

The Economic Role of Paid Child Care in the U.S.

A Report Series — Part 4: Child Care Data in the Current Population Survey, A Primer

June 2022



This report series was commissioned by the Committee for Economic Development of The Conference Board and produced by RegionTrack, Inc., an economic research firm, with generous support from the W.K. Kellogg Foundation. We also express special thanks to Grace Reef, President of Early Learning Policy Group, for her review and contributions. The report series provides an overview of the role of paid child care in the U.S. economy and across states. The series analyzes paid child care usage over the past two decades and its impact on workforce participation, including a focus on women and mothers.

The Economic Role of Paid Child Care in the U.S.

A Report Series — Part 4: Child Care Data in the Current Population Survey, A Primer



Table of Contents

Child Care Data in the Current Population Survey	5
CPS Survey Design	6
CPS Database Structure	7
Child Care Questions in the CPS	8
Accessing and Extracting CPS Child Care Data	9
Child Care Variables in the IPUMS-CPS	9
Child Care Sample Sizes	12
Producing Summary Results	13
Choosing a CPS Sample Weight	13
Child Care Queries	14
Derived Child Care-Related Measures	15
The Decision to Use CPS Child Care Data	16
Why Use CPS Child Care Data?	16
Other Federal Sources of Child Care Data	16
Technical Issues for CPS Data Users	17
COVID-19 Period	18
Confidence Intervals and Hypothesis Testing	18
2014 Change to Survey Panel Design	20
Measuring Income in the CPS	20
Annual Population Controls	21
Summary	22
Appendix	23
Glossary	24
Abbreviations and Acronyms	31

Table of Figures

Figure 1. Unweighted Survey Responses and Imputations	10
Figure 2. Unweighted CPS ASEC Sample Sizes - Child Care Variables	12

Child Care Data in the Current Population Survey

This data primer is the fourth report in a four-part series related to the use of paid child care in the United States. This final report in the series provides an introductory guide to child care-related data in the Current Population Survey (CPS). The CPS is the leading source for labor force data in the U.S., but its coverage also extends to data on paid child care usage and expenditures.

Child care data in the CPS has many advantages, including a long and continuous survey history, consistent survey methodology, wide coverage of children by age, and substantial state-level data coverage. The child care data in the CPS can also be analyzed alongside extensive economic and demographic data collected within the survey. This CED report series represents a deep analysis of the CPS data that has not yet been completed prior in the field about paid child care usage in the U.S. and seeks to provide guidance for researchers and others to increase the effectiveness and efficiency in working with these specific CPS data sets.

Despite some known limitations with the data, the CPS remains a versatile source of U.S. child care data for researchers, policymakers, and others studying the U.S. child care system. The following sections of the report discuss the design of the survey, the data captured by the survey, sources for the data, best practices when using the data, and technical issues likely encountered along the way.



CPS Survey Design

The CPS is a monthly survey of U.S. households sponsored jointly by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics (BLS). The CPS traces its roots back to 1940 when the Works Progress Administration initiated the Monthly Report of Unemployment following the Great Depression era.

The CPS, also commonly referred to as the household survey, is a sample-based survey of about 60,000 eligible households.¹ The survey is generally administered during the calendar week that includes the 19th day of the month, with questions referring to activity during the calendar week that includes the 12th day of the month. Responses are collected using a combination of live telephone and in-person interviews.

The core of the monthly CPS survey is designed to capture the labor force status of the civilian noninstitutional population ages 15 and older in the 50 states and the District of Columbia. Published labor force estimates typically include only those persons ages 16 and older. Active-duty members of the Armed Forces are excluded from the monthly survey, along with persons residing in institutions (e.g., correctional institutions or long-term care facilities).

The CPS initially focused on collecting labor force statistics but has far broader topical coverage today. The basic monthly survey now offers users a comprehensive historical body of monthly data on the U.S. labor force plus a range of economic and demographic characteristics for each person in the household, including children.² This allows for robust analysis of labor force data by sex, race, income, education, parental status, and other measures.

