There are relationships between these variables:

- Some places are described in the general part of the questionnaire (the residence and habitual workplace),
- Trip length has to match with origin and destination (trip length must be greater than crow-flight distance, with a tolerance of 5 km unless origin and destination are located in two neighboring municipalities),
- Door-to-door average speed (calculated as the ratio of trip length to trip duration) must stay beyond reasonable limits (see Table 3): by car for instance, between 2 km/h and the maximum authorized on motorways—which is 130 km/h in France.

### TABLE 3 Controlling Data Mode by Mode

<table>
<thead>
<tr>
<th>Speed (in km/h)</th>
<th>Minimum</th>
<th>Medium</th>
<th>Maximum</th>
<th>Trip/Crow-flight Distance (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking</td>
<td>1</td>
<td>3</td>
<td>10</td>
<td>0.3</td>
</tr>
<tr>
<td>Bicycle</td>
<td>1</td>
<td>8</td>
<td>30</td>
<td>1.0</td>
</tr>
<tr>
<td>Motorcycles</td>
<td>2</td>
<td>15</td>
<td>130 (2)</td>
<td>1.2</td>
</tr>
<tr>
<td>Car, truck, taxi</td>
<td>2</td>
<td>24</td>
<td>130 (3)</td>
<td>1.3</td>
</tr>
<tr>
<td>Bus</td>
<td>2</td>
<td>12</td>
<td>110</td>
<td>1.3</td>
</tr>
<tr>
<td>Rail urban transport</td>
<td>3</td>
<td>12</td>
<td>75</td>
<td>1.1</td>
</tr>
<tr>
<td>Train</td>
<td>10</td>
<td>54</td>
<td>150 (4)</td>
<td>1.2</td>
</tr>
<tr>
<td>Aircraft</td>
<td>100</td>
<td>400</td>
<td>1000</td>
<td>1.1</td>
</tr>
<tr>
<td>Seacraft</td>
<td>1</td>
<td>10</td>
<td>75</td>
<td>1.1</td>
</tr>
</tbody>
</table>

Source: INSEE-INRETS 1981-82, and 1993-94 NPTS.

(1) It is crow-flight distance between different municipalities. Thus, this coefficient is low for short distance modes (especially for walking, and also for bicycle and urban transport), since some of those trips just cross the boundary between two neighboring municipalities. This coefficient is smaller for long trips (for instance, made by air) than for medium-distance trips.

(2) Only 70 km/h for moppets.

(3) We have admitted a few verified exceptions, up to 140 km/h door-to-door.

(4) Up to 250 km/h for TGV (high-speed train).
Let us first show how we have completed nonresponses and made data consistent in the last French NPTS (1993-94). Like most surveys, there are almost no nonresponses on origin and destination locations: only 10 out of 100,000 trips could not be coded. Thus, we have used crow-flight distances to fill nonresponses on trip length (1300 cases) or to replace responses leading to an unreasonable mean speed (400 cases). Generally, the crow-flight distance is multiplied by a coefficient specific to each mode (for instance, 1.3 for private car).

In order to estimate missing or dubious values for duration, we have used a regression technique by calibrating the relationship between average speed and trip distance. For motorcycles and cars, this equation is

\[ \text{SPEED} = 1.4 + 14.6 \log(DIST+1) \]

For the 1993-94 car diary, four different estimates of this equation have been made on correctly described trips

- If origin and destination are in a city-center
  \[ \text{SPEED} = 1.54 + 15.25 \log(DIST+1.3) \quad R^2 = 0.474 \]
  
  (9.3) (185.7)

- If origin or destination is in a city-center
  \[ \text{SPEED} = 2.46 + 15.72 \log(DIST+1.3) \quad R^2 = 0.467 \]
  
  (14.9) (219.5)

- If origin and destination are not in a city-center
  \[ \text{SPEED} = 4.39 + 15.64 \log(DIST+1.3) \quad R^2 = 0.445 \]
  
  (31.0) (246.6)

- If this information is missing for origin or destination
  \[ \text{SPEED} = 1.74 + 15.90 \log(DIST+1.3) \quad R^2 = 0.511 \]
  
  (4.7) (102.0)

In 1981-82, the previous form of this question was about the use of a motorway during the trip. This information did not provide significantly different equations of speed as a function of trip distance.

As walking trips usually have their origin and destination in the same municipality, crow-flight distance between municipalities does not allow computing trip distances. For this mode, we have supposed that the mean speed is 3 km/h, either to estimate trip lengths (500 cases), or to fill the few missing data on duration.

Using these techniques, we succeeded in getting totally consistent data on locations, distance, duration, mean speed and mode. There are very few missing values left: 2 on trip distance, 6 on trip duration, plus 11 cases where trip duration has been filled, but arrival and departure times are unknown (see Table 4).
### TABLE 4 Validation and Correction of Daily Trips Data

<table>
<thead>
<tr>
<th>Number of cases and %</th>
<th>1993-94 Survey</th>
<th>1981-82 Weekly Diary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Original file</td>
<td>After correction</td>
</tr>
<tr>
<td><strong>UNKNOWN PLACES:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Origin</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>- Destination</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td><strong>DISTANCE:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Unknown</td>
<td>1812</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1.9%</td>
<td>0.0%</td>
</tr>
<tr>
<td>- More than 5 km inferior to crow-flight distance</td>
<td>299</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.3%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Unknown DURATION</td>
<td>59</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>0.1%</td>
<td>0.0%</td>
</tr>
<tr>
<td><strong>SPEED:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Too Fast</td>
<td>194</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.2%</td>
<td>0.0%</td>
</tr>
<tr>
<td>- Too slow</td>
<td>292</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.3%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Unknown transport mode</td>
<td>74</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>0.1%</td>
<td>0.1%</td>
</tr>
</tbody>
</table>

Sources: INSEE-INRETS National Transportation Surveys.

- Weekly stage diary in 1981-82. It has been converted into trip diary for the comparison with 1993-94 survey (see the two columns at the righthand side).
- Previous day and last weekend trips in 1993-94.

For daily trips in the French 1993-94 NPTS, hot-deck is not appropriate, because trips are described for typical days (Saturday, Sunday and a weekday). In 1981-82, similar trips were more frequent, since trips were reported in a weekly diary; thus, hot-deck inside a diary can be used in order to fill nonresponses or to make data consistent.

After matching origin-destination and trip distance, hot-decks are run to fill nonresponses, first on the transport mode, and then on trip duration. Criteria used to find a correctly described trip similar to one with inconsistent or missing information are geography (origin and destination in the same municipalities), trip purpose to fill the mode and trip distance to fill the duration. The results are not as good as for the 1993-94 survey: out of 66 000 trips, 81 missing values are left on trip length, and 161 on trip duration.

In car diaries, the driver has to check the odometer at the beginning of each day and at the end of each trip. This information is highly structured (mileage must increase all along the diary and gives an objective measurement of trip distance), but there are several missing values. In order to fill them, we try a hot-deck structured by origin-destination and duration. If it fails, we compute mileage proportional to trip duration or to crow-flight distance, while ensuring that the mean speed stays beyond reasonable limits. Finally, we fill nonresponses on trip duration with a hot-deck run on geographical and distance criteria. At the end, there are no missing values left, either on mileage, or on trip duration, but departure and arrival times are still missing and 105 trips out of 58,000 in 1981-82, and 2485 out of 200,000 in 1993-94.
A PROMISING APPROACH: ARTIFICIAL INTELLIGENCE

Common imputation methods take into account only few variables. We feel that a more comprehensive understanding of the data is needed in order to improve imputation methods. For the TEST (Technology for a European Survey on Travel Behavior) project, funded by DG VII for the 4th Framework Research Program, we have proposed to explore two new approaches: FUNDP will develop a parser, and INRETS a neural network-based imputation system.

To correct and/or validate the data from travel diaries, their semantic structure has to be known. We must develop a scheme to split the answers into their different semantic components so that the correction techniques can be developed and applied for each particular element.

The first idea is to base the analysis on techniques related to the LEX and YACC parser tools present in the Unix environment. With LEX, we can associate tokens to different lexical elements. Then, with YACC, we represent the structure of the data based on these tokens. So, we can parse the diary and analyze its semantic according to these structures. YACC also allows us to associate treatment to each element of the structure. In fact, from LEX and YACC description of the data, a C program is automatically constructed. This program reads the file with the data from the diary, parses them according to the rules described in YACC language (using the tokens defined in the LEX syntax) and deals with the treatments associated, in the YACC description of the data structure, with each parsed element. These treatments will include validation techniques. The results of such an analysis would be a formal representation of the data structures allowing computer-aided correction techniques. This study can be based on the work done during the development of the PACSIM model for the achievement of a parser for all the data files (see PACSIM Users’ Manual, E. Cornelis, GRT, FUNDP, 1996).

