

The Role of Demographic Factors in the Analysis of Survey Versus Diary Purchase Reporting Accuracy

EDWARD R. BRUNING and MICHAEL Y. HU¹

ABSTRACT

In this article the authors evaluate the relative performance of survey and diary data collection methods in the context of the long-distance telephone communication market. Based on an analysis of 1,530 respondents, the results indicate that two demographic variables, sex and income, are important in explaining the difference in survey reporting and diary recording of usage data.

KEY WORDS: Survey; Diary; Data collection.

1. INTRODUCTION

A perusal of the marketing literature underscores our lack of knowledge regarding the relative accuracy of survey and diary methods for collecting consumer expenditure data. Clearly, the resolution of this issue has ramifications for researchers as well as those for whom the research is conducted. Wind and Lerner (1979) stress the need to appropriately evaluate the two methods and to identify the characteristics of those reporting purchase behavior accurately versus those that have a high discrepancy between reported and actual consumption. To be sure, an analysis of the discrepancy focuses attention on the data collection instrument, for the choice of instrument could affect management decisions relating to "product positioning and market segmentation strategies, advertising media and copy research, and concept/product testing." (Wind and Lerner 1979).

The purpose of our article is to assess empirically the relationship between several demographic variables and the two expenditure reporting methods from a single sample of respondents in the U.S. long-distance telephone market. We present additional evidence on the issue initially posed by Wind and Lerner. First, the current state of knowledge regarding the nature of the two instruments is surveyed. Then the research methodology is described and findings from the long-distance telephone market are reported. We conclude with a number of implications relevant to both providers and users of consumer expenditure data.

2. LITERATURE REVIEW

The two prominent methods for recording household consumption expenditures are survey (recall) methods, whereby household members are asked to recall expenditures made during a predefined period, and the diary method, whereby a daily or weekly log is maintained which identifies specific expenditures. Neter (1970) provides case examples and empirical studies which address the relative advantages and disadvantages of the two expenditure collection devices but do not compare their relative accuracies. In general, the survey approach possesses advantages in economy while simultaneously possessing a number of disadvantages relative to the diary method. Because of time and resource constraints, most researchers utilize the survey method even with the multitude of measurement problems.

¹ Edward R. Bruning and Michael Y. Hu, Graduate School of Management, Kent State University, Kent, Ohio, 44242.

It is commonly believed that diary methods have advantages over survey approaches principally because diarists have the opportunity to record the event within a short period after it has occurred. For this reason, Sudman and Ferber (1971) have all but discredited the survey approach for collecting expenditure data and have suggested the exclusive use of diaries. But the diary method is not problem free. The authors evaluated households in the Chicago area in 1972 and found evidence of underreporting by the survey method with respect to the number of purchases. They also found that respondents had difficulties in separating purchases into specific item categories with the survey recall method.

A number of writers report that the diary approach is appropriate only for certain expenditure categories (Pearl 1968; Grootaert 1986; Wind and Lerner 1979; Stanton and Tucci 1982). Pearl (1968) has stated that individual diaries are to be preferred because of reporting thoroughness. For large ticket items the method is preferred; however, reporting frequency declines for small valued purchases. Grootaert (1986) adds to this prescription by suggesting that all eligible household members keep diaries to reduce omissions in expenditure reporting. Wind and Lerner (1979) and Stanton and Tucci (1982), in separate studies on expenditure reporting for specific food items, substantiate the superiority of the panel method relative to surveys.

The construction and design of the diary instrument poses collection problems (Kemsley 1961; Kemsley and Nicholson 1960; Lewis 1948; Sudman 1964a, b; Sudman and Ferber 1971; Walsh 1977). Kemsley (1961) and Kemsley and Nicholson (1960) evaluated record books kept on consumer expenditures over a three week period in 1953. They found that significant variations occurred in expenditure recording over the three week period by type of expenditure and by season of the year. Lewis (1948) evaluated the accuracy of weekly versus monthly diary recording of grocery and clothing expenditures. The author found a 16% reduction in monthly reporting in comparison to weekly expenditure reporting. Sudman (1964a) and Sudman and Ferber (1974) studied alternative means of obtaining consumer expenditure data. They evaluated the role of compensation, training of respondents, and method of reporting. In the studies they conducted, compensation was significant in improving respondent cooperation and accuracy, and direct training aided in respondent reporting accuracy. The frequency of purchase and the construction of the reporting form were also important in reporting accuracy.

