TOWARD THE DEVELOPMENT OF AN ORDERED SET OF DATA QUALITY CATEGORIES

Clyde Tucker, Bureau of Labor Statistics
600 E Street, N.W., Washington, D.C. 20212

1. Introduction

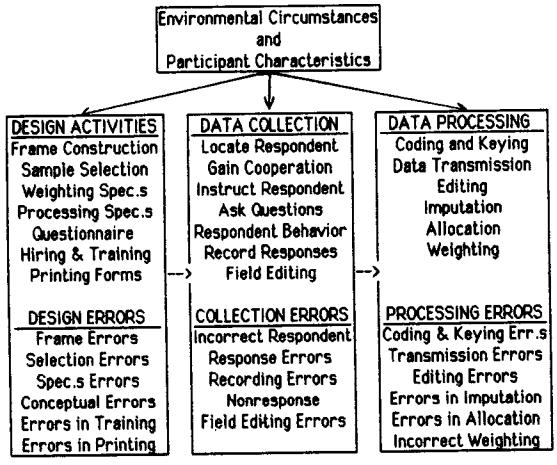
Traditionally, the measures of a survey's overall quality with respect to nonsampling error have been the response rate and, occasionally, a comparison to either an independent source for the aggregate survey statistics or the summary results of reinterviewing or record checks. A more comprehensive measure is needed, and this paper represents a first step toward developing such a measure. This new variable consists of a set of ordered categories, and each sampled unit is assigned to a category based on an assessment of the quality of the data received from it. The categories are derived by taking account of the pattern of nonresponse from a potential respondent (sample unit in scope) in conjunction with indirect, micro-level measures of response errors. An analysis of data from various demographic subpopulations in the Bureau of Labor Statistics Consumer Expenditure Diary Survey serves to illustrate the utility of this new measure.

2. The Survey Process, Human Behavior and Nonsampling Errors

2.1 Behavior and Data Quality

Quality in the survey process must be defined in terms of the behaviors of the individuals participating in that process. Both Morris Hansen (1987) and Janet Norwood (1987) described the quality assurance programs of their respective organizations in terms of particular behaviors of organization personnel. In essence, quality in the final survey product depends on the behaviors that were performed in the "manufacturing" of that product. Defining quality behavior is not a simple matter. For instance, we often do not know which question wording or ordering will produce the best data. Furthermore, the quality of the final product is not easily judged. Measurement of nonsampling errors accomplishes, in a sense, both of these tasks. Not only can these errors indicate the level of quality in the final product, but they also can point to problems with specific behaviors in the survey process.

Fig. 1. The Survey Process



2.2 The Survey Process and Nonsampling Error

Figure 1 depicts the survey process as a system of interrelated components. Each component of the system (design, collection, and processing) is a constellation of behaviors including a number of social interactions. Errors result directly from incorrect behaviors which, in turn, are products of previous behaviors in the system, environmental circumstances and participant characteristics.

Representation error occurs in the design phase through faulty frame specification or construction and the miscalculation of weights. Systematic measurement error (measurement bias) and measurement fluctuation (measurement variance) occur at one of three points--response emission, the recording of the response or the processing of the response. The causal factors, however, often are found in the design phase. A number of errors are possible in the processing phase, and the more complex the survey, the more likely it is that these errors will occur.

2.3 Measurement Error and Nonresponse

2.3.1 Types of Nonresponse

In studies of nonresponse, two types are commonly identified--unit and item. Unit nonresponse is the case where no (or very little) useable data are received from the respondent. Item nonresponse is defined as the failure to obtain data for individual items. Another type of nonresponse is partial nonresponse. In this case the respondent terminates the interview in the middle, failing to answer any further questions. Unit nonresponse and item nonresponse are analyzed separately more often than not. As Kalsbeek (1980) points out, however, when looking at the response rate for a particular item, "... nonresponse conceptually reduces to the item level ...."

2.3.2 The Relationship Between Measurement Error and Nonresponse

A majority probably hold the view that, in the case of unit nonreponse, nonresponse produces a sampling bias. The primary method used to overcome the effects of unit nonresponse is a weighting adjustment designed to restore the correct probabilities of selection for significant population subgroups (See Kalton and Kasprzyk, 1986.). In the case of item nonresponse, some form of imputation is the usual solution (Kalton and Kasprzyk, 1986). It is more common in this situation to see a discussion of measurement error, particularly response bias, than with unit nonresponse (See Platek and Gray, 1983.). The likelihood of error resulting from imputation is increased when the nonresponse is "nonignorable" (Rubin, 1976; Little, 1980, 1982; Greenlees, Reece and Zieschang, 1982).

