

RESPONSE PERFORMANCE IN THE CONSUMER EXPENDITURE DIARY SURVEY

Adriana R. Silberstein

Bureau of Labor Statistics, 600 E Street, N.W., Room 5217, Washington, D.C. 20212

KEY WORDS: age, education, diary completion, recall

INTRODUCTION

Keeping a diary demands considerable ability and motivation by respondents, often encompassing retrieval of information from other individuals in the household. Respondent burden is the most frequently used explanation for declining reports with increasing time of diarykeeping. Expenditure means from the Consumer Expenditure (CE) Diary in the U.S. show, on average, a 20 percent drop from first to second day of the first week, and 10 percent drop from first to second week (Silberstein and Scott, 1991). These day and week effects are measured by analyzing diaries kept by respondents during the week. Greater effects are expected as expenses are reported by recall during the pickup interview. This technique is used in order to include items missed during diarykeeping. About ten percent of the diaries are completed through *partial recall*. Another 16 percent of the diaries are filled out entirely through interviewer-aided recall, or *total recall*, when respondents fail to make entries in the diary.

Response performance includes two components, one dealing with compliance, the other with data quality. Examples of compliance issues in a diary survey are whether recordkeeping starts on the first day or is delayed, and whether cooperation is maintained to the end of the diary period or is truncated. Inadequate understanding of what has to be recorded, an incorrect time frame, or failure to gather information from other family members are some of the data quality issues. Groves (1989) describes techniques that can be used to identify levels of performance and measurement error. A technique applicable to diaries relies on "interviewer observations" about placement and pickup visits. This auxiliary information may include an assessment of the conformity to procedures and a judgement of respondent cooperation. The respondent may be asked directly to provide input about completion levels and difficulties with the recording process. Tucker (1988) studied response quality by analyzing an assessment survey in connection with the 1984 CE Diary. He found that young and single respondents are more likely to compile diaries of lower quality than other respondents.

Routine observations are included in ongoing diary surveys in order to obtain minimum information on the field experience. The present study is an analysis of response performance using this type of information from the 1987-88 CE Diary. The data pertain to 12,444 consumer units and 23,747 diaries. Whether respondents *kept* the diary during the week or responded by recall is a response condition

analyzed. Another condition examined is whether respondents provided usable diaries for two weeks or only one. Most responses yield diaries for two weeks, but 10 percent result in only one usable diary week. Findings from this study indicate that these less than desirable responses are associated with certain respondent characteristics. These responses are found in considerable proportions for all segments of the population, however, and, for this reason, the effects on expenditure estimates are sizable.

Selected field procedures are explained in section 1, followed by a description of summary statistics relevant to diary performance. The study includes a categorical data analysis of respondent characteristics (Section 2) and comparisons of expenditure means (Section 3). A discussion of the findings can be found in Section 4.

1. THE CE DIARY

1.1 Selected field procedures and definitions

The reporting unit is the *consumer unit* (CU). The CU comprises those members of a household who are either related or share responsibility for major types of expenses, such as food and housing. An annual sample of approximately 5,000 respondent CUs keep a diary for two one week periods. All expenses are reported, with emphasis on grocery expenses. (A separate sample of respondents is included in a panel interview expenditure survey.) The placement schedule allows for even representation of year-round expenses, and equal numbers of sample units are assigned to be placed each week day. Field procedures call for three visits to each sampled address for diary placements and pickups. Data on income and employment of CU adult members are collected on the third visit.

Instructions to interviewers are to place the first-week diary the day before the start of recordkeeping. This procedure is followed in most cases. The second-week diary should be placed on the day the first week diary is picked up, i.e., the day after the end of the first week, but there is deviation from this procedure in about 40 percent of the cases: 12% are placed together with the first week, 15% are placed early but not together with the first week, and 10% are placed one or more days late.

Recall status codes are assigned by interviewers for each diary week, indicating whether the diary was completed by respondents prior to the pickup visit, or was filled out by partial or total recall instead. Total recall indicates that all reported expenses were entered on the diary during the pickup visit, since respondents agreed to participate but failed to record data. The use of partial recall results from a variety of response conditions: efforts may be made to add expenses

for a skipped day, or additional expenses may be remembered during the diary check; another reason for partial recall is due to late placements of second-week diaries. Expenses for missed days are collected by recall since the two weeks must be contiguous.

