

COMPARING CERTAIN EFFECTS OF REDESIGN ON DATA  
FROM THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

Elaine C. Hock and Franklin Winters, U.S. Bureau of the Census<sup>1</sup>  
Elaine C. Hock, U.S. Bureau of the Census, Washington, DC 20233

KEY WORDS: CAPI, PAPI

## I. MOTIVATION

The Survey of Income and Program Participation (SIPP) is completing implementation of a major redesign. Starting with the 1996 panel, we will be changing the pattern of how we interview households. Instead of overlapping panels with a new panel being introduced each year, we will be changing to a non-overlapping design with a new panel being introduced every 4 years. The sample size for each panel will also be increased. The 1996 panel will contain about 37,000 household compared to 20,000 for previous panels. The 1996 panel will also include an oversample of the low income population.

The change in design supports the primary objectives of the SIPP which is producing longitudinal estimates of income and program participation, paying most attention to improving the information for people who are economically at risk, and improving the capability to respond to current policy needs in topical areas.

The redesign embraces all aspects of the SIPP program including sample design, questionnaire design, a move to computer assisted personal interviewing (CAPI), and automation. This paper will focus on the effects of the questionnaire changes, the move to CAPI, and the automation enhancements.

As part of the CASIC, computer assisted survey information collection, initiative at the Census Bureau, we will be using a CAPI questionnaire for the 1996 panel. Rather than simply automate the paper documents, we changed some question wording and content, moved some data consistency checks from post collection processing to the CAPI questionnaire, automated survey management and redesigned the data processing system. The areas of concentration were the labor force questions, follow-up on amounts in the assets income questions, and clarification/enhancement of general income questions and health insurance questions (Singh and Huggins, 1994).

The Census Bureau conducted a Dress Rehearsal (DR) during 1995 to prepare for full-scale implementation in 1996. The DR represents our first opportunity to compare data collected from the automated SIPP/CAPI

instrument with data collected on a paper and pencil questionnaire for concurrent time periods.

This paper will identify the incidence and estimate the magnitude of data differences between the 1995 DR data collected from the SIPP/CAPI and data collected by paper and pencil interviewing (PAPI) in the concurrent 1993 panel. Our effort here is only to identify data differences because the transition of SIPP from paper and pencil to CAPI is overshadowed by other important changes in the survey design. Our reference to CAPI throughout the paper will refer to all aspects of the survey redesign. Also, our reference to PAPI will refer to the 1980 SIPP design. Results from this research will provide some information about the redesign's effect on the SIPP data and expectations for 1996. Therefore, the subject of study is the redesign of SIPP.

## II. METHODOLOGY

### A. Sample Selection

The SIPP/CAPI DR sample of 6,600 housing units was selected from 200 Primary Sampling Units (PSUs) and 30 additional counties. The PSUs cover both urban and rural areas. We selected PSUs and counties specifically because they were in both the 1980 and 1990 SIPP sample designs and had the same geographic borders. This selection process should improve the reliability of the comparisons. The sample is not nationally representative. The nationally representative redesign sample has a total of 322 PSUs.

### B. Study Design

The primary objective of the DR was to test all components of the software, systems, and procedures that simulate to the greatest extent possible the full 1996 SIPP/CAPI production environment. Approximately 130 Supervisory Field Representatives (SFRs) and 360 FRs participated in administering the DR. All the SFRs and about one-half of the FRs were already experienced CAPI users.

Training was tailored to three specific groups: experienced SIPP FRs with no CAPI experience, experienced SIPP FRs with CAPI experience, newly hired FRs. The training used generic modules which applied to all CAPI surveys (i.e., laptop use, typing skills, laptop case management system, etc.), with supplemental modules specific to SIPP (Quasney, 1994). The FRs were

---

<sup>1</sup> This paper reports the general results of research undertaken by the Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

also provided training in the area of basic interviewing skills, concepts, and gaining respondent cooperation.

The SFRs did not have separate assignments. They were responsible for nonresponse follow-up and conducting reinterview for data quality assessment. The FRs conducted wave 1 interviews in person at the respondent's home. In wave 2, the FRs conducted personal or telephone interviews.

The first interview of the 1995 DR occurred during February, March, April, or May of 1995. The interviews in the four months create subsamples called rotation groups. Rotation groups are used to distribute interviewing workloads. One round of interviewing for the sample covering all four rotations is called a wave. The 1995 DR consisted of two waves. Therefore, each household had the potential of being interviewed twice, the second interview occurring four months after the initial interview. At each interview the reference period was the four months preceding the interview month.

The 1995 DR collected data for all four rotations of wave 1 and 2, but interviews from rotations 1 and 2 were used for training purposes. Rotations 1 and 2 allowed the field staff to become familiar with the survey, case management, the instrument, procedures, and the equipment. We produced estimates only for rotations 3 and 4 of the 1995 DR.