Rather then treat unit and item nonresponse differently, I believe that every type of nonresponse--unit, item, partial--should be considered measurement error rather than sampling bias. This particularly seems appropriate when viewing all nonresponse from the item level. The measurement error resulting from nonresponse is usually some form of response error. After all, most cases of nonresponse involve some form of respondent behavior. It is just that the behavior is something other than a valid response to the survey or survey item. Methodologies for dealing with nonresponse are not dependent on the category of nonsampling error in which nonresponse is placed.

Treating nonresponse as a special type of measurement error focuses attention on the behaviors in the survey process largely responsible for nonresponse in the first place. Although there are certain features of the environment beyond the control of survey designers, the failure of the survey procedures to obtain valid responses must be addressed. Certainly, a change in the sample design would not solve the problem.

2.3.3 Fixed and Variable Response Probabilities

Kalsbeek differentiates between the "deterministic" and the "stochastic" views of nonresponse. In the deterministic case, potential respondents fall into one of two fixed strata--respondents and nonrespondents--with respective response probabilities of 1 and 0. From the stochastic viewpoint, response probabilities vary between 0 and 1, which makes for a more complex response model. To avoid confusion with the imputation methods known also as deterministic and stochastic depending on the absence or presence of an error term, the two views of response probabilities will be referred to, henceforth, as "fixed" (deterministic) and "variable" (stochastic).

Kalsbeek prefers the variable approach to setting response probabilities because he believes chance is a factor in the determination of response probabilities. From the standpoint of a scientific analysis of the survey process, however; chance might be given a less important place in the production of measurement errors, including nonresponse. Nonresponse and other forms of measurement error should be viewed deterministically. This does not mean that the prediction of measurement errors will not be subject to error because some causal variables have yet to be taken into account, but the critical task is to find these missing variables and not to ignore them.

2.3.4 The Patterns of Nonresponse

Variable response probabilities are useful to the extent that they can serve as weights to correct for nonresponse which is not random (Little and Rubin, Ch. 4, 1987). The probabilities, however, are not always as accurate as one would like. Often a potential respondent is assigned an aggregate probabilitiy based on modelling fixed probabilities. These probabilities are simply response rates for similar individuals. To truly have micro-level, variable probabilities requires repeated measurements under the same conditions (Lessler, 1983). Unfortunately, the conditions are seldom exactly the same, especially when compared to the initial measurement.

In developing procedures which prevent or, at least, reduce the amount of nonresponse, attention must be paid to the differences in the characteristics of potential respondents. An attractive feature of variable response probabilities, especially those formed from observations at the micro-level, is that they provide a way of differentiating between respondents not possible in the dichotomous, fixed probability case. There is a way, however, to gain greater differentiation and, at the same time, use fixed probabilities. Instead of limiting the response strata to two, respondents and nonrespondents, define more strata by viewing the decision to respond to the survey as a series of decisions or behaviors. This notion now has been applied to the analysis of response to surveys by Thran, Marder and Willke (1986) who developed a three category measure of responsiveness. Response strata can be defined by the patterns of item nonresponse.

2.3.5 Taking Account of Both Nonresponse and Other Measurement Errors

The set of response (or nonresponse) strata is only part of the information needed to form an ordered set of data quality categories for the Diary Survey. Measurement errors from other sources also must be taken into account. The estimates of the total measurement error for a case in any survey including the Diary will depend on the following principles no matter what the method used to adjust for nonresponse:

1. The probabilities of response are fixed after being conditioned on all causal variables (0 and 1).
2. Given that the design is not interpenetrated, measurement bias and measurement variance cannot be separated.
3. Measurement errors should be, in general, greater for nonrespondents than for respondents.
4. An idea of the relative magnitude of measurement errors can be obtained at the micro-level from information within the survey itself.
5. The impact of missing data varies by item content, and this impact can be estimated.
6. Indicators of the relative magnitude of measurement errors (taking into account both the extent and the impact of nonresponse) can be joined together to produce an ordered set of data quality categories.

Using these principles and adapting results presented by Kalsbeek (1980) and Platek and Gray (1983), a model can be specified which combines the effects of measurement errors due to either response or nonresponse. Given that response has been conditioned on the causal factors, the model is

$$x_{ij} = p_{ij}\,(x_{ij} + {}_{R}e_{ij}) + (1 - p_{ij})\,(x_{ij} + {}_{NR}e_{ij}) \quad (1)$$

where x_{ij} is the estimate of characteristic j for respondent i, p_{ij} is the probability of response to item j for respondent i (1 or 0), $x_{ij} + {}_{R}e_{ij}$ is the true value plus measurement error if $p_{ij}=1$, and $x_{ij} + {}_{NR}e_{ij}$ is the true value plus measurement error if $p_{ij} = 0$.