In the same way, we plan to analyze the structures of the errors and, more precisely, of the errors the parser would be able to correct. The errors can be of several kinds: missing item, misspelled item, wrong number, etc. These errors must be detected during the validation phase. To do this, we must use the semantics of the data (for example, to detect missing items), but also rules based on common sense (for example, to give bounds on some data) or dictionaries (for example, to detect misspelled items).

Associated with the data structures described in the YACC and LEX syntaxes, we will develop treatments to deal with the data validation and correction. These techniques will also be based on rules. We have already experimented with use of rules in the IMAURO project (DRIVE I program) for the representation of users’ behavior. An example of such a rule could be if the destination of a trip is missing, try to find in previous data another trip with the same characteristics, and copy its destination. The techniques to recover missing items could be to use previous data to search for an answer with the same characteristics, and deduce from it the missing information. Sometimes, common sense can also be useful (e.g., if the mode is missing for a trip starting from a railway station, we can suppose that the train was used). Geographical knowledge must also be exploited (for example, when trip length is missing). For misspelled items, searches in dictionaries must be undertaken. In this area, a particular emphasis must be put on language dependent issues. We can also make a context-sensitive analysis to determine which are the correct data.

In summary, the parsing techniques derived from YACC and LEX will be used to analyze the data, validate them and detect errors. Then, rules-based techniques will be employed to recover from errors. These rules will rely on previous recorded data, common
sense, geographical knowledge and dictionaries (taking language into account). Advantage can be taken of FUNDP’s experience with the validation and the correction of the data of the household survey (from the energy survey realized for the SPPS in 1986-87).

In a parallel and complementary way, INRETS plans to investigate a second method for correction and imputation problems, based on another type of AI technique: neural network models. Neural network models are commonly and successfully used as data analysis tools, especially for uncertain or noised data, for high-dimensionality data and for nonlinear analysis.

We plan to use the Kohonen’s Self-Organizing Map algorithm (the SOM algorithm) to design correction and missing data completion methods for the daily trips data. Theoretical analysis of this algorithm still has to be carried out, but it can be interpreted as a non-linear extension of principal component analysis. A self-organizing map provides a mapping between a data definition space and a set of prototypes that are arbitrarily arranged in a 2-dimensional grid called the map. The self-organization of the map consists in slightly moving the prototypes in the data definition space, according to the data set distribution and to the prototypes’ localization in the map. The resulting mapping is said to preserve topological relations between data: two points that are close in the data definition space are associated with two prototypes that are close on the map.

The analysis of the map organization brings to the fore a selective influence among the data dimensions: some of these dimensions play a preponderant part during the organization. It has been shown experimentally that, under certain conditions, the map can be used to estimate missing values in those dimensions for partial data. We plan here to investigate this property and design a correction and missing data completion tool based on this principle. Let us mention that treatment and completion of missing data with the SOM algorithm has already been tackled.

These methods will be tested and compared with classical statistical methods on data from the French NPTS (description of long-distance journeys, car-diary, etc.) and from pilot surveys conducted for the MEST project (Methodology for a European Survey on Travel Behavior).

Could rounded responses or data coded in brackets be considered as an intermediate case between precise and missing data?

Interviewees are generally unable to describe their mobility with the accuracy suggested in the questionnaire: 1 min for departure and arrival time, 1 km for car annual mileage and daily mobility (even 100 m for trips under 2 km in the French NPTS). For the same reasons, plus confidentiality, income is coded in brackets in most surveys. For most analysis, we do not need so much accuracy, but we have to be aware that roundings modify variables’ distributions.

In order to measure the accuracy of responses, the duration of trips (calculated as the difference between departure and arrival time) provides a good indicator. As roundings can only be compared for rather homogeneous data, we will study separately car trips and long-distance travel. It seems that people remember the duration of their trip (consistent with its length) and departure time. As most car trips are short, arrival time (deduced from departure time and trip duration) appears paradoxically to be less rounded than departure time. In weekly car diaries 20% of departure times correspond to o’clock hours, 25% in 1981-82 personal weekly trip diaries, 30% in the 1993-94 interview about trips made the
day before, and up to 39% for those made during the last weekend (Table 1). For 1993-94, Table 5 shows that this proportion is 48% for long-distance trips described in the daily mobility questionnaire (interview on trips made the day before and during the last weekend), and up to 65% in the retrospective long-distance part (trips made during the three months before the interview). About trip duration, the accuracy of data is better than 5 minutes for 20% of trips in the 1993-94 car diary (18% in 1981-82), 10% for car trips in the 1981-82 mobility diary, 6% for car trips made the day before the interview in 1993-94 (5% for those made during the last weekend), and less than 3% for long-distance trips.

Summarizing the main results, it appears that

- Time variables are less rounded when reported in diaries than when collected by interview,
- Fortunately, memory effects affect time (of departure or arrival) more than duration, which needs to be known more accurately, especially for modeling,
- The deterioration due to memory obviously increases when the facts reported have occurred a long time before the interview (during the last weekend or three months ago),
- The car diary is more accurate than the other methods, probably because of the clock that is displayed on most car boards,
- Roundings do not seem to affect the estimation of mean values: the average trip duration is a little longer when it is measured by car-diaries (15.8 mn instead of 15.5 mn for trip diaries in 1981-82, 16.4 mn instead of 16.2 mn for interviews in 1993-94), but this difference is not very significant, when tested with a 5% threshold.

**TABLE 5  Rounding for Departure Time of Car Driver Trips**

<table>
<thead>
<tr>
<th>(Column %)</th>
<th>Car Diary</th>
<th>Stage Diary</th>
<th>Interview on:</th>
</tr>
</thead>
<tbody>
<tr>
<td>o’clock (0 mn)</td>
<td>20</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>half past (30 mn)</td>
<td>16</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>quarter (15 or 45)</td>
<td>17</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>ten (10, 20, 40 or 50)</td>
<td>22</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>five (5, 25, 35 or 55)</td>
<td>15</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>more accurate</td>
<td>10</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

N.B. Weekly diaries filled by the household, except for daily mobility in 1993-94, for which it was a direct interview on trips made the day before (a weekday) and during the last weekend.

The measurement of trip distances is also an important issue. Table 6 shows that roundings does not depend on survey instrument for long-distance travel.

In order to avoid introducing artificial accuracy, imputations have been rounded to:

- 1 km for trip distance in car diaries (checked by odometer readings),
- 5 mn for trip duration in car diaries,
- 10 mn for trip duration in the trip diary, where this information is less accurate than in car diaries.
TABLE 6 Rounding on Time Data for Long-Distance Trips (1)

<table>
<thead>
<tr>
<th>(Column %)</th>
<th>Trip duration</th>
<th>Departure time</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Daily</td>
<td>Long-distance</td>
<td>Daily</td>
<td>Long-distance</td>
<td>Uniform</td>
</tr>
<tr>
<td>o’clock (0 mn)</td>
<td>48</td>
<td>65</td>
<td>31</td>
<td>45</td>
<td>1</td>
</tr>
<tr>
<td>half past (30 mn)</td>
<td>29</td>
<td>23</td>
<td>30</td>
<td>33</td>
<td>1</td>
</tr>
<tr>
<td>quarter (15 or 45)</td>
<td>11</td>
<td>6</td>
<td>19</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>ten (10, 20, 40 or 50)</td>
<td>8</td>
<td>4</td>
<td>10</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>five (5, 25, 35 or 55)</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>more accurate</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>80</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99</td>
</tr>
</tbody>
</table>


(1) Trips over 100 km, whose origin or destination are more than 80 km crow-flight away from the residence of the household.

N.B. Face-to-face survey on trips made
- during the day before the interview and the last weekend for “Daily Trips”: sample of 1978 trips,
- during the three months before the interview, for “Long Distance” (only one month for the most mobile persons): sample of 41 774 trips.

When information is collected in brackets (for instance, income), many results can be obtained by attributing an average value to each class (generally its middle, with some difficulties to determine it for the upper and the lower one). When a continuous distribution is needed, it can be computed using simulated residual methods (Lolivier and Verger, 1988), but you need to be careful when implementing it. For instance, using the same variable in the simulation and in the analysis should be avoided: car ownership is certainly useful for the simulation of an income distribution, but motorization income elasticity cannot be estimated with those data.

CONCLUSION

This resource paper has listed some issues on item nonresponse and illustrated them with examples drawn from the French NPTS. Let us now summarize some questions to be discussed during the sessions (there are certainly missing items in this list!).

First, are there still uncertainties about the field of reweighting (much information is missing: for instance, missing trips) and the field of imputation (missing data for single variables)?

For reweighting, in the examples presented in this paper, response models have been determined by what could appear as an “a priori” grouping of data (however, based on the experience and on a careful analysis of response behavior). Are there methods (logit model, segmentation,...) that could optimize this grouping of data? Is this step of grouping absolutely necessary, or is it possible to work directly on individual data?

