ALTERNATIVE CONTINUING POINT OF PURCHASE SAMPLE DESIGNS

Cathryn S. Dippo and William D. McCarthy, Bureau of Labor Statistics

I. INTRODUCTION

The Point of Purchase Survey (POPS) is a household survey used to obtain the names and addresses of outlets where the relevant Consumer Price Index (CPI) populations purchase items to be priced in the CPI. For the CPI revision just completed, the POPS was conducted in 1974. Approximately 23,000 sample addresses were selected in the 85 CPI PSU's (primary sampling units) for interviewing. Two different questionnaires or checklists were used. Three-fourths of the selected units were asked the names and addresses of outlets where they had purchased a list of 73 groups of items including: prescription drugs, fees for participant sports and beer and ale for home use within the last two weeks; household linens, toys, and garden supplies in the last three months; various clothing items and household appliances in the last six months; hospital care and furniture in the last year; and stereo equipment and used cars in the last two years. The other quarter of the households were asked about 34 items including groceries and gasoline in the last week, some clothing and other items in the last three months, and a subset of the two year recall items asked on the other questionnaire. Along with the names and addresses of outlets, respondents reported their expenditures for the particular category at the outlet in the given reference period.

In 1974, the POPS sample was a clustered sample of approximately five housing units per enumeration district (ED), where about 8 ED's were selected from secondary sampling units (SSU's) consisting of about 2500 housing units. 1/ For most of the self-representing (SR) PSU's, the total sample size was approximately 480. New York, Chicago and Los Angeles had larger samples selected and Anchorage, Honolulu, N.E. Penna, and Portland had only 240 units selected. In the non-self-representing (NSR) PSU's 160 sample units were selected.

In order to keep the outlet sample for the CPI up-to-date, BLS has begun a program for updating the outlets by conducting CPOPS in one-fifth of the PSU's each year. The samples for the 1977 and 1978 CPOPS have been designed to provide data to answer a number of questions, including the feasibility of merging CPOPS with CCES or the Continuing Consumer Expenditure Survey. CCES, which will be very similar to the 1972-73 CES with respect to questionnaire content, will be initiated in late 1979 in the 85 urban CPI PSU's plus Fairbanks and 16 PSU's selected to represent the rural areas of the United States. The CCES sample will consist of approximately 9600 usable 1-week diaries per year and 4800 bounded usable interviews per quarter. The housing units will be selected in a systematic unclustered manner.

The long-range goal of BLS is to optimize the entire CPI survey program including CPOPS and CCES with respect to cost and reliability. One aspect of this optimization is the examination of the merger - either partially or totally - of the two household surveys. The issues involved in this examination are by no means simple and straightforward for the factors to be considered are numerous and interdependent. For example, a complete merger of the two surveys could overburden respondents. This could result in large non-interview rates and, therefore, less reliable expenditure data, which in turn effects the reliability of the cost-weights and the index. As CCES is presently designed, a complete merger would also imply outlets would be obtained from a much smaller and more geographically disperse sample. Will such a sample provide the necessary number of outlets for CPI pricing? Is the cost of this "supposedly" more disperse sample within the budget limitations for CPI pricing? In order to obtain data to help answer these complex issues, two studies are presently being conducted:

(1) A pretest for the CCES, being conducted this summer in 3 PSU's, is testing two methods of obtaining outlet name and address information on the diary questionnaire against a control without outlet information as in 1972-73. The sample sizes for the pretest will limit our examination to gross response rates and highly aggregated expenditure levels.

(2) More importantly, the household samples for the first two years of CPOPS have been selected using four different methods - one method per PSU. These are: (a) Use the same households as selected in 1974. (b) Select an independent sample of housing units from the same SSU's as selected in 1974. (c) Select a new sample using 1974 procedures, i.e., make an independent selection of SSU's and then ED's and housing units within. (d) Select the housing units in a systematic unclustered manner corresponding to the sample selection procedure for CCES.

II. ANALYSIS OF 1974 POPS DATA

As a preamble to comparing data between 1977 and 1974, three aspects of the 1974 data have been analyzed to date: expenditure levels, response rates, and the geographic location of the reported outlets.

A. Expenditure Levels

All the outlets reported for a particular category by all the households in a PSU in either the U (all urban) or W (urban wage-earner and clerical workers) CPI population are used to create the outlet sampling frame for one or more CPI ELI's (entry level items). The measure of size associated with each outlet is the total amount of expenditures reported at that outlet by the appropriate households for the particular category. Separate frames are created for each population, category and PSU.