The nonresponse rate has averaged 15 percent. Nonresponse includes refusals, noncontacts, and diaries reclassified as nonresponse after computer screening for minimal acceptance. Each diary week is screened independently, resulting in approximately 4 percent of the placed diaries reclassified each year. Several combinations of four criteria have been used since 1985 for the screening: recall status (partial or total recall), number of entries (less than 10), food and beverage expenses reported for the whole week (less than \$5 or \$10, depending on income levels), and incomplete reporting of income information. Incomplete reporting refers to refusal to disclose major sources of income.

1.2 Placement, recall, and weekly participation

Total-recall diaries represent 16% of diaries placed according to standard procedures, 12% of diaries placed together in the first week, and 22% of late placement diaries. (See Table 1.) Total recall is twice as frequent for CUs with one usable week, compared to CUs with two usable weeks. This is due in part to reclassification, since half of one-week diaries had the previous or following week reclassified. One quarter of one-week diaries results from ineligibility for one diary week and one quarter is due to nonresponse in one of the two weeks.

Total recall and one-week participation are more frequent for young and old respondents, especially if living alone, as shown by percentages in Table 2. These characteristics are often mentioned when references are made to responses of lower quality (Tucker, 1988), and to nonrespondents (Kemsley, 1980; Groves, 1989; Éltető, 1991). Lyberg (1991) notes that one-person households have greater nonresponse in Sweden, and better cooperation to expenditure surveys is gained from families with children. Harrison (1991) cites higher nonresponse for households with fewer members in

Table 1. Percent Diaries By Use of Recall

	(%CUs)	Use of Recall		
		None	Partial	Total
All Diaries	(100)	74	10	16
Two-week	(90)	75	10	15
One-week	(10)	52	15	33
Week 2 Placement Procedure:				
Standard	(79)	75	9	16
With Week 1	(11)	79	9	12
Late	(10)	62	16	22

the Australian Household Expenditure Survey. A multivariate study of nonrespondents to the U.K. Family Expenditure Survey shows that response tends to increase for households with children, and is lowest for households with more than one adult and no children (Elliot, 1991). In this survey, partial participation is considered nonresponse and recall is not allowed.

2. CATEGORICAL DATA ANALYSIS

2.1 Method

The log-linear model used is:

$$\log(F_{ijk}) = \lambda + \lambda_i^I + \lambda_j^J + \lambda_h^H + \lambda_k^K + \lambda_{jh}^{JH} + \lambda_{jk}^{JK} + \lambda_{hk}^{HK} + \lambda_{jhk}^{JHK} + \lambda_{ij}^{IJ} + \lambda_{ih}^{IH} + \lambda_{ik}^{IK} \quad (1)$$

where i is the dependent variable, j , h , and k are independent variables. The model measures the interaction of each independent variable with the dependent variable, after eliminating interaction effects between the independent variables. Higher order interactions were not necessary for a good fit.

The model was fitted using the computer program CPLX (Contingency Table Analysis for Complex Surveys), developed by Fay (1987). CPLX computes maximum likelihood estimates of the parameters and their standard errors, and is designed for analysis of data from complex sample surveys. One of the options in CPLX is the use of balanced repeated replication methods. There are 44 half-sample replicates for the 1987-88 data. The overall fit of the model and the contribution made by additional parameters are tested with jackknifed chi-square tests developed by Fay (1985).

Univariate distributions for selected variables are given in

Table 2. Percent CUs By Total Recall and Weekly Participation

	Two-week Participant			One-week Participant
	By Total Recall			
	Neither Week	One Week	Both Weeks	
All CUs	73.9	5.6	10.6	9.9
Less than 27 1 person	67.1 66.6	6.7 5.5	12.3 13.2	13.8 14.7
27-44 1 person	75.4 70.7	6.0 6.4	8.7 10.6	9.6 12.3
45-69 1 person	76.4 72.1	5.3 5.1	9.8 10.6	8.4 12.2
70 and over 1 person	68.4 49.1	4.2 3.5	16.7 19.5	10.7 13.2

Table 3. Results from CPLX given in Tables 5 and 6 show standardized values of the parameters for the last three terms in (1). An additional interaction term is displayed for the model in Table 4.

