

THE EFFECTS OF FORMAT CHANGES ON REPORTING IN THE 1991 CONSUMER EXPENDITURE DIARY SURVEY

Clyde Tucker, Bureau of Labor Statistics
2 Massachusetts Ave., NE, Rm. 4915, Washington, DC 20212

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 for some time (Pearl and Levine 1971; Fluck, Waksberg, and Kaitz 1971) because they are particularly good instruments for collecting small, inexpensive items. 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 and Waksberg 1965; Stanton and Tucci 1982; Silberstein and Scott 1992). As Grootaert (1986) has pointed out, the results from the studies are inconclusive. One method may be superior over the other for some expenditures, but the reverse seems to be the case for other expenditures. Silberstein and Tucker (Silberstein and Scott 1991; Silberstein 1991; Tucker 1992) have looked at various measures of error in the U.S. Consumer Expenditure Diary Survey (CE Diary).

Variations in diary procedures also have been examined. One group of studies has dealt with the effects of placing multiple diaries in a household as opposed to only one diary. Generally speaking, providing a diary to every family member over a certain age produces better reporting than having one member keep a diary for the entire family (Kemsley and Nicholson 1960; Sudman and Ferber 1971). On the other hand, in addition to the difficulty in obtaining cooperation from all participating family members in the multiple-diary situation, Grootaert (1986) found that proxy reporting was better where elderly respondents were concerned.

As for the other aspects of diary methodology which have been investigated, they include length of the reporting period, the format of the diary itself, and the impact of incentives (Turner 1961; Sudman and Ferber 1971; Walsh 1977; Nasholm, Lindstrom, and Lindkvist 1989). Respondents typically report more items at the beginning of the reporting period than at the end. This probably reflects a loss of interest in keeping the diary due to the tediousness of the task.

There is evidence that diaries organized according to commodity categories produce the best results and that incentives can increase response rates.

The experiment reported here continues in the tradition of those which have examined variations in diary methodology. It builds upon an earlier study conducted by the BLS and the Census Bureau to test several methods for collecting consumer expenditure information using a diary. 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 here by examining the effects from a variety of perspectives, including both the level and the distribution of expenditure reports for different commodities. The results are drawn together to arrive at overall conclusions about the relative merits of each methodology.

The next section provides some general information about the BLS Consumer Expenditure Diary Survey. Section 3 describes the previous study as well as the design for the present one. Various measures of performance are defined in Section 4, and Section 5 describes the statistical methodology used. Sections 6 and 7 provide results and conclusions, respectively.

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; as mentioned previously, 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. Data also are collected on the income, work experience and demographic characteristics 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. 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. 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.

The CE Diary sample is designed to be representative of the national, noninstitutional civilian population. Besides the population residing in regular housing, persons residing in selected group quarters, such as college dormitories, also are represented. Five to six thousand consumer units are interviewed each year from a nationwide two-stage, clustered design with 101 PSUs (primary sampling units). Because the design is complex, a set of 44 balanced, half-sample replicate weights is constructed for use in variance estimation. The weights are assigned to each consumer unit in the survey in order to provide estimates for the U.S. population. For further a description see U.S. Department of Labor (1986).

DESIGN OF THE 1991 CE DIARY TEST

Previous research on item reporting rates indicated that explicit references to particular products in the diary can have a positive effect on the reporting of these items, especially if the reporting rates are low to begin with (Silberstein 1983 and Tucker 1984). In 1985, a field test was undertaken to evaluate two experimental diary formats which provided more explicit information about what commodities should be reported. These diaries covered fewer expenditure categories than the diary that was being used. In particular, the apparel section and some other nonfood categories were eliminated. Respondents also were not required to specify the quantity and weight of items, as was the case in the current production diary.

Another feature of the experimental diaries was a new method for collecting recalled expenditures. Interviewers had been recording these expenditures directly into the diary using unscripted procedures and also asking a series of follow-up questions called "diary-check items" about some specific commodities often forgotten by respondents. The new procedure involved a scripted recall section for each expenditure category and was contained in the household characteristics questionnaire. Thus, recalled expenditures could be identified, and the rather cumbersome check items were eliminated.

What distinguished the two experimental diaries from one another was the specificity of the item descriptions in each commodity section. One diary, referred to as the "nonspecific," had blank lines for recording purchases under each of the section headings, and the headings contained relatively

extensive descriptions of items in that section. The second diary, the "specific," had only category titles in the headings, but the lines beneath each heading had specific items printed on them. Respondents just checked if an item was purchased and recorded the price.