The seventh and eighth waves of the SIPP 1993 Panel were in the field during the same time period of 1995. The 1993 Panel is the last panel from the 1980 SIPP design. To produce comparable data to the DR, we took a subset of the 1993 panel waves 7 and 8 and called it Bridge Data. To do this, we first pulled off the same PSUs and counties as the DR and second we processed only the two rotations corresponding to the DR (rotations 1 and 4).

The value of having data from the same time period which were collected using the different data collection modes (CAPI vs. PAPI) should be realized in our ability to accurately measure the impact of the redesign on the SIPP data. The 1995 DR (waves 1 and 2) and the 1993 Panel Bridge Data (waves 7 and 8) were both in the field at the same time: April and May of 1995 (wave 1 of the DR and wave 7 of the 1993 panel) and August and September of 1995 (wave 2 of the DR and wave 8 of the 1993 panel).

This paper includes analysis of data from the 1995 DR wave 1 and Bridge Data wave 7.

#### C. Weighting

The usual weighting scheme for SIPP is described in (Waite, 1988a) and (Waite, 1988b). The procedure makes two adjustments to the base weights, where the base weights are the reciprocals of the probabilities of selection. The adjustments attempt to compensate for nonresponse and undercoverage, using variables thought to be highly correlated with SIPP variables of interest.

The first adjustment is a noninterview adjustment and has two different forms: the first is for wave 1 and the second is for wave 2 and subsequent waves. The cells are defined by characteristics of people who were eligible for the appropriate wave sample. The second adjustment is a raking procedure using data from the Current Population Survey as controls. This is done to reduce the MSE of estimates, primarily the variance component.

We developed a simplified weighting scheme to test the impact on estimates from the CAPI instrument and other procedural changes that were implemented during the DR. We applied the same weighting procedure to both samples. We simplified our standard weighting procedure by taking away some criteria for both the wave 1, household noninterview adjustment and the wave 2+, person level noninterview adjustment. In the second adjustment of weighting, we ratio adjusted to population controls at the total level. This adjustment was simplified by 1) simpler tables, 2) in the standard weighting, we do not ratio adjust, we perform an iterative raking adjustment, and 3) ratio adjusting to Population Estimates based on Census counts. To further complicate the weighting, we had to treat wave 7 of the bridge data as if it were wave 1, for comparison purposes. This means that wave 7 went through both the wave 1 and wave 2+ noninterview adjustment.

#### D. Variables

Since the major changes to the questionnaire were in the areas of general and asset income, labor force and health insurance questions, variables from these sections were selected for comparison. Also, it was important to examine the effects of the redesign on the program participation questions. The following variables were selected for comparison.

1. Education Attainment Distributions;
2. School Enrollment Distributions;
3. Persons covered by: any kind of health insurance, Medicare, and Medicaid;
4. Recipients of: Social Security, Supplemental Security income, Aid to Families with Dependent Children (AFDC) and other cash public assistance, food stamps, and Women Infants and Children (WIC).

#### E. Variances

We used VPLX to compute the estimates and variances of the variables above. VPLX is a computer program written by Robert Fay of the Census Bureau, which calculates the estimates and variances for totals, means, and proportions through replication methods. The system shares techniques of several standard methods of variance estimation and combines them together. (For more information on VPLX, see Fay [1990].)

### III. RESULTS

We include here only results from analysis of the 1995 DR wave 1 and Bridge Data wave 7. Wave 8 of the

Bridge Data is still being processed and we will continue our investigation into the effects of the SIPP redesign once the processing is complete. Indications of whether the implementation of the redesign efforts were successful cannot be fully determined at this point in time. The results thus far, however, seem to indicate that the redesign efforts had no overall adverse effects on the SIPP data and point to some potential programming problems in processing the data, recoding, etc.

In Table 1, Wave 1 Nonresponse Rates, the data indicates that the DR sample wave 1 nonresponse rate is higher than that of previous panels. There is some indications that some of the FRs might have viewed the DR as only practice and did not put forward their "best" effort. There is also evidence that some FRs needed additional training on gaining respondent's cooperation, especially the new hires. Another factor which could have affected the response rate for the DR was the less intensive effort to convert refusals. Table 2, Response Rates by Mode of Interview, indicates that the Bridge Data response rate is lower than that of the DR, 83.6% to 87.1% respectively, which was expected since the sample had already been in the field for 2 years. However, we expected a bigger difference between the two response rates than what was actually observed. The DR results indicate the possibility of a higher nonresponse rate for the redesigned SIPP.

The key variables chosen for comparison were delineated in section IID.