Concerning imputation, what data have to be left missing or rounded, and what data have to be replaced? For instance, the answer “unknown” to a question about opinion surely gives some information. Since we know that this kind of answer generally come from old persons and, in more recent surveys, from young unemployed persons, a hot-deck will give “good results”. But is it interesting to implement it, or is it better to keep the original information?
Issues of variance and of confidence interval calculations have not been discussed. Some imputation methods artificially reduce the variance (for instance, replacing all missing data by the average value calculated on respondents), while other methods increase it (for instance, hot-deck when the choice at random of the individual supplying the non-missing data is added to the overall variance of the distribution).

ACKNOWLEDGMENT

The work reported has benefited from the scientific support of Jean-Claude Deville (INSEE). It also contains some of the first results of the EU-funded 4th Framework project, ‘Methods for European Surveys of Travel Behavior, involving the Institut fr Straaenbau und Verkehrsplanung, Leopold-Franzens-University Innsbruck, Statistics Netherlands, Bro Herry (Wien), University of London Centre for Transport Studies, Imperial College (London), Deutshe Versuchsanstalt fr Luft-und Raumfahrt (Kln), INRETS (Arcueil), Transportes inovao e Sistemas (Lisbon), TOI (Oslo), Transport Studies Group Facultés Universitaires Notre Dame de la Paix (Namur), Socialdata (Munchen), Statistics Sweden, and TU Delft. All the conclusions drawn are solely those of the authors, and not the CEC or the Consortium.

BIBLIOGRAPHY

application de la méthode des résidus simulés, in Mélanges économiques: essais en
l’honneur de Edmond Malinvaud, *Economica*.


Armoogum, J., Han, X.L., Polak, J., and Madre, J-L., MEST Deliverable 8 for EEC:
Improved methods for weighting and correcting of travel diaries.

Polak, J.W., and Ampt, E.S. (1996) An analysis of response wave and nonresponse effects
in travel diary surveys. Paper presented at the 4th International Conference on Survey

Paper presented at the 7th Conference of the International Association for Travel
Behavior Research, Santiago, Chile, July.

Planning*, Eucalyptus Press, Melbourne.


ABSTRACT

This paper reports on issues pertaining to item nonresponse in travel surveys. It presents a typology of item nonresponse errors, identifies missing data problems, and provides solutions for preventing and/or reducing data gaps at the item level. The authors conclude with recommendations for post-survey corrections to missing data, given the fact that it is impossible to achieve perfect, complete data in surveys.

INTRODUCTION

In recent years, with an increased desire for better urban transportation systems arising from environmental and traffic congestion concerns, there has been a need for more sophisticated modeling in the transportation planning process. In turn, the need for increasingly complex transportation planning models has focused attention on model inputs, specifically, travel survey data. Much of the emphasis has been on the introduction of modeling processes that use comprehensive travel survey data sets and that include newer and more complicated analyses. Thus, designers, conductors, and users of travel survey data are being challenged to obtain more and better data for modeling purposes in a time when surveys, in general, are becoming increasingly difficult to conduct.

Travel surveys compile information voluntarily provided by projectable samples of trip-makers about themselves and the trips they make. In other words, travel surveys seek to measure trip-making behavior and to answer very specific questions about who makes trips, how people make trips, and where people go. Thus, as reports of behavior, and not as observations of events, the information compiled during travel surveys is susceptible to various types of errors. The survey research literature has developed a well-accepted typology of such errors, their effects on survey quality, and approaches for diminishing or compensating for such errors. Most approaches base the typology on statistical models for estimating the effects of errors.

Borrowing on Groves’ (1) excellent overview of the components of survey quality, we can classify errors in travel surveys into four main types:

1. Coverage errors typically refer to the exclusion of some members of the study population from the sampling frame. The most common instance of noncoverage error in household travel surveys in the United States is the exclusion of households without telephones. Other examples of coverage error might include the exclusion of persons with unlisted telephone numbers, if the survey is administered by telephone, or of people living in hospitals, hotels, prisons or other institutions, if these persons are defined as members of the target population. Groves (1) has suggested that “coverage error is the forgotten child among the family of errors to which surveys are subject.” Indeed, coverage error is rarely addressed in the transportation research literature.
2. Sampling errors refer to the estimating quality of sample statistics and are mainly a consequence of sample sizes and sample design. Probability sampling techniques are geared mostly to reducing sampling error. Numerous approaches to use weighting for unequal selection probabilities are well-known, but rarely applied in travel surveys. For example, the unadjusted inclusion of households with multiple telephone lines (and, thus, higher chances of selection) in random generated telephone samples is commonplace.

Travel survey experts (2) have noted that a primary goal of most travel surveys is to collect data that allow analysts to estimate a range of travel-related and socioeconomic parameters. The fact that sampling errors can affect the precision of the parameter estimates is well-documented in the literature.

3. Nonresponse errors cover the fact that certain individuals selected in a sample do not participate in the survey or fail to answer an item in the interview. Most of this type of error is being covered in a separate paper at this conference. However, item nonresponse, as defined in this paper, is partially a problem of nonresponse error and must be factored as such in techniques for amelioration or for post-adjustment.

A review of the literature by survey experts at the National Opinion Research Center of the University of Chicago has indicated that the primary factors contributing to high levels of missing data in travel surveys are comprehension errors and privacy concerns (3).

4. Measurement errors are those that cover the discrepancy between an individual’s true characteristics or actual behavior and the individual’s responses in a survey interview. There are two basic kinds of error that affect empirical measurement: random error and nonrandom error. The amount of random error is inversely related to the degree of reliability of the measuring instrument. The amount of random error may be large or small, but it is universally present to some extent (as evidenced by the fact that two sets of measurements of the same features of the same individuals will never exactly duplicate each other). Nonrandom error, on the other hand, lies at the very heart of validity. Nonrandom error prevents indicators from representing what they are intended to measure.

Richardson, Ampt, and Meyburg (4) use the analogy of sharpshooters firing at a target to convey the concepts of accuracy and precision in travel surveys. To relate this analogy to measurement error, random error would be indicated if the shots were as likely to hit above the target as below it, or as likely to hit to the right of the target as to its left. On the other hand, if those shots aimed at the bull’s-eye hit approximately the same location, but not the bull’s-eye, then some form of nonrandom error has affected the targeting of the rifle.

Item nonresponse, as used in this paper, is defined as “performance gap” or the failure to attain “true” and complete data from each survey respondent. We are concerned with both “missing” data and “biased” data (i.e., deviations from some underlying true value for the statistic). Thus, this paper deals with item nonresponse as it relates to the last two types of survey errors—nonresponse errors and measurement errors. Nonresponse errors are of interest because they lead to missing data. Measurement errors, while not usually associated with item nonresponse, are a focus of this paper, because they deal with data that are present but are inaccurate or imprecise. In our discussion of measurement errors, we focus on nonrandom errors.
The travel surveys we will discuss in addressing issues of nonresponse and measurement errors are those that ask questions about behavior or facts about respondents. These questions deal with characteristics of people or their households, activities they have performed, or events that have happened to them. In contrast, questions about attitudes, beliefs or prospective behavior are unverifiable. Certain transportation-related questions can deal with these latter types of questions, such as satisfaction questions in origin and destination surveys, or questions about hypothetical behavior, as in stated preference surveys. However, these types of questions and surveys are not the focus of the current paper.

One final introductory note needs to be made regarding the scope of this paper. Rather than attempt one more reiteration of the already well-documented conceptual classification of survey errors, this paper seeks to explore more practical, prescriptive issues involving nonresponse and measurement errors. Throughout the paper, the focus is on the key players—the respondents, the survey designers, and the data collectors—and their roles.

RESPONDENTS: THE ESSENTIAL PLAYERS

We believe the respondents are the essential players when addressing issues of item nonresponse because they ultimately make the decision to answer or not answer a survey question (5). First, the respondents decide whether or not to report that they have engaged in the type of behavior being asked about. This decision is influenced by the facts of the situation and by their own feelings about whether they want to tell the interviewer about it. If this decision is positive, the form of the subsequent questions and the persons asking the questions (if applicable) play a very important role in influencing how much behavior is reported. Thus, in this two-step process, the first step appears to be primarily influenced by the characteristics of respondents, their actual behavior, and their personalities or beliefs about social norms.

Virtually all issues associated with measurement error and item nonresponse error in surveys can be fully described, most thoroughly understood, and best solved by focusing on characteristics, perceptions, and motivations of respondents. At the same time, full appreciation of respondents on the part of the researcher, and the question design and data collection tactics that are formed by this recognition, are almost certain to produce significantly higher quality data.

TYPOLOGY OF ITEM NONRESPONSE ERRORS

From the perspective of the prominence of respondents, item nonresponse is a result of five major phenomena. Each of these may result in missing or misreported data. These phenomena are related to respondents’ general characteristics, motivations and perceptions. Before exploring each in detail, we should briefly review them. They are:

- **Knowledge and recall.** Invariably, we frequently ask questions for which some respondents simply do not know the answer. More frequently, respondents suffer memory lapses that are not adequately recognized and produce missing or low-quality data. Memory lapses can include the forgetting of minor events, such as certain side trips or stops on a trip, or incorrect recall. Events, such as trips or activities, may be well-recalled,
but the lapses affect the sequence or order of events, or their exact timing. This phenomenon results in both missing data problems and gaps in data accuracy.