Since the probability of selecting an outlet is dependent on its reported expenditures, comparisons of total expenditures for a category have been made to data obtained from the 1972-73 CES. Unfortunately, numerous qualifications had to be made on all the comparisons due to definitional and operational differences between the surveys.

For example, expenditures reported for 1974 POPS were restricted to purchases made while the family lived in the PSU and did not include purchases made while out-of-town on vacation. (The latter problem has been eliminated in CPOPS by including such purchases.) However, a few general conclusions could be drawn.

1. There was no statistical difference between the estimated expenditures for the two surveys for about 60% of the categories where it was felt a comparison could be made. Therefore, for a majority of categories the reported expenditures are within tolerance limits.

2. In two-thirds of the instances (10 of 15) where a significant difference was found, the POPS estimate was greater than that from CES. Included in this group were some very broad categories such as: over the counter drugs and medicine, other groceries and prepared foods, and cleaning and laundry supplies; three of the four alcoholic beverage categories; only two of 16 clothing categories; and two of the four more specific food categories.

B. Response rates

Three separate types of response rates were examined: household response, outlet response per household and outlet-category response.

1. The noninterview ratio ranged by PSU from 9 to 46 percent with 21 PSU's having noninterview ratios greater than 20 percent. However, only eight PSU's actually fell short of the interview goals and none were short by more than two percent.

2. The overall average number of gross outlets, including military, private individuals not in business and "don't knows", reported per interview was 19 for the longer form and 17 for the shorter. As you would expect the following trends were observed: Owners report more outlets than renters, CU's with white heads of household report more outlets than those with non-white heads of household, the number of outlets reported increased with income and the number of outlets increased with family size. No apparent significant differences could be detected between urban and rural CU's or central city and balance urban.

3. The most important aspect of examining the response to 1974 POPS was determining whether or not enough outlets were reported for each category to satisfy the sampling needs for CPI. Out of 87 PSU's, 67 PSU's did not have enough outlets reported to support outlet sampling for the W population in one or more categories. In some cases, no outlets for the W population were reported for the categories. In 36 categories, considered to be problem categories, 16 had no outlets reported for the W population in a total of 44 PSU's. For these reasons the sample sizes in the B, C and D areas were increased and some reference periods extended for the 1977 CPOPS.

Additional analysis has been completed on the 1974 data for the 18 PSU's included in 1977 CPOPS. These additional facts are:

- In terms of net outlets at the PSU level, the average number of outlets returned per interview in 1974 was about 7 and the average number outlet-categories per interview was about 14.
- At the SSU level, the average number of households visiting an outlet was 1.6, and the average number of categories purchased per outlet was 2. The average number of households reporting the same, outlet-category at the SSU level was 1.2.

C. Geographic Location of Reported Outlets

The geographic location of the reported outlets is an important factor in determining the travel costs for CPI pricing. The more disperse the outlets are the higher the travel costs for CPI will be and one would suspect a more disperse household sample would yield a more disperse outlet sample. However, our goal is to try to merge the CPOPS and the CCES as much as possible and the optimum CCES sample would be systematic unclustered. Immediately one sees the problem. How much more will it cost to collect prices from outlets obtained from an unclustered household sample and how much will you save in household data collection costs? Along with cost considerations the relative reliability of the data obtained must be examined.

At the heart of the problem is the assumption that an unclustered household sample will yield a more disperse outlet sample than a clustered one. The primary purpose for the alternate CPOPS sample design is to determine if this assumption is true.

Our first step in this process has been to code each outlet reported in 1974 for the 18 PSU's in 77 CPOPS to a 1970 Census tract. This has been a very long and tedious process, since the quality of a majority of the addresses was inadequate, requiring most of the addresses to be looked up in telephone books. After keying the tracts, the number of net outlets, gross outlets, net outlet-categories and gross outlet-categories have been tabulated at the SSU level for certain groups of categories.

Table 1 shows the average percentage of outlets per SSU located inside the same SSU as the interviewed households. Although overall there does not appear to be large differences associated with PSU type, 2/ large differences exist in all PSU types between SSU's -- the more rural the SSU, the fewer the outlets within the SSU.