Annual Social and Economic Supplement. The Annual Social and Economic Supplement (ASEC), conducted along with the monthly survey every February, March, and April, is a widely used supplement to the CPS.³ The supplement extends well beyond the core monthly survey and gathers information on a range of topics including health insurance coverage, work experience, income from all sources, receipt of noncash benefits, poverty, migration, geographic mobility, and paid child care usage and expenditures.

The Census Bureau has collected ASEC data, formerly known as the CPS March Supplement, annually since 1947. The ASEC represents a tradeoff from monthly to annual data but provides a broader sample and larger universe than the basic CPS. The ASEC sample exceeds 75,000 households.

Persons living in noninstitutional group quarters (e.g., military barracks, missions, shelters, and group homes) are included in the expanded sample if it is their usual residence. The ASEC sample also includes Armed Forces personnel if they reside off post or with at least one other civilian on post.

The added households provide for more reliable coverage and improved precision of the ASEC estimates for many household types. The expanded sample also supplements coverage in the smallest states where the sample size is relatively small.

The ASEC uses a substantial sample but is far smaller than samples used in the American Community Survey (ACS) program.⁴ The ACS typically begins with a pool of 3.5 million addresses, nearly 50 times larger than the CPS ASEC pool. Census recommends the use of the ACS rather than ASEC for many state-level economic and demographic measures, particularly income. However, while the ACS provides a wide range of economic and demographic measures, it provides no direct measures of child care usage or expenditures. Other federal sources of child care data are discussed later in the report.

CPS, ASEC, and the Supplemental Poverty Measure. Data on child care expenditures are collected in the ASEC as part of the Census Bureau's Supplemental Poverty Measure (SPM) program.⁵ Beginning in 2011, the Census Bureau, in collaboration with the Bureau of Labor Statistics (BLS), began publishing an alternative measure of the official poverty rate that includes a broader range of income sources. The new poverty measure includes income received from many government programs designed to assist low-income families and individuals that is not included in the official poverty measure.⁶ The new poverty measure also considers work-related expenses, one of which is the cost of paid child care to enable a parent to work.

1 The October 2019 Survey Design and Methodological Technical Paper for the Current Population Survey is available online at: <https://www2.census.gov/programs-surveys/cps/methodology/CPS-Tech-Paper-77.pdf>.

2 A helpful discussion of issues surrounding the longitudinal features of the CPS is found here: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4477847/>.

3 A major reason for conducting the ASEC around the month of March is to obtain better income data, given its proximity to tax season.

4 For a detailed comparison of the design of the CPS-ASEC and ACS household surveys, see: <https://www.census.gov/topics/income-poverty/poverty/guidance/data-sources/acs-vs-cps.html>.

5 For more details on the SPM program and survey methodology, see: <https://www.census.gov/topics/income-poverty/supplemental-poverty-measure.html>.

6 For details on items collected for the SPM program, see: <https://cps.ipums.org/cps/resources/spm/p60-241.pdf>.

The need for estimates of child care costs in calculating the new SPM poverty measure prompted the introduction of questions related to expenditures on paid care beginning in the 2010 survey. A comprehensive report on the prevalence of poverty in the United States is produced annually using the official poverty measure along with the SPM measure based on information collected in the ASEC.⁷

CPS Database Structure

Users of CPS data must work within the data structure underlying the CPS data files. Datasets are structured using three basic data, or record, types: household, family, and person.

A household is defined as a set of related persons living in a housing unit and is the basic structural component of data produced in the CPS. The concept of a household is closely tied to the postal address of individual housing units. The number of households in the CPS will equal the number of occupied housing units.

There is a household record in the CPS files for each household or group quarters in the survey. All people not living in an individual housing unit (i.e., a house, apartment, mobile home, or rented room) are living in group quarters.⁸

A household or group quarters can comprise one or more families. The Census Bureau defines a family as “a group of two people or more (one of whom is the householder) related by birth, marriage, or adoption and residing together; all such people (including related subfamily members) are considered as members of one family.”⁹ There is a family record for each family unit within a household or group quarters in the CPS files. Families are also identified in the data for use with SPM program data and are referred to as SPM family units. Child care expenditure data are tracked for each SPM family unit in the CPS ASEC.

