George H. Gray, U.S. Bureau of the Census Marvin M. Thompson, U.S. Bureau of the Census

INTRODUCTION

The Office of Education of the Department of Health, Education, and Welfare (HEW) has for a number of years distributed funds authorized by Title I of the Elementary and Secondary Education Act of 1965, utilizing a formula that includes the estimate of the number of children 5 to 17 years of age in poverty families in each State. Since 1972, the estimate used has been the number of poor children in 1969, according to the 1970 Census of Population and Housing. As we move further in time from the census, the interstate relationships for children in poverty are likely to be changed because of changes in population growth, family formation and dissolution, and economic activity. Since 1970, national estimates of children in poverty have been available from the Current Population Survey (CPS). However, CPS estimates were not sufficiently reliable on a State basis to substitute for the census figures.

Accordingly, Congress in enacting the Educational Amendments of 1974 (Public Law 93-380) provided in section 822(a) that, "The Secretary of Commerce shall, in consultation with the Secretary of Health, Education, and Welfare, expand the current population survey (or make such other survey) in order to furnish current data for each State with respect to the total number of schoolage children in each State to be counted for purposes of section 103(c)(1)(A) of title I of the Elementary and Secondary Act of 1965." Pursuant to this legislative requirement, the Bureau of the Census in cooperation with agencies of HEW, mounted the Survey of Income and Education (SIE) and carried it out between April and July 1976 at a sample of approximately 190,000 designated addresses.

The SIE was also designed to satisfy the requirements of section 731(c)(1) of the Bilingual Education Act, Title VII, ESEA as amended by Public Law 93-380, which authorizes the Commissioner of Education to estimate from a survey the number of children and other persons in the States who, because of limited English-speaking ability, are in need of bilingual education, guidance, and counseling.

Finally, at HEW's request, the opportunity presented by such a large survey was used to gather some additional income-related information such as receipt of food stamps, housing costs for homeowners and renters, and estimated cash assets. Also, information relevant to a number of HEW programs was collected, including data on education, disability, health insurance coverage, and institutionalized persons.

SURVEY DESIGN

The primary objective of the Survey of Income and Education was to determine for each State the

number of children 5-17 years of age in poverty. In discussions with HEW and the Congressional staffs involved, it was agreed that the criterion to be used for providing equity among the States was an estimated coefficient of variation (C.V.) of 10 percent for the count of poverty children. A preliminary sample design was created to yield this reliability for that statistic. Since we were also interested in obtaining reliable estimates of persons with limited English-speaking ability, additional cases had to be added. While we were able to achieve an estimated coefficient of variation of 10 percent or better on persons with limited English-speaking ability for most States, the estimated C.V. for 12 States was above this level and ranged up to an estimated C.V. of 20 percent.

SAMPLE DESIGN

The sample was designed to be State representative and was to be completely independent of other Census samples, such as the CPS. The sample for SIE was a stratified multistage noncompact cluster design. For the first stage of selection, each State was divided into areas called Primary Sampling Units (PSU's). These areas were either a Standard Metropolitan Statistical Area (SMSA) or a group of geographicallyneighboring counties or independent cities. The PSU's were then grouped in strata based on estimates of like characteristics derived from the 1970 Census. The primary determination of strata classification was the proportion of children 5-17 years of age living in poverty, based on 1970 Census data. PSU's with large populations in relation to the sampling rate for the State formed strata by themselves and came into sample with certainty.

In eight States (Connecticut, Delaware, Hawaii, Maryland, Massachusetts, New Hampshire, Rhode Island and Vermont) and the District of Columbia, every PSU was selected for sample with certainty. In the remaining States, two PSU's were selected from each strata that were not large enough to be in sample with certainty.

Within each PSU, the majority of the sample of housing units and group quarters were selected from the list of units in the 20-percent sample of the 1970 Census. The 20-percent sample file was used because it provides the information on income and poverty which determined the stratification of the sample.

In order to represent persons living in units completed since the 1970 Census, a sample was selected from the building permits issued since 1970 in those areas under the jurisdiction of building permit offices. This represents the majority of this type of unit. For the remaining areas (those without a building permit office), a sample of units built since 1970 was obtained by

selecting such units in the area segments from recently-retired CPS samples.