- **Comprehension.** Too often, the questions we ask are difficult to understand and to answer. As Sheatsley (6) observes:

> Because questionnaires are usually written by educated persons who have a special interest in and understanding of the topic of their inquiry, and because these people usually consult with other educated and concerned persons, it is much more common for questionnaires to be overwritten, overcomplicated, and too demanding of the respondent than they are to be simpleminded, superficial, and not demanding enough.

Writing sufficiently clear and simple questions requires following four important principles: simple language, common concepts, manageable tasks, and widespread information. The need for heavy-duty work on the part of question writers is exacerbated because of the disproportionate distribution of comprehension gaps (e.g., among low literacy levels or immigrants who do not speak the national language). In addition, idiosyncratic interpretation of questions on the part of respondents is a significant factor, mainly because it is so difficult to document or estimate, and therefore cannot be corrected in post-hoc analysis.

- **Perceived and/or real burden.** Stopher and Metcalf (7) point out that the trend in household travel surveys in the 1990s has been an increasing level of detail in what people are asked to report. They ask “how much more can people be asked to report or even whether the point of asking for too much information has already been passed.” In an age of increasing demands on peoples’ time from all types of sources (e.g., family, job, community), perceptions about the time burden inflicted by travel surveys are an important factor.

- **Desire for privacy or concerns about personal information.** Questions about a household’s trip-making behavior can be as threatening to respondents as questions about gambling, drinking alcohol, or sexual activities. These questions are considered threatening for a variety of reasons, including fear of consequences of divulging data or distrust of the person asking the questions.

- **Deliberate misreporting.** Motivation can be fear of consequences and, also the wish to present oneself in a favorable light or to give a good impression. Most frequent is the desire to report socially acceptable behavior or to not report socially sanctioned behavior.

As we stated, this paper seeks to explore practical, prescriptive issues involving nonresponse and measurement errors. Toward this end, we focus on ways to reduce item nonresponse as related to the five phenomena described above.

**Knowledge Gaps**

Travel surveys typically ask respondents to answer questions of personal fact—bearing on their own experience or characteristics. Therefore, one would assume that it would be easier for respondents to answer such questions than to respond to questions about opinions or attitudes. However, these “facts” may be prey to the same kinds of ambiguities
of meaning and frames of reference that attitude questions are. Thus, travel survey data include “don’t know” responses.

There is much debate in the survey literature whether “don’t know” responses are legitimate responses (8). Respondents may legitimately not know an answer because they never knew the information or can no longer remember it. On the other hand, respondents may be taking the easy way out with a “don’t know” response, rather than taking time to search memory, to refer to records, or to ask someone else who may process the needed information. In the area of household travel behavior surveys, “don’t know” may be given as a response during the recruitment or screener phase, in which demographic characteristics of the household or its members are collected, or during the retrieval of activity or trip information, in which trip end locations are collected.

Illustrative Problem #1

Survey field-workers and respondents have surprising difficulty with precise addresses for locations that are traveled to frequently. Ironically, the difficulty is getting the exact address without respondents’ engaging in reluctant research, for the street address or even the intersection of such places as friends’ residences, schools their children attend, retail outlets, and even places of employment. Because people rarely have to know the exact address for locations other than their own residences, those places that they travel to frequently are perceived as known locations, even though geocodable information may not be forthcoming. Incomplete data of this type can be avoided by increasing use of electronic databases that compile exact addresses for places by name—with them, a respondent’s knowledge gap and unwillingness to “look up” the address can be ameliorated. Another solution is to collect information about “habitual destinations” at the time a household is recruited to participate in a survey. Then, a second opportunity to verify or clarify location information can be secured at the time of actual trip/activity retrieval.

Illustrative Problem #2

Household-income missing data, as will be noted in several parts of our paper, can be the result of multiple factors. One common causal factor is a simple knowledge gap. In certain income groups and among some immigrants, the householder who is employed may be the only one who knows the full income of the household. The spouse may know only the actual available household budget. Also, in multi-earner households, or in households of extended families or with unrelated individuals, no single respondent may be able to provide an answer to the single question on aggregate household income. Because the household income variable may be used to link with Census data, it is important to recognize that the travel survey item does not measure the same concept as the Census Bureau household income. The ultimate solution would be to ask for household income in exactly the same format as the Census Bureau does—a long series of questions for every single household resident about the amount of income from each of about eight sources. This would be perceived as unacceptable burden and could not be implemented. Other solutions that have been successfully implemented involve asking the household income question multiple times. This can produce two different responses, with
the average being adopted as the “true” value, or can reduce refusal incidences from 15+% to less than half of that.

**Recall Barriers**

Among the most challenging problems surrounding item nonresponse is inaccurate recall. This is a serious problem not only because it can generate polite refusals and “don’t remember” responses, but also because responses politely given may have validity problems. If travel survey designers were guided entirely by a concern for valid and complete trip and activity data, they would ideally focus on real-time measurement so that the respondents’ responses are consistently about strict record-keeping of what they are doing. Handheld computers and geographic positioning and tracking systems may provide a partial, albeit expensive, solution. However, a great deal more can be done to diminish the effect of memory lapses on data.

The social psychology literature provides valuable insights regarding the recall problem. Clearly, attention needs to be focused on the cognitive demands of the interview. While there are many examples of early research on problems of recall for health events (11), for expenditure data (12), and for victimization data (13), only recently have these problems been linked to theories from cognitive psychology (14). This has taken the form of methods of cuing the retrieval of memories of past behavior that have found temporal landmarks useful for reminding people of the time of occurrence of events. It has also focused on the effort of current mental states on measures of past mental states, on the mechanisms by which context effects come about, and on the nature of reconstructing past events for reporting in the interview. The aim of this line of research is to reduce measurement error through the use of questions that prompt memory retrieval more efficiently.

Converse and Presser (9) cite circumstances that lead to recall barriers: the decision was made almost mindlessly in the first place, the event was so trivial that people have hardly given it a second thought since, or the questions require the recall of many separate events. Three strategies to deal with recall barriers are

- Bounded recall—establishing a baseline measure in an initial survey. If one is concerned about underreporting of trip-making behavior, a question could be asked in the recruitment interview about how many trips (well-defined) the respondent took “yesterday,” and if this was a typical travel day. Then, in the subsequent retrieval interview, the respondents’ trip-making reports can be “validated” using the earlier response.
- Narrowing the reference period—Travel surveys typically assign a travel day to the household on which members of the household keep a travel diary. It is common practice to begin attempts to retrieve the travel data on the day following the travel day, and to continue trying to retrieve data for many days subsequent to the travel day. A limitation on the retrieval attempt period should be maintained at 5 days or less.
- Memory cues—to stimulate recall by presenting a variety of associations. It recognizes the fact, that because human memory uses a variety of coding schemes to store information, what appears to be a “forgotten” trip may be perfectly accessible if the correct storage file is tapped. Instead of asking respondents where they went next, for example,
the interviewers could ask if they made any stops at the drug store, post office, etc., after leaving work.
The best way to address the range of difficulties of recall is to review several concrete elements of the problem, and how they are generally solved.

Illustrative Problem #3

In travel surveys that focus on trips that are being taken at the time of the interview, memory lapses are greatly reduced. Good examples of such surveys are transit use interviews that intercept the transit traveler in the act of using transit and elicit responses about the current trip being taken. However, in travel surveys that require recall of detailed events, even a single day before the interview, accurate recall begins to atrophy. The household travel survey, even when focused strictly on yesterday’s travel, must address many barriers of recall. Solutions adopted in many travel surveys include strict procedural rules regarding deadlines for data retrieval, limitations on proxy responses, and the use of effective mnemonic devices. Strict deadlines for collection are important for diminishing the damage of poor recall. This is extremely difficult because willing respondents are busy and frequently not accessible without multiple attempts to contact.

Illustrative Problem #4

The most frequent outcome of how inaccurate recall affects data quality is the reporting of departure and arrival times for trips. Respondents frequently give rounded time responses which reflect a combination of failure to maintain real-time recording of information and poor recall. Analysts have noted the pattern of high frequency of times that are on the hour, the half hour, and the quarter hour. One easy-to-implement solution to correctly disperse responses on the quarter hour involves interviewer training to use probing with the query, “Was it a little before or a little after 8:15?” and then moving those responses to 8:10 or 8:20. For responses on the hour or the half hour, like 8:30, the probe can be “Was it closer to 8:15 or 8:45?” and then moving the answer accordingly. Simplified diaries and other mnemonic devices that emphasize the keeping of exact start and end times are another solution.