The final record type captures each person in the survey, both adults and children. Households and families comprise one or more persons, with one person record in the dataset for each person identified in the survey.

Survey Sampling and Weighting. The CPS survey uses a probability-based sample designed to produce estimates that are representative of the national and state labor forces. An independent sample is prepared for each state and the District of Columbia using the unique demographic and labor market conditions present in each state. The sample design requires approximately 60,000 housing units to meet minimum Census reliability criteria.¹⁰

The CPS uses the weighted sample to draw inferences about the full population of persons ages 16 and over in the civilian noninstitutional population. Creating summary estimates with the data involves weighting the data for each person in the sample. The base weight used for each person in the sample is a rough measure of the number of actual persons that the sample person represents in the national or state population. Persons in the same state will generally have the same base weight, but weights vary across states.

After administering the survey, the final step in preparing the database is the assignment of a series of weights to each household, family, and person in the survey. The CPS weights enable users of the CPS to produce their own weighted demographic and labor force estimates that are statistically representative of the nation, states, and other regions. The process of selecting and applying weights when forming summary measures from the data is discussed later in the report.



7 The 2020 SPM report is available online at: <https://www.census.gov/library/publications/2021/demo/p60-275.html>.

8 <https://www.census.gov/topics/income-poverty/poverty/guidance/group-quarters.html>.

9 For a more detailed definition of family and family unit, see: <https://www.census.gov/programs-surveys/cps/technical-documentation/subject-definitions.html#family>.

10 Census states that: “Sufficient sample is allocated to maintain, at most, a 1.9-percent coefficient of variation on national monthly estimates of unemployment level, assuming a 6-percent unemployment rate. This translates into a change of 0.2 percentage point in the unemployment rate being significant at a 90-percent confidence level.” See: <https://www.census.gov/programs-surveys/cps/technical-documentation/methodology/sampling.html>.

Child Care Questions in the CPS

Users of child care data in the CPS should review and understand the underlying set of questions used to collect the data. For reference, the full set of interviewer questions and instructions underlying the child care usage and expenditure data series is included in the Appendix.

Paid Care. The defining characteristic of child care data in the CPS is that it specifically covers the use of paid child care services. Parents are asked whether or not they paid for child care while they worked and how much they spent on care in the prior year. The expenditure questions ask respondents about their costs across all child care arrangements for all children in the household under the age of 15.

CPS ASEC first surveyed paid child care usage in 2001 and expenditures on child care in 2010. The wording of child care-related questions has remained highly consistent over the survey history. IPUMS-CPS tracks changes in the questions over time that affect the harmonized version of the data.¹¹

Interview Process. The initial child care question asks about all persons aged 15 and over with children in the household and their use of paid child care to enable them to work:

Did (you/anyone in this household) PAY for the care of (your/their) (child/children) while they worked last year? (Include preschool and nursery school; exclude kindergarten or grade/elementary school)?

The focus of the question on the use of paid care while working eliminates responses of paid care use for nonwork purposes, including leisure. The question design also excludes all forms of unpaid care, including care provided by parents, neighbors, grandparents, siblings, and other common care providers. The question is asked for each child under age 15 in the household. No questions are asked on the frequency of child care usage or the type of care used.

If question 1 is answered yes for a child in the household, the interviewer then follows a multi-step process to determine an estimate of annual expenditures on paid care. The respondent is first asked to provide both the periodic amount paid for care and the frequency of payments. These two components are multiplied to form an estimate of total annual expenditures. The respondent is then asked to confirm whether the estimate sounds correct to them.

Then (you/they) paid \$(amount) altogether in child care while (you/they) worked during 2020. Does that sound about right?

Yes
No

If the question on the correctness of total expenditures is answered no, they are asked to provide their best estimate of the annual amount paid while working.

What is your best estimate of the correct amount (you/they) paid for child care while (you/they) worked in 2020?

The total expenditure questions are asked for all children within each SPM family unit rather than for the entire household, which may include multiple families. Because CPS ASEC uses the SPM definition of “family” in reporting child care expenditures, it is used as the standard definition of family throughout all sections of the CED child care report.