Finally, the SIE sample included units selected from (1) a list of special places, such as rooming and boarding houses, communes, flop houses, military installations (excluding military barracks), agricultural workers' dormitories, etc., and (2) a list of mobile homes in mobile home parks established since the 1970 Census.

QUESTIONNAIRE CONTENT

Public Law 93-380 amends section 103 of the Secondary and Elementary Education Act of 1976 to read, "... in determining the families which are below the poverty level, the Commissioner (of Education) shall utilize the criteria of poverty used by the Bureau of the Census in compiling the 1970 decennial census." In the years since 1970, the same definition has been used in the Current Population Survey's March Income Supplement to determine poverty status though it is updated annually to reflect changes in the Consumer Price Index. As previously noted, section 822(a) of Public Law 93-380 specifically mentions expansion of the current population survey as an acceptable method of determining the number of poor children. In addition, the existence of a processing system based on CPS made it possible to meet the stringent deadlines imposed by the Congressional mandate. Finally, very serious consideration was being given to combining the SIE and CPS to provide a larger sample for estimates of the count of poor children. For these reasons, it was decided that SIE would replicate exactly the March CPS questionnaire content though it would be expanded to cover additional subject matter. Therefore, the core questions on current labor force status, last year's work experience and money income, together with such demographic variables as age, sex, marital status, family membership, household membership, veteran status, educational attainment and ethnic origin, are asked and recorded in the same manner on both questionnaires.

The items on foreign birth, language or languages spoken in the household and language spoken in the home when the sample person was a child are screening questions to determine if the questions on English Language Proficiency should be asked. These last questions (what language the sample person speaks, how well the person speaks and understands English, what language does he usually speak to friends, and what language does he usually speak to his children, or in the case of children speaks to his brothers or sisters) are used as a Measure of English Language Proficiency (MELP). This series was developed by the Center for Applied Linguistics under a contract with the National Center for Education Statistics.

For the foreign born, there are questions to determine when they came to the United States to stay and where they were born. All sample persons are asked how long they have lived in the State and, for movers, what State they lived in before moving to the State of residence at the time of interview. These questions will be used to

develop measures of immigration to the United States and measures of internal migration.

Additional items asked for all persons, though screened on appropriate age groups, include questions on school enrollment, disabilities that limit the person's ability to attend school, limit or keep the person from working at a job, or limit the amount or kind of housework they can do. For those with a limitation, it is determined how severe that limitation is, the cause of the limitation and who diagnosed it. Finally, for each person, questions are asked concerning their coverage by health insurance plans or other programs that provide health benefits or services and whether they received any of these benefits or services in the past year.

For the household as a whole, information was collected on food stamp recipiency in 1975 and 1976, cash assets, mortgage or rent payments and if a rental unit, whether or not it was subsidized. While the data from these questions and those on education, disability and health insurance will be used to meet the needs of various programs sponsored by HEW, they are more specifically to be used in the estimation of costs and caseloads under various alternative assumptions about eligibility for programs such as food stamps and AFDC. In addition, they will be used to analyze the impact of the inclusion of such in-kind costs and assets on alternative definitions of poverty.

Finally, there are a set of questions designed to determine household membership during the reference year (1975). These questions will be used in research concerning the effect of changing household membership on the income and size of the family and hence on their poverty status.

DATA COLLECTION

Interviewing was begun in late April 1976 and extended through July of that year. Approximately 95 percent of the workload was completed during the months of May and June.

The 191,459 assigned households were located in approximately 1,800 counties and independent cities. To complete this task required 2,500 interviewers, of whom 1,600 or 63 percent were new to Census operations. About one-fifth of the interviewers had worked on CPS (including March 1976) and the remainder were working on other Census surveys at the time they were assigned to SIE. In addition, about 200 persons were hired as crew leaders whose primary function was as reinterview specialists though they performed other tasks such as aiding new interviewers begin their work and observing and helping those who needed additional training. The crew leaders also assisted in reducing the number of refusals and other noninterviews, especially in areas with high noninterview rates. The data collection effort was coordinated through the Bureau's 12 regional offices, where the regular staff was supplemented by supervisory and clerical help to perform the extensive reviewing of the questionnaires required.

