

Clyde Tucker, Bureau of Labor Statistics and Claudette Bennett, Census Bureau  
Clyde Tucker, 600 E Street, NW, Washington, DC 20212

## 1. Introduction

Diaries have been used extensively to collect data in fields as diverse as transportation and health (Roghman and Haggerty, 1972; Thompson, et. al., 1977; Harkins, 1979; Verbrugge, 1980). Diaries also have been an important source of information on consumer spending (Pearl and Levine, 1971; Flueck, Waksberg and Kaitz, 1971). The Bureau of Labor Statistics (BLS) in conjunction with the Bureau of the Census, conducted a consumer expenditure survey that included a diary during 1972 and 1973 in the United States. A similar survey has been ongoing in the U.S. since 1980.

Much research has been devoted to the topic of consumer expenditure diary methodology. Several studies have compared the differences in the estimates from personal interviews involving recall and those from diaries (See Neter, 1970; Neter and Waksberg, 1965; Stanton and Tucci, 1982.). Variations in diary procedures also have been examined (Kemsley and Nicholson, 1960; Sudman and Ferber, 1971; Grootaert, 1986). The study reported here continues in the tradition of those which have examined variations in diary methodology. In 1985, the BLS and the Census Bureau conducted a field test of several methods for collecting consumer expenditure information using a diary. These methods differed with respect to the level of structure in the diary and the prior survey experience of the interviewer. Almost four thousand consumer units were included in the experiment.

Although survey procedures can have dramatic effects on response quality, these effects often are complicated and, thus, not easily measured. This problem is dealt with in this instance by examining the effects from a variety of perspectives, including both the level and the distribution of expenditure reports for the different commodities, and the amount of interviewer and respondent burden. Measurements are made which require the use of a number of statistical techniques such as analysis of variance and contingency table analysis.

## 2. The Consumer Expenditure Diary Survey

The Consumer Expenditure (CE) Diary Survey is conducted by the Bureau of the Census for the BLS and provides, along with the CE Quarterly Interview Survey, the information needed to construct the cost weights for the Consumer Price Index. The data also are used for economic analysis. Although the diary was designed to collect all daily expenditures made within the primary sampling unit (PSU) (or locally) over a two-week period, it is especially effective for gathering information about small, frequently purchased items which are normally difficult to recall over an extended period. These expenditures include grocery items, meals eaten out, household supplies and personal care products and services. In addition to the expenditures, data also are collected on the income, work experience and demographic characteristics of family members using a household characteristics questionnaire.

The unit of analysis in the CE Diary, and the level at which most data are collected, is the consumer unit (CU). A CU 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. To be considered financially independent, at least two of the three major expense categories (housing, food and other living expenses) have to be provided by the respondent.

## 3. Research Design

Previous research on the Consumer Expenditure Diary Survey has shown that a number of factors influence the response (Tucker 1985, 1986, 1987). Respondent (or consumer unit) characteristics, in combination with environmental circumstances and the intervening survey procedures (including interviewer characteristics), affect both the respondent's

attitudes toward the survey and his or her record-keeping behaviors. The attitudes and record-keeping behaviors are collateral in that they occur at approximately the same time. Attitudes are not responsible for the outcome of the survey process; the record-keeping behaviors are. Attitudes and record-keeping behaviors will coincide in many cases, but the research cited above has shown that this is not always true.

The analysis of the CE Diary Survey process up to this point has focused on the contributions of respondent and environmental characteristics in the shaping of the response. Survey procedures have been ignored for two reasons. In the first place, the survey forms have been the same for all respondents. Secondly, although there undoubtedly are differences in the ways in which the survey is administered, these differences have not been measured. The purpose of the Diary Operational Test is the evaluation of the effectiveness of different survey procedures, but these procedural effects must be considered in light of the other causal variables.

The effects of two procedural variables are examined in a design which uses data from both a special research sample and the regular diary sample to make comparisons. One variable measures the effect of the current practice of having the diary and the more extensive CE Quarterly Interview done by the same interviewers. Since the quarterly survey involves so much more of the interviewer's time than the diary (both in terms of length and number of respondents), it is suspected that less emphasis has been placed by the interviewers on the quality of data gathered with the diary. To evaluate the extent to which this is true, two interviewer conditions have been used. In the regular sample, interviewers continued to conduct both surveys. In the research sample, interviewers, new to CE, worked only on the diary.