Both our experience and other researchers’ reports indicate that faulty recall problems are most severe because the indicators are not just missing data (nonresponse or “don’t knows”), but are more likely responses of indeterminate validity. The ideal tool kit for dealing with these problems includes rigorous interviewer training in the area of probing, mnemonic devices, and selected redundancy for problematic questions. Further research involving validation and evaluation can provide refined techniques for diminishing the impact of poor recall.

Comprehension Barriers

Item misreporting is related to a respondent’s not understanding exactly what the interviewer wants to know. This situation is most often caused by the way in which the interviewer asks and/or clarifies the survey question. The importance of the exact wording of the questions on a questionnaire seems obvious and hardly worth dwelling on. The fact that seemingly small changes in wording can cause large differences in responses has been
well-known to survey practitioners since the early days of surveys. Wording is everything, as illustrated in the story of two priests, a Dominican and a Jesuit, on whether it is a sin to smoke and pray at the same time. Unable to reach a conclusion, each priest goes off to consult his superior. Subsequently, they reconvene, and the Jesuit reports, “My superior said it is all right,” but the Dominican says, “Mine called it a sin.” They soon discover that the Dominican’s query to his superior was “Is it OK to smoke while praying?” but the Jesuit asked, “Is it OK to pray while smoking?”

Comprehension barriers deal not only with how a question is worded but also with how a question is “spoken.” Questions should be written in “spoken” language in which conventions of the written, formal language may be violated if the pure construction sounds stilted or pretentious. Questions should not only be easily understandable but also use a common frame of reference. What the researcher offers the respondent as a frame of reference may not be the one that the respondent commonly uses and it may be difficult for the respondent to put on the researcher’s angle of vision. What “people living in your household” is to mean, for instance, may have to be specified, lest respondents vary from the census definition in their frame of reference. The Valentine and Valentine study (15) of low-income black communities shows that reluctance to report the existence of a household member was related to the frequency of his presence there.

One way of overcoming comprehension barriers is to relax the requirement that a questionnaire script be read “exactly as written.” The similarities between a survey interview and an ordinary social conversation have been noted frequently. The opportunity to meet and talk with a variety of people appears to be one of the major attractions for the interviewers. By the same token, a major motivation for respondents appears to be the opportunity to talk about a number of topics with a sympathetic listener. There appears to be recent sentiment to rethink the structure of the survey interview. For years, most interviewing practice has focused on ensuring consistency of questionnaire administration. The implied goal of the effort is the standardization of the measurement instrument in hopes of achieving a consistent product from each respondent. Researchers of social interaction and discourse have noted that standardization of question wording does not necessarily imply constancy of meaning. Instead, using concepts from conversational analysis, they note that many of the normal mechanisms of ensuring clear communication, of correcting mis impressions, and of addressing the questions of the listener have been stripped away from the “standardized” interview. The effects of this may have been to minimize the interviewer variance but to increase bias, due to poor comprehension or minimal memory search for relevant information (1, 9, 10).

**Illustrative Problem #5**

In many sociodemographic segments, particularly lower-income and immigrant groups, the notion of “total, annual (before taxes) household income from all sources” is a difficult one, because almost all income is perceived as weekly or monthly take-home income. The translation from this into a true measure of total household income is exceedingly difficult, even with well-trained interviewers. Most alternatives provide little improvement or are even more unwieldy and threatening to respondents. The only realistic option for the comprehension barrier regarding household income is a slightly longer question, with added explanation about what is intended. After securing a response, interviewers can follow up by asking whether the income of everyone was included, whether income from
all sources (perhaps giving examples) was included, and whether all of this was for the full
preceding year and was before taxes. Obviously, these follow-up queries are, at best,
cumbersome and, most likely, very threatening to maintaining the cooperation of a
respondent. We must therefore be resigned to whatever a particular household member
may be able to provide us.

**Illustrative Problem #6**

Household composition questions can become complex enough to generate some
comprehension difficulties. In many cases, conceptual decisions, question wording, and
interviewing rules are set based on specific inclusion or exclusion of certain types of
people. Students away from home may frequently be considered members of the
household by their parents. Divorced parents sharing custody of minor children may both
claim a child as a resident. Visitors may be counted, but each survey may have unique
rules depending on length of stay or on provenance of the visitors. Young adult males in
urban cores may have uncertain or inconsistently reported residences. In short, the
complexity of inclusion and exclusion rules for categories of residents may be too
cumbersome for easy administration in telephone interviews, and thus produce responses
of questionable validity. The best solutions we have found involve simplicity of wording
and follow-up probing when the household composition is first retrieved, and then a
verification by providing the household with its composition listing, as part of the travel
diary package, with a subsequent probe about whether anyone was excluded or incorrectly
included.

**Perceived Burden and Respondent Tactics To Ease Burden**

In dealing with item nonresponse, we are working with households who have indicated a
willingness to participate. Willingness to participate on the part of one or more household
members does not automatically mean that each person is prepared to absorb any burden
related to providing full data. Because travel surveys are generally multistage processes,
with two to four contacts with a household, the residents will have insights about what is
expected of them. Thus, we encounter “shortcutters” or respondents who figure out that
each vehicle they have generates three or more questions about it, or that each trip they
make is followed by a series of detailed questions about it, or that each time gap in their
day is followed by probing questions about trips or activities that are missing.
Respondents can quickly learn the basic rules of what is expected of them; most of the
time, this is what a good interviewer’s intentions and goals are, and this is used to great
advantage. However, other times, respondents figure out the process and decide to take
shortcuts in order to shorten the interview and reduce the burden of participating. How do
we foil this type of item missing data by short-cutters? Let’s examine a couple of
examples.

**Illustrative Problem #7**

The most frequent problem is skipped or unreported trips. Although some of these may be
simply forgotten, and some mnemonic queries can bring them out, other skipped trips may
be purposely omitted by respondents attempting to shorten the interview. The
unconsciously forgotten trips are a problem of memory lapses and comprehension failures and have been discussed previously. Underreporting of trips as a burden-avoidance tactic is another matter altogether. There are several ways to ameliorate the problem of this type of unreported trip. In the earliest stages of respondent contact—advance-notification mailings—it is important to stress the significance of compiling a complete picture of a traveler’s day. This can be done effectively in the short project brochures that we send to prospective respondents. It can also be stressed in the instructions and other materials included in the mailing of diaries. During the actual retrieval of data, the “hidden trip” can be extracted by the probes about going anyplace between the completion time of one trip and the beginning of the next trip. A variety of other types of probes to explore activities, trips, or stops can be inserted into the flow of a data retrieval instrument.

**Illustrative Problem #8**

Another problem, directly related to the underreporting of trips, is the identification and targeting of households or persons reporting zero trips. We cite this under the category of burden-reducing respondent tactics because survey data users have identified this particular category as a result of nonresponse or polite avoidance by respondents. In other words, these are considered prime candidate examples of missing data. Although the rationale for concern about trip underreporting is well-founded, it is probably unwise to target only zero-trip cases for challenge. It is just as likely that persons who report more than zero trips are also underreporting trips. The solution is therefore not to discard suspect cases of respondents with zero-trip days, but instead, to adopt thorough procedures for validation and more extensive probing for retrieving unreported trips. The suggestions just mentioned apply in this case as well.

The solution to retaining the focus of a respondent and to discouraging shortcutting is to make explicit, at the outset, the level of burden that is required, and to make disclosures that enhance the perceived relevance of the survey and the practical importance of complete data.

**Intrusions on Respondents’ Perceived Privacy and Security**

The recently completed analysis of travel surveys by the National Opinion Research Center concluded that perceived intrusion into privacy and security was one of the principal causes of item nonresponse and outright refusals. This is not surprising, particularly in the United States, where a public perception of over-surveying is becoming widespread. Real and imagined abuses associated with surveys, or with telemarketing inappropriately seen as the same as surveys, have had a serious negative impact on people’s attitudes about surveys and willingness to participate in them.

**Illustrative Problem #9**

In addition to outright refusals, there are indications that some respondents may intentionally give inaccurate responses to questions regarding their children’s activities or trips, location of their children’s schools, exact details of their schedules (or exact times not at home), possibly illicit trips, and other destinations that respondents may be reluctant to disclose. Other than honest and explicit assurances of confidential treatment of data, and clearer explanations about the purpose of exact-location information, little can be done
to persuade a respondent to provide this information. Other approaches we have found effective in getting responses adequate for geocoding involve immediate agreement with the respondent about the importance of holding certain information secure, and then asking for approximate location, in the form of the closest, or even a nearby, street intersection. As in many other situations, the effectiveness of interviewers and their creativity in making respondents comfortable are the optimal tool.

Illustrative Problem #10

Household income is by far the most prevalent problem in item nonresponse, and the most frequent instance of respondent’s refusal to answer is in the variable of household income. For refusals on the grounds of privacy, an alternative approach is to ask two or three “over or under” specified-level questions, or to ask the household income question on two or even three different occasions.