Interviewers and office clerks completed a 4-hour home-study which introduced them to the survey and

the forms to be used. They were then given 3 days of classroom training on the concepts to be applied and procedures to be used in interviewing. During their training they were led through several practice interviews to familiarize them with the content and skip patterns on the questionnaire. Following this, they completed a 6-hour post-classroom home-study which gave additional training and tested them on the training already received. All newly-hired interviewers were given 2 days of on-the-job training during which they were accompanied by more-experienced personnel who demonstrated interviewing techniques and observed them perform several interviews before leaving them on their own.

Interviewing for SIE was conducted by personal visit to the assigned address. Any responsible adult, that is someone who was knowledgeable about the work patterns and income of the family, could act as the respondent for the entire household. While technically, anyone over 14 years of age could be a household respondent, in practice, teenagers were accepted only as a last resort. In most cases, the respondents were the head of the household or the head's spouse. In any case, the interviewers were encouraged to make extensive callbacks either by phone or personal visit to obtain more precise information when not available from the household respondent. While the average time required to complete an interview was about 45 minutes, some households took much longer, especially when callbacks were required.

QUALITY CONTROL MEASURES

Throughout the period of interviewing, the questionnaires received from the interviewers were closely monitored to determine the number and type of noninterviews each was reporting. If the number seemed excessive, a crew leader or supervisor contacted the interviewer to explore the problem and help reduce the noninterviews. During the latter part of June and July, weekly reports were made by the Regional Offices, setting forth the noninterviews by type for each State in their regional area. A target of 5 percent for noninterviews at occupied households and 20 percent for all types was set for each State. During July, crews of experienced interviewers and supervisors visited those States with rates above the targets and attempted to reduce them. While their efforts met with considerable success in most places, a few States remained above the target noninterview rates when field work was closed out.

Nationwide, the noninterview rate for occupied households was 4.6 percent, which is identical to the like rate for CPS in April, May and June of 1976. The noninterview rate for all types of assigned addresses (occupied, vacant, demolished, condemned, etc.) was 21.0 percent, which compares to 20.3 percent for CPS.

While the noninterview rate for occupied households exceeded 5 percent in 15 States, this rate exceeded 6 percent in only 5. The highest rates were posted in the District of Columbia--13.5 percent, Alaska--8.1 percent, and Nevada--7.5 percent. On the other hand, 11 States recorded noninterview rates for occupied households below

3 percent. The lowest was 1.7 percent for Arkansas.

The other major control on the quality of interviewing was an extensive and detailed review of the questionnaires as they were turned in by the interviewers. The first of these was at the Regional Office level at which time the first 25 of the questionnaires returned by the interviewer were reviewed in their entirety. If certain critical items were mishandled or left blank, the Regional Office contacted the interviewer to correct the error or directly called the respondent for missing information. As the family income is an important determinant of poverty status, the Regional Office continued to review the income items on all remaining questionnaires beyond the first 25 from each interviewer and continued to try to obtain any information missing in that area.

When the questionnaires arrived in the Census Bureau's Processing Center in Jeffersonville, Indiana, the review of the questionnaires was repeated; this time on every item on every questionnaire. The Regional Offices were notified of any systematic errors. While it was now too late to call upon the respondent for missing information, the Regional Office could contact the interviewer and correct the problem for the remaining interviews. The single greatest gain from the Processing Center review was the correct marking of the machine-readable data and numbers on the questionnaire. For example, in the income area both a write-in space and machine-readable numbers are used to record each response. The most common interviewer error was failure to mark the machine-readable numbers. On items for which records were kept, this type of error was reduced from 1 percent of the entries to .3 percent for any one item. In light of the large number of newly-hired interviewers, this compares favorably with the .2 percent blank rate per item for the Current Population Survey.

In addition to the close supervision of interviewers and extensive review of the questionnaires, two other procedures were used to control the quality of the interviewing. The first was a telephone recheck of the interviewer's work. The rechecker verified with the interviewed household the list of household members and then re-asked five items that pertained to the household as a whole and five items that had been asked for each household member. The recheck responses were compared to the original and differences were reconciled. Any differences attributable to the original interviewer were discussed with him and remedial training provided where necessary. The first three interviewed households returned by the interviewer were rechecked and thereafter, one interview was rechecked every 2 weeks the interviewer continued working. On the average, seven interviews were checked out of the total workload of approximately 80 assigned addresses per interviewer.