The response pattern from the sample of potential respondents (those in scope) can be represented by $\underline{D}$ which is an n x q matrix (n respondents and q items) where the entries are the p_{ij}'s. Ideally, the entries in $\underline{D}$ would be inserted in place of the p_{ij} to form $\underline{E}$, the matrix of errors in responses or estimated responses in the cases of nonresponse. Because rough estimates of the e_{ij}'s for only a portion of the items or groups of items initially will be available, the elements of $\underline{D}$ are the proportions of sets of items which contain valid responses and are used in conjunction with values which estimate the extent of measurement error in each set of items to procduce $\underline{E}$. Then, the vector $\underline{s}$, the rank of each group of items in terms of substantive importance, is used with $\underline{E}$ to determine data quality.

3. The Study Design

3.1 Description of the Data Set

Data quality categories are formed and used to evaluate the Consumer Expenditure Diary Survey conducted by the Bureau of the Census for the Bureau of Labor Statistics. The information collected during the second quarter of 1984 is used for this purpose. Although the Diary collects data on all expenditures made locally during a period of two consecutive weeks, it actually was designed to provide information about small, frequently-purchased items which are often difficult to recall such as grocery items. In addition, demographics and income information are collected. During the second quarter of 1984, a supplement to the survey was administered to all sampled units with a diary for the second week in order to ascertain the respondent's attitude toward the Diary and the way in which the Diary was kept.

The unit of analysis in the CE Diary, and the level at which most data are collected, is the consumer unit. A consumer unit is defined as one of the following: (1) the collection of all members of a household who are related by blood, marriage, adoption, or other legal arrangement; (2) a person living alone or sharing a household with others or living as a roomer in a private home or lodging house or in a permanent living quarters in a hotel or motel, but who is financially independent; or (3) two or more persons who live together and pool their incomes to make joint expenditure decisions. For further description of the CE Diary Survey, including the sample design and weighting procedure, see U.S. Department of Labor, 1986.

Potential respondents during the second quarter of 1984 are considered to be the 1510 units in scope during either week of the diary period. Of this number, 1303 units participated both weeks, and 1339 participated for at least the second week. There were 176 respondents

participating during the second week who did not complete the supplemental questionnaire. The only weight comparable for respondents and nonrespondents is the product of the original probability of selection weight and a subsampling factor, referred to here as the selection weight. When all potential respondents are analyzed, this weight is used, and tests are performed at the .01 level to account for the design effect. Often two-week diary respondents who completed the supplement are analyzed separately. In these cases, an average of the final weights (including various adjustments such as one for noninterviews) for the two weeks is used within the framework of balanced half-sample replication in order to include the design effects in tests at the .05 level.

3.2 Estimation and Analysis

Creating $\underline{s}$ involves evaluating the relative importance of the various parts of the CE Diary Survey in terms of their value for economic statistics. $\underline{D}$ is developed by determining the status of the response made by every potential respondent to each section of the survey considered in creating $\underline{s}$. This status is measured in terms of the proportion of section items with a valid response. To create $\underline{E}$, internal indicators of the accuracy of the data from each potential respondent are used.

Once all of the information is combined, and the set of quality categories is available, the distribution of the subpopulations across these classes are examined. Most of the analyses are done using log-linear models. In the cases where the replicate structure can be exploited, the log-linear models are estimated using Robert Fay's CPLX program (Fay, 1982 and 1985).

4. Identifying and Rating the Importance of the Sections in the CE Diary Survey

The CE Diary is a complicated survey which contains two distinct parts, a personal interview and a diary, with several sections within each of these parts. The personal interview is used to gather information about household characteristics, but the diary requires the respondent to take a more active role. The separate sections of the diary evaluated to create the elements of $\underline{s}$ are (1) household demographics used in weighting (race, household size, and home ownership), (2) demographics of the CU's reference person (comparable to head of household), (3) a set of consumer unit characteristics which includes housing information, vehicle ownership, a description of the frequency and content of grocery store purchases, and some demographics, (4) income and work experience information used to compute the total income for the CU, (5) expenditure information in two weekly diaries, (6) check-item sections for recall information, (7) the supplement, and (8) a record of house guests and CU members away during the two diary weeks.

The impact on both univariate estimates and multivariate relationships are considered in rating the importance of the different sections. Ratings of the importance of each diary section are in Table 1. These ratings are based on a ten-point scale with ten (10) being "essential" and one (1) being "unnecessary."