Estimates were analyzed using t scores, which were calculated using the following formula:

$$\frac{| \text{CAPI Estimate} - \text{PAPI Estimate} |}{\sqrt{S.E.^2_{\text{CAPI}} + S.E.^2_{\text{PAPI}}}}$$

If the ratio is less than 1.645, the difference is not significant at the .10 significance level. If the ratio is equal to or greater than 1.645, the difference is significant at the .10 significance level. At the overall level displayed in Table 3, there was no systematic tendency for one method to produce higher estimates than the other for

any of the variables examined and there is therefore no reason to suspect that the differences reflect anything other than random effects.

We also looked at age and gender breakdowns for several variables, see Tables 4, 5, 6, 7, and 8. The data in Tables 7 and 8 indicates some possible processing problems with the redesign data (DR) for the higher gender by age breaks. More investigation is necessary. Specifically, in Table 8, we identified a problem with full time workers 75+ that is possibly due to a recode problem. Further investigation is taking place.

#### IV. CONCLUSION

The present research was an attempt to 1) document our knowledge gained from implementing the DR, use the DR to 2) identify processing problems to eliminate in 1996, and 3) improve the CAPI questionnaire for future cycles of the survey.

The transition of SIPP to CAPI for the 1996 Panel has been successful thus far due in large part to the 1995 DR. Research continues in the area of determining the improvement in data quality due to the Survey's major redesign efforts.

#### REFERENCES

- Fay, R. E. (1990), "VPLX: Variance Estimates for Complex Samples," Proceedings of the Survey Research Methods Section, American Statistical Association, pp. 266-271.
- Quasney, A. (1994), "Dress Rehearsal Meeting Documentation." Internal U.S. Bureau of the Census Memorandum from Quasney to SIPP Redesign Distribution List, June 20, 1994.
- Singh, R. J. and Huggins, V. J. (1994), "An Overview of SIPP Redesign and Research Activities."
- Waite, P. J. (1988a), "SIPP 85+: Cross-Sectional Weighting Specifications for Wave 1." Internal U.S. Bureau of the Census Memorandum from Waite to Walsh, July 6, 1988.
- Waite, P. J. (1988b), "SIPP 85+: Cross-Sectional Weighting Specifications for the Second and Subsequent Waves -- Revision." Internal U.S. Bureau of the Census Memorandum from Waite to Walsh, August 8, 1988.

**Table 1. Wave 1 Nonresponse Rate**

Panel	Noninterviews
Dress Rehearsal	12.9%
1993	8.9%
1992	9.2%
1991	8.4%

**Table 2. Response Rates by Mode of Interview**

	CAPI - DR Wave 1	PAPI - 1993 Panel Wave 7
Interviews	87.1%	83.6%
Noninterviews	12.9%	16.4%

**Table 3. Person Totals for the Wave**

	PAPI		CAPI		t score
	Estimate	S.E.	Estimate	S.E.	
Social Security	19879198	5344358	19740736	3689002	0.021
SSI	3408827	1091077	2625425	603318	0.628
AFDC	2803865	1115975	2407489	706623	0.300
WIC	2918112	1260094	2125461	531987	0.580
Food Stamps	5200707	2022748	5510929	1582232	0.121
Medicare	18553648	4943712	18082911	3428424	0.078
Medicaid	17287762	6278014	16019459	4571834	0.163
Any kind of Health Insurance	123622966	36385413	120807713	27557446	0.062
Total Workers	70136882	21364390	70905546	17655587	0.028
Full Time Workers	55344260	17315152	46603853	11646001	0.419

**Table 4. Social Security Recipients**

	Age	PAPI		CAPI		t score
		Estimate	S.E.	Estimate	S.E.	
Males	<45	398419	121680	509694	69490	0.794
	45-54	327753	53538	267004	34045	0.957
	55-59	204978	64673	125068	26735	1.142
	60-64	895027	233998	861371	195852	0.110
	65-69	2129299	630847	2039264	316392	0.128
	70-74	1706802	392127	1793656	551184	0.128
	75+	2387549	655861	2343784	244941	0.063
Females	<45	507376	217150	871243	216996	1.185
	45-54	216345	47385	280664	45793	0.976
	55-59	280768	88423	313410	50676	0.320
	60-64	1329516	432729	1106924	177933	0.476
	65-69	2743027	684001	2542700	591537	0.222
	70-74	2343966	466266	2355201	447554	0.017
	75+	4408365	1282059	4330748	771130	0.052