Other studies have focused more explicitly on consuming unit cooperation. (Kemsley and Nicholson 1960; Pearl 1968; Sudman and Ferber 1974). Kemsley and Nicholson (1960) report that the size of the individual purchase has a significant effect upon the degree to which respondents cooperate in reporting expenditures. Pearl (1968) and Sudman and Ferber (1974) emphasize the incentive payments in terms of amount and duration in generating cooperative expenditure reporting.

An additional concern with the diary method is the extent of panel mortality (Sandage 1956; Sodol 1959; Sudman 1964a, b) and panel decay (McKenzie 1983; Sandage 1956; Sodol 1959; Sudman 1964 a, b). Sandage (1956) investigated whether consumer panels develop bias as a result of being interviewed. Based on three separate investigations on Indiana farm households over the period 1947-1954, the author found that bias was not a significant concern with panel collection methods. Sudman (1964a, b), however, found mortality tended to be greater for male respondents. In addition, the degree of effort involved in recording appeared to have no impact on accuracy or mortality rates for respondents involved in panel recording of expenditures. In terms of panel decay, McKenzie (1983) reported that greater attrition occurred with longer panel periods while Sandage (1956) found that repeated use of a given panel did not result in a bias in reporting accuracy.

Parfitt (1967) argues that housewives in surveys recall accurately only purchases for frequently bought products in the most recent past. Thus, the diary recording of past purchases

yields a more reliable and accurate measure than survey reporting. In surveys, respondents typically are asked to report purchases over a long time period or to engage in a mental averaging exercise to arrive at an expenditure figure for a typical week or month. As a consequence, Parfitt (1967) concludes that a strong likelihood exists for respondents to exaggerate the amount and frequency of purchases and to oversimplify the complexity of the expenditure decision.

As indicated in an earlier section of this paper, our research focuses on the accuracy of survey versus diary purchase reporting. Only a few articles address this issue empirically. Wind and Lerner (1979) analyze the validity of survey versus diary approaches in accounting for consumer expenditures. Their data are taken from a sample of 450 housewives serving on a MRCA consumer diary panel. The housewives completed a mail survey questionnaire and were instructed to maintain a record of their expenditures of various brands of margarine for a six month period. The results indicate a discrepancy in the relative accuracy of the two reporting methods between the aggregate and the individual consumer response level. At the aggregate level, survey and diary instruments are consistent in predicting the rank-ordering of brand market shares. Major discrepancies are detected, however, at the consumer level as survey responses are less accurate as compared to diary reporting. The authors attribute this inaccuracy as resulting from ignorance, forgetfulness, poor survey questioning, reporting errors, falsification, and interviewer bias.

Stanton and Tucci (1982), following the work reported by Wind and Lerner (1979), sample 7,945 participants in the National Food Consumption Survey (1977-78). Personal interviews are used as the reporting vehicle for food expenditures which occurred in the previous twenty-four hour period. The participants were asked to maintain diaries of all food and beverage expenditures for two days following the interview. Their results indicate that, at aggregate levels, personal interviews provide information which is as accurate and reliable as diary reports. They were not able to address the relative accuracy of the two approaches at the consumer level because of the nature of the data.

The apparent discrepancies in results reported by Wind and Lerner (1979) and Stanton and Tucci (1982) may be attributable to the differences in the time frames within which consumers operated in reporting expenditures. In Wind and Lerner's study, respondents were requested to report the brand most often purchased. Questions of this nature require a greater amount of recall since the time reference is over an extended period. In Stanton and Tucci's study, however, the reporting period is restricted to the previous twenty-four hours. Parfitt (1967) indicates that respondents are more effective in reporting recent purchases. In this light, Stanton and Tucci's conclusion is not truly surprising and, furthermore, does not contradict the results of Wind and Lerner's analysis since recall for both the survey and diary recording methods was high.