Extensive analyses (Tucker and Bennett 1988; Sliwa 1988; Tucker 1992) showed that both of the experimental diaries produced gains in expenditure reports for groceries, but the specific diary gave the greatest overall improvement. The additional recall section proved to be worthwhile not only in uncovering forgotten expenditures but also in reducing interviewer errors common to the check-item section.

Before a new diary format could be implemented, several issues still needed to be addressed. Any new diary would have to incorporate the categories not covered on the experimental diaries. It was unclear how these additional categories would affect reporting by increasing respondent burden. Furthermore, although the specific diary seemed to be the better of the two, the number of printed lines would be very large if all nonfood categories were included. Also, follow-up work done in a laboratory suggested that respondents had difficulty correctly selecting the appropriate line for recording some expenditures in the specific diary (Tucker, Vitrano, Miller, and Dobby 1989).

For these reasons, it was decided that a second field test would evaluate a diary similar to the current production diary and an expanded version of the nonspecific diary which included all expenditure categories. Diary A, the one almost identical to the production diary, had two pages for recording expenditures each day. Diary B, the expanded nonspecific diary, had six pages per day containing a number of sections with headings describing which items should be recorded in them. The recall section was used with both, and, again, no quantity or weight information was collected. To save money, a separate test was not done. Instead, these two diaries replaced the diary currently being used with the production sample.

Both diaries were in the field for calendar years 1991 and 1992, and each consumer unit was asked to keep the same diary for two consecutive weeks. Because Diary A was so similar to the production diary, 80% of the sample used it. Diary B was given to the other 20%. Over 4500 consumer units received Diary A in 1991, yielding 9170 weekly diaries treated as independent observations. There were 2264 weekly B diaries. The analysis is based on the 1991 data.

OUTCOME MEASURES

Several measurements are taken in order to evaluate the results of the test from a variety of perspectives.

To facilitate comparisons between the diaries using these variables, commodities are grouped into expenditure categories. They are **Food at Home, Food Away From Home, Apparel, Other Nonfood Items**.

Perhaps the most important respondent performance measure is the reported mean weekly expenditure for each category. Mean weekly expenditure is computed in the following way for each expenditure category:

$$WEM_c = \sum_{i=1}^n \sum_{r=1}^{s(c)} w_i \cdot e_{irc} / \sum_{i=1}^n w_i \quad (1)$$

where $w_i (i = 1, \dots, n)$ is the weight for a particular consumer unit i and e_{irc} is a single expenditure $r (r = 1, \dots, s(c))$ from category c for that consumer unit.

Another performance measure, percent of respondents reporting an expenditure in a category, is computed similarly:

$$PER_c = \left(\sum_{i=1}^n p_{ic} \cdot w_i / \sum_{i=1}^n w_i \right) \cdot 100 \quad (2)$$

where p_{ic} is an indicator variable which is either 0 or 1 depending upon whether or not consumer unit i reported a purchase from category c .

A third measure is the mean number of reports in a category. It is computed as below:

$$NRE_c = \sum_{i=1}^n \sum_{r=1}^{s(c)} w_{irc} / \sum_{i=1}^n w_i \quad (3)$$

The noninterview rate for each diary is also computed. Noninterviews are coded into three categories: (1) no one home, (2) refused and (3) other. The two rates of interest are the following:

$$NIR = \left(\frac{NI_1 + NI_2 + NI_3}{I + NI_1 + NI_2 + NI_3} \right) \cdot 100 \quad (4)$$

$$RR = \left(\frac{NI_2}{I + NI_1 + NI_2 + NI_3} \right) \cdot 100 \quad (5)$$

where NIR is the noninterview rate, RR is the refusal rate, I is the number of completed interviews and $NI_t (t = 1, 2, 3)$ are the types of noninterviews.

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 reporting in the first week to that in the second. Previous research (Sudman and Ferber 1971; Silberstein and Scott 1991) has shown that diary respondents become less diligent as time goes on. In fact, Silberstein and Scott show that the largest drop in reporting occurs after the first day and, thus, is termed "first-day bias."