11 For more details on changes to the usage questions, see: https://cps.ipums.org/cps-action/variables/KIDCNEED#questionnaire_text_section. For more details on changes to the expenditure questions, see: https://cps.ipums.org/cps-action/variables/SPMCHXPNS#questionnaire_text_section.

Accessing and Extracting CPS Child Care Data

Raw CPS data files provided by the Census Bureau often present significant access challenges for first-time users.¹² Along with their unwieldy size, CPS files are not pre-structured for longitudinal analysis and typically require the user to do extensive data cleaning and manipulation to produce summary measures of the data.¹³ Census provides only a limited online interface for simplifying access to raw CPS data.

Because the use of raw CPS files is challenging, many users instead create more manageable data extracts using resources such as the IPUMS-CPS database maintained by the Minnesota Population Center at the University of Minnesota.¹⁴ IPUMS-CPS is a harmonized version of the Census Bureau's CPS database that allows for continuity of use and consistent definitions of survey items across multiple years of data. The IPUMS-CPS online interface provides full access to both the monthly CPS and the annual ASEC datasets and makes smaller custom extracts available for easy download and use.

All CPS data used to form summary measures throughout the first three sections of the CED report series, *The Economic Role of Paid Child Care in the U.S.*, were extracted in tabular form from the IPUMS-CPS database. All discussion of CPS data in the remainder of this report refers to data extracted from the IPUMS-CPS dataset and uses the associated variable names and definitions unless identified otherwise.



Child Care Variables in the IPUMS-CPS

Three child care-related variables are found in the IPUMS-CPS database.¹⁵

1. KIDCNEED – children ages 14 and under who are reported to need paid child care while their parent(s) worked.

Each person record in the dataset is assigned a value for KIDCNEED of either 0, 1, or 2. All persons ages 15 and over are assigned a value of 0. All children under the age of 15 are assigned a value of either 1 or 2, where 1 denotes a child not in paid care and 2 denotes a child in paid care.¹⁶ Not all children ages 14 and under in a household or family will have the same entry for KIDCNEED. Some children under 15 in a household are in paid care, and some are not. Counts of the number of children in paid child care at the national or state level are formed using person records where KIDCNEED=2.

2. SPMCHXPNS – total annual family child care expenditures for each SPM family unit.

SPMCHXPNS captures the dollar amount of total annual child care expenditures for all children in paid care in each SPM family unit. The same dollar value is reported in the person record of all family members. When no child in the family is in paid care, a value of zero is reported in the person record of all family members. SPMCHXPNS is reported in current dollars but is easily adjusted for inflation using a Consumer Price Index adjustment factor.¹⁷

It is critical for users to note that the same total child care expenditure amount is repeated for each person record associated with a family if there is one or more children in paid care. Not accounting for duplicate entries can result in an overestimate of total child care expenditures in a household when more than one child is present. Users should select only one person record per SPM family unit when calculating child care expenditures.

¹² Raw CPS files are available online from the Census Bureau at: <https://www.census.gov/programs-surveys/cps/data/datasets.html>.

¹³ A typical file for the monthly CPS survey is approximately 120MB in size, while individual annual ASEC files are approximately 300-400MB in size.

¹⁴ IPUMS-CPS files are available online at: <https://cps.ipums.org/cps/>.

¹⁵ Child care variable names used in IPUMS-CPS differ from those used in the original CPS files: the paid child care usage variable is HCHCARE_YN and the expenditure variable is HCHCARE_VAL. There is also a variable for the uncapped annual amount of child care expenditures by SPM family unit named SPM_ChildcareXpns.

¹⁶ The values for 1 and 2 are reversed in the raw CPS files: HCHCARE_YN is assigned as 1=yes and 2=no.

¹⁷ For CPI adjustment factors in IPUMS-CPS, see: <https://cps.ipums.org/cps/cpi99.shtml>.