Dependent variable

The quality of diary reports can be assumed to be lower when total recall is used. The situation is not as clear cut in the case of partial recall, since the reason for using partial recall is not stated. The detail provided by respondents gives indications of the report's quality, as shown by Silberstein and Scott (1991). The aggregate percentage of edited expenses for food and beverage items increases from 22% for diaries with no recall to 36% for diaries with partial recall and 58% when total recall is used (1987-88 data). Increments are observed for other commodities as well (8%, 12%, and 17%, respectively). Partial-recall diaries appear to be midpoint between no-recall diaries and total-recall diaries, and, for this reason, are treated in two ways, in separate models.

One dependent variable classifies recall status (none, partial, and total) using the second week. This permits including placement timeliness effects. Another variable considers the use of recall in both weeks simultaneously: (1) total recall neither week, (2) total recall one week, and (3) total recall both weeks. Diaries completed by partial recall are grouped with diaries with no recall (level 1), provided neither week had total recall. In so doing, partial recall is considered an *extended diary method* rather than a different method. A fourth level of this variable identifies CUs with one-week usable diary. The recall status for these responses is not needed, because they already identify large proportions of total recall, as shown in Table 1. Responses with recall status unknown (6 percent of the diaries) are included with diaries completed by respondents, since an analysis of expenditure means shows greater similarity between these two groups than with others.

Independent variables

Age, education, and other demographic characteristics of the diarykeeper are believed to be important factors influencing the ability to keep a diary. Demographic characteristics of the *reference person* are used since it is not

known who the diarykeeper was in a given CU. The reference person is the first person listed by respondents when asked "Start with the person or one of the persons who owns or rents the home." The age groups chosen focus on lifestyle issues: under 27, 28 to 44 (highest frequency of families with children), 45 to 69, and 70 and over.

Recordkeeping habits, personal organization, and lifestyle can influence diarykeeping quality. People often not at home tend to find the recording task harder, since the CE diary is neither portable nor personal. The frequency and amount of weekly expenses may influence diarykeeping: the diary may be forgotten when there are few expenses or become too burdensome when there are many. Economic level, size of family, and home tenure affect the amount of purchases made, but also capture lifestyle factors. It is hypothesized, for instance, that home owners are generally more involved in keeping records of expenses, hence would have higher performance in diarykeeping. Renters, instead, tend to be younger and more often away from home. Incomplete reporting of income is indicative of lower cooperation, since it is either related to greater mistrust by respondents or results from a missing interview at second-week diary pickup.

Region and size of locality define broad differences in respondents and interviewers across the country. The four regions used also group together a number of local offices administering the survey, thus defining potentially different field practices prevalent in an area.

2.2 Findings

Placement Timeliness

Differences in placement practices and recall status are tested in a model that includes home tenure. Results are displayed in Table 4. Standard placement procedures tend to be followed relatively less in the South and the Northeast than in other regions. Placement of the two diaries together appears to be more frequent in the Northeast; the West would have been a more likely candidate for this outcome, due to greater number of remote areas.

Standard placement procedures tend to result in diary completion by respondents. Placing both diaries at the same time, although not a standard procedure, does not seem to affect reporting levels nor increase the use of total recall. Pickup of these diaries together (at the second-diary pickup) occurs in about half these cases. This procedure raises some concern for first-week diary pickup, since it is a week late.

Late placements lead to the use of not only partial recall but also of total recall. Total recall usage is 18% when placement is 1 day late and 27% when placement is 2 or more days late. In addition to delays caused by interviewers, respondents often do not make themselves available for a visit, suggesting lack of cooperation. Late placements are more likely in the South and West.

Table 3. Percent Distribution of Selected Variables

CU size:	100	Home Tenure:	100
1 person	28	Renters	37
2 or 3 persons	47	Owners	63
4 or more persons	25		
Reference person's age:	100	Region:	100
Less than 27	13	Northeast	21
27-44	39	Midwest	25
45-69	35	South	34
70 and over	13	West	20

Use of recall

The variable combining use of recall and weekly participation appears more effective in describing overall performance. Results of two models using this variable are shown in Tables 5 and 6. (Results of a model including region, home tenure and size of place are not shown.)

CU income and home tenure have the largest test values. There is a strong association of home owners and respondents in higher income brackets with no-recall diaries. Incomplete reporting of income is associated with the use of total recall and one-week participation. CUs living in larger places show a greater propensity to use total recall, but only one of the two weeks, compared to CUs living in smaller places, who tend to keep the diary both weeks. There are significant differences by region. Total recall for one or both weeks is more likely in the Northeast than in other regions, and completion by respondents is more likely in the West.