An examination of the location of the outlets by reference period and category type indicates the more frequently purchased items of groceries, gasoline, and meals are purchased closer to home than clothing and household categories with 3 and 6 months reference periods. In turn even smaller proportions of the categories with one and two year reference periods are purchased within the SSU.

In D areas, urban areas outside SMSA's which are in general untracted, the second stage of selection was the ED instead of SSU. For analysis purposes, groups of ED's have been formed. Table 2 shows percentage distribution of outlets for groups of ED's within the main town of the D PSU's -- Grand Island, NB, Beaufort, SC, Corvallis, OR and Alamorordo, New Mexico. The columns indicate the household location and the rows the outlet location. Although in some cases the percent of outlets in the area of the

households is greater than for other areas, the general distributions between SSU's is very similar. This is not true in A, B or C PSU's where the households tend to shop in the 2 or 3 main shopping areas close to home. However, since the households are not close, the overlap between SSU's is minimal. For example, the number of outlets reported by all SSU's range from 7 to 12 in B PSU's and 7 to 19 in C PSU's with an average of 9 in the B's and 15 in the C's. The number of outlets reported by more than 10 housing units averages 12.5 in B areas, 17.5 in C's, and 28.5 in D's.

III. Analysis of 1977 CPOPS Data

So far only the outlets for a few PSU's have been coded to tract for 1977; however, a few trends can be seen.

In C PSU's, changing SSU's seems to cause significant differences in the distribution of the outlets outside the central city of the SMSA. For example, in the Johnstown SMSA only two of the four SSU's in 74 were in the city of Johnstown. The other two were in Somerset County. As a result about one-third of the outlets were in the city of Johnstown and another third in Somerset County. In 1974 the six SSU's included areas in Northern Cambria County along with the city of Johnstown and Somerset County north of Johnstown, an area where no outlets were reported in 74, 20% in Johnstown and 25% in Somerset County. Although overall only 13% of the 1977 outlets were also reported in 1974, 42% of the 77 outlets in the main downtown shopping area were also in 74.

In D PSU's, there is little difference in the distribution of outlets within the PSU in either method 1 or method 3 PSU's. See Table 3. However, while all the outlets in 74 were reported within the SSU, only about 50% of the outlets reported in 77 are within the SSU. This is true in both method 1 and method 3 PSU's. For example over 22% of the outlets for Beaufort are in Savannah and Charlestown and over 25% of the outlets for Alamogordo are in El Paso, Albuquerque and Las Cruces. Since the same households were interviewed in Alamogordo, this difference has to be explained by changes in procedures rather than sample selection.

At this time we are still refining and coding the outlets for the large cities, so it is impossible to even speculate on any preliminary conclusions. As for more definitive answers to our basic questions we have a long way to go. The investigations mentioned so far are all being made on the universe sampling frame. They will be repeated on samples selected via various methods and with varying sample sizes. At present the outlet sample for the U population is selected using a Keyfitzing procedure to the sample selected for the W population. We will be examining various ways of Keyfitzing the two new U and W samples to the old U and W samples. Some of the questions we will be asking are:

Would Keyfitzing the new CPI outlet samples to the old compensate cost wise for a more disperse universe? And if so, which of the numerous options for Keyfitzing provides the most gains? Should we use a clustered sample in some areas while switching to an unclustered sample in others? If we use an outlet frame developed from CCE, how many years interviewing are necessary to provide sufficient outlets? Then one must consider the death rate of outlets over the time period which probably varies by category and type of business.

1/ A more complete description of the 1974 POPS sample design is given in "Point of Purchase Survey and its Use in the Consumer Price Index Revision", Proceedings of ASA, Social Statistics Section, Atlanta, 1975.

2/ There are four PSU types: A - self-representing, B - SMSA > 400,000, C - SMSA < 400,000 and D - urban areas outside SMSA's.