Intentional Misreporting

Item misrepresentation is a final category of difficulties with item nonresponse or item validity. Although deception may be too strong a word in certain ways, many respondents will consciously and knowingly give misleading or incorrect responses on certain questions. In general, respondents are motivated to be “good respondents” and to provide the information that is asked for. At the same time, they are motivated to be and to appear to be good people. They will try to represent themselves to the interviewer in a way that reflects well on themselves. This problem (social-desirability bias or response-delivery error) is a significant one in survey research. Many questions are about social desirable or undesirable behaviors. If respondents have acted in ways that they feel are not the socially desirable ones, they are placed in a dilemma: They want to report accurately as good respondents; at the same time, they want to appear to be good people in the eyes of the interviewer. Techniques for helping respondents resolve this dilemma on the side of being good respondents are required.

Socially conditioned responses that suggest or imply socially approved behavior, or hide the implication of socially sanctioned behavior, have been cited in the survey research literature extensively.

Illustrative Problem #11

Although no documentation exists, there is some anecdotal evidence from interviewers that some respondents in travel surveys tend to be very solicitous and cooperative. This tendency sometimes shows up with respondents’ seeking to find out what the “correct” answers might be. Social correctness may vary from one situation to another, but in questions regarding transport mode, there may be some inclination to overreport ride sharing and to underreport the use of an auto for very short trips.

POST-SURVEY SOLUTIONS: A BRIEF SUMMARY

This paper has concentrated on identifying missing data problems and proposing solutions for preventing and/or reducing such data gaps at the item level. Regardless of what is done to prevent or reduce item nonresponse, it is impossible to achieve perfect, complete data in
surveys. Further, not all variables or data items are created equal—for numerous reasons discussed in this paper, certain questions will tend to have more nonresponse and lower response validity. In addition to the many preventive measures discussed, there are three additional procedures that transportation survey researchers can use to compensate for item nonresponse: (a) item substitution; (b) independent allocation; and (c) statistical imputation.

**Item Substitution**

After a survey has been completed and analysts find unacceptably high rates of item missing data, a possible solution to make the overall survey data more useful is to merge in a substitute measure for the item with levels of missing data. Household income, again, is the most frequently cited problem item. In addition to high rates of refusal and other nonresponse, household income measures in transportation surveys do not always secure the same measure as in the decennial Census. Thus, substitute items for household income could be very useful, if they can be reliable and accessible measures of wealth.

In selected jurisdictions in the United States, a potentially effective substitute for household income is the actual value of the residence. In some places, the value of property can be accessed in the public domain; in other areas, it is possible to establish interagency arrangements that provide controlled access to the data for research purposes. In states or local jurisdictions where data on home values are available, it is possible to merge this measure and use it in lieu of household income. It is necessary, but relatively easy, to also implement an adjustment to take into account residential tenure (owning or renting) and type of dwelling (single-household unit, duplex, apartment in small complex, apartment in large complex, etc.), so that the independent, substitute measure of wealth better reflects the circumstances of the occupant household.

We have mentioned other variables in travel surveys that may not have much missing data but which have very poor external validity. A good example is respondent reporting of parking costs for specific locations. Frequently, parking costs in commercial locations and workplaces have cost components, particularly subsidies, that are hidden to most users. Thus, the reporting of parking costs (as opposed to actual payment for parking at a specific location on a specific trip) by respondents may be mostly a measure of perception and knowledge, rather than of actual, real values. It may be better to avoid this question and to substitute a more systematic measure that can be secured from other surveys or from available administrative records. Ultimately, the ideal substitution may be a measure drawn from the actual land value of the specific parking location; this might be the purest measure of the full, unsubsidized cost of using that parcel for parking.

**Imputation**

The most common approach for compensating for missing data in surveys is statistical imputation. Several widely used techniques are available, and their use depends on technical resources available, anticipated usage of the imputed data, and even on the comfort levels of the research consumers. There are fairly well-developed approaches that can be placed into three main categories. The first and simplest is to impute a value for a nonresponse item equal to the mean value for that item among a class of cases. The class of cases is defined according to specific demographic or behavioral indicators for the
specific case, with missing data on that variable. The various approaches for this type of imputation for a stratified mean vary by the number of variables that are used for the calculation of the class mean.

A second and more powerful calculation of a value to be imputed to a missing variable involves the use of regression models. Selected variables that predict the variable with missing data are calculated by using cases with full responses. The regression results are then used to determine an equation that can be used for producing a value for a case that has missing data on the variable in question.

Another approach to imputation is the “hot-decking” model. This approach duplicates a value on the variable in question from one case (the donor case) to the case with nonresponse on the specified variable. The determination of the donor case and the rules that are used for determining or limiting the use of a single case as a donor rely on the use of several variables that are thought to be associated with the variable with missing data. The goal is to find the case with a non-missing value that is most like the case with a missing value. Then, the non-missing value in the former is imputed (or donated) to the case with the missing value.

**Independent Allocation**

Another approach that we would suggest for the post-survey adjustment for missing data is the allocation of values on missing data variables that are drawn from sources independent of the survey. This approach has its richest potential for filling in missing data on household income. If full residential geocoding has been conducted for participating households, and they can be allocated to the smallest available Census geography—the block or a block group—characteristics of households in the respective geography can be linked to the individual households. Then, with procedures similar to imputation, the independent variables are added to the mix for finding a best-fit case to serve as donor (in hot-decking), or for calculating means or regression equations in the other approaches.

In any consideration of item substitution, independent allocation, or imputation, it is essential that the type of item and its purpose be carefully considered. When the item is a status attribute (household income, tenure, age, gender), these techniques will be more acceptable. When the item is a geographic variable, the techniques can be applied, but the results need to be thoroughly reviewed and edited so that the value generated “makes sense” or is plausible for the case in question.

**PROJECT BUDGETS: CONSTRAINTS AND SOLUTIONS**

It is necessary to address one additional issue that bears significantly on survey data quality—survey costs. Although this discussion has not dealt with the cost of alternative approaches, it is nevertheless a critically important issue. Anyone who collects survey data knows that one of the dominant influences on design decisions is the available financial resources for the survey. Contrasting with the emphasis on quality and reduction of error in survey research methods is the frequent disregard of those prescriptions, under the constraints of cost. The reduction of error requires the expenditure of scarce resources. The easiest example is the reduction of sampling error by increasing the sample size, requiring responses from more members of the population.
Similarly, from an item-based data problem perspective, construct validity is increased by increasing the number of indicators measuring the same underlying concept. This means more questions in the survey; more questions imply more time to complete the questionnaire. In surveys, as in business, time is money. Given the link between errors and cost, many new ideas require spending money to reduce an error. Given budget constraints, the reduction of one error often leads to the increase in another. Acknowledgment that surveys are inherent compromises is always a solid starting point for adopting optimal designs.

Recent concerns about data quality in the United States have stimulated welcome initiatives to conduct methodological projects. These efforts to test survey alternatives will likely produce new solutions to the nonresponse problem. Much more must be done, however. For example, efforts to find substitute, independent measures for respondent-reported household income and development of rigorous, multivariate imputation models would certainly deliver useful assistance to travel survey efforts.

CONCLUSION

Missing and invalid data in travel surveys, at the item level, are a complex and widespread problem. Defined broadly as item nonresponse, this problem is particularly intractable because many distinct causes and factors generate it. Although travel survey methodologists and users have not adequately addressed this problem, several fields involved in survey research have produced a rich and valuable literature that informs this topic quite well.

To understand the general problem of item nonresponse, it is useful to examine it within a classification system of causes, such as the approach presented in this paper. Then, within an acceptable model of causes of nonresponse, strategies and tactics for amelioration of nonresponse can be identified, tested and disseminated for wide adoption. It is hoped that the dozen or so illustrative examples discussed in this paper are a starting point for further development.

Finally, this consideration of the problem of item nonresponse has generated a clear awareness that much more needs to be done in pooling together the many practical field difficulties and creative solutions from many sites. This exchange would then require the empirical evaluation of innovations and/or alternatives to the way travel surveys have been conducted in the past. The momentum of the past several years, at meetings and conferences such as this one, needs to be further encouraged and more formally focused on specific approaches to data quality problems and solutions.

REFERENCES

2. Travel Survey Manual.


INTRODUCTION

Nonresponse is a common problem in household travel surveys. This nonresponse can take one of three major forms; specific item nonresponse, nonreported trips, and unit nonresponse. Specific item nonresponse occurs where the respondent has provided answers to most questions, but has failed to answer a specific question (e.g., the time of arriving at a particular destination). Nonreported trips are more extreme in that the respondent fails to tell us anything about specific trips or activities. This often occurs with respect to short-duration trips by non-motorized modes (e.g., walking to the sandwich shop at lunchtime). Unit nonresponse occurs where, for example, an entire household fails to respond to the survey. Workshop 3 was charged with looking at the first two of these forms of nonresponse, which were collectively titled “item nonresponse” (INR) for the purposes of the Workshop.