The second procedure used as a quality control was the reinterview of a 5-percent systematic sample of the households assigned. The reinterview was conducted by a staff of interviewers who were more thoroughly trained than the average SIE interviewer and which had a higher proportion of interviewers from CPS and other Bureau programs. While the questionnaire used by the reinterviewer differed markedly from that used in the original interview and there was a time lag between the two visits to the address necessitated by the sampling procedure, nevertheless, the reinterview did uncover some gross errors on the part of the interviewers and these were fed back through the Regional Office staff. The reinterview, together with the check for units missed in the Census is, of course, far more important as part of the overall evaluation of the quality of the data than as an interviewer control.

PROCESSING THE DATA

After reviewing the questionnaires for errors and correcting those for which the information was available, the Processing Center personnel entered codes for all industry and occupation responses, grouped all members of the household into families according to their relationship to the head of the household and entered codes where necessary, and where appropriate, coded State of previous residence for movers.

All clerical reviewing and coding was rechecked on a 100-percent basis to assure an acceptable level of quality. The clerical review, coding and verification operations took place during June, July and August of 1976 and required the services of approximately 40 persons working full time. Approximately 160 mandays of overtime were also required to meet the deadline.

The SIE questionnaires were then microfilmed and the data transferred to computer tape by means of the FOSDIC process. FOSDIC (Film Optical Sensing Device for Input to Computers) is a programmable machine that scans the developed film to ascertain the presence or absence of a mark in the coded dots or numeric figures on the questionnaire and transfers this information to a magnetic computer tape.

This computer tape is then run through the computer and processed by a Data Acceptance Program. The Data Acceptance Program checks the filming operation to assure that all required pages have been filmed and that index marks used in the FOSDIC program have been properly recorded. It also verifies that certain critical data have been correctly entered, such as the Household Identification Number, the Interview/Noninterview status of the household and if noninterview, the type of noninterview. If a questionnaire fails one of these or any of the other checks in the Data Acceptance Program, it is rejected. The error then has to be corrected, the questionnaire refilmed and recycled through the Data Acceptance Program. Most questionnaires are accepted on the first pass. However, approximately 11,700 or 5.5 percent of the SIE questionnaires were recycled, a few for more than one time. After all questionnaires had been through the Data Acceptance Program at least once, some 75 were dropped from the file as they were rejected again and time had run out in late October.

The accepted records were then passed through a series of programs to edit the labor force, work experience in 1975 and income questions. Those programs were the same as used in producing the March 1976 CPS file. They not only edit the data but create a number of recodes used in tabulations, and impute missing data, including income. These programs were used to produce, as closely as possible, data that would fulfill the Congressional requirement to use the same poverty definition as was used for the 1970 Census.

The remaining data on the SIE questionnaire were subjected to a consistency edit that made only those changes that could be inferred from the data themselves and did not impute for any missing information.

Each stage of the editing and imputation programs provided for printouts of actual data or counts of changes so that the operations could be reviewed.

After the editing and imputation had been reviewed and the file accepted, it was then weighted to represent the population as a whole. Initially, each record was assigned a base weight that was the reciprocal of the probability of sample selection. Next, factors were applied to adjust for occupied households that were not interviewed. Adjustments were then made to account for differences between the sample areas chosen and the strata from which they came. The resultant weights were summed and compared to independent estimates of the national population in 116 agesex-color categories. Factors derived from this comparison were then applied to the individual weights. Finally, the weights were again summed and factors applied for three age groups (5-17 years old, 65 years old or older, and all other ages) for each State and the District of Columbia. To bring these last two groups of estimates into closer agreement, the adjustments were iterated a total of three times.

During the weighting process, factors for the national age-sex-color controls are calculated and at that time ratios of the coverage of the population in those various groups are produced. These revealed that SIE had a coverage ratio of 93 percent as opposed to 96 percent for CPS. This is in addition to the undercoverage experienced by the 1970 Census, as the independent estimates of population used as controls for both surveys are derived by updating the census counts by taking into account births, deaths and migration since that time. SIE had coverage of the population equal to or better than CPS for Blacks and other races. It was appreciably lower for Whites, both males and females in almost every age category.