The measure of decline within a diary week, called first-day bias, compares the mean expenditure for the first day of the week to the average daily expenditure for that week. This is done for respondents who completed both diaries. Recalled expenditures are eliminated from the calculations and those respondents whose entire expenditure report was obtained from recall are not included. The first-day bias is computed as follows:

$$FDBIAS_c = \frac{\left(\sum_{i=1}^n \sum_{r=1}^{s(c)} w_i \cdot e_{irc1} / \sum_{i=1}^n w_i \right)}{\left(\sum_{i=1}^n \sum_{r=1}^{s(c)} w_i \cdot e_{irc} / 7 \cdot \sum_{i=1}^n w_i \right)} \quad (6)$$

where e_{irc1} is an expenditure record in category c from the first diary day for consumer unit i .

The measure of decline from first week to second week, first-week bias, also is computed in the following manner on respondents with two diaries after eliminating recalled expenditures:

$$FWBIAS_c = \frac{\left(\sum_{i=1}^{n1} \sum_{r=1}^{s(c)} w_{i1} \cdot e_{irc1} / \sum_{i=1}^{n1} w_{i1} \right)}{\left(\sum_{i=1}^{n2} \sum_{r=1}^{s(c)} w_{i2} \cdot e_{irc2} / \sum_{i=1}^{n2} w_{i2} \right)} \quad (7)$$

where the numerator is based on data from the first diary week, and the denominator is based on data from the second.

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. It will be important to know how the decision on a new diary would affect the CPI. Relative importances are calculated for the four expenditure categories. The relative importance of category c is computed as follows:

$$RIM_c = \frac{\sum_{i=1}^n \sum_{r=1}^{s(c)} w_i \cdot e_{irc}}{\sum_{c=1}^j \sum_{i=1}^n \sum_{r=1}^{s(c)} w_i \cdot e_{irc}} \quad (8)$$

where the summation in the denominator is done across all categories.

ANALYTICAL METHODS

The most important survey performance measures are mean expenditures and response (noninterview) rates, but relative importances are of at least equal interest since changes in them could affect the CPI. Therefore, treatment differences in all three indicators are subjected to significance testing at the .10 level. Given the computationally intensive nature of the

replicate design and their lesser importance, differences for the other measures are compared by inspection only. In general, differences of 10% or more are of interest, especially if there is a trend of that magnitude in one direction or another.

The refusal and noninterview rates are compared using a difference of sample proportions test based on simple random sampling. To account for possible design effects, α is set at .05. The T^2 statistic (Johnson and Wichern 1982) 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 the relative importances. With this statistic, a simultaneous test of the values for all expenditure categories can be made so that the type-one error rate is held constant. The sample for each diary is assumed to be independent, and the variance-covariance matrices are assumed to be equal. The pooled estimate of variance is given as:

$$S_{\text{pooled}} = \frac{\sum_{j=1}^{44} (\bar{x}_{1j} - \bar{x}_1)(\bar{x}_{1j} - \bar{x}_1)' + \sum_{j=1}^{44} (\bar{x}_{2j} - \bar{x}_2)(\bar{x}_{2j} - \bar{x}_2)'}{n_1 + n_2} \quad (9)$$

where j specifies the replicates, \bar{x}_{1j} and \bar{x}_{2j} are the vectors of replicate statistics (means and relative importances) for the two treatments, and \bar{x}_1 and \bar{x}_2 are the vectors of statistics for the entire sample. S_{pooled} is calculated with $n_1 + n_2 = 88$ rather than $n_1 + n_2 - 2$ when using balanced half-sample replicates.

T^2 is computed as follows:

$$T^2 = (\bar{x}_1 - \bar{x}_2)' \cdot [S_{\text{pooled}}]^{-1} \cdot (\bar{x}_1 - \bar{x}_2) \quad (10)$$

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 (commodity categories) in the vectors \bar{x}_1 and \bar{x}_2 . If T^2 is significant (a two-tail test with $\alpha = .10$) for a particular treatment comparison, simultaneous confidence intervals for the differences between pairs of elements are established as follows:

$$\ell'(\bar{x}_1 - \bar{x}_2) \pm \sqrt{\frac{(n_1 + n_2) \cdot p}{(n_1 + n_2 - p + 1)} \cdot F_{p, (n_1 + n_2 - p + 1)}^\alpha \cdot \ell'(S_{\text{pooled}})\ell} \quad (11)$$

where ℓ contains a 1 for the particular category comparison and the rest 0's.

Diary B represents a greater departure from current procedures than Diary A. For that reason, a demonstrable improvement in the outcomes would be needed to justify choosing Diary B over Diary A. This improvement would include increases in expenditure means and, presumably, the other reporting measures (number of reports and reporting rates). At the same time, the response rates should be

reasonably comparable to Diary A. Differences in relative importances that can be tied to improvements in reporting will be viewed positively.