Figure 1. Unweighted Survey Responses and Imputations

Survey Year	Total Records		KIDCNEED - Children in Paid Care		
	KIDCNEED=0,1,2	KIDCNEED=0	KIDCNEED=1	KIDCNEED=2	KIDCNEED=1,2
	All Persons	Not of Child Care Age	Not in Paid Care	In Paid Care	Both not in and in Paid Care
2001	218,269	161,662	45,964	10,643	56,607
2002	217,219	161,412	45,886	9,921	55,807
2003	216,424	161,270	45,994	9,160	55,154
2004	213,241	159,618	44,967	8,656	53,623
2005	210,648	157,804	44,138	8,706	52,844
2006	208,562	157,114	43,079	8,369	51,448
2007	206,639	155,954	40,761	9,924	50,685
2008	206,404	156,633	40,045	9,726	49,771
2009	207,921	158,048	40,116	9,757	49,873
2010	209,802	159,609	41,336	8,857	50,193
2011	204,983	156,849	38,571	9,563	48,134
2012	201,398	154,647	37,501	9,250	46,751
2013	202,634	155,751	37,466	9,417	46,883
2014	199,556	153,998	36,593	8,965	45,558
2015	199,024	153,921	36,248	8,855	45,103
2016	185,487	143,626	33,528	8,333	41,861
2017	185,914	144,640	33,054	8,220	41,274
2018	180,084	140,562	32,023	7,499	39,522
2019	180,101	141,251	31,227	7,623	38,850
2020	157,959	124,806	26,495	6,658	33,153
2021	163,543	129,110	28,670	5,763	34,433

Table continues on next page

3. QKIDCNEE – denotes values of KIDCNEED that are imputed.

Imputed values are Census Bureau estimates of missing responses traced to two sources – nonresponding households and item nonresponse.¹⁸ Some households provide no responses at all while others do not respond to certain items in the survey. The Census Bureau notes that most of the imputations in the CPS are due to nonresponding households.

Evaluating the potential effect of imputed values on the reliability of CPS child care data requires a deeper evaluation of the responses for KIDCNEED. Figure 1 details the prevalence of imputed values in KIDCNEED in the 2001 to 2021

survey years. Measured relative to the total of 163,543 total person records in 2021, Census imputed 9,241 (or 5.7 percent) of the reported values. The share of imputations rises steadily from about 3 percent to 6 percent across the 2001 to 2021 survey years.

While this may seem a relatively modest share of overall imputations, users should note that values are imputed only for children reported as either in paid care or not in paid care (KIDCNEED=1 or 2) and not for the broader group of adults and older children (KIDCNEED=0) which comprises most entries. The share of imputed values measured using the subgroup of children under age 15 (KIDCNEED=1 or 2) is far larger. Across the 2001 to 2021 survey years, 16.5 per

¹⁸ For more discussion of imputation methods used, see: <https://www.census.gov/programs-surveys/cps/technical-documentation/methodology/imputation-of-unreported-data-items.html>.