The Workshop essentially looked at two sides of the question of item nonresponse. First, it examined preventative actions that could reduce INR. However, while attention to good design and follow-up quality-control techniques can be used to minimize item nonresponse, there will always be some respondents who fail to provide complete information for all questions in the interview or questionnaire. Therefore, the Workshop also examined corrective actions once INR had been detected.

PREVENTATIVE ACTIONS

In considering preventative actions for INR, it was considered that a thorough understanding of the causes of INR were necessary. As noted by Zmud and Arce, in their Workshop resource paper, item nonresponse is the result of five major factors, namely,

- **Knowledge and Recall.** Sometimes, respondents are asked questions to which they simply do not know the answer. More frequently, respondents suffer memory lapses that produce missing or low-quality data. Memory lapses can include the forgetting of minor events, such as certain trips or activities, or the incorrect recall of these events. Events may sometimes be well-recalled, but the lapses affect the sequence or order of events, or their exact timing.

- **Comprehension.** Often, questions are difficult to understand and to answer. Writing clear and simple questions requires attention being paid to four important principles: simple language, common concepts, manageable tasks, and widespread information. The need for intensive work on the part of question-writers is exacerbated because of the disproportionate distribution of comprehension gaps (e.g., among those from non-English-speaking backgrounds).

- **Perceived and/or Real Burden.** Travel surveys in the 1990s have seen an increasing level of detail being asked from respondents. In an age of increasing demands on people’s time from all types of sources (e.g., family, job, community), perceptions about the time burden inflicted by travel surveys are an important factor.
• **Desire for Privacy or Concerns About Personal Information.** Questions about a household’s trip-making behavior can be as threatening to some respondents as questions about gambling, drinking alcohol, or sexual activities. These questions are considered threatening for a variety of reasons, including fear of consequences of divulging data, or distrust of the person asking the questions. This may be a special concern for persons living alone, who may feel threatened by divulging when they are at home, or when their house is empty.

• **Deliberate Misreporting.** For a variety of reasons, respondents may be tempted to give deliberately inaccurate answers. One motivation, as outlined above, can be the fear of consequences. In addition, the desire to present oneself in a favorable light or to give a good impression may be a strong motivation for some respondents. Most frequent is the desire to report socially acceptable behavior or to not report socially sanctioned behavior.

**CORRECTIVE ACTIONS**

Attention to the above five points in the design of the survey instrument and process can reduce the extent of INR. However, given that there is some level of INR in the data set, the question remains as to what can be done to account for this nonresponse. Essentially, there are four courses of corrective action that can be taken to deal with INR:

• **Ignore Missing Data.** The simplest option, and the one used most often, is to simply ignore the missing values on a case-by-case basis, that is, for each analysis (such as a frequency distribution, a cross-tabulation or a regression model) a record is ignored if it has a missing value for any of the required variables. This has the side effect that totals of distributions and cross-tabulations will be different because different records will have been omitted from each calculation. Unless the level of item nonresponse is high, or unless a large number of variables are used in the analysis, this effect will not be significant.

• **Remove Records with Any Missing Data.** Because it is easier to deal with a “clean data matrix” (i.e., one that does not have any missing data), one way of achieving this is to remove all records with any missing data, thereby ensuring that the data matrix contains no missing data. This form of data editing is, however, rather extreme and wasteful of data. An even more important consideration is the biasing effect that this process has on the remaining data. Since any household with item nonresponse is omitted from the final data set (which may then be used for data analysis or modeling), it stands to reason that, if item nonresponse is distributed randomly through the data, those households with more people and those people making more trips are more likely to be omitted because they have a higher chance of item nonresponse (simply because they are providing more data from which something may be missing).

• **Re-Weigh the Data.** As noted above, in the first method of dealing with item nonresponse, ignoring the missing data on a case-by-case basis when performing the analysis will result in different totals being obtained in the marginals of any tables. This applies when performing calculations on the sample data, but also when performing analysis on data that has been weighted to allow for expansion to the total population (e.g., by comparison with Census data). This is because the expansion weights have usually been calculated by comparing a cross-tabulation of the sample data (e.g., persons, by age and sex) with the population data to calculate the expansion weights. Any missing data in the sample cross-tabulation are treated in a specific manner, such as assigning the weight that
corresponds to the average of the missing variable. These weights are then attached to the records in the data set. In later analyses, however, different records will be ignored depending on which variables are being analyzed, and hence the sum of the weights of those records included in the analysis will not always be equal.

One way around this problem is to recalculate the expansion weights for every specific analysis conducted. Thus, the analysis sample is first determined by removing records with missing values for the variables in question, and then the weights are calculated before the analysis is performed. In this way, the population estimates will always agree with the totals in the population data set. This method can become unwieldy, however, since every new analysis creates a new set of weights. Very soon, there are more weights in the data set than there are real data!

• **Impute Missing Data.** The fourth method of dealing with item nonresponse is to impute (estimate) values for the missing data based on some other source of information. This method has the advantage that all data in the existing data set are used (i.e., no data are discarded), the imputation is done only once (compared with the multiple recalculations of weights using the re-weighting method), and a clean data matrix is obtained for future analysis. For these reasons, imputation is the preferred method of dealing with item nonresponse.

As noted by Armoogum and Madre, in their Workshop resource paper, there are a number of different methods of imputation that can be used with household travel survey data, namely:

• **Deductive Imputation.** This method allows a missing value to be replaced by a perfect prediction, based on a logical conclusion drawn from other data in the data set. This is often the case when redundant questions are asked in a survey, where missing responses to one question can be replaced by information derived from the other questions.

• **Overall Mean Imputation.** In this method, the missing value is replaced by the mean of that variable across all respondents in the sample. For example, a missing income would be replaced by the mean income of the respondents in the sample. This can be a dangerous method, unless the extent of item nonresponse is very small, because the method leads to reduced estimates of the variance (because all the imputed values are at the mean of the distribution), and hence leads to invalid confidence intervals.

• **Class Mean Imputation.** This method overcomes some of the problems of overall mean imputation by first dividing the sample population into strata, based on other variables in the data set, and then calculating the mean of the variable to be imputed within each stratum. The observation requiring imputation is then assigned to one of these strata, based on the values of the stratifying variables, and the mean of the variable within the stratum is assigned to the missing value. There will still be some reduction in variance using this method, but far less than would have occurred using overall mean imputation.

• **Hot-Deck Imputation.** In hot-deck imputation, missing responses are obtained by finding a record within the data set that is similar in all respects to the record with the missing value. The value of the variable (e.g., income) for this record is then substituted for the missing value. A variety of hot-decking procedures have been proposed, including random overall hot-deck imputation (whereby a set of records with similar characteristics
is formed, and the value to be imputed is obtained by random sampling from this set), random imputation within classes, sequential hot-deck imputation (where imputed values are obtained from the set of records by selecting each record in sequence) and hierarchical hot-deck imputation (where a set of records is developed with exact or non-exact matches to the target record, and then the better matches are used preferentially as the source of imputed data).

- **Cold-Deck Imputation.** Whereas hot-deck imputation uses information from the data set of the current survey, cold-deck imputation uses data from sources other than the current survey. In most other respects, cold-deck imputation is very similar to hot-deck imputation.

- **Regression Imputation.** In this method, a regression equation is estimated from the data set, and then used to predict the variable to be imputed from other variables within the data set. This method is useful when the use of class mean imputation stratification may result in a large number of empty cells within the stratification. Regression imputation allows these cells to be filled with information from neighboring cells.

- **Multiple Imputation.** In all the above methods, a single value of the imputed variable is obtained and substituted into the data matrix. With multiple imputation, a number of different values are imputed to create a number of “clean data matrices”, which are then analyzed as different representations of the complete data set.

Within this framework of preventative and corrective actions for INR, the Workshop considered a number of specific issues, as follows.

**Zero-Trippers**

A particularly severe case of INR is what was termed “zero-trippers”, i.e., people who said they did not travel at all on the survey day. It was considered important to distinguish between genuine zero-trippers (i.e., those who really did not travel on the survey day) and non-genuine zero-trippers (i.e., those who said they did not travel, merely to avoid having to complete the rest of the survey). It was considered, from a review of known surveys, that a normal value of zero-trippers was between 10% and 20% of the population (although those at the high end of this range probably contain a reasonable number of non-genuine zero-trippers). However, there did not seem to be any established diagnostic tools for the detection of non-genuine zero-trippers. One practice that was becoming standard was to ask zero-trippers for the reason for not traveling, although it was recognized that there were no standard codes established for this question. It was considered that another alternative was to ask zero-trippers when they last did actually make a trip. In this way, one could distinguish between the long-term housebound person and the normally active person who just happened not to travel on the survey day. It might also provide a diagnostic tool for statistically determining the extent of non-genuine zero-trippers, by comparing the intervals from their previous trips with their stated non-travel on the survey day.