Table 1 - RATINGS OF THE IMPORTANCE OF DIFFERENT TYPES OF INFORMATION IN THE DIARY ($\underline{s}$)

Section	Rating
Expenditure Information (Diary)	10
Expenditure Information (Check-Item Sct.s)	9
Weighting Demographics (Household Level)	8
Income and Work Experience	8
Reference Person Demographics	7
Other Consumer Unit Characteristics	6
The Diary Supplement	5
Record of House Guests and CU Members Away	3

5. Measuring Nonresponse in the CE Diary Survey

5.1. Nonresponse in the Sections of the CE Diary

Table 2 provides a list of the eight nonresponse indicators and, except for INCWPROP, the number of items used in their creation. INCWPROP is based on an estimate of the amount of useful data in the income and work experience sections. Where proportions for each week were calculated, the indicator value is the average of the two proportions unless the unit was out-of-scope for one of the weeks. In these cases, the proportion calculated for the in-scope week is used as the indicator value.

Table 2 - THE INDICATORS OF NONRESPONSE IN THE SECTIONS OF THE CE DIARY ($\underline{D}$)

Indicator	Section	No. of Items
EXPNPROP	Expenditure Info. (Diary)	1
CHKPROP	Expenditure Info. (Check Items)	29
WTPROP	Weighting Demo.s (Hsld Level)	3
INCWPROP	Income and Work Experience	1
REFPROP	Reference Person Demo.s	6
CHARPROP	Other Consumer Unit Char.s	14
SUPPROP	The Diary Supplement	36
MGPROP	Record of Guests and Mbrs Away	3

5.2 Relevant Respondent and Consumer Unit Characteristics

One household characteristic which has been consistently associated with nonresponse is family size (Paul and Lawes, 1982; Burt and Cohen, 1984; Silberstein, 1986). Instead of household size, I previously used a measure which combined aspects of household size with a description of the relationships between family members. This measure, consumer unit composition, also will be used. Age and the ethnic origin of the respondent also have been associated with response quality (Ferber, 1966; Lowenstein, Colombotos and Elinson, 1975; Weaver, Holmes and Glenn, 1975; Garner and Blanciforti, 1987). Even though education has not been linked closely to data quality, it will be included along with age and ethnic origin in this study. Finally, although a previous attempt to identify important environmental variables was not successful (Tucker, 1986), another is made using region and degree of urbanization.

5.3 The Creation and Analysis of the Response Strata

Creation of the response strata takes two simple steps. The first step, stated formally in Eq. 2, yields $\underline{r}$, the vector of n weighted sums of the eight nonresponse indicators. These values range from zero to a complete response score of 56.

$$\underline{r} = \underline{D}\,\underline{s} \qquad (2)$$

In the second step, the values in $\underline{r}$ are grouped into classes to form the response strata variable, RSUM. The five categories of RSUM along with their frequencies are given in Table 3. Stratum 5 includes 346 respondents with a score of 56, but there are only seven cases which have a zero score.

Table 3 - THE RESPONSE STRATA

Stratum (RSUM)	Range of Values for r_i	Weighted Frequency
5 (High Response)	$r_i >= 55$	721.9
4	$50 <= r_i < 55$	354.5
3	$40 <= r_i < 50$	248.9
2	$20 <= r_i < 40$	48.9
1 (Low Response)	$r_i < 20$	135.8

Although $\underline{s}$ and $\underline{D}$ are to be used along with $\underline{E}$ to produce a set of data quality categories; RSUM, in itself, is a measure of response quality and can be analyzed as

such. Tables 4A and 4B (following the text) provide information on the relationships between consumer unit characteristics and RSUM for the entire data set and the smaller group of 1163, respectively. The fact that the only significant relationships in Table 4A involve degree of urbanization and region may be the result of missing data in the other demographic variables. In any case, residents of the Northeast and central cities appear to be the least responsive while respondents in rural areas are the most responsive.

Turning to Table 4B, notice that only the three highest values of RSUM are present in the subset. The chi-square tests based on simple random sampling (SRS) show that region, degree of urbanization and age of the reference person are significantly related to RSUM. When the effects of the complex design are taken into account, however, only region and age (marginally in the latter's case) are related to RSUM. Western respondents are the most responsive, and those in the Northeast the least. Region and age were entered into a log-linear model together to explain RSUM, and each still had a significant effect while controlling for the other.

6. Measurement Error in the Sections of the CE Diary

6.1 Error in the Expenditure Reports

The calculation of the Re_{ij}'s and the NRe_{ij}'s is much more difficult than defining $\underline{s}$ and $\underline{D}$, and only an intermediate solution to the problem is used here. The analysis from this point on will be confined to the data set containing the 1163 respondents who completed both diaries and the supplement. None of the 1163 failed to complete the diaries, even if some were done through recall; therefore, no NRe_{ij}'s are needed for the diary sections. Estimates of the Re_{ij}'s in the expenditures for food for home consumption were developed in an earlier paper; however, included in the information used to arrive at these estimates was a dichotomous measure of income nonresponse. In addition, these Re_{ij}'s were for the combination of expenditure reports for food from the diaries and the check-item section. New estimates of response error for these expenditure reports which are independent of income nonresponse were generated, but these estimates still apply to the combination of the diary and check-item reports. They will serve as the response error estimates not only for the combination of food reports but also for the entire expenditure profile given in the diaries and the check-item sections.

