Appendix D. Data Quality

Two principal determinants of the quality of data collected in household surveys are the magnitude of the imputed responses and the accuracy of the responses that are provided. This appendix provides information on the imputation rates for selected education items in the Survey of Income and Program Participation and covers some of the problems encountered in collecting data on education from the respondents in the survey.

Imputed responses refer either to missing responses for specific questions or "items" in the questionnaire or to responses that were rejected in the editing procedure because of improbable or inconsistent responses. An example of the latter is when a person with 6 years of regular school completed has also said they have obtained a Ph.D degree.

The estimates shown in this report are produced after all items have been edited and imputed whenever necessary. Missing or inconsistent responses to specific questions are assigned a value in the imputation phase of the data processing operation. The procedure used to assign or impute responses for missing or inconsistent data in SIPP is commonly referred to as the "hot deck" imputation method. The process assigns item values reported in the survey by respondents to nonrespondents. The respondent from whom the value is taken is called the "donor." Values from donors are assigned by controlling for demographic and labor force data avilable for both donors and nonrespondents.

Imputation rates for some of the major items discussed in this report are shown in table D-1. The imputation rates are calculated by dividing the number of missing responses by the number of persons who should have legitimately responded to the item. The rate of 16.5 percent for high school courses is based on the imputation of any of the seven different kinds of courses; in this context, the rate may be no worse than the values of around 7 percent for most of the other items. Over 90 percent of the sample had no or only one item imputed in this section of the questionnaire.

Another means of determining data quality is by comparison of the weighted survey estimates to other data, either from elsewhere in the questionnaire, a different survey, or known administrative estimates. Comparison of the educational attainment data to data from several other sources indicates that the estimates of highest degree attained are reasonable, given the limitations of the comparative data. Detailed information concerning high school courses and programs taken by the population while in school is not available. The relative proportions of persons taking specific courses by type of track, however, are internally consistent (that is, persons in vocational tracks were more likely to have reported taking 2 or more years of vocational courses than were persons in academic tracks).

Table D-1. Imputation Rates for Selected Education and Training History Items

Item	Rate
High school program	7.7
High school courses (any of seven)	16.5
Highest degree obtained	7.4
Field of highest degree	6.9
Participation in Federally sponsored work training program	7.4

In the 1970 census, about 34 million persons reported that they had ever completed a "vocational training program"; SIPP estimates 32 million positive responses to the question: "Has...ever received training designed to help people find a job, improve job skills, or learn a new job?" Since the SIPP question is much more global in nature, the SIPP estimate may not truly reflect the actual number of persons who could legitimately respond positively to the item.

Estimates of participation in specific job programs vary considerably from available administrative estimates (table D-2). While the estimate for WIN participants was about 25 percent higher than the program estimate for this time period, the estimates of persons receiving training from Job Corps or CETA/JTPA are less than one-half the numbers reported by these programs. Some of the discrepancies may be due to administrative reporting problems and less than exact comparability of reporting periods. It is unlikely, however, that such large shortfalls are due solely to inadequacies of the administrative data.

Table D-2. SIPP and Administrative-Based Estimates of Training Program Participants

Program	SIPP estimates	Program estimates
JTPA/CETA	671,000	1,450,000
WIN	163,000	122,000
Job Corps	109,000	280,000
Trade Adjustment	20,000	20,463