Figure 1. Unweighted Survey Responses and Imputations *(continued)*

Survey Year	QKIDCNEE – Imputed Values for KIDCNEED				Imputed Share of Total Records			
	No Imputed or not in Universe	Imputed QKIDCNEE=1	Imputed QKIDCNEE=2	Imputed QKIDCNEE=1,2	QKIDCNEE=1,2 And KIDCNEED=0,1,2	QKIDCNEE=1 And KIDCNEED=1	QKIDCNEE=2 And KIDCNEED=2	QKIDCNEE=1,2 And KIDCNEED=1,2
	Not of Child Care Age	Not in Paid Care	In Paid Care	Both not in and in Paid Care	Not of Child Care Age	Not in Paid Care	In Paid Care	Both not in and in Paid Care
2001	161,662	45,964	10,643	56,607	161,662	45,964	10,643	56,607
2002	161,412	45,886	9,921	55,807	161,412	45,886	9,921	55,807
2003	161,270	45,994	9,160	55,154	161,270	45,994	9,160	55,154
2004	159,618	44,967	8,656	53,623	159,618	44,967	8,656	53,623
2005	157,804	44,138	8,706	52,844	157,804	44,138	8,706	52,844
2006	157,114	43,079	8,369	51,448	157,114	43,079	8,369	51,448
2007	155,954	40,761	9,924	50,685	155,954	40,761	9,924	50,685
2008	156,633	40,045	9,726	49,771	156,633	40,045	9,726	49,771
2009	158,048	40,116	9,757	49,873	158,048	40,116	9,757	49,873
2010	159,609	41,336	8,857	50,193	159,609	41,336	8,857	50,193
2011	156,849	38,571	9,563	48,134	156,849	38,571	9,563	48,134
2012	154,647	37,501	9,250	46,751	154,647	37,501	9,250	46,751
2013	155,751	37,466	9,417	46,883	155,751	37,466	9,417	46,883
2014	153,998	36,593	8,965	45,558	153,998	36,593	8,965	45,558
2015	153,921	36,248	8,855	45,103	153,921	36,248	8,855	45,103
2016	143,626	33,528	8,333	41,861	143,626	33,528	8,333	41,861
2017	144,640	33,054	8,220	41,274	144,640	33,054	8,220	41,274
2018	140,562	32,023	7,499	39,522	140,562	32,023	7,499	39,522
2019	141,251	31,227	7,623	38,850	141,251	31,227	7,623	38,850
2020	124,806	26,495	6,658	33,153	124,806	26,495	6,658	33,153
2021	129,110	28,670	5,763	34,433	129,110	28,670	5,763	34,433

Notes: All record counts reflect the number of unweighted survey responses each year. KIDCNEED is the number of children in paid care. QKIDCNEE is the number of imputed values for KIDCNEED.
Source: IPUMS CPS - University of Minnesota, U.S. Census Bureau, and RegionTrack calculations

cent of all responses for children under the age of 15 were imputed. The share increased steadily over the full period, from 11.6 percent of imputations in 2001 to 26.8 percent in 2021. The imputation share has averaged approximately 27 percent since the 2018 survey period. There is no noticeable increase in the share of imputations in 2020 and 2021 during the COVID-19 pandemic, despite surveying difficulties reported during the pandemic and a decline in the overall ASEC sample size of more than 10 percent from the 2018/2019 surveys to the 2020/2021 surveys.

Those children imputed as being in paid care are also assigned a value for child care expenditures. No estimate of expenditures is required for those imputed as not in paid

care. While there is a far greater number of imputations for children not in paid care, the share of imputations is closely balanced over time between both groups across the 2001 to 2021 survey years.

Users must determine the best treatment of imputed values when forming estimates of children in paid care. While the imputed values are typically used in full, the effect of the imputations on the statistical properties of the estimates is unknown. Users may also drop the imputed estimates and proceed with a smaller sample of known responses, but this approach may distort the internal sample weights underlying the full dataset.

Child Care Sample Sizes

Figure 2 provides an overview of the unweighted sample sizes underlying the overall CPS ASEC and child care-related data the past two decades.

2021 Survey Year. In the 2021 survey year, the 62,850 households in the survey comprised 65,620 SPM family units and 163,543 individual persons. Among the SPM family units in the unweighted sample, 3,862 (5.9 percent) reported paying for child care services in the prior year.

Among all persons in the unweighted sample, 129,110 (79 percent) were ages 15 and over and 34,433 (21 percent) were children under age 15. Among children under age 15, more than two-thirds (24,440) are ages 5 to 14. Fewer than 10,000 children under age 5 were captured in the 2021 sample.

Only 5,763 children (16.7 percent) under the age of 15 are reported as using paid care in the unweighted sample. There are slightly more (3,136) older children ages 5 to 14 in

paid care in the sample relative to younger children under age 5 (2,627).

Declining Sample Size. Unweighted sample sizes used in the CPS ASEC declined steadily the past two decades (see Figure 2). The unweighted number of households in the survey declined by more than 10,000 (12.5 percent) between 2001 and the pre-pandemic years of 2018 and 2019.

Unweighted person counts dropped from 218,000 in 2001 to only about 180,000 (17.5 percent decline) in both 2018 and 2019, prior to the pandemic. Even larger percentage declines in unweighted counts are found the past two decades for the number of SPM families with children in paid care, children under age 15, and children under age 15 in paid care.