The respone error estimates are generated from a latent structure analysis of the relationships among indicators of response error developed from information in the survey itself (See Goodman (1974), Clogg (1977) and Tucker (1985) for a description of this procedure.). This procedure assumes response error in the diary reports comes largely in the form of underreports. In this particular case, three of the indicators from the original analysis contained in the 1986 paper were used. These indicators are (1) the discrepancy between the respondent's estimate of typical food expenditures provided prior to keeping the diary and the food expenditures reported during the two-week diary period, (2) the difference between the food expenditures reported in each of the diary weeks, and (3) a measure of respondent style developed from the respondent's answers to the questions in the diary supplement.

The chi-square test of the relationship between the three internal indicators of error prior to the creation of the latent variable is significant. The latent variable is, from a theoretical standpoint, a complex, unobserved measure of response error which accounts for the interactions among the three manifest error indicators. Once this latent variable is considered, the relationship between the three, original variables is no longer significant. Respondents are assigned to one of three error classes--low, moderate and high--based on modal probabilities for each cell of the original three-way table.

To use the new latent variable as the representative of the Re_{ij}'s for the expenditure reports, the categories must be assigned appropriate numeric values. The same scale used for nonresponse is used here, but a value of 1 is never achieved. The lowest error category is given a value of .89, the moderate-error category a value of .78, and the highest error category a value of .50. This scale relies on the assumption that every expenditure report will have at least some error.

6.2 Error Measures for the Other Diary Sections

The estimate of the e_{ij}'s for the other diary sections will rely on both logic and several assumptions. The most important assumptions are (1) there is often some measurement error in any given set of responses and (2) a response will usually contain less error than a nonresponse which must be imputed.

It is unlikely that many errors will occur in the responses to the weighting demographics. No respondent in this subset had more than one missing value; therefore, the following error values are assigned to the responses in this section:

No missing values	.98
Missing race	.92
Missing tenure	.84
Missing family size	.75

The following values for measurement error were assigned for income and work experience:

INCWPROP = 1	.90
INCWPROP = .8	.72
INCWPROP = .5	.45
INCWPROP = .4	.36
INCWPROP $\leq$ .3	.27

Each of the last four scores are calculated by multiplying .9 by the value of INCWPROP. The lowest value is set at .27 because it is assumed that an effective imputation procedure will restrict the amount of error to some extent.

The following method for assigning measurement error values was used for reference person demographics:

No missing values .92

Beginning with .92:

If sex is missing, subtract	.02
If race is missing, subtract	.03
If education is missing, subtract	.06
If ethnic origin is missing, subtract	.06
If age is missing, subtract	.10
If occupation is missing, subtract	.12

There are fourteen other consumer unit characteristics. The eight concerning home and vehicle ownership are relatively straightforward and are not likely to have much response error. The remaining four variables measuring grocery buying habits are subject to a fair amount of response error, and imputation would just increase the problem. Based on the above, the following method of assigning measurement error was used:

No Missing Value .88

Beginning with .88:

If one of the eight variables dealing with home and vehicle ownership is missing, subtract .03.

If the CU size variable is missing, subtract .04.

If the employment pattern variable is missing, subtract .08

If the three-category variable regarding grocery trips is missing, subtract .03.

If the open-ended variable regarding grocery trips is missing, subtract .09.

If the estimates for different purchases are missing (either of two variables), subtract .14.

The questions in the diary supplement are subject to different types of response error. On the other hand, if the amount of missing data is great, the direction of the values for these variables might be assumed. Given this, the following method for assigning levels of measurement error was used:

SUPPROP = 1	.85
1) SUPPROP) .50	.85 x SUPPROP
SUPPROP$\leq$ .50	.85 x .5

The following method for assigning measurement error to for the house guests and CU members away items:

MGPROP = 1	.95
1 > MGPROP ≥ .50	.95 x MGPROP
MGPROP ≤ .50	.95 x .5

7. Development and Demographic Analysis of the Data Quality Categories

7.1 Development of the Quality Categories

Using the information in D' (a reduced form of D excluding the rows for the additional 347 potential respondents) and the error calculations described in the previous section, E' (again, the reduced form) is specified. The variables summarizing the error calculations are listed in Table 5. Notice that EXPNERR is used for both the diary and the check-item sections. Once E' is defined, Equation 3 is used to produce q', the vector of quality scores for each of the 1163 respondents.

$$q' = E' s \qquad (3)$$

At this point, the values in q' (ranging from 32.005 to 50.77) are grouped into convenient categories. The categorization scheme is described in Table 6, and the resulting categorical variable is named QUALITY.