3. THE STUDY

During the years 1978 and 1979, AT&T (American Telephone and Telegraph Company) initiated a major data collection effort with the objective of providing information for corporate market planning and strategy formulation in its residential long-distance telephone market. A nationally projectable sample of roughly 4,000 households were recruited and asked to participate on a panel for a period of twelve months. The sample was demographically balanced with respect to six variables: population density, income, marital status, age, sex and geographical region of domicile.

The entire panel responded to a pre-assessment survey instrument administered through the mail in January 1978. Once completed, each panel member was instructed to fill out a weekly diary over the next twelve months. In the pre-assessment phase, respondents were asked to

respond to the question: "During an average or typical month, how often do you communicate for non-business reasons with relatives and friends who reside at least 50 miles from your home?" This measure is referred to as [PERCEIVED 1]. Also, each panel member in the pre-assessment phase responded to the question: "Would you consider yourself a heavy, medium, light or non-user of long distance calling?" (Refer to this measure as PERCEIVED 2). So that comparisons could be made between panel and survey data collection methods, panel respondents were asked to record information on the frequencies of long-distance communication by day of the week. This measure is referred to as REPORTED 1.

Throughout the entire study every attempt was made to conceal the sponsor of the project. Moreover, the positions of the response categories were randomized in order to remove any possibility of position bias. A sample of 2,350 respondents was retained after twelve months of reporting. Panel attrition was perceived to be a potential problem in this study because attrition rates may vary substantially among demographically defined subgroups. In order to resolve this problem, a sample balance program was developed and used to randomly select a subsample of participants from the pool of 2,350 respondents which would be demographically balanced. After editing and sample balancing, 1,530 panel members who had completed the pre-assessment and the twelve-month diaries were used in this study.

4. DATA ANALYSIS

An important question in the pre-assessment survey asked the respondents to report their "perceived" usage for a typical month [PERCEIVED 1]. In order to obtain consistency in the unit of measurement, weekly diary recorded usage [REPORTED 1] is aggregated to twelve monthly totals for each respondent. Refer to the aggregated diary reported measure as REPORTED 2. Matched differences between "perceived" usage reported in the pre-assessment survey [PERCEIVED 1] and "actual" usage extracted from diaries [REPORTED 2] are calculated for each respondent for twelve monthly periods as well as for the average of the twelve months. A one-way ANOVA design is employed monthly and for the twelve month average to detect if significant variations exist with respect to the matched differences across levels of several demographic variables: sex, income, education, and age. An a posteriori contrast test is performed to compare all possible pairs of level means for each demographic variable. Finally, to evaluate the effects of interactions among the four demographic variables, a four-way ANOVA procedure is employed using the twelve-month average scores.

5. RESULTS

5.1 Survey and Diary Average Reported Usage

Table 1 reports the average number of long distance telephone communications extracted from respondent diaries for each of the twelve months as well as the usage for a typical month [PERCEIVED 1] taken from the pre-assessment survey. Interestingly, this "perceived" usage reported in the pre-assessment survey is substantially greater than actual recorded usage [REPORTED 2] for each month of the analysis.

The diary averages indicate the presence of seasonality in the usage. December 1978 usage of 4.123 is the highest among the twelve reported months. Even though the pre-assessment survey requested the respondents to report usage for an average or typical month, it is quite likely that they would use December 1977 as the basis for response since the pre-assessment

survey was administered in January 1978. A one-sample t-test indicates that the average of the paired-difference between pre-assessment and the December diary usage, 0.235, is significantly different from zero (p -value = 0.001). By the same token, t-test results for the other eleven averages are statistically significant. These results imply that the respondents have indeed over-estimated in the pre-assessment survey as compared to the diary reported usage.

A potential concern is that the reported usage in the pre-assessment survey could be influenced by the unusually high usage in December 1977. If so, then it is argued that the results of our study are subject to seasonality bias. In addressing this issue, the authors have examined the difference between the reported usage in the pre-assessment survey and the December 1978 diary. Comparing the same months over a year of time could help to eliminate the seasonality factor. As indicated in Table 1, this difference is statistically significant. This difference, however, can be due to the difference in the data collection method and to a trend factor since the comparison involves two different years. Assuming a positive trend in the usage of services over time, the reported usage in December 1978 should be higher than that of December 1977. The data from Table 1 indicates quite the contrary. Usage in December 1977 was significantly higher than that reflected in December 1978. Thus, this evidence leads us to conclude that there is indeed a significant difference due to the data collection method. Respondents in our study had over-estimated their usage in the pre-assessment survey as compared to their estimates reported in the diary.