**Table 5. Medicare Recipients**

	Age	PAPI		CAPI		t score
		Estimate	S.E.	Estimate	S.E.	
Males	<45	291760	133519	366799	113946	0.427
	45-54	262642	43262	240965	19500	0.457
	55-59	168450	31621	81628	13053	2.538
	60-64	282215	77981	262522	71886	0.186
	65-69	2353561	711349	2183278	345495	0.215
	70-74	1783700	406975	1875974	562749	0.133
	75+	2508221	725548	2419372	248887	0.116
Females	<45	262628	111155	254697	44024	0.066
	45-54	131248	20122	123562	17764	0.286
	55-59	183561	38405	219368	25849	0.773
	60-64	282618	124320	225212	35535	0.444
	65-69	2940553	729192	2763391	639266	0.183
	70-74	2505675	498829	2490370	493357	0.022
	75+	4596809	1316587	4575765	845693	0.013

**Table 6. Medicaid Recipients**

		PAPI		CAPI		t score
		Estimate	S.E.	Estimate	S.E.	
Males	<15	4144041	1465244	3751438	1179818	0.209
	15-24	884932	378650	900693	269431	0.034
	25-34	465566	209286	565830	176481	0.366
	35-44	673665	315540	370427	106342	0.911
	45-54	215903	67096	289197	46117	0.900
	55-64	313740	86276	270923	99715	0.325
	65-74	217864	69694	274510	81172	0.529
	75+	121844	63076	174372	12938	0.816
Females	<15	3680082	1402107	3294634	910464	0.231
	15-24	1488252	570755	1659798	537429	0.219
	25-34	1739021	685011	1642077	583412	0.108
	35-44	1151556	456110	784173	170895	0.754
	45-54	447242	97914	423744	65318	0.200
	55-64	647908	194177	496375	111648	0.677
	65-74	580288	98398	595784	135639	0.092
	75+	515851	132780	525475	105403	0.057

**Table 7. SSI Recipients**

	PAPI		CAPI		t score
	Estimate	S.E.	Estimate	S.E.	
<45	1307617	546155	1011452	287727	0.480
45-54	258303	52356	357343	33549	1.593
55-59	356746	104218	206798	78131	1.151
60-64	239731	83465	172370	37380	0.737
65-69	290745	71336	252843	64895	0.393
70-74	360956	56110	191758	40454	2.446
75+	594726	186034	432857	69630	0.815

**Table 8. Total Workers**

	Age	PAPI		CAPI		t score
		Estimate	S.E.	Estimate	S.E.	
Males	15-24	5873419	1912474	6002321	2101418	0.045
	25-34	11418770	4645338	11029206	3828659	0.065
	35-44	9949349	2728725	9780941	2120630	0.049
	45-54	6208841	1372802	6045007	686460	0.107
	55-64	3614439	886318	3739460	774756	0.106
	65-74	977575	276837	955130	235485	0.062
	75+	138919	42208	185789	13283	1.059
Females	15-24	5498571	1955730	5668740	1752872	0.065
	25-34	8954113	2978317	9176332	3019163	0.052
	35-44	8570502	2332288	8545975	1666391	0.009
	45-54	5457745	1350909	5574906	686252	0.077
	55-64	2713105	690921	3225752	608146	0.557
	65-74	655684	187630	819058	189587	0.612
	75+	94676	20265	156923	14477	2.499

**Table 9. Full-Time Workers**

	Age	PAPI		CAPI		t score
		Estimate	S.E.	Estimate	S.E.	
Males	15-24	3552823	1324187	2859770	1097470	0.403
	25-34	10541310	4307789	8544422	3061896	0.378
	35-44	9491692	2591269	8011515	1641849	0.483
	45-54	5833131	1315762	4824643	531781	0.711
	55-64	3067722	751874	2831122	579345	0.249
	65-74	452569	127608	386950	72999	0.446
	75+	23973	5003	41687	2743	3.104
Females	15-24	2694429	1099560	2110155	747842	0.405
	25-34	7008068	2388778	5665601	1962363	0.400
	35-44	6297460	1802581	5329999	1099333	0.422
	45-54	4216983	1054271	3757147	438550	0.371
	55-64	1965654	512424	1883388	364757	0.121
	65-74	195205	46664	318956	76495	1.273
	75+	3234	355	38492	3688	8.774

**Table 10. Educational Attainment (age 25+)**

	PAPI		CAPI		t score
	Estimate	S.E.	Estimate	S.E.	
0 to 8 yrs complete	7177100	1794544	7188620	1683641	0.005
9-12 yrs complete	39149031	11807352	34971028	7684338	0.297
1+ yrs college	41471672	12081256	47473452	10810637	0.370

**Table 11. School Enrollment (ages 15-24)**

	PAPI		CAPI		t score
	Estimate	S.E.	Estimate	S.E.	
full-time	11240496	3594900	10710648	2835827	0.116
part-time	1117293	372309	1450510	526524	0.517
not enrolled	8210686	3225021	8407319	3038577	0.044