Table 5 - INDICATORS OF THE MEASUREMENT ERROR FROM RESPONSE OR NONRESPONSE (E)

Section	Indicator
Expenditure Info. (Diary)	EXPNERR
Expenditure Info. (Check Items)	EXPNERR
Weighting Demo.s (Hsld Level)	WTERR
Income and Work Experience	INCWERR
Reference Person Demo.s	REFERR
Other Consumer Unit Char.s	CHARERR
The Diary Supplement	SUPERR
Record of Guests and Mbrs Away	MGERR

Table 6 - THE DATA QUALITY CATEGORIES

QUALITY	Range of Values for q_i	Weighted Frequency
4 (High Quality)	$q_i >= 50$	212.6
3	$45 <= q_i < 50$	439.4
2	$40 <= q_i < 45$	420.3
1 (Low Quality)	$q_i < 40$	90.7

7.2. The Quality Categories: A Demographic Analysis

Table 7 (following the text) contains the results of the crosstabulations of QUALITY and the six demographic characteristics considered previously. Testing the significance of these relationships at the .01 level; age, ethnic origin and region are associated with the data quality measure. In addition the relationship between QUALITY and CU composition is of borderline significance.

The youngest age group has a particularly large percentage of its members in the bottom two categories. Almost half of the elderly also are found in these categories. The two middle age groups appear to have better quality data. Blacks and hispanics have somewhat lower quality data than other ethnic groups. As in the analysis of RSUM, the respondents in the West are the best ones. Those in the Northeast, however, are no worse than the other two regions. It is interesting that, as in earlier studies (Tucker, 1985 and 1986) and also in the findings for RSUM, education is not related to the quality of response.

8. Discussion and Recommendations

8.1 Discussion

The primary achievement of this paper has been the development of a measure of nonresponse which can be used in conjunction with other indicators of measurement error to evaluate the overall quality of survey responses. In addition, the nonresponse measure, by itself, provides important information. For instance, it seems clear that the social, or perhaps cultural, environment plays a significant role in determining the level of nonresponse. This may be a reflection of more relaxed or less threatening surroundings, or it may simply say something about the people who live in these surroundings. The type of interviewers in different areas of the country also could be a factor. Since individual characteristics are not necessarily randomly distributed with respect to geography, significant interactions may be present.

The development of the data quality categories still leaves much to be desired. This is partly because not enough is known about response error for sections other than expenditures. Estimates for response error for these different sections were not based on micro-level data. The effects of nonresponse on measurement error also is largely a matter of speculation. With respect to this latter problem, a method for estimating these effects at the aggregate level has been suggested by Rubin and Schenker (1986). The method, multiple imputation, can provide not only estimates of the increase in variance, but it also can give an indication of the impact of different assumptions about the relationships between nonresponse and the values of the item in question.

Other problems exist with the methodology employed in this paper. The weights for the relative importance of the diary sections and the measurement error scales were fairly arbitrary. Also, there appear to be large design effects in this survey; and the size of the design effects depend on the variables involved.

8.2 Recommendations

1. Conduct experiments to examine the effects of different methodologies in both urban and rural areas in different parts of the country.
2. Conduct studies concerning the relationship between geography and interviewer styles.
3. Conduct research similar to that here but with larger sample sizes.
4. Use multiple imputation to evaluate the effects of nonresponse.
5. Create measures of response error for the other diary sections.

References

Burt, Vicki L. and Cohen, B. (1984), Levels of Item Nonresponse in the National Medical Care Expenditure Survey, Paper presented to the Stat. Sect. Amer. Pub. Heal. Assoc.

Clogg, C.C. (1977), Unrestricted and Restricted Maximum Likelihood Latent Structure Analysis -- A Manual for Users, Working Paper No. 1977-09, Population Issues Research Office, Penn. St. Univ.

Fay, Robert E. (1982), Contingency Table Analysis for Complex Designs: CPLX, Proc. Sect. Surv. Res. Meth. Amer. Stat. Assoc., 44-53.

Fay, Robert E. (1985), A Jackknifed Chi-Squared Test for Complex Samples, Jour. Amer. Stat. Assoc., 80, 148-157.

Ferber, Robert (1966), Item Nonresponse in a Consumer Survey, Pub. Op. Quart., 30, 399-415.

Garner, Thesia I. and Blanciforti, Laura A. (1987), Reporting of Household Income: Complete Versus Incomplete Response, Paper delivered at the 3rd Ann. Cens. Res. Conf., March 29-April 1, Baltimore, Md.