Prior to our analyzing the relationship between the difference in survey versus diary data collection methods and the several demographic variables, it is important to evaluate the role played by actual usage in explaining this difference. Our reasoning for this test is that if the difference between survey reporting and diary recording is due to the absolute level of usage, then further analysis would prove suspect since experience (learning) would tend to bias our dependent variable (McKenzie 1983). On the other hand, if no statistical significance is attributable to the differences in collection methods and absolute usage levels, then the analysis with the demographic variables would be of greater validity.

Table 1
Average Absolute Number of Long-Distance Telephone
Communications and Pre-assessment Survey Estimates

Month	Average Absolute Number of Communications
February	3.516
March	3.878
April	3.486
May	3.610
June	3.414
July	3.604
August	3.606
September	3.250
October	3.426
November	3.518
December	4.123
January	3.891
Preassessment Survey Estimate	4.358
	$n = 1530$

Table 2
One-Way ANOVA Results Relating the Degree of Long-Distance Telephone Usage and the Difference Between Survey and Diary Reporting (12 month average)

	Degree of Usage				<i>P</i> -Value
	Heavy	Medium	Light	Non-User	
Mean Difference (survey-diary)	0.762	0.799	0.795	0.580	0.9905
<i>n</i>	316	605	547	45	

Table 2 reports the results of the analysis of the relationship between the difference in survey [PERCEIVED 1] and diary [REPORTED 2] reportings and the degree of absolute usage [PERCEIVED 2]. McKenzie evaluated the form of both response and recording bias involving the collection of telephone call details by diary methods. Response rate was found to vary with customer usage. Furthermore, telephone usage recording rates tended to decrease with usage as well. Thus, telephone call data collected by diary methods are subject to several biases. Our study focuses on the difference in survey versus diary collected data and customer usage where the emphasis lies with the discrepancy between “perceived” and “actual” consumption/purchase and the level (degree) of usage. Even though recording biases exist with both methods, nonetheless, the *difference* between the two recordings is not related to usage.

In addition, the validity of using PERCEIVED 2 as a categorization variable can be examined by correlating this measure with REPORTED 2 and PERCEIVED 1. REPORTED 2 and PERCEIVED 1 measurements were first categorized into heavy, medium, light, and non-user employing different cut-off levels. Cross-tabulations were then conducted between PERCEIVED 1 and these two categorical measures. Significant statistical relationships were detected in all cases.

Our dependent variable is the difference in the survey [PERCEIVED 1] and diary usage recordings [REPORTED 2] and the independent variable is the degree of usage divided into four levels: heavy, medium, light and non-user [PERCEIVED 2]. The results from the one-way ANOVA procedure using the least-squares estimation procedure indicate that the degree of usage is not statistically significant ($p = .9905$) in explaining the recorded usage difference between the survey and diary methods. A one sample t-test of each of the four individual group means showed that each mean was statistically different from zero at the 0.01 significance level. Therefore, the results imply that with respect to each of the four usage groups the positive mean values represent that respondents tend to over-estimate usage in the pre-assessment survey relative to the diary recording method.

5.2 Relationship Between Survey and Diary Reported Usage Differences and Selected Demographic Variables

In Table 1 we reported the existence of a substantial difference between survey and diary collection methods for the same respondents over a twelve month period. An interesting question is: what accounts for the perceptual bias in survey reporting of purchase data? To answer this question, a number of demographic factors are evaluated. Several levels of each factor are specified and a one-way ANOVA procedure is employed to account for the reporting differences. Tables 3 through 7 report the results of the analyses.