TABULATIONS AND TAPE FILES

Counts of the children 5-17 years of age in poverty for each State were produced in December 1976 and after a review and analysis of these data, a preliminary report was forwarded to Congress on February 18, 1977. A final report incorporating the results of the evaluation is expected to be sent forward in October 1977.

Tabulations have been produced and forwarded to various groups at HEW, Department of Labor, and Census, covering food stamp recipiency, public assistance, child care and labor force status of mothers, characteristics of families and unrelated individuals, income, characteristics of persons with language difficulties, school enrollment, the educationally handicapped, health insurance, work experience in 1975 and labor force status for a number of geographical areas.

Computer tapes have been provided to HEW, the Congressional Budget Office, and the Civil Rights Commission Age Discrimination Study to aid in analyzing the impact of various alternative changes to the welfare system.

A tape is being prepared for general public use that will carry all the information collected by the SIE. All 50 States and the District of Columbia will be identified on the tape. In addition, 122 SMSA's will be identified and within the limits of the Bureau's confidentiality restrictions, the central city of the SMSA, the remainder of the SMSA and the nonmetropolitan areas of the State. The tape will contain individual records for 336,405 persons 14 years old or older, including 2,769 members of the Armed Forces and records for 104,410 children 0-13 years of age. There are summary records for 160,973 families or unrelated individuals as well as 151,170 records for the interviewed households. The tapes and information concerning them can be obtained from:

Customer Services Branch Data User Services Division Bureau of the Census Washington, D.C. 20233