Table 6 shows results of a model including age and education of the reference person and CU size. Older respondents (70 and over) fail to keep the diary both weeks in greater proportions than respondents of other ages. The use of total recall is more pronounced when older persons are living alone (20% compared to 17% for all older respondents). This age significance persists even though CU

Table 4. Model of Use of Recall, Region, Week 2 Placement, and Home Tenure (Standardized Values)

	Use of Recall in Week 2		
	None	Partial	Total
Interaction of use of recall and region			
Northeast	-4.1	0.2	2.3
Midwest	1.1	-0.3	-0.5
South	-0.3	0.5	-0.5
West	3.6	-0.6	-2.6
Interaction of use of recall and week 2 placement			
Standard	3.2	-3.3	0.9
With week 1	5.5	-1.6	-2.3
Late	-8.1	3.8	2.3
Interaction of use of recall and home tenure			
Renters	-4.6	-0.6	4.1
Owners	4.6	0.6	-4.1
	Week 2 Placement		
	Standard	With week 1	Late
Interaction of week 2 placement and region			
Northeast	-1.4	2.3	-2.7
Midwest	1.8	0.3	-1.5
South	-3.1	-0.0	2.0
West	1.8	-3.8	3.9
Jackknifed chi-square test:			
	Test value		df
Region	2.7 **		6
Placement	8.0 ***		4
Home tenure	6.5 ***		2

** 0.5% level, *** 0.1% level.

size is a factor in the model. CU size is also significant, and so is education. Single person CUs and less educated respondents, in general, show a greater propensity to use total recall both weeks. Respondents aged 45 to 69 tend to complete both diaries without recall. Greater use of total recall for one of the two weeks is found for respondents aged

Table 5. Model of Performance, Region, CU income, and Unit structure (Standardized Values)

	Two-week Participant			One-week Participant
	By Total Recall			
	Neither Week	One Week	Both Weeks	
Interaction of performance and region				
Northeast	-1.7	2.8	3.1	-5.4
Midwest	1.3	0.4	-1.2	-0.3
South	-1.4	-0.3	-1.6	3.6
West	2.1	-2.2	-1.4	2.6
Interaction of performance and CU annual income				
LT 20,000	2.8	-0.2	2.0	-3.3
20,000 or more	16.3	1.1	-5.5	-4.5
Incomplete	-14.9	-0.6	3.4	8.1
Interaction of performance and unit structure				
Single unit home	6.6	-1.6	-1.0	-2.9
Other	-6.6	1.6	1.0	2.9
Jackknifed chi-square test:				
	Test value			df
Region	3.1 **			9
CU income	22.4 ***			6
Unit structure	7.8 ***			6

Table 6. Model of Performance, Age and Education of Reference Person, and CU size (Standardized Values)

	Two-week Participant			One-week Participant
	By Total Recall			
	Neither Week	One Week	Both Weeks	
Interaction of performance and age				
Less than 27	-4.2	1.3	-0.5	1.6
27-44	1.2	2.0	-3.0	0.1
45-69	4.1	0.6	-1.7	-1.7
70 and over	0.5	-3.3	5.3	-0.5
Interaction of performance and education				
LE high school	-7.5	0.4	4.3	-0.4
GT high school	7.5	-0.4	-4.3	0.4
Interaction of performance and CU size				
1 person	-4.5	-2.5	2.5	3.9
2+ persons	4.5	2.5	-2.5	-3.9
Jackknifed chi-square test:				
	Test value			df
Age	6.0 ***			9
Education	7.6 ***			3
CU size	7.6 ***			3

27 to 44, compared to other respondents in the same education and CU size class.

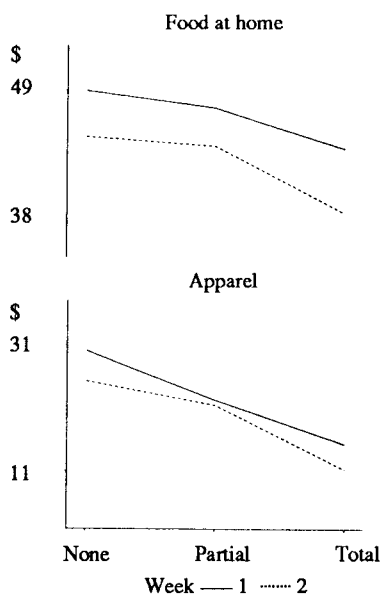
One-week diaries

Younger respondents show greater propensity to produce only one usable diary (14% compared to 10% for all respondents), although the difference is not significant when age is tested together with CU size. The latter factor is significantly associated to one-week participation, as shown in Table 6. Specifically, single person CUs (of any age) are more likely to produce only one usable diary. One-week diaries are more prevalent for renters and in the South and West. CUs living in single unit homes are the least likely to participate only one week, but CUs living in buildings of 2 or more units are more likely to do so.