COVID Period Declines. Unweighted sample sizes for the CPS ASEC were far smaller in both the 2020 and 2021 survey years relative to pre-pandemic levels. Figure 2 highlights the decline in the number of household and family records in the sample in both the 2020 and 2021 survey

Figure 2. Unweighted CPS ASEC Sample Sizes - Child Care Variables

Survey Year	Households, Family Units, and Persons				Total Children			Children in Paid Care		
	Total Persons	Total Households	Total SPM Family Units	SPM Family Units with Child Care Expenses	Ages 0-14	Ages 0-5	Ages 5-14	Ages 0-14	Ages 0-5	Ages 5-14
2001	218,269	78,054	-	-	56,607	17,303	39,304	10,643	4,893	5,750
2002	217,219	78,265	-	-	55,807	17,012	38,795	9,921	4,618	5,303
2003	216,424	78,310	-	-	55,154	16,975	38,179	9,160	4,120	5,040
2004	213,241	77,149	-	-	53,623	16,447	37,176	8,656	3,948	4,708
2005	210,648	76,447	-	-	52,844	16,357	36,487	8,706	3,880	4,826
2006	208,562	75,939	-	-	51,448	15,923	35,525	8,369	3,897	4,472
2007	206,639	75,477	-	-	50,685	15,936	34,749	9,924	4,465	5,459
2008	206,404	75,872	-	-	49,771	15,710	34,061	9,726	4,363	5,363
2009	207,921	76,185	-	-	49,873	15,814	34,059	9,757	4,430	5,327
2010	209,802	76,260	79,938	5,953	50,193	15,947	34,246	8,857	3,993	4,864
2011	204,983	75,188	78,598	6,442	48,134	15,276	32,858	9,563	4,327	5,236
2012	201,398	74,383	77,552	6,195	46,751	14,513	32,238	9,250	4,087	5,163
2013	202,634	74,821	78,290	6,334	46,883	14,209	32,674	9,417	4,229	5,188
2014	199,556	74,170	77,449	6,058	45,558	13,594	31,964	8,965	3,921	5,044
2015	199,024	74,257	77,472	5,972	45,103	13,546	31,557	8,855	3,808	5,047
2016	185,487	69,484	72,551	5,611	41,861	12,362	29,499	8,333	3,553	4,780
2017	185,914	69,957	73,178	5,559	41,274	12,313	28,961	8,220	3,573	4,647
2018	180,084	67,909	71,009	5,086	39,522	11,647	27,875	7,499	3,220	4,279
2019	180,101	68,345	71,517	5,158	38,850	11,632	27,218	7,623	3,397	4,226
2020	157,959	60,460	63,092	4,464	33,153	9,771	23,382	6,658	2,958	3,700
2021	163,543	62,850	65,620	3,862	34,433	9,993	24,440	5,763	2,627	3,136

Notes: Each entry reflects the raw unweighted record count in the CPS-ASEC dataset. SPM-based expenditure data are only available beginning in the 2010 survey year.
Source: IPUMS CPS - University of Minnesota, U.S. Census Bureau, and RegionTrack calculations.

years. The drop in the number of person records exceeded 12 percent in the 2020 survey year before increasing slightly by 3.5 percent in 2021. The number of person records in 2021 remained more than 9 percent below the 2019 count.

The number of SPM family units reporting children in paid care dropped in both 2020 and 2021 by a cumulative

25.1 percent. The number of children in paid care similarly dropped in both 2020 and 2021 by a cumulative 24.4 percent.

Further discussion of surveying issues experienced by the Census Bureau during the COVID-19 pandemic is contained in the Technical Issues section of the report.

Producing Summary Results

Summary measures used most commonly in the CED child care report series include counts, sums, ratios, and averages of one or more individual data series. A range of common data processing and statistical tools allow CPS users to create summary measures of the child care data. The most frequently used tools include either statistical software packages (e.g., SAS, Stata, R, and EViews) or traditional database management systems (DBMS) using structured query language (SQL).¹⁹

The summary measures presented throughout the CED child care report series are prepared using SQL queries applied to data tables within MySQL, a widely used DBMS. IPUMS makes several instructional documents available to aid users in extracting data and creating summary estimates from IPUMS-CPS data.²⁰ IPUMS also makes some analytical tools available online for creating summary measures.²¹

The remainder of this section of the report describes the overall process used to summarize both child care usage and expenditure data in the CED child care report. The examples are not specific to any individual software package but are intended to highlight the basic process followed when using any software tool to summarize the data.