	1. Total	2. Interviewed	ewed 3. Occupied Type A Nonint. Type B Nonint.					Type C N	onint.	Type A+	B+C NI's
	Hhlds	Hhlds	Hhlds	4. Number	5. Rate	6. Number	7. Rate	8. Number	9. Rate	10. Number	11. Rate
			(2+4)		(4:3)		(6:1)		(8:1)	(4+6+8)	(10:1)
UNITED STATES	191,459	151,170	158,475	7,305	4.6	24,600	12.8	8,384	4.4	40,289	21.0
NEW ENGLAND:	26,970	20,754	21,604	850	3.9	4,501	16.7	865	3.2	6,216	23.0
Maine	3,123	2,189	2,240	51	2.3	734	23.5	149	4.8	934	29.9
New Hampshire	5,884	4,261	4,434	173	3.9	1,265 822	21.5	185	3.1	1,623	27.6
Vermont Massachusetts	3,752 4,614	2,723 3,664	2,796 3,879	73 215	2.6 5.5	616	21.9 13.4	134 119	3.6 2.6	1,029	27.4 20.6
Rhode Island	4,193	3,386	3,509	123	3.5	546	13.0	138	3.3	950 807	19.2
Connecticut	5,404	4,531	4,746	215	4.5	518	9.6	140	2.6	873	16.2
MIDDLE ATLANTIC:	16,506	13,459	14,323	864	6.0	1,662	10.1	521	3.2	3,047	18.5
New York	5,276	4,211	4,521	310	6.9	585	11.1	170	3.2	1,065	20.2
New Jersey	5,684	4,694	5,007	313	6.3	518	9.1	159	2.8	990	17.4
Pennsylvania	5,546	4,554	4,795	241	5.0	5 59	10.1	192	3.5	992	17.9
EAST NORTH CENTRAL:	25,797	20,933	21,905	972	4.4	2,913	11.3	979 184	3.8	4,864	18.9
Ohio Indiana	5,508 4,820	4,501 3,965	4,766 4,083	265 118	5.6 2.9	558 550	10.1 11.4	187	3•3 3•9	1,007 855	18.3
Illinois	5,480	4,499	4,776	277	5.8	474	8.6	230	4.2	981	17.7 17.9
Michigan	5,744	4,450	4,669	219	4.7	810	14.1	265	4.6	1,294	22.5
Wisconsin	4,245	3,518	3,611	93	2.6	521	12.3	113	2.7	727	17.1
WEST NORTH CENTRAL:	25,592	20,448	21,230	782	3.7	3,198	12.5	1,164	4.5	5,144	20.1
Minnesota	4,238	3,485	3,579	94	2.6	496	11.7	163	3.8	753	17.8
Iowa	4,694	3,879	4,000	121	3.0	479	10.2	215	4.6	815	17.4
Missouri	3,088	2,343	2,463	120	4.9	450	14.6	175	5.7	745	24.1
North Dakota	3,644	2,922	3,007 1,846	85 81	2.8	493	13.5	144	4.0	722	19.8
South Dakota Nebraska	2,365 3,624	1,765 2,932	3,075	143	4.4	371 427	15.7 11.8	148 122	6.3 3.4	600 692	25.4 19.1
Kansas	3,939	3,122	3,260	138	4.2	482	12.2	197	5.0	817	20.7
SOUTH ATLANTIC:	22,052	17,098	18,031	933	5.2	3,042	13.8	979	4.4	4,954	22.5
Delaware	3,001	2,310	2,455	145	5.9	444	14.8	102	3.4	691	23.0
Maryland	3,262	2,714	2,869	155	5.4	326	10.0	67	2.1	548	16.8
Dist. of Columbia		1,578	1,824	246	13.5	249	11.5	99	4.6	594	27.3
Virginia	2,478	2,036	2,122	86	4.1	238	9.6	118	4.8	442	17.8
West Virginia	2,073	1,671	1,709	38	2.2	234	11.3	130	6.3	402	19.4
North Carolina South Carolina	1,997 1,895	1,555 1,380	1,613 1,441	58 61	3.6 4.2	310	15.5	74	3.7	442	22.1
Georgia	1,937	1,534	1,582	48	3.0	323 242	17.0 12.5	131 113	6.9 5.8	515 403	27.2 20.8
Florida	3,237	2,320	2,416	96	4.0	676	20.9	145	4.5	917	28.3
EAST SOUTH CENTRAL:	8,057	6,361	6,552	191	2.9	982	12.2	523	6.5	1,696	21.1
Kentuck y	1,970	1,517	1,587	70	4.4	275	14.0	108	5.5	453	23.0
Tennessee	2,185	1,736	1,791	55	3.1	253	11.6	141	6.5	449	20.5
Alabama	2,055	1,653	1,686	33	2.0	231	11.2	138	6.7	402	19.6
Mississippi	1,847	1,455	1,488	33	2.2	223	12.1	136	7.4	392	21.2
WEST SOUTH CENTRAL:	11,531	9,158	9,511	353	3.7	1,357	11.8	663	5•7	2,373	20.6
Arkansas	1,925	1,505	1,531	26	1.7	259	13.5	135	7.0	420	21.8
Louisiana	2,065	1,659	1,735	76	4.4	196	9.5	134	6.5	406	19.7
Oklahoma Texas	2,429	1,896 4,098	1,989 4,256	93 158	4.7	287	11.8	153	6.3	533	21.9
MOUNTAIN:	33,755	26,383	27,773	1,390	3.7 5.0	615 4,447	12.0 13.2	241	4.7 4.5	1,014	19.8 21.8
Montana	3,963	3,034	3,190	156	4.9	538	13.6	1,535 235	5.9	7,372 929	23.4
Idaho	5,879	4,568	4,773	205	4.3	843	14.3	263	4.5	1,311	22.3
Wyoming	4,536	3,569	3,741	172	4.6	565	12.5	230	5.1	967	21.3
Colorado	3,782	3,014	3,174	160	5.0	478	12.6	130	3.4	768	20.3
New Mexico	2,589	2,077	2,164	87	4.0	307	11.9	118	4.6	512	19.8
Arizona	2,705	2,042	2,160	118	5.5	447	16.5	98	3.6	663	24.5
Utah Nevada	5,110	4,136 3,943	4,309 4,262	173	4.0	616	12.1	185	3.6	974	19.1
PACIFIC:	5,191 21,199	3,943 16,576	17,546	319 970	7.5	653 2,498	12.6 11.8	276	5.3 5.4	1,248	24.0 21.8
Washington	4,406	3,567	3,743	176	5.5 4.7	487	11.0	1,155 176	4.0	4,623 839	19.0
Oregon	4,841	3,944	4,141	197	4.8	486	10.0	214	4.4	897	18.5
California	5,067	4,202	4,432	230	5.2	465	9.2	170	3.4	865	17.1
Alaska	3,677	2,360	2,568	208	8.1	668	18.2	441	12.0	1,317	35.8
Hawaii	3,208	2,503	2,662	159	6.0	392	12.2	154	4.8	705	22.0