3. COMPARISONS OF EXPENDITURE MEANS

Expenditure means show declines with increasing use of recall from partial to total. Both diary weeks are affected, as shown in Fig. 1 for food-at-home and apparel expenses.

Fig. 1. Weekly Expenditure Means by Use of Recall



While declines are observed for overall means, different relationships result when means are analyzed for groups of respondents. Multivariate comparisons by age and CU size are shown in Table 7 for two expense classes: *food* (food and beverages) and *apparel and other* diary expenses. The comparisons refer to respondents that used the same procedure both weeks, i.e., (1) partial or no recall versus (2) total recall, and take into consideration selected respondent characteristics found to be associated with diary performance. The tests are multiple comparisons based on the Hotelling T² for differences between mean vectors derived from the two sets of respondents. The covariance matrix was obtained by

the balanced repeated replication method. The contrasts chosen identify main effects and some interactions. (Johnson and Wichern, 1982; Kock and Lemeshow, 1972).

Overall, greater effects are found for *apparel and other* expenses: both age and CU size show significant test values. Only the test for age is significant for *food*. Comparisons for both expense classes suggest an interaction between age and size, but this is not significant due to large standard errors.

Only declines (often greater than 50%) are observed for *apparel and other* expenses. The comparisons for *food* show

Table 7. Comparisons of Weekly Expenditure Means

	Food and Beverage Expenses		Apparel and Other Expenses	
	Total Recall Neither Week	Both Weeks	Total Recall Neither Week	Both Weeks
OVERALL	\$ 75	\$ 66	\$ 49	\$ 19
By Age and CU size:				
Less than 27				
1 person	\$ 40 (2)	\$ 43 (5)	\$ 25 (2)	\$ 16 (5)
2+ persons	64 (2)	78 (10)	38 (3)	14 (3)
27-44				
1 person	52 (2)	63 (10)	39 (3)	28 (9)
2+ persons	91 (1)	90 (10)	63 (2)	28 (4)
45-69				
1 person	46 (2)	39 (3)	32 (3)	8 (3)
2+ persons	91 (1)	77 (3)	60 (2)	22 (3)
70 and over				
1 person	32 (1)	27 (2)	16 (1)	6 (2)
2+ persons	64 (2)	57 (7)	29 (2)	15 (4)
Hotelling T²				
(df=3)				
Age effects				
LT 45 vs. GE 45	7.7 *		0.3	
LT 27 vs. 27-44	0.4		0.9	
45-69 vs. GE 70	1.2		12.4***	
CU size effects				
(df=1)				
1 vs. 2+	0.3		17.8***	

Standard error in parenthesis. "Other expenses" includes small home furnishings and entertainment expenses. Simultaneous confidence intervals derived using the Bonferroni method with percentile $t_n(.05/2df)$. * 5% level, *** 0.1% level.

different results. Declines are apparent for the two older age groups, but not for younger respondents and respondents not quite as young (up to about 40 years of age) living alone. Total recall tends to produce greater means there, as pointed out by the significant contrast shown in Table 7. Telescoping effects are not likely. There are other reasons that seem more plausible. First, the response method of total recall may isolate, within the same group of people, greater proportions of individuals consuming more expensive items. Second, responses by total recall are not reported in great detail, and the cost is likely to be for *groceries* rather than *food*. This interpretation is supported by findings from comparative research on food estimates derived from different methods. Data obtained by interview are responses to *global* questions, e.g., pertaining to a whole month, and generally provide higher estimates than diary based estimates. Nevramount (1991), for example, found this relationship in the Canadian Expenditure Survey; in addition, the difference between the estimates was more accentuated for single person CUs.