The other procedural variable concerns the physical layout of the diary. Research on item reporting rates from both the 1972/73 and 1980/81 diaries indicates that explicit references to particular products in the diary increase the likelihood that these items will be reported, especially if the reporting rates are low to begin with (Jacobs, 1983 and Tucker, 1984). To evaluate diary formats which provide more explicit instructions as to the commodities to be reported, two experimental diaries were developed.

What distinguishes the two experimental diaries from one another is the specificity of the item descriptions within each section. In experimental diary A, the non-specific diary, only blank lines for recording purchases are provided under each of the section headings, just like the current diary; but, in contrast to the latter, the section headings contain more complete descriptions of the items to be reported. Experimental diary B, the specific diary, has only category titles; however, the lines beneath each heading have specific items printed on them. Respondents need only check whether an item was purchased and record the price.

In addition to the two experimental diaries, the current diary is used as a control. These three diaries were administered at random to consumer units in the research sample, and the interviewers in this sample did not work on the quarterly survey. The fourth cell contains the on-going "production" diary where interviewers, experienced with CE, conducted both surveys.

Two other features of this experiment need to be mentioned. The first is that an additional section was added to the household characteristics questionnaire in order to assess the respondent's attitudes toward the diary and his or her record-keeping behavior at the end of the two weeks. The interviewer also records his or her impressions of the respondent's performance. The other feature is a new method of collecting recalled expenditures. Currently, the interviewer records these expenditures directly into the diary using unscripted procedures and also asks a series of follow-up questions called diary check items about specific commodities which the respondent may have forgotten to report. The new procedures, used with the two experimental diaries, involves a scripted recall section contained in the household characteristics questionnaire.

To examine the effect of interviewer experience, the control diary is compared to the production. The format effect is measured by differences between the three research cells. The production is not compared directly to the experimental diaries, but effects can be inferred assuming no interaction between treatment conditions.

The research sample used in the Diary Operational Test was a retired Current Population Survey (CPS) sample from 1979. Households in this sample had not been in the CPS since January or February of 1979. For the Diary Operational Test, 22 self-representing PSU's (large cities) were included, and 2586 CU's were surveyed. The CE production sample consisted of approximately 500 sample units per month over the 22 PSU's.

Weights were assigned to consumer units participating in either sample. These weights were a product of the probability of selection of the housing unit, a factor adjusting for subsampling in the field, and a noninterview adjustment. They were calculated monthly for diaries beginning within the month, and each week's interview was weighted separately. Variances which include design effect are calculated using 24 balanced, half-sample replicates in each data set.

#### 4. The Analytical Design

##### 4.1 The Dependent Variables

As mentioned previously, several measurements are taken in order to evaluate the results of the test from a variety of perspectives. Most of these dependent or outcome variables measure respondent performance. To facilitate comparisons between the diaries using these variables, commodities are grouped into categories referred to as "expenditure classes." All possible expenditure classes are not considered because the experimental diaries were not designed to collect information on all commodities. A list of the expenditure classes by general category included in the analysis are found in Table 1. These are the classes for which respondents were asked to report expenditures in the specific diary.

One group of performance variables provide information about the level of expenditure reporting in each of the four diaries. The most important of these are the mean weekly expenditure for each expenditure class and the percentage of weekly diaries reporting the purchase of at least one item from the class.

Also measuring level of reporting is the number of reports in a class. This variable is of limited utility because each line of the specific diary, unlike a line in the other diaries, can represent the total expenditure for several items; but, at least, the other diaries can be compared using this variable. Both noninterview and refusal rates are other measures of response quality. Refusal rate is a subset of the noninterview rate.

Another measure of response quality is a respondent typology (Tucker, 1986) which combines indicators of the respondent's attitudes toward the diary and his or her record-keeping behavior. This measure also takes into account the interviewer's opinion of the respondent's performance. These data came from the Diary Assessment section of the Household Characteristics Questionnaire.

Attitude and behavior scales are formed and then collapsed into dichotomies and cross-tabulated. A respondent style typology (ATTBEH) based on their relationship is developed. The first category of this typology, "resisters," is composed of respondents with both unfavorable attitudes and undesirable record-keeping behavior. On the other extreme, in category four, are the "accommodators" who have both positive attitudes and behavior. Category two ("misleaders") is made up of respondents who exhibit positive attitudes, but their behaviors do not coincide with these attitudes. Respondents in category three have desirable behavior but unfavorable attitudes and are called "complainers."