RESULTS

To be sure that differences between the diaries are not simply a reflection of differences in sample characteristics, three variables were examined-- consumer unit size and income and the percentage of incomplete income reporters. The first two are closely related to expenditures, and the last gives some indication of level of respondent cooperation. Table 1 provides the information about these indicators in both samples. Statistical tests using variances taking into account the complex design showed no differences between the samples. Table 2 gives the nonresponse rates for both diaries. Again, there are no statistically significant differences.

The expenditure means for the four categories in Table 3 are higher in Diary B in every case, although statistically significant differences are limited to Food At Home and Other Nonfood Items. Table 4 provides information on the other reporting measures. There is no consistent pattern among the reporting rates, with only the difference for Apparel approaching the 10% mark. A clearer trend in favor of Diary B might have been expected given the differences in expenditure means. There is more than a 10% difference in number of expenditure records for Food at Home. This difference coincides with the expenditure differences reported in Table 3. On the other hand, the number of records for Food Away From Home is greater for Diary A. The numbers in the other two categories are almost identical.

Turning to the other important measure, relative importances, Table 5 gives these for the four expenditure categories. No statistically significant differences exist; and, in fact, there are none of any substantive interest. The proportion for Food Away From Home is slightly smaller (about 8%) for Diary B because the increase in its mean expenditure was so much smaller than those for the other categories.

Table 6 contains the data on first-day and first-week biases, and they show the same pattern as in the Diary Operational Test. There is generally a greater first-day bias in Diary B; but the first-week bias is, if anything, in the opposite direction. This is particularly true for Apparel. These results may indicate that first-day bias is simply a matter of higher reports at the beginning of each diary week for Diary B as opposed to a larger decline in effort later in the week.

CONCLUSIONS

The results of this study indicate that Diary B performs better than Diary A in prompting expenditure reports in most categories. Furthermore,

the response rates in the two diaries are comparable; and the relative importances for the different expenditure categories would remain about the same if Diary B were used. Thus, it is recommended that Diary B be used in the future; however, a more indepth analysis should be undertaken. In particular, the different response pattern for Food Away From Home should be examined more closely. Also, a better understanding of the different responses to the two diaries by various subpopulations is needed.

REFERENCES

- Fluek, J.A., Waksberg, J., and Kaitz, H.B. (1971), "An Overview of Consumer Expenditure Survey Methodology," Proceedings of the Business and Economic Statistics Section, American Statistical Association, 238-246.
- Grootaert, C. (1986), "The Use of Multiple Diaries in a Household Expenditure Survey in Hong Kong," Journal of the American Statistical Association, **81**, 938-944.
- Harkins, E. (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, 38-55.
- Johnson, R.F., and Wichern, D.W. (1982), Applied Multivariate Statistical Analysis, Englewood Cliffs, NJ: Prentice Hall, Inc.
- Kemsley, W.F.F., and Nicholson, J.S. (1960), "Some Experiments in Methods of Conducting Family Expenditure Surveys," Journal of the Royal Statistical Society, Ser. A., **123**, Part 3, 307-328.
- Nasholm, H., Lindstrom, H.L., and Lindkuist, H. (1989), "Response Burden and Data Quality in the Swedish Family Expenditure Surveys," Proceedings of the Fifth Annual Research Conference, U.S. Bureau of the Census, 501-514.
- Neter, J., and Waksberg, J. (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, R.B., and Levine, B. (1971), "A New Methodology for a Consumer Expenditure Survey," Proceedings of the Business and Economic Statistics Section, American Statistical Association, 254-259.
- Roghmann, K.J., and Haggerty, R.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.
- Silberstein, A.R. (1983), "Diary Reporting Levels-1972/73 and 1980/81," U.S. Bureau of Labor Statistics, Memorandum to Curtis A. Jacobs.
- Silberstein, A.R. and Scott, S. (1991), "Expenditure Diary Surveys and Their Associated Errors," in: Measurement Errors in Surveys, (Eds P.Biemer, R.M.Groves, L.E.Lyberg, N.A.Mathiowetz, S.Sudman), New York: Wiley, 303-326.
- Silberstein, A.R. (1991), "Response Performance in the Consumer Expenditure Diary Survey," Proceedings of the Section on Survey Research Methods, American Statistical Association, 338-343.
- Silberstein, A.R. and Scott, S. (1992: In Press), "Seasonal Effects in the Reporting of Consumer Expenditures," Proceeding of the Section on Survey Research Methods, American Statistical Association.
- Sliwa, G.E. (1988), "The CE Diary Operational Test-Differences in Reporting Levles," Proceedings of the Section on Survey Research Methods, American Statistical Association, 250-255.
- Stanton, J.L., and Tucci, L.A. (1982), "The Measurement of Consumption: A Comparison of Surveys and Diaries," Journal of Marketing Research, **19**, 274-277.
- Sudman, S., and Ferber, R. (1971), "Experiments in obtaining Consumer Expenditures by Diary Methods," Journal of the American Statistical Association, **66**, 725-735.
- Thompson, W.L., et al (1977), "Improving the Quality of Data from Monthly Gasoline Purchase Diaries," Proceedings of the Social Statistics Section, American Statistical Association, 498-503.
- Tucker, C. (1984), "Decision-Making in Diary Research Planned for 1985," U.S. Bureau of Labor Statistics, Memorandum to Curtis A. Jacobs.
- Tucker, C. (1992), "The Estimation of Instrument Effects on Data Quality in the Consumer Expenditure Diary Survey," Journal of Official Statistics, **8**, 41-61.
- Tucker, C., Vitrano, F., Miller, L., and Doddy, J. (1989), "Cognitive Issues and Research in the Consumer Expenditure Diary Survey," Paper Presented at the Annual Meeting of the American Association for Public Opinion Research, St. Petersburg, FL, May 20-23.
- Tucker, C., and Bennett, C. (1988), "Procedural Effects in the Collection of Consumer Expenditure Information: The Diary Operational Test," Proceedings of the Section on Survey Research Methods, American Statistical Association, 256-261.
- Turner, R. (1961), "Inter-week Variations in Expenditure Recorded During a Two-week Survey of Family Expenditure," Applied Statistics, **10(3)**, 136-146.