Table 3
One-Way ANOVA Results Relating Sex of Respondent and
the Difference Between Survey and Diary Reporting of Data

Month	Differences by Sex (Survey — Diary)		ANOVA p-value
	Male	Female	
February	0.412	1.135	0.006*
March	-0.015	0.818	0.005*
April	0.379	1.201	0.002*
May	0.310	1.304	0.008*
June	0.562	1.205	0.016**
July	0.376	1.008	0.018**
August	0.395	0.987	0.031**
September	0.927	1.225	0.258
October	0.605	1.149	0.042**
November	0.593	1.003	0.129
December	-0.112	0.464	0.041**
January	0.164	0.675	0.075**
Mean ^a	0.380	0.990	0.010*
<i>n</i>	617	911	

^a Twelve month average.

* Significant at the 0.01 level.

** Significant at the 0.05 level.

5.3 Sex

The relationship between the difference in survey and diary recordings of usage and sex of the respondent is depicted in Table 3. The one-way ANOVA p-values are statistically significant for 9 of the 12 months at the 0.05 level or below and significant at the 0.01 level for the twelve month average. Thus the results indicate that both male and female respondents over-estimate their actual usage of long distance telephone service and that females over-estimate to a greater degree than do males.

5.4 Income

In Table 4 we present the difference between survey and diary usage reports in relation to respondents' household income level. For 6 of the 12 months the one-way ANOVA p-values are statistically significant at the 0.05 level or better and the 12-month average is significant at the 0.037 level. Furthermore, the results of Tukey's Studentized t-test indicate that respondents with annual household income in the Category 1 range (\$5,000 or less) are statistically distinct from respondents earning incomes within the range of \$10,001 to \$20,000.

An obvious anomaly in the findings reported in Table 4 is that for respondents within the lowest income category (\$5,000 or less), estimated average monthly usage is below the actual monthly usage in 9 of the 12 periods. Furthermore, with increasing household income a definite persistence to over-estimate usage occurs although this process begins to subside at the highest income category. At lower income levels consumers may perceive long-distance telephone service as a luxury item with respect to the other modes as well as with regard to other consumer expenditures. Consequently, when asked to report expected usage, as in a survey, respondents from this income strata tend to discount their perceived usage because of the belief

Table 4
ANOVA Results Relating Respondent's Income Level and
the Difference Between Survey and
Diary Reporting of Long-Distance Telephone Usage

	Differences by Income (Survey — Diary)					<i>p</i> -value
	0-\$5,000 (1)	\$5,001 – 10,000 (2)	\$10,001 – 15,000 (3)	\$15,001 – 20,000 (4)	Over 20,000 (5)	
February	-0.010	-0.583	1.180	1.120	0.571	0.110
March	-0.480	-0.738	0.780	1.009	-0.062	0.019**
April	-0.337	0.851	1.188	1.258	0.550	0.031
May	-0.327	0.560	0.928	0.991	0.636	0.220
June	0.102	0.911	1.027	1.331	0.756	0.249
July	-0.439	0.500	0.895	1.050	0.694	0.128
August	-0.408	0.512	1.021	1.235	0.498	0.036**
September	0.306	0.798	1.298	1.367	0.976	0.301
October	-0.469	0.542	1.231	1.413	0.720	0.009*
November	0.010	0.494	0.941	1.214	0.741	0.248
December	-1.010	0.060	0.209	0.792	0.101	0.050**
January	-0.633	-0.339	0.654	0.956	0.392	0.030**
Mean ^{a,b}	-0.308	0.517	0.946	1.145	0.548	0.037**
<i>n</i>	98	168	373	341	536	

^a Twelve month average.

^b Tukey's Contrast Test: (1) and (4) and (1) and (3) are different at the $p = 0.05$ level.

* Significant at the 0.01 level.

** Significant at the 0.05 level.

that limited monies should be spent elsewhere. At the actual point of consumption, however, relative values may have changed since the urgency of the situation may dictate a long-distance telephone call is indeed the low-cost option relative to alternative communication means. Thus, survey reporting of planned usage may deviate from diary recordings of actual usage because of situational factors that intervene during the time of consumption.

As respondents' household incomes increase, long-distance telephone use is still perceived as a superior good; however, whereas respondents in lower income levels perceive long-distance telephone use as an expendable (and perhaps frivolous) purchase, wealthier respondents "expect" to employ the telephone more often than the other modes. Thus, when surveyed as to their "expected" usage, wealthier respondents tend to overestimate the number of long-distance telephone communications since in most situations it is their preferred method of communicating.