Two other variables which have to do with respondent performance are examined for all four diaries. The first is a measure of the decline in reporting which takes place during a diary week, and the other compares the amount of

Table 1. Expenditure Classes Include in the Analysis

- I. Food at Home
  - a. Flour, cereal and other grain products
  - b. Bakery products
  - c. Beef
  - d. Poultry
  - e. Pork
  - f. Other meats
  - g. Fish and seafood
  - h. Eggs
  - i. Dairy products
  - j. Fruits and fruit juices
  - k. Vegetables and vegetable juices
  - l. Sugar, sugar substitutes and sweets
  - m. Fats, oils and dressings
  - n. Nonalcoholic beverages
  - o. Miscellaneous food at home
  - p. Combined food and nonalcoholic beverages at home
- II. Food Away From Home
  - a. Breakfast/brunch
  - b. Lunch
  - c. Dinner
  - d. Snacks and nonalcoholic beverages
  - e. Combined food and nonalcoholic beverages away from home
- III. Alcoholic Beverages at Home.
- IV. Alcoholic Beverages Away From Home
- V. Nonfood Items
  - a. Tobacco products and smoking supplies
  - b. Personal care products and services
  - c. Housekeeping supplies
  - d. Gasoline, motor oil and additives
  - e. Nonprescription drugs and medical supplies
  - f. Other nonfood items

reporting in the first week to that in the second. The measure of decline within a diary week is called first-day bias which compares the mean expenditure for the first day of the week to the average daily expenditure for that week. This is done only at the general category level for respondents who completed both diaries. Recalled expenditures are eliminated from the calculations. Expenditures from the diary check items also are excluded. The measure of decline from first week to second week, first-week bias, also is computed on respondents with two diaries after eliminating recalled expenditures. Thus, both measures of decline are calculated on the same subsets of respondents so that they can be used in conjunction with one another to arrive at conclusions. It is hypothesized that the experimental diaries will exhibit less decline.

For the two experimental diaries, the proportion of expenditures obtained through recall is investigated. The assumption is that the more data recalled, the less the respondent took the time to record. This may not be a valid assumption, so, in addition, the value of the expenditures recorded in the diary itself must be considered. Thus, the means for recalled expenditures at the general category level are compared to the same means from the diary reports.

As already stated, one of the primary reasons for collecting the CE data is the development of the cost weight or relative importance (i.e., proportion) of each commodity group in the average consumer's budget for use in the Consumer Price Index. Relative importances are calculated for the expenditure classes listed in Table 1 from data produced in each of the treatment conditions, and the distributions are compared. The differences of greatest interest are those between the production distribution and the other three. If one of the experimental diaries is chosen, it will be important to know how this decision would affect the CPI.

Another set of outcome variables measures the interviewers' assessments of the different diary formats along several dimensions, including interviewer and respondent burden and the

accuracy of reports. These assessments are examined with respect to interviewer experience. The data were collected with a debriefing questionnaire at the end of the test.

Although the variances of the expenditure estimates across the conditions are expected to be similar, they, nevertheless, are compared at the expenditure class level to establish the differences in precision from one diary to another. However, because the design of the production survey is different from the design for the research cells, no attempt is made to compare the production to the control.

#### 4.2 The Independent Variables

The design variables are, of course, the focus of this study, but their effects must be evaluated in conjunction with the other independent variables (respondent and environmental characteristics). At present, these other independent variables are limited to those studied in previous analyses of data quality in the diary (Tucker, 1985, 1986, 1987, 1988). These variables are presented in Table 2.

