

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 child care items in the Survey of Income and Program Participation and covers some of the problems encountered in collecting data on child care expenses 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 14 year old child is said to be cared for in a nursery school during the time his parent is at work.

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 most responses for missing or inconsistent data for SIPP is commonly referred to as the "hot deck" imputation method. This 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 available for both donors and nonrespondents. The control variables used for child care items generally included the age of the child for whom there was missing data, the parent's marital status and whether the parent was employed part time or full time.

Imputation rates for both primary and secondary child care arrangements (items 1a and 1e in the questionnaire shown in Appendix E) for the respondents' three youngest children are shown in table D-1. The imputation rates are calculated by dividing the number of missing or inconsistent responses by the total number of responses that should have been provided based on the number of children in the household who required child care responses. In general, the level of imputation for child care arrangements in SIPP was about 5 percent, a level comparable to those reported in prior Census Bureau child care surveys.

Table D-2 shows imputation rates for items concerning time lost from work due to failures in child care arrangements and cash payments made for child care arrangements. Of the 1,586 respondents who were to answer the item if they or

Table D-1. Imputation Rates for Items on Primary and Secondary Child Care Arrangements

(For the three youngest children under 15 years old)

Arrangement and order of child	Unweighted number of children	Percent of responses imputed
Primary arrangement:		
First child	3,462	5.3
Second child	1,703	4.7
Third child	438	5.0
Secondary arrangement:		
First child	846	3.7
Second child	498	2.8
Third child	123	5.7

their spouse lost any time from work during the last month, 7.5 percent had their response imputed. Another 7.6 percent failed to answer the question if any cash payment was made for child care services, but for those women who were determined to have made a cash payment, only 2.1 percent failed to report on the amount of the payment.

An evaluation of the quality of the responses on SIPP is limited because of the general lack of data sets on child care at the national level. Wherever appropriate in the text of this report, comparisons have been made with data sources to evaluate the distribution of child care arrangements of preschoolers, the amount and frequency of cash payments made by families, time lost from work due to failures in child care arrangements, and estimates of unmarried males living with their children.

Table D-2. Imputation Rates for Time Lost From Work Because of Failures in Arrangements and for Cash Payments Made for Arrangements

Item	Unweighted number of respondents	Percent of responses imputed
Time lost from work ¹	1,586	7.5
Was cash payment made? ¹	1,586	7.6
Amount of cash payment ²	1,044	2.2

¹Limited to respondents who for any of their three youngest children, one or more of the following primary or secondary child care arrangements were used: grandparent, other relative of child (excluding family members), nonrelative of child, day/group care center, nursery school or preschool.

²Limited to respondents who were determined to have made a cash payment for child care arrangements.

Estimates of weekly child care payments presented special data collection problems. The data in SIPP represent the total child care expenses for all children in the household who were cared for by grandparents, other non-family relatives, nonrelatives, group/day care centers, nursery schools or preschools. Because of the above collection procedures, cost estimates for specific child care arrangements can only be ascertained if there was only one child in the household and if that child used only one type of arrangement. This procedure, however analytically limiting, was chosen because it became apparent when this questionnaire was pretested that the desired detail could not effectively be given by the respondents.

Unlike many other services purchased by individuals, the scope of duties and hours of child care services are not uniformly defined across households. Several types of

problems were encountered by the respondents. One such problem was that respondents often hired child care providers to work in their home who also performed other duties such as household cleaning, cooking, and marketing as part of their total cash payment. Thus, the respondent could not determine the actual cost incurred by the child care component out of the total cash payment.

Another typical problem arose when the respondent made a single cash payment to a caretaker who provided child care services for more than one child in a household. Often, it was not possible for a respondent to prorate the costs per child as child care providers may spend different amounts of time looking after children of different ages. Thus, it would be incorrect to assume that child care costs for two children in different age groups would be the same.