Appendix D. Data Quality

Two principal indicators of the quality of data collected in household surveys are the magnitude of imputed and modified responses, and the accuracy of the responses that are provided. This appendix provides information on the imputation rates for selected items in the Education and Training History module from the second wave of the 1987 panel of the Survey of Income and Program Participation, and covers some of the problems encountered in collecting these data.

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 they were improbable or inconsistent. An example of the latter is when a person with only 6 years of regular schooling completed also said they had a Ph.D.

The estimates shown in this report are produced after all items have been edited and imputed wherever 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 is referred to as the "hot deck" imputation method. This process assigns item values reported in the survey by respondents to those who to not respond. 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 available for both donors and nonrespondents.

Imputation rates for some of the major items 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 respondend to the item. Some items are imputed because a respondent did not respond to the entire module (or wave interview). About 4 percent of those persons eligible for the Education and Training History module did not respond to any question in the module (About half of these were nonrespondents for the entire interview.) As table D-1 shows, about 5 percent of the highest degree responses, and 7 percent of the field of degree answers, were imputed.

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

Item	Rate
Highest degree	5
Field of highest degree	7
Received job training	1
Specific training program	1-3
Uses training on current job	2

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 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. Degree awards for recent years, both in terms of numbers and the fields they are received in, compares favorably to administrative estimates.

While about 38 million persons reported that they had received work training of some kind at some time, there is no way to determine the quality of this estimate. In fact, given the general nature of the question, the SIPP estimate may not truly reflect the number of persons who could actually respond positively to this item. Comparison of SIPP estimates to available administrative figures for specific job training programs also indicate large differences between the two. At least part of the discrepancy may be due to problems in the administrative estimates. However, it is also likely that many individuals who participate in government programs do not know the "official" name of the program, and are not able to give the "correct" answer when asked.