Table 2. Distributions for Selected Consumer Unit Characteristics by Diary

	Prod	Cntl	Nonsp	Spec
Characteristics	%	%	%	%
<b>Age of Reference Person</b>				
Under 25	10.8	9.4	6.8	6.9
25 - 44	43.3	44.5	48.9	44.4
45 - 64	28.8	28.5	27.5	29.3
65+	17.1	17.6	16.8	19.4
<b>Education of Reference Person</b>				
Less than H.S.	21.9	22.4	22.9	25.9
High School	30.4	29.6	28.4	29.8
Post High School	47.7	48.0	48.7	44.3
<b>Ethnicity of Reference Person</b>				
Other	81.3	81.4	80.7	81.1
Black or Hispanic	18.7	18.6	19.3	18.9
<b>Composition of CU</b>				
Husband/Wife	53.0	52.3	54.0	52.0
Single Parent/Single	36.4	37.9	33.2	36.9
Other	10.6	9.8	12.8	11.1
<b>Degree of Urbanization</b>				
Central City	37.0	34.1	36.4	35.8
Other in SMSA	63.0	65.9	63.6	64.2
<b>Region</b>				
Northeast	14.1	17.0	16.0	15.1
North Central	36.0	34.9	34.8	34.6
South	21.1	20.7	22.0	21.8
West	28.8	27.4	27.2	28.5
<b>CU Tenure</b>				
Owner	57.5	57.7	57.4	58.2
NonOwner	42.5	42.3	42.6	41.8
Weighted Weekly Diaries	2014.0	1830.0	1802.0	1761.0

#### 4.3 Statistical Methodology

Given the number of dependent and independent variables measured at different levels of precision, a variety of statistical techniques are needed. The  $T^2$  statistic (Johnson and Wichern, 1982, Ch. 6) is used to evaluate, at both the population and subpopulation levels, the overall differences between the diaries with respect to the vectors of the means for expenditures and number of reports, the percents reporting, and the relative importances. With

this statistic, a simultaneous test of the values for all expenditure classes can be made so that the type-one error rate is held constant. The differences between treatments are calculated separately, and the sample from each diary is assumed to be independent of the other samples. Furthermore, the variance-covariance matrices are assumed to be equal across treatments. Thus, a pooled estimate of the variance-covariance matrix for treatment differences calculated from the replicates is employed to test the equality of the vectors.

$T^2$  is computed as follows:  
 $T^2 = (\bar{X}_1 - \bar{X}_2)' \cdot [Spooled]^{-1} \cdot (\bar{X}_1 - \bar{X}_2)$ , (1)  
 and  $T^2 / [(n_1 + n_2) \cdot p / (n_1 + n_2 - p + 1)]$  is distributed as  $F_p, (n_1 + n_2 - p + 1)$  where  $p$  is the number of elements in the vectors  $\bar{X}_1$  and  $\bar{X}_2$ . This number of elements is either 29 or 5 for the means and percents reporting depending on whether the expenditure classes or general expenditure categories are being considered. In the case of relative importances, the numbers are 28 and 4 because the last element is a linear combination of the others. If  $T^2$  is significant (a two-tail test with  $\alpha/2=.05$ ) for a particular treatment comparison, simultaneous confidence intervals for the differences between pairs of elements are established as follows:

$$\Delta' (\bar{X}_1 - \bar{X}_2) \pm \sqrt{\frac{(n_1 + n_2) \cdot p}{(n_1 + n_2 - p + 1)} \cdot \frac{F(\alpha/2)}{p, n_1 + n_2 - p + 1}} \cdot \Delta' (Spooled)^{1/2} \Delta$$
, (2)

where  $\Delta$  contains a 1 for the particular category comparison and the rest 0's. The same procedures as above are followed for calculating first-day/first-week bias and the proportion of expenditures collected from recall. In these cases, the five general categories of expenditures are examined and then only for mean expenditures.

The comparison of variances based on a direct computation of the variance of the variances is not feasible given that the sample sizes in the replicates are already fairly small. A further subdivision might destroy their integrity and would be both time consuming and tedious. Therefore, the variances of the mean expenditures in the three research cells are compared using an F-test.

The refusal and noninterviews rates are compared using a difference of sample proportions test which assumes simple random sampling. To account for the design effects  $\alpha$  is set at .01. The rates also are compared by region and degree of urbanization -- the only demographic variables certain to be present for the noninterviews. Monthly and quarterly rates are examined in addition to the ones based on the entire survey period.

The respondent typology is developed for each of the research cells, and these typologies are compared simultaneously at the .05 level using a log-linear model. The respondent typology is the dependent variable, and the different diaries are the categories of the independent variable. Interactions between this independent variable and the various demographic variables also are considered. To carry out this analysis, CPLX is used (Fay, 1983 and 1985).

Selected results from the interviewer debriefing questionnaire are analyzed with the aid of contingency tables or frequency distributions. No assumptions are made about the distribution of interviewers, and no significance testing is done.