5.5 Age

Table 5 reports that respondents at every age level tend to over-estimate their "perceived" usage relative to "actual" usage as recorded in diaries. Although the one-way ANOVA p -values indicate the nonexistence of a significant relationship between measurement methods and age of the respondent; nonetheless, in 10 of the 12 months respondents less than 31 years of age incurred the lowest difference relative to older respondents. In addition, average differences for respondents between 31 and 40 and over 50 were lower than the average for the less than

Table 5
 One-Way ANOVA Results Relating Respondent's Age and
 the Difference Between Survey Reporting and
 Diary Recording of Long-Distance Telephone Usage

Month	Differences by Age (Survey — Diary)				<i>p</i> -value
	Below 31 (1)	31-40 (2)	41-50 (3)	Over 50 (4)	
February	0.632	0.749	1.026	0.949	0.310
March	0.016	0.348	0.837	0.709	0.210
April	0.413	1.083	1.174	0.889	0.209
May	0.305	0.706	1.085	0.923	0.217
June	0.525	0.845	1.570	0.989	0.080
July	0.535	0.706	1.226	0.667	0.371
August	0.507	0.807	1.070	0.667	0.578
September	0.924	1.003	1.459	1.109	0.580
October	0.789	0.816	1.307	0.903	0.583
November	0.632	0.805	1.415	0.741	0.240
December	0.337	0.203	0.574	-0.030	0.494
January	0.603	0.519	0.922	0.069	0.197
Mean ^a	0.518	0.716	1.139	0.715	0.385
<i>n</i>	383	374	270	495	

^a Twelve month average.

* Significant at the 0.01 level.

** Significant at the 0.05 level.

31 group in each of the twelve periods. Thus, the relationship between differences in survey and diary usage reports and age of respondent is a monotonically increasing function up to age 50 where the difference, although still positive, declines after age 50. Again, the average differences across the various age levels are not statistically significant based on the one-way ANOVA or the Tukey Studentized *t*-tests.

5.6 Education

The relationship between the difference in survey versus diary reported usage and respondents' level of education is depicted in the statistics found in Table 6. As reported in the table, a general tendency to over-estimate usage in surveys is characteristic of respondents at all education levels. Respondents with the least amount of formal education tend to over-estimate usage in survey reporting to a lesser extent than respondents with more formal education. The greatest tendency to over-estimate usage in surveys occurs for respondents who have completed high school followed by those who have had some college. The results of the one-way ANOVA and Tukey Studentized *t*-tests, however, indicate that the differences across education levels are not statistically significant at the $p = 0.05$ level.

5.7 Four-Way ANOVA Results

The main and interaction effects of the demographic variables as explanations of the difference between survey and diary purchase data reporting are presented in Table 7. It is reported that income, sex, and their interaction are the variables with statistically significant *p*-values. All other main and interaction affects are insignificant in explaining variations in the difference variable.

Table 6
One-Way Anova Results Relating Respondent's Education Level and the
Difference Between Survey and Diary Reported Usage

Month	Differences by Education (Survey-Diary)				<i>p</i> -Value
	Some High School (1)	Completed High School (2)	Some College (3)	Completed 4-Yr. Col.Deg (4)	
February	0.790	1.015	0.951	0.578	0.546
March	0.290	0.853	0.592	0.059	0.117
April	0.556	1.275	0.979	0.445	0.078
May	0.685	1.134	0.756	0.345	0.139
June	0.548	1.158	1.111	0.696	0.368
July	0.194	0.931	1.021	0.467	0.195
August	0.347	0.891	0.845	0.620	0.681
September	1.040	1.137	1.190	1.018	0.959
October	0.468	1.195	0.826	0.878	0.475
November	0.508	1.119	0.896	0.592	0.383
December	0.081	0.500	0.244	0.061	0.558
January	-0.097	0.626	0.842	0.129	0.138
Mean ^a	0.438	0.986	0.854	0.491	0.294
<i>n</i>	124	476	431	490	

^a Twelve month average.

* Significant at the 0.01 level.