A related issue was raised for surveys when the survey period was longer than one day. In particular, it was raised in connection with a fuel-use survey, where respondents were asked details about “fill-ups” they had made within a specified period. About 30% of all respondents stated that they made no fill-ups within this period. The question then was whether to assume that they use no fuel at all, or to re-weight or use imputation methods on the data. A similar problem exists in surveys of long-distance travel, where respondents
are typically asked details about long-distance trips they have made within a two- or three-month period. In such surveys, up to 50% of respondents have (truthfully or otherwise) reported that they made no trips. A possible solution to this problem is to change the questioning technique and ask when they last filled up or made a long-distance trip, thereby obtaining information from all respondents by not allowing an easy way out of answering the question.

**Activities as a Basis for Reporting**

It was considered that a useful method of reducing nonreported trips was to use activities as the basis for reporting of travel behavior. Several studies have shown increases in trip rates by asking first about the activities undertaken during the day, and then obtaining details about the travel used to connect those activities. Such an emphasis on activities also facilitates the collection of stage-based data, whereby access and egress trips are collected along with the “main mode” used on a trip.

Other techniques for reducing nonreported trips were to put the main activities (e.g., work) and modes of travel (e.g., car) at the bottom of lists, thereby forcing the respondent to read through the list at least once. In this way, they saw, along the way, that the survey also included trips to bus stops (as a change-mode activity) and by the slow modes (e.g., walk and bicycle).

In considering the activity-based approach, attention was focused on situations were non-reporting has always been a problem. How are “multiple-activity” situations handled (such as using a laptop computer on the train ride to the city), and how are multiple activities within the one site to be recorded? For example, a trip to the shopping mall may entail “trips” to several shops, the bank, a cinema and a restaurant. If these destinations were freestanding buildings along a street, they would probably be recorded as separate trips, connected by walking. But because they are within a single enclosed space, many people think of them as one trip. This is probably OK if the interest is only in the travel associated with the outing, but not if the interest is in the activities that gave rise to the demand for the outing. Developing a means of recording multiple activities with a single “site” was seen to be an essential prerequisite to obtaining data sets that reflected the full range of activities undertaken.

**MISSING DATA STANDARDS**

In discussing the issue of INR within the Workshop, it became clear that there was a need for documentation of the levels of INR that are acceptable in various types of surveys. For example, it was generally accepted that income had one of the highest levels of INR, but this ranged from 7% up to 20% and beyond, depending on the question design and the type of survey instrument. Other variables had levels of INR in the range of 1% to 5%. More care in documenting INR was seen as essential, and a meta-analysis of INR was seen as a highly useful project.

**Missing Income**

As noted above, income suffers from one of the highest levels of INR. It may also suffer from a high level of misrepresentation, since the definition of income was often far from clear in the survey instrument. One participant reported on inconsistent answers being
obtained when the same income question was asked twice in a survey. Another noted that the U.S. Census Bureau uses 25 questions to obtain estimates of income, thereby casting doubts on our ability to obtain it in a single question in our travel and activity surveys. Several participants asked why we wanted income information, and the answers to this question strongly suggested that INR for income could be reduced by asking for income categories, while still obtaining income estimates that were sufficiently accurate for most policy and modeling purposes.

**Missing Distances and Locations**

Discussions about the problems of INR in connection with trip distance and duration led to a fruitful discussion about the need to collect distance information at all, in these days of widespread use of GIS. Unless the analyst was specifically interested in the perception (and misperception) of trip distance by respondents, it was considered more useful to collect more detailed geographic information about the location of activities, and thereby (using GIS techniques) obtain more precise and reproducible estimates of trip distance, with less potential for INR. This then led to methods of minimizing INR for activity locations in an age of GIS. While most GIS packages require full street addresses for accurate geocoding of locations, most respondents do not know the full street addresses of the places they visit.

Therefore, acting on the conference motto of “respondents are customers”, and knowing that “the customer is always right”, the design of the activity location question should ensure that it obtains the best information from the respondents, in the format that they know best. Thus, for example, one should ask for the name of the shop they visited, rather than the address. By assembling a GIS database of shop names and x-y coordinates, one can then proceed directly to geocoding without needing a full street address. An example of such a system in Melbourne, Australia, whereby over 50,000 such “landmarks”, including all shops, were compiled, mainly by geocoding of the addresses given in the electronic version of the telephone books, was described. In this way, the INR for activity locations, and hence the trip-distance INR, was effectively reduced to zero.

**Required Precision of Imputation**

In discussing the various methods of imputation, it became clear that different situations needed different levels of precision of imputation. For example, in imputing missing ages, one may need an estimate to the nearest year or five years for some purposes, while in other situations, it may be sufficient to just distinguish between a child and an adult. Similarly, arrival times may be required to the nearest five minutes, or just to morning or afternoon. In such situations, it may be more advisable to simply create a new variable (such as child/adult) and impute a value for the new variable, rather than for the original variable.

**Imputation—How Far Can You Go?**

While this Workshop was concentrating on item nonresponse, it was clear that there was considerable overlap between this type of nonresponse and other types of nonresponse. The question therefore arose as to how much could be imputed. If it was acceptable to impute for an item nonresponse rate of 5%, was it acceptable to impute for an item
nonresponse rate of 50%? Was it acceptable to impute entire nonreported trips? Was it acceptable to impute missing persons within a household? Was it acceptable to impute missing households? While this latter proposition seemed to be going too far, it did raise the important difference between imputation and weighting for missing data. The problem with weighting is that it gives more weight to those items of data that have been reported, without correcting for data that have not been reported. Imputation gives the possibility of “additive” weights, rather than the conventional “multiplicative” weights, thereby adding data that is missing rather than multiplying data that is not missing. The interrelationships between item nonresponse and unit nonresponse were seen to be worthy of further investigation.

**Documentation of Imputed Variables**

No matter how the imputation of missing data was performed, it was seen to be essential that the imputation process be fully documented. It is important that the analyst be fully aware of which data came originally from the respondent, and which data were imputed by the survey analyst. One method of doing this was by means of a triplicate data set; one set of data contains the original data, one contains the final data, including imputations, and the third contains (for each value of each variable) the method by which the original data were transformed into the final data set. A complimentary document would describe in full the imputation methods listed in the third data set.

**Biases from Imputation and Multiple Imputation—Care with Modeling**

If data sets, including imputed values, are used in later modeling exercises, care needs to be taken to be aware of and allow for any biases introduced as a result of the imputation process. In many cases, a “model” is used as part of the imputation process (for example, with regression-based imputation), and any later modeling exercises may simply be reflecting this model, rather than the underlying data. A useful study was seen as consisting of a simulation exercise in which different imputation methods were used before the development of various models, to see whether the imputation methods significantly affected the outcomes of the modeling.

**Imputation Systems**

While most of the discussion focused on the merits and mechanics of imputation for single variables, there was some discussion on the process of imputation for an entire survey. While most analysts concentrate on imputation for those variables with the highest extent of missing values (e.g., income), it was suggested that it might be more sensible to reverse this process by using a three-stage imputation process, as follows:

1. Use deductive imputation to correct for the most obvious missing values.
2. Use one of the various imputation methods to correct for those variables with low levels of missing data.
3. Use one of the imputation methods to correct for those variables with high levels of missing data, perhaps using imputed values from step 2 as inputs into the construction of the imputation models for the poorly reported variables.
It was also suggested that there exists a whole range of more complex imputation processes, such as expectation maximization, which use an iterative combination of modeling and imputation processes to derive data sets that allow for ignorable and unignorable nonresponse effects. Further research needs to be performed on real data sets to see how these methods work in practice, and the effect they may have on later full-scale modeling efforts.

**SUMMARY OF INR ISSUES**

As a result of the discussions in the Workshop, the following issues were seen to be most important and most in need of further attention:

- The relationships between item nonresponse and unit nonresponse.
- The desirability of a meta-analysis of INR, to document acceptable ranges of INR for different variables and survey methods.
  - The preference for imputation of missing values, rather than re-weighting.
  - The desirability of developing additive weights rather than multiplicative weights.
- The importance of seeing imputation as a process for the entire survey, rather than as a variable-by-variable correction.
- The desirability of using imputation schemes (such as probabilistic imputation) that preserve the inherent variability in the data, rather than reducing the variability (as with mean imputations).
- The need to investigate the implications of imputation for modeling, perhaps through various simulation studies.
- The biasing effects of simple techniques of dealing with INR, such as removing all records with missing data.
- The need to develop full documentation of the imputation process employed in a survey, such that users of the data are fully aware of the imputation methods used.
- The need to distinguish between genuine and non-genuine zero-trippers, and to develop survey methods to minimize the extent of non-genuine zero-trippers.
- The value of activity-based travel diaries in reducing the extent of nonreported trips.
- The need to take advantage of developments in GIS technology, which will assist in minimizing the extent of INR.