### 5. The Analysis of the Performance Measures

#### 5.1 Mean Expenditures

Although the  $T^2$  statistics show significant differences between the vectors of the 29 expenditure means for all of the treatment comparisons, there are no significant differences when contrasting the individual expenditure classes. This is consistent given that the  $T^2$  statistic considers all possible linear contrasts and not just those involving one class at a time. The differences between the five broad expenditure categories, like the 29 classes, are computed directly using the replicate structure, and the  $T^2$  statistics for these tests are also significant. The expenditures for these categories are presented in Table 3.

Table 3. Weekly Mean Expenditures

Class	Production	Control	Nonspecific	Specific
Food at Home	\$37.92	\$34.69	\$40.71	\$41.57
Food Away from Home	23.19	20.98	19.51	21.10
Alcoholic Beverages at Home	3.65	3.52	3.03	3.47
Alcoholic Beverages Away from Home	2.49	2.41	1.43	2.12
Nonfood Items	38.62	34.53	31.72	39.61

Unlike the smaller expenditure classes, significant differences are found between the individual general categories in some treatment comparisons. In particular, both of the experimental diaries have significantly larger expenditures in "Food at Home" than does the control diary. Also, the expenditure class differences within this general category, while not significant, are consistently in favor of the experimental diaries. Expenditure means also are different for the general category of "Nonfood Items." The control and nonspecific diaries are relatively comparable, but the specific outperforms each of them. Except for the nonfood category, differences between the specific and nonspecific are usually small and somewhat inconsistent.

These results make sense given the formats of the three diaries. Both the specific and nonspecific diaries provide a number of cues (compared to the control) in the "Food at Home" sections, even if they are more general in the nonspecific than the specific. The "Food Away from Home" section, on the other hand, does not have as many additional cues. In the case of the "Nonfood Items" section, there are a large number of possible commodities, and better cuing for these items exists in the specific, compared to the other two diaries, especially with respect to personal care and housekeeping products and medical supplies.

As for the comparison of the production to the control, no significant differences were found at either the expenditure class level or the general category level using a one-tail test where the control is considered the better diary under the alternative hypothesis. On the other hand, the direction of the differences consistently favors the production which is the opposite of what was expected. There are probably two reasons for this result. Although the condition to be tested was whether or not the interviewer worked on both surveys, the actual test was between experienced and inexperienced interviewers since all of the interviewers in the research cells were new to the CE survey. A second reason is that many of the respondents in the control cell, having already participated in the CPS, may have been less likely to be cooperative respondents in another government survey. The most conservative interpretation is that the interviewer condition made no difference. It is likely that the nonspecific and specific diaries would perform even better than they have using production interviewers and production samples.

At the subpopulation level, significant effects again appear only at the general category level. Where these differences occur, they are always in the same direction as those for the population. A few of the subpopulation results are noteworthy. Young respondents may prefer the specific diary over the other two, but elderly respondents clearly have higher mean expenditures for "Food at Home" in the nonspecific compared to the control. There is little difference in these means for the elderly when the nonspecific and the specific are compared. Blacks and hispanics may be somewhat more inclined toward the specific diary.

## 5.2 Percents Reporting and Number of Reports

Table 4 gives the percents reporting at the general category level. Only the percent reporting difference between the control and nonspecific diaries for "Food at Home" is significant. It is true that the difference between the control and specific in this category is about the same as in the control-nonspecific comparison, and this difference probably would be significant if a less stringent test were used. In both cases, the directions of the differences between the expenditure classes in this category are consistent.

The fact that there are no differences between the control diary and the two experimental diaries in the nonfood category is explained largely by the reports in the "Gasoline, Motor Oil and Additives" class. There is a special check item for both the control and production diaries asking respondents to recall any forgotten gasoline expenditures if none were reported in the diary. Differences between the production and control diaries, while not significant, generally favor the production.

Table 4. Percents Reporting

Class	Production	Control	Nonspecific	Specific
Food at Home	87.7%	83.3%	87.9%	87.4%
Food Away from Home	78.1	73.5	72.2	70.6
Alcoholic Beverages at Home	29.1	27.6	27.0	27.5
Alcoholic Beverages Away from Home	23.3	24.3	17.8	21.1
Nonfood Items	91.0	87.6	86.2	86.1

At the subpopulation level, the overall tests for difference in percents reporting are, like the mean expenditures, less likely to be significant for comparisons of the production and control or the specific and the nonspecific. Very few differences are found, and those that do exist are only in the general categories. These differences are either in the same places and in the same directions as those for the total population (most commonly the difference between the control and nonspecific for "Food at Home") or in different locations than in the total population, but the differences are in the same direction.