4. DISCUSSION

The use of total recall is linked to a number of respondent and environmental characteristics, such as age and region. Older respondents are more likely to use total recall both weeks, implying a greater need for this procedure. Other groups tend to use total recall either one or both weeks. Inadequate cooperation rather than inability to keep the diary may be the cause. The hypothesis surrounding the use of total recall is whether respondents with fewer expenses are more likely not to keep the diary. This hypothesis is rejected, for two reasons. First, total recall is not used only in unusual circumstances, e.g., elderly respondents living alone and unable to keep the diary. The procedure is used by all types of respondents. The frequencies for owners, for instance, are lower than average but still relatively high (5% one week, 9% both weeks). Second, this analysis shows that expenses reported by recall are non-trivial, and sometimes greater than by diary method.

The use of total recall affects overall estimates downward. It is apparent that non-food items are more affected, suggesting that less efforts are placed on those sections. Food expenses tend to be reported relatively more, although with less than desirable detail.

There are a number of approaches for improving data collection in the CE Diary. Field procedures for using recall have already been modified in 1991, including a structured interview by expenditure section. At the same time, two diary formats are being tested in an "embedded" experiment using 80%-20% subsamples of the ongoing sample. Another embedded experiment, not currently being looked at, could test the auxiliary use of personal pocket diaries for individual spenders in the CU. These booklets could focus on special categories of expenses often overlooked in the diary, such as

apparel and food consumed away from home. A different approach is to test a separate diary for non-food items. A prototype of an apparel diary has been designed by Westat under contract to BLS (Cantor et al., 1990).

Other research should focus specifically on the use of total recall. There are cognitive effects, such as potential differences by age and type of expense. There are also differences in field practices across the country. Specific guidelines should be developed for total recall, e.g., if records are consulted or if accompanied by some receipts. Eliminating total recall as a fall-back procedure would be desirable. However it is not a viable option, since a large increment in nonresponse would result. Instead, new recording methods should be developed in order to make diarykeeping more easily accomplished by all respondents.

REFERENCES

- Cantor, D. et al. (WESTAT), and Rose, M. et al. (American Institute for Research) (1990), "Cognitive Research to Design an Apparel Diary Field Test," Final Report under contract to BLS.
- Elliot, D. (1991), "Weighting Expenditure and Income Estimates from the U.K. Family Expenditure Survey to Compensate for Nonresponse," Paper presented at the Workshop on Diary Surveys, Statistics Sweden.
- Éltető, Ö. (1991), "Nonresponse Rates and a Way of Adjusting for Nonresponse in Household Expenditure Surveys in Hungary," Paper presented at the Workshop on Diary Surveys, Statistics Sweden.
- Fay, R.E. (1985), "A Jackknifed Chi-Square Test for Complex Surveys," *Journal of the American Statistical Association*, Vol. 80, pp. 148-157.
- Fay, R.E. (1987), "Contingency Table Analysis for Complex Surveys (CPLX), Program Documentation," *Washington Statistical Society Short Course*, Washington, D.C.
- Groves, R.M. (1989), *Survey Errors and Survey Costs*, New York: Wiley, pp. 337-341.
- Harrison, R. (1991), "Respondent Burden and Response Fatigue in the 1988-89 Australian Household Expenditure Survey," Paper presented at the Workshop on Diary Surveys, Statistics Sweden.
- Johnson, R.A., and Wichern, D.W. (1982), *Applied Multivariate Statistical Analysis*, 2nd Edition, Englewood Cliff: Prentice Hall.
- Kemsley, W.F.F., Redpath, R.U., and Holmes, M. (1980), "Family Expenditure Survey Handbook," Office of Population and Surveys, London, Her Majesty's Stationary Office.
- Koch, G.G., and Lemeshow, S. (1972), "An Application of Multivariate Analysis in Complex Surveys," *Journal of the American Statistical Association*, Vol. 67, pp. 780-782.
- Lyberg, L. (1991), "Reducing Nonresponse Rates in Household Expenditure Surveys by Forming Ad Hoc Task Forces," Paper presented at the Workshop on Diary Surveys, Statistics Sweden.
- Nevramount, U. (1991), "Evaluation of the Canadian Food Diary Survey," Paper presented at the Workshop on Diary Surveys, Statistics Sweden.
- Silberstein, A., and Scott, S. (1991), "Expenditure Diary Surveys and Their Associated Errors," in: *Measurement Errors in Surveys*, Wiley (In Press).
- Tucker, C. (1988), "Estimating Measurement Errors in Surveys," Paper presented at the Annual Conference of the American Association for Public Opinion Research.