U.S. Department of Labor (1986), Consumer Expenditure Survey: Diary Survey, 1982-1983, Bulletin 2245, Bureau of Labor Statistics.
 Verbrugge, L.M. (1980), "Health Diaries," Medical Care, **18**, 73-95.

Walsh, T.C. (1977), "Selected Results from the 1972-1973 Diary Surveys," Journal of Marketing Research, **14**, 344-352.

Table 1. Demographic Characteristics by Diary Type

	<u>Diary A</u>	<u>Diary B</u>
Consumer Unit Size	2.56	2.60
Consumer Unit Income	\$26,997	\$27,533
Incomplete Income Reports	19.3%	20.1%

Table 2. Nonresponse Rates by Diary Type

	<u>Diary A</u>	<u>Diary B</u>
Noninterview Rate	13.6%	14.4%
Refusal Rate	8.7	8.4

Table 3. Expenditure Means by Diary Type

<u>Category</u>	<u>Diary A</u>	<u>Diary B</u>	<u>% Increase</u>
Food At Home	\$53.43	\$60.08	12.4*
Food Away From Home	26.88	27.39	1.9
Apparel	30.51	34.60	13.4
Other Nonfood Items	246.40	277.66	12.7*

* p<.10

Table 4. Percent Reporting and Number of Records by Diary Type

<u>Category</u>	<u>Diary A</u>		<u>Diary B</u>	
	<u>Percent</u>	<u>Number</u>	<u>Percent</u>	<u>Number</u>
Food at Home	91.5	17.6	92.0	20.5
Food Away From Home	74.9	4.2	72.4	3.7
Apparel	42.6	1.2	46.1	1.3
Other Nonfood Items	93.6	10.0	92.1	10.6

Table 5. Relative Importances of Expenditures by Diary Type

<u>Category</u>	<u>Diary A</u>	<u>Diary B</u>
Food At Home	.150	.150
Food Away From Home	.075	.069
Apparel	.085	.087
Other Nonfood Items	.690	.694

Table 6. First-Day and First-Week Biases by Diary Type

<u>Category</u>	<u>Diary A</u>		<u>Diary B</u>	
	<u>First-Day</u>	<u>First-Week</u>	<u>First-Day</u>	<u>First-Week</u>
Food At Home	1.31	1.08	1.48	1.08
Food Away From Home	1.12	1.04	1.19	1.04
Apparel	1.15	1.07	1.32	0.88
Other Nonfood Items	1.34	1.04	1.31	0.97