The only significant difference in number of reports is that favoring the nonspecific over the control in "Food at Home", which reinforces earlier findings. The means for the nonspecific are consistently greater than those for the control in the expenditure classes of this category. The differences between the production and control are inconsistent. As in the total population, the comparison of the nonspecific to the control at the subgroup level favors the nonspecific. The difference, however, is significant for only about half of the demographic subgroups.

## 5.3 Noninterview and Refusal Rates

Table 5 contains the information about the comparisons of noninterview and refusal rates for the entire survey period (May -October). Because a positive one-tail test has been performed on the differences between the production and control, no significant effects appear. Obviously, there are a number of important differences which favor the production diary. Given the previous findings, this was expected.

In general, the control and nonspecific have similar noninterview rates. These rates are 40-50 percent larger than the production rates. Although the rates for the specific are not significantly different from the control and the nonspecific, the rates for the specific may be a little higher than these other two. The specific rates are approximately 60% higher than production. In most cases,

noninterview and refusal rates are higher in the central cities when compared to the outlying areas of the PSUs. When it comes to region, both the production and specific have the greatest non-response problems in the Northeast. The control diary has greater nonresponse in the South.

One might have expected the noninterview rates to be high for the first month of interviewing, which they are, but the same is also true for production.

The noninterview rates for the production and two experimental diaries seem to have a similar seasonal component. This seasonal component is not as evident for refusal rates. Part of the reason may be that the noninterview rate is more reliable because the classification of type of noninterview is not a factor. There is little variation in the rates from quarter to quarter in each diary.

Table 5. Noninterview and Refusal Rates

Rate	Production	Control	Nonspecific	Specific
Noninterview	12.8%	19.2%	17.6%	20.4%
Refusal	9.7	13.1	12.9	15.1

#### 5.4 Relative Importances

Turning to Table 6, differences between the relative importances from the production and control are, as expected, practically nonexistent. There are, however, a number of differences between the control and the two experimental diaries. Differences also exist between the two experimental diaries themselves.

The largest differences are between the control and the nonspecific. The relative importance of "Food at Home" expenditures in the nonspecific is over six percentage points greater than that in the control, and the direction of this difference holds for all classes in this category. The opposite is true for the relative importances of "Alcoholic Beverages Away from Home" and "Nonfood Items". Together, these two categories have a greater relative importance of four percentage points in the control. In the comparison of the control and specific, the specific has a higher weight for "Food at Home" and a correspondingly lower one for "Food Away from Home". As for the differences between the specific and nonspecific, the relative importance for "Food at Home" is almost four points lower in the specific when compared to the nonspecific. The opposite is the case for "Nonfood Items". Analysis of relative importance at the subpopulation levels indicates that significant differences are found only at the general category level. Most of these differences are the same as those in the total population level.

Table 6. Relative Importance

Class	Production	Control	Nonspecific	Specific
Food at Home	.358	.361	.422	.385
Food Away from Home	.219	.218	.202	.196
Alcoholic Beverages at Home	.034	.037	.031	.032
Alcoholic Beverages Away from Home	.024	.025	.015	.020
Nonfood Items	.365	.359	.329	.367

#### 5.5 Other Measures

It has been hypothesized that the production and control diaries would display more first-day and first-week bias than the two experimental diaries. This was based on the belief that large biases indicate poor data quality resulting from a decline in reporting over time, and it was felt that the experimental diaries would limit this decline.

As it turns out, the results are conflicting. First of all, the better diary in each treatment comparison based on the measures previously analyzed is the one with the greater first-day bias. The picture is quite different for first-week bias,

especially in the case of the specific diary. There are no large differences as in first-day bias, but the signs of the differences involving the specific tend to be reversed from those for first-day bias. Thus, the specific has greater first-day bias but less first-week bias.

These results provide information for advancing a new theory of bias. Previous work on the diary (Tucker, 1985, 1986, 1987) suggested that the best respondents report about the same amount of expenditures in each of the diary weeks; however, no examination of first-day bias was done. It seems quite plausible that the best diary (and, therefore, the one with the best respondents) encourages good reporting at the beginning of each week, but there is an inevitable decline. Other diaries (or respondents) do not even have this high level of reporting at the beginning of the week. Thus, the specific diary, overall, has the highest mean expenditures when compared to the control and nonspecific. Because the numbers of cases are small, a demographic analysis has not been undertaken at this time.