** Significant at the 0.05 level.

Table 7
Four-Way ANOVA Results Relating
Demographics (Sex, Education, Age, and Income)
to the Differences Between Survey and
Diary Recording of Long-Distance Telephone Usage

Variable	Df	Sum of Squares	F-Value	<i>p</i> -Value
Sex	1	131.082	6.48	0.011**
Education	3	79.001	1.30	0.272
Age	3	58.465	0.96	0.409
Income	4	210.077	2.60	0.035**
Sex & Education	3	77.629	1.28	0.280
Sex & Education	4	220.032	2.72	0.028**
Sex & Age	3	47.311	0.78	0.506
Ed. & Income	12	263.931	1.09	0.367
Ed. & Age	9	81.083	0.45	0.911
Income & Age	12	211.718	0.87	0.576

* Significant at the 0.01 level

** Significant at the 0.05 level.

6. CONCLUSION

The findings of our study indicate that, at the individual respondent level, survey data are very inaccurate in measuring the respondents' actual usage of long-distance telephone communication. Our results support the earlier conclusions of Parfitt (1967), Sudman (1964) and Wind and Lerner (1982) who analyzed this issue with respect to non-service related consumer products. We cannot report, however, that our results either support or refute those of Stanton and Tucci (1984) since the time frames, and thus the recall periods, are considerably different in the two studies.

The importance of our findings extends beyond simply confirming the results of previous studies and extending the range of product types to include the analysis of a consumer service item. Our findings identify the fact that the over-reporting that occurs in surveys varies along two important demographic dimensions: respondents' household income and sex. Respondents who report very low household income tend to under-estimate usage in survey reporting while wealthier respondents do the opposite. Furthermore, this relationship tends to increase monotonically with increases in income levels and then declines. Female respondents tend to over-estimate usage in surveys by a considerably greater magnitude relative to male respondents. Taken together the findings suggest a strong possibility for measurement problems occurring if purchase data are collected using the survey method.

ACKNOWLEDGEMENTS

We are indebted to the anonymous referees for providing valuable suggestions which have improved the clarity and precision of our work.

REFERENCES

- 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.
- HIRSHLEIFER, J. (1984). *Price Theory and Applications* (3rd.Ed.). Englewood Cliffs, New Jersey: Prentice-Hall, Inc.
- KEMSLEY, W.F.F. (1961). The Household Expenditure Enquiry of the Ministry of Labour: Variability in the 1953-54 Enquiry. *Applied Statistics*, 10, 117-135.
- KEMSLEY, W.F.F., and NICHOLSON, J.L. (1960). Some experiments in methods of conducting Family Expenditure Surveys. *Journal of the Royal Statistical Society, Series A*, 123, 307-328.
- LEWIS, H. F. (1948). A comparison of consumer responses to weekly and monthly purchase panels. *Journal of Marketing*, 12, 449-454.
- McKENZIE, J. (1983). The accuracy of telephone call data collected by diary methods. *Journal of Marketing Research*, 20, 417-427.
- NETER, J. (1970). Measurement errors in reports of consumer expenditures. *Journal of Marketing Research*, 7, 11-25.
- PARFITT, J. (1967). A comparison of purchase recall with diary panel records. *Journal of Advertising Research*, 7, 16-31.
- PEARL, R.B. (1968). Methodology of Consumer Expenditure Surveys. Technical Working Paper 27, Washington D.C.: U.S. Bureau of the Census.
- SANDAGE, C.H. (1956). Do research panels wear out? *Journal of Marketing*, 20, 397-401.

- SODOL, M.G. (1959). Panel mortality and panel bias. *Journal of the American Statistical Association*, 59, 52-68.
- 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. (1964a). On the accuracy of recording of consumer panels: I. *Journal of Marketing Research*, 1, 14-20.
- SUDMAN, S. (1964b). On the accuracy of recording of consumer panels: II. *Journal of Marketing Research*, 1, 69-83.
- SUDMAN, S., and FERBER, R. (1974). A comparison of alternative procedures for collecting consumer expenditure data for frequently purchased products. *Journal of Marketing Research*, 11, 128-135.
- WIND, Y., and LERNER, D. (1979). On the measurement of purchase data: Surveys versus purchase diaries. *Journal of Marketing Research*, 16, 39-47.