TABLE 1

Average Percentage Per SSU of Net Outlets Located Within Some SSU as Households Reporting Them - 1974

	A213 Cincinnati	A426 Honolulu	B102 Providence Pawtucket	B213 Indianapolis	B332 Memphis	B436 San Bernardino Riverside	C104 Binghamton	C108 Johnstown	C321 Huntsville	C327 Baton Rouge	A	B	C
#SSU's	12	6	4	4	4	4	4	4	4	4			
Overall	27.1	16.1	25.3	15.5	22.0	19.2	16.1	15.4	36.6	30.3	23.4	20.5	24.6
Min/Max	1.5/75.4	4.6/32.6	13.8/38.0	12.2/19.1	14.3/28.8	1.0/45.6	10.2/24.0	4.3/23.6	2.4/70.2	19.3/48.7			
By recall period													
1 Week	38.5	31.6	34.9	27.4	45.3	31.4	26.2	26.1	47.1	43.6	36.2	34.7	35.7
2 Weeks	39.5	22.0	37.9	28.3	28.9	34.0	24.0	18.7	41.0	45.8	33.7	32.3	32.2
3 Months	20.3	13.8	18.8	8.1	17.8	14.4	14.3	13.8	39.3	21.8	18.2	14.8	22.3
6 Months	22.1	12.6	23.5	6.6	15.6	13.5	9.3	11.8	31.2	19.8	18.9	14.8	18.0
1 Year	11.0	5.8	4.8	7.8	3.5	23.6	6.9	11.3	28.5	10.9	9.3	9.9	14.4
2 Years	12.4	5.4	14.9	11.2	15.3	8.4	7.2	9.6	28.9	22.4	10.1	12.5	17.0
By category groups													
Gasoline	36.1	37.5	33.9	10.6	38.4	40.9	43.2	45.6	57.1	53.1	36.6	30.9	49.8
Groceries	43.5	28.9	42.6	13.3	41.7	33.3	39.7	20.5	40.0	20.0	38.6	32.7	30.1
Meals and alcohol away from home	37.3	20.7	24.7	23.4	20.6	21.6	15.9	21.0	26.6	18.7	31.8	22.6	20.6
Clothing	18.8	6.9	18.4	9.3	14.0	11.0	7.2	8.4	34.7	24.5	14.8	13.2	18.7
Medical	21.7	17.2	21.8	5.2	21.5	6.4	13.5	14.7	33.2	18.3	20.2	13.7	19.9

TABLE 2

Percentage of Outlets in Groups of ED's in D PSU's - 1974

	D211 Grand Island, NB				
	1	2	3	4	5
1	26.8	5.5	4.8	3.9	4.1
2	7.0	10.9	8.2	2.6	5.1
3	15.9	21.1	21.4	14.5	11.2
4	9.6	10.2	12.6	15.8	16.3
5	8.9	14.1	13.1	19.7	21.4
6	24.2	28.9	27.2	30.3	32.6
Subtotal	92.4	90.7	87.3	86.8	90.7

	D319 Beaufort, SC		
	1	2	3
1	48.8	40.1	40.2
2	16.1	22.8	22.7
3	1.6	1.8	7.8
4	21.2	23.4	19.7
Subtotal	87.7	88.1	90.4

	D425 Corvallis, OR		
	1	2	3
1	26.8	19.4	21.1
2	11.3	14.3	8.2
3	57.7	48.8	61.5
Subtotal	95.8	82.5	90.8

D427
Alamogordo, NM

	1	2	3	4	5	6	7
1	4.0	3.8	4.6	3.1	2.2	2.8	1.6
2	2.0	2.3	0	0	0	0	.8
3	1.3	.8	0	1.5	0	0	.8
4	7.4	7.6	3.7	3.1	1.1	8.3	4.1
5	56.4	56.1	63.0	52.3	71.9	52.8	65.6
6	4.7	3.8	4.6	4.6	4.5	16.7	9.0
7	10.1	9.1	10.2	18.5	12.4	2.8	10.6
Subtotal	85.9	83.5	86.1	83.1	92.1	83.4	92.5

TABLE 3

Percentage distribution of outlets within PSU - 1974 and 1977 - for D319 & D427

D 3 1 9		
Location	74	77
1	2.6	1.1
2	10.1	7.0
3	13.6	8.1
4	19.5	14.6
5	1.9	.6
6	17.5	20.3
7	0	.2
8	10.4	7.8
9	0	.4
10	0	1.5
11	3.6	3.5
12	4.5	6.3
13	1.0	.9
14	1.0	1.3
15	14.3	20.3
16	0	3.05
Jasper Co	0	3.05

D 4 2 7		
Location	74	77
1	8.8	8.2
2	3.1	4.9
3	2.8	1.5
4	6.3	9.3
5	.6	.8
6	4.1	5.1
7	10.7	7.0
8	18.6	16.9
9	17.0	19.2
10	2.8	2.5
11	3.8	3.6
12	14.8	14.2
13	2.5	2.1
14	3.8	4.0
15	.3	.6