In the comparisons of the recall expenditure proportions from the nonspecific and specific diaries, the  $T^2$  statistic indicates a significant relationship in at least one contrast, but none of the comparisons of the individual category proportions are significant at the .05 level. Although a less stringent test would find differences, the interpretation of these differences in terms of data quality would be difficult. Earlier it was hypothesized that large proportions of expenditures coming from the recall section would reflect negatively on the ability of the particular diary format to encourage respondents to record expenditures during the week. To be sure that this was the case, however, the means of expenditures just from the diary had to be taken into account. As it turns out, not only are the differences in both directions, but the largest difference, that for "Alcoholic Beverages at Home," occurs when the larger diary mean is accompanied by the larger proportion of expenditures coming from recall.

Generally speaking, the proportions for "Food Away From Home" and "Alcoholic Beverages Away From Home" are the largest ones at the subpopulation level.

"Alcoholic Beverages at Home" often has a large value in the specific diary. It stands to reason that items consumed away from home are less likely to be recorded in the diary, especially if purchased in the last day or so of the diary week. Furthermore, the person keeping the diary often will not know about these expenditures when made by other CU members. They may only be recalled when the interviewer returns to pick up the diary. The recall proportions for these

expenditures are particularly large for subgroups that have a large amount of their expenditures in these categories-- young and single respondents.

Comparisons of the variances from the research cells for the estimates of mean weekly expenditures at both the expenditure class and the general category levels reveal few significant differences. About a tenth of the F-ratios are significant, but there does not appear to be any pattern. It appears that the variances are similar enough so that pooled estimates of variance used to compute  $T^2$  are legitimate.

The most remarkable finding with respect to the respondent typology is that the distributions across the four categories are almost identical. They differ from the one reported in Tucker, 1986 in that the proportions of "complainers" and "misleaders" are reversed. In this instance, behaviors are somewhat less desirable than in the earlier case. On the other hand, respondents are more apt to have positive attitudes now than before.

Some of the differences might be explained by the relatively minor changes in the construction of attitude and behavior scales. Yet, these changes did not affect the two extreme categories. The poorer behaviors may be due, in part, to the fact that a number of the respondents, having already been in the CPS, were asked to participate in another government survey. Furthermore, they were interviewed by less experienced personnel. These factors do not, however, address the increase in positive attitudes.

When the relationship between diary and respondent typology is examined for the demographic subpopulations using CPLX, there is a significant three-way interaction between age, diary, and

typology. These findings about the interaction with age, however, provide little information for choosing any one diary.

Based on the results from the interviewer debriefing questionnaire, the control and specific diaries were rated as more accurate than the nonspecific. The fact that a greater number of interviewers view the specific as accurate compared to the control is balanced by the fact that the specific is also more likely to be viewed as inaccurate. The more experienced interviewers are particularly likely to favor the specific and believe the nonspecific is the most inaccurate. The control is the least likely to be viewed as inaccurate regardless of the experience of the interviewer.

With respect to the amount of explanation required, the interviewers point to the control diary as the one needing the most explanation at both the first and second interview. Yet, the largest number of interviewers claim the diaries are all about the same. The specific is somewhat less likely than the nonspecific to be cited as requiring the most explanation at the second interview.

## 6. Discussion

Although both experimental diaries have performed better than the control, the specific diary appears to be somewhat better than the nonspecific based on the results of the Diary Operational Test produced thus far. The specific is particularly effective in gathering expenditure information on nonfood items. The noninterview rate is highest for the specific, but this rate should decline if the specific were put into production. It might still be somewhat higher than the current production or a production nonspecific diary. The use of the specific should have relatively small effects on the current cost weights. Clearly, using interviewers who had no experience with the quarterly survey was not at all effective. Unfortunately, the combination of the old CPS sample and inexperienced CE interviewers, who had to administer three different versions of the questionnaire, outweighed the experimental condition.

Before final conclusions are drawn, further analysis will be undertaken. One study will involve creating micro-level measures of response error and nonresponse (Tucker, 1988). The impact of the survey process on these data quality variables will be evaluated. A second study will involve laboratory investigations of problems encountered by respondents in classifying items on the specific form. The preliminary findings concerning the relative importances suggest that, for the most part, the classification system used by respondents did not differ from the one used at BLS; however, more analysis of the individual commodity classes is needed.

