

## 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. The fall 1986 data include the combined 1985 Wave 6 and 1986 Wave 3 panels conducted from September to November, 1986. Fall 1987 data consist of the combined 1986 Wave 6 and 1987 Wave 3 panels conducted from September to November 1987.

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 or her 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 items 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 the 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 edited 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 full or part time or attending school.

Imputation rates for both primary and secondary child care arrangements 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 primary child care

**Table D-1. Imputation Rates for Items on Primary and Secondary Child Care Arrangements for Employed Women: Fall 1986 and 1987**

(Numbers represent actual numbers of children. Data shown are for combined panels)

Survey data and arrangement	Number of children	Percent imputed
FALL 1987		
Primary arrangement:		
First child .....	3,314	4.1
Second child .....	1,624	3.6
Third child .....	454	5.3
Secondary arrangement:		
First child .....	835	5.9
Second child .....	502	8.0
Third child .....	140	15.0
FALL 1986		
Primary arrangement:		
First child .....	3,331	4.7
Second child .....	1,606	3.9
Third child .....	457	5.3
Secondary arrangement:		
First child .....	813	6.4
Second child .....	458	6.8
Third child .....	112	12.5

arrangements for employed women in the SIPP panels in this report was about 4-5 percent. Higher imputation rates were found for secondary arrangements (from 6 to 15 percent).

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 female respondents who were to answer the item if they or their spouse lost any time from work during the month prior to the survey date because of a failure in child care arrangements, about 7 to 8 percent had their responses imputed both in the surveys in fall 1986 and 1987. Another 8 percent in both time periods failed to answer the question if any cash payment was made for child care services, but for those who were determined to have made a cash payment, only about 4 percent failed to report on the amount of the payment.

Estimates of weekly child care payments presented special data collection problems. Because of the questionnaire format, information on specific child care costs

**Table D-2. Imputation Rates for Time Lost From Work Because of Failures in Arrangements and for Cash Payments Made for Arrangements: Fall 1986 and 1987**

(Numbers represent actual numbers of respondents. Data are shown for combined panels)

Type of payment	Number of respondents	Percent imputed
FALL 1987		
Time lost from work <sup>1</sup> .....	1,591	7.4
Was cash payment made? <sup>1</sup> .....	1,591	7.6
Amount of cash payment <sup>2</sup> .....	1,095	4.3
FALL 1986		
Time lost from work <sup>1</sup> .....	1,529	7.8
Was cash payment made? <sup>1</sup> .....	1,529	8.2
Amount of cash payment <sup>2</sup> .....	1,031	4.4

<sup>1</sup>Limited to respondents who used grandparents, other relatives (excluding family members), nonrelatives, day/group care centers, or nursery schools/preschools for primary or secondary child care arrangements for any of their three youngest children.

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

for individual children or types of arrangements cannot

be ascertained. Costs refer to expenditures for all children in the household. 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 can be anticipated in quantifying cost estimates for child care services. One such problem is that respondents often hire child care providers, who in addition to providing child care services, also perform other services such as household cleaning, cooking, and marketing. The total cash payment to the child care provider, therefore, occasionally includes payments for these other services which have a market value in addition to child care services. Thus, a respondent could not determine the actual cost incurred solely for the child care component out of the total cash payment.

Difficulties are also encountered in data interpretation when a single cash payment is made to a caretaker who provides child care services to more than one child in the household. Often, it is not possible for a respondent to prorate costs per child. Child care providers may spend varying amounts of time looking after children of different ages and would charge differential rates. Thus, it would be incorrect to assume that child care costs for individual children in the same household but in different age groups would be the same.