Goodman, L.A. (1974), The Analysis of Systems of Qualitative Variables When Some of the Variables are Unobservable: Part I --A Modified Latent Structure Approach, Amer. Jour. Soc., 1179-1259.

Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953), Sample Survey Methods and Theory, Vol. II, New York: John Wiley.

Hansen, Morris (1987), Remarks on How Large Survey Organizations Build Quality into Their Operations, 3rd Ann. Cens. Res, Conf., Baltimore, Md.

Kalsbeek, William D. (1980), A Conceptual Review of Survey Error Due to Nonresponse, in A Taxonomy of Error Sources and Errors Measures for Surveys: Final Report,(Eds. Judith T. Lessler, and Ralph E. Folsom), Project RTI/1791/00-03F, NSF Grant #SOC-780 4597, Research Triangle Park, NC: Research Triangle Institute.

Kalton, Graham and Kasprzyk,Daniel (1986), The Treatment of Missing Survey Data, Surv. Method., 2, 1-16.

Lessler, Judith T. (1983), An Expanded Survey Error Model, in Incomplete Data in Sample Surveys, Vol. 3: Proceedings of the Symposium, (Eds. William G. Madow and Ingram Olkin), New York: Academic Press, 259-270.

Little, R.J.A. (1980), Superpopulation Models for Non-Response I: The Ignorable Case, and II: The Non-Ignorable Case, in Non-Response in Sample Surveys: The Theory of Current Practice, Part V, Panel on Incomplete Data, National Academy of Sciences, Washington, D.C.

Little, R.J.A. (1982), Models for Nonresponse in Sample Surveys, Jour. Amer. Stat. Assoc., 77, 237-250.

Little, R.J.A. and Rubin, Donald B. (1987), Statistical Analysis With Missing Data, New York: John Wiley.

Lowenstein, Regina, Colombotos, John, and Elinson, Jack (1962), Interviews Hardest to Obtain in an Urban Health Survey, Proc. Soc. Stat. Sect. Amer. Stat. Assoc., 160-166.

Norwood, Janet L. (1987), What is Quality? Shirley Kallek Memorial Lecture, 3rd Ann. Res. Conf., Baltimore, Md., March 30.

Paul, Elizabeth Clayton and Lawes, Murray (1982), Characteristics of Respondent and Non-Respondent Households in the Canadian Labour Force Survey, Surv. Method., 8, 48-85.

Platek, Richard and Gray, Gerald B. (1983), Imputation Metholodogy, in Incomplete Data in Sample Surveys: Theory and Bibliographies, Vol. 2, (Eds. William G. Madow, Ingram Olkin, Donald B. Rubin), New York: Academic Press, 255-294.

Rubin, Donald B. (1976), Inference and Missing Data, Biometrika, 63, 581-592.

Rubin, Donald B. and Schenker, Nathaniel (1986), Multiple Imputation for Interval Estimation From Simple Random Samples with Ignorable Nonresponse, Jour. Amer. Stat. Assoc., 81, 366-374.

Silberstein, Adriana (1986), Nonresponse Rates by CU Size, Tenure, and Race: Diary and Interview Surveys, 1980 to 1983, BLS memorandum to Paul Hsen.

Thran, Sara L., Marder, William D., Willke, Richard J. (1986), Probability of Response: A Multivariate Probit Analysis, Paper presented to the Amer. Med. Assoc., September.

Tucker, Clyde (1985), Identifying Measurement Error in Consumer Expenditure Reports, Proc. Sect. Surv. Res. Meth. Amer. Stat. Assoc., pp. 473-481.

Tucker, Clyde (1986), An Analysis of the Dynamics in the CE Diary Survey, Proc. Soc. Stat. Sect. Amer. Stat. Assoc., 18-27.

U.S. Department of Labor (1986), Consumer Expenditure Survey: Diary Survey, 1982-1983, Bulletin 2245, Bureau of Labor Statistics.

Weaver, Charles N., Holmes, Sandra L., and Glenn, Norval D. (1975), Some Characteristics of Inaccessible Respondents in a Telephone Survey, Jour. App. Psych., 60, 260-262.

TABLE 4A. RELATIONSHIPS BETWEEN RSUM AND DEMOGRAPHICS (POTENTIAL RESPONDENTS N=1510)