On the surface, there do not appear to be any significant differences in the results from the separate analyses conducted by Census and BLS. That is, different weighting and editing procedures did not have a great effect. A comparison of the variances using six-month weighting and the one-month weighting still has to be done. Finally, some thought should be given to problems involved in generalizing the results of this test to all PSU's, especially given the high noninterview rates in the South for the control and in the Northeast for the specific and the production.

## References

- Fay, Robert E. (1985), "A Jackknifed Chi-Squared Test for Complex Samples," Journal of the American Statistical Association, 80, 148-157.
- Fay, Robert E. (1983), "CPLX -- Contingency Table Analysis for Complex Sample Designs, Program Documentation," Technical Report, U.S. Bureau of the Census.
- Flueck, John A., Waksberg, Joseph and Kaitz, Hyman B. (1971), "An Overview of Consumer Expenditure Survey Methodology," Proceedings of the Business and Economic Statistics Section, American Statistical Association, pp.238-246.
- Grootaert, Christian (1986), "the Use of Multiple Diaries in a Household Expenditure Survey in Hong Kong," Journal of the American Statistical Association, 81, 938-44.
- Harkins, Elizabeth (1979), "An Evaluation of the Reliability of Data Gathered From Three Primary Care Medical Specialties Using a Self-Administered Log Diary." National Center for Health and Human Services, Public Health Service, pp.38-55
- Johnson, Richard F. and Wichern, Dean W. (1982), Applied Multivariate Statistical Analysis, Englewood Cliffs, NJ: Prentice Hall, Inc.
- Jacobs, Curtis A. (1983), "1983 CES-Analysis of Reporting Completeness," BLS Memorandum to John L. Marcot.
- Kemsley, W.F.F., and Nicholson, J.L. (1960), "Some Experiments in Methods of Conducting Family Expenditure Surveys," Journal of the Royal Statistical Society, Ser. A, 123, Part 3, 307-328.
- Neter, John (1970), "Measurement Errors in Reports of Consumer Expenditures," Journal of Marketing Research, 7, 11-25.
- Neter, John, and Waksberg, Joseph (1965), Response Errors in the Collection of Expenditure Data by Household Interviews: An Experimental Study," Technical Paper 11, Washington, DC: U.S. Bureau of the Census.
- Pearl, Robert B. and Levine, Daniel B. (1971), "A New Methodology For a Consumer Expenditure Survey." Proceedings of the Business and Economic Statistics Section, American Statistical Association, pp.254-259.
- Roghmann, Klause J. and Haggerty, Robert J. (1972), "The Diary as a Research Instrument in the Study of Health and Illness Behavior: Experience with a Random Sample of Young Families." Medical Care, 10, 143-163.
- Stanton, John L. and Tucci, Louis A. (1982), "The Measurement of Consumption: A Comparison of Surveys and Diaries," Journal of Marketing Research, 19, 274-277.
- Sudman, Seymour and Ferber, Robert (1971), "Experiments in obtaining Consumer Expenditures by Diary Methods," Journal of the American Statistical Association, 66, 725-735.
- Thompson, Wendel L., et al (1977), "Improving the Quality of Data from Monthly Gasoline Purchase Diaries," Proceedings of the Social Statistics Section, American Statistical Association, pp.498-503.
- Tucker, Clyde (1986), "An Analysis of the Dynamics in the CE Diary Survey," Proceedings of the Social Statistics Section, American Statistical Association, pp.18-27.
- Tucker, Clyde (1984), "Decision-Making in Diary Research Planned for 1985," BLS Memorandum to Curtis A. Jacobs.
- Tucker, Clyde (1988), "Estimating Measurement Error in Surveys," Paper Presented at the Annual Meeting of the American Statistical Association for Public Opinion Research, Toronto, Canada, May 19-22.
- Tucker, Clyde (1985), "Identifying Measurement Error in Consumer Expenditure Reports," Proceedings of the Survey Research Methods Section, American Statistical Association, pp.473-481.
- Tucker, Clyde (1987), "Toward the Development of an Ordered Set Data Quality Categories," Proceedings of the Survey Research Methods Section, American Statistical Association, pp.558-563.
- Verbrugge, Lois M. (1980), "Health Diaries," Medical Care, 18, 73-95.