	WGTD N	LOW RESPONSE 1	2	RSUM 3	HIGH RESPONSE 4	5	PROB
REGION							0.00
Northeast	304	10.7%	2.8%	19.7%	31.9%	35.0%	
North Central	385	6.9	3.0	20.0	16.9	53.2	
South	518	8.4	3.4	13.9	24.4	49.8	
West	304	11.0	3.8	13.2	21.8	50.3	
DEGREE URBANIZATION							0.00
Central City	506	12.1	2.6	16.1	27.2	42.1	
Other in SMSA	601	8.6	4.5	17.7	22.9	46.4	
Outside SMSA	403	5.7	2.2	15.2	19.7	57.2	
b							
ETHNICITY							0.54
Black or Hispanic	248	0.5	3.7	19.6	28.8	47.4	
Other	1132	0.6	3.2	17.7	25.0	53.4	
b							
EDUCATION							0.44
Less than H. S.	422	0.2	3.0	15.7	26.9	54.3	
H. S.	394	0.7	3.9	16.9	24.8	53.8	
Post H. S.	563	0.9	3.2	20.6	25.4	49.8	
b							
AGE							0.38
Under 25	141	0.0	3.7	22.9	31.2	42.2	
25-44	542	1.0	3.5	15.7	25.4	54.5	
45-64	420	0.3	3.2	18.0	26.1	52.4	
65+	277	0.7	2.9	20.3	22.8	53.2	
CU COMPOSITION							0.24
Husband/Wife	781	0.7	2.9	17.0	24.2	55.2	
Single Parent/Single	469	0.7	4.4	19.4	28.2	47.3	
Other	129	0.0	1.9	19.6	25.6	53.0	

a. Frequencies often do not add to 1510 because of missing data. A small number of values(approximately 40) for the demographics are imputed.
b. Of reference person (similar to head of household).

TABLE 4B. RELATIONSHIPS BETWEEN RSUM AND DEMOGRAPHICS (RESPONDENT SUBSET N=1163)

	WGTD N	3	RSUM 4	5	PROB (SRS)	PROB (CPLX)
REGION					0.00	0.00
Northeast	239	21.4%	31.6%	46.9%		
North Central	317	19.7	16.0	64.4		
South	380	11.4	21.0	67.6		
West	227	5.0	24.0	71.0		
DEGREE URBANIZATION					0.00	0.12
Central City	366	12.3	27.3	60.4		
Other in SMSA	480	16.9	22.5	60.7		
Outside SMSA	317	13.4	16.6	70.0		
b					0.13	----
ETHNICITY						
Black or Hispanic	199	12.7	27.8	59.5		
Other	964	14.9	21.3	63.9		
b					0.64	----
EDUCATION						
Less than H. S.	342	12.0	23.2	64.8		
H. S.	341	15.3	22.2	62.5		
Post H. S.	479	15.7	22.0	62.4		
b					0.00	0.04
AGE						
Under 25	110	7.6	34.1	58.4		
25-44	472	13.7	21.2	65.1		
45-64	359	14.3	24.1	61.7		
65+	222	19.9	16.5	63.6		
CU COMPOSITION					0.88	----
Husband/Wife	678	14.4	21.4	64.2		
Single Parent/Single	369	14.7	23.4	62.0		
Other	116	14.2	25.2	60.6		

a. Frequencies often do not add to 1163 because of missing data. A small number of values(approximately 10) for the demographics are imputed.
b. Of reference person (similar to head of household).

TABLE 7. RELATIONSHIPS BETWEEN QUALITY AND DEMOGRAPHICS (RESPONDENT SUBSET N=1163)

	WGTD N	LOW 1	QUALITY 2	3	HIGH 4	PROB (SRS)	PROB (CPLX)
REGION						0.00	0.08
Northeast	239	6.6%	38.9%	39.9%	14.6%		
North Central	317	10.5	34.3	32.8	22.4		
South	380	8.6	39.6	36.3	15.6		
West	227	4.1	30.0	45.1	20.9		
DEGREE URBANIZATION						0.07	----
Central City	366	7.0	41.8	36.0	15.2		
Other in SMSA	480	9.2	31.7	38.6	20.5		
Outside SMSA	317	6.6	36.3	38.7	18.4		
b						0.01	0.03
ETHNICITY							
Black or Hispanic	199	8.2	33.9	39.1	18.8		
Other	964	5.9	47.1	31.3	15.7		
b						0.72	----
EDUCATION							
Less than H. S.	342	6.4	37.6	37.6	18.4		
H. S.	341	7.2	36.4	39.4	17.0		
Post H. S.	479	9.3	34.9	36.7	19.1		
b						0.00	0.02
AGE							
Under 25	110	7.4	53.9	29.4	9.3		
25-44	472	6.8	34.7	38.1	20.3		
45-64	359	6.6	32.0	41.1	20.3		
65+	222	11.9	37.0	35.8	15.2		
CU COMPOSITION						0.02	0.12
Husband/Wife	678	7.7	32.2	39.6	20.5		
Single Parent/Single	369	9.0	41.7	34.6	14.6		
Other	116	4.5	41.4	37.0	17.2		

a. Frequencies often do not add to 1163 because of missing data. A small number of values(approximately 10) for the demographics are imputed.
b. Of reference person (similar to head of household).