Choosing a CPS Sample Weight

Because a weighted sample is used to produce CPS survey data, all summary measures of CPS data should use an appropriate sample weight. Both the Census Bureau and IPUMS provide detailed instruction on the appropriate use of weights to produce summary measures from CPS survey data.²²

Many users of CPS data unknowingly use unweighted sample data (see Figure 2) to produce summary measures for the underlying population. The use of unweighted data is appropriate only for the sample at hand and is not representative of the broader U.S. population. The Census Bureau provides pointed caution to users of unweighted CPS data: “*Unweighted counts can be very misleading and should not be used in demographic or labor force analysis.*”²³

The appropriate weight to use with a given data series is determined by both the source and type of data being analyzed. Person-based measures such as population and the number of children in paid care should use person-based weights. For example, the appropriate weight for the number of children in paid care is ASECWT, the primary person-level weight for the ASEC.²⁴

For household or family-based measures from the ASEC such as child care expenditures, the appropriate weight is SPMWT, a family-based weight for family-level analysis with SPM-related data in the ASEC.

Weights are generally multiplied by the unweighted values in the survey to produce weighted values. This is a simple task when calculating a total across all persons, or a group of persons, in the survey. Each raw person value in the survey is multiplied by the weight for each person value and summed to create a weighted total. To calculate an appropriately weighted mean of a CPS data series (e.g., average age), each unweighted observation on age is multiplied by the appropriate weight, summed together, and then divided by the sum of the weights.

19 Frequently used database management systems include MS SQL, Oracle, MariaDB, MySQL, SQLite, and PostgreSQL.

20 https://pop.umn.edu/sites/pop.umn.edu/files/final_review_-_cps_stata_exercise_1_o.pdf.

21 See the available IPUMS-CPS online tools at: <https://cps.ipums.org/cps/sda.shtml>.

22 From IPUMS-CPS, see: https://cps.ipums.org/cps/sample_weights.shtml. From the Census Bureau, see: <https://www2.census.gov/programs-surveys/cps/methodology/Source%20and%20Accuracy.pdf>.

23 <https://www.census.gov/programs-surveys/cps/technical-documentation/methodology/producing-summary-statistics.html>.

24 For a more detailed discussion of the use of ASECWT, see: https://cps.ipums.org/cps-action/variables/ASECWT#description_section.

The process of using weights is rarely immediately transparent to users but becomes more familiar through examples:

- To calculate the total number of children in paid care, sum the weights (ASECWT) for each person record where a child is in paid care.
- To calculate the total number of families with children in paid child care, sum the weights for each distinct SPM family unit where child care expenditures are greater than zero.
- To calculate total child care expenditures in all U.S. families, multiply the child care expenditures of each family unit in the survey by the weight (SPMWT) and sum them.
- To calculate average child care expenditures per SPM family unit, divide total child care expenditures calculated above by the number of families with children in paid care calculated above.

Many other measures of child care-related summary estimates are possible using a similar process.

Child Care Queries

Producing useful summary estimates of CPS child care-related data often entails the use of non-child care specific data series available within the survey. Other data commonly used include the year, age of child, state, and parent demographics. The following examples describe the methodology used to create expanded queries for both child care usage and expenditures. These summary measures appear throughout the CED child care report series.

Children in Paid Care Query. To create a count of the number of children ages 0 to 14 in paid care (KIDCNEED) in the U.S. beginning in the 2001 survey year:

- Count the number of person records where KIDCNEED=2; or count the number of children not in paid care using the number of records where KIDCNEED=1.
- More detailed extracts typically use these additional parameters:
 - Use YEAR to restrict the years used in the analysis and to group the number of children by year
 - Use AGE to select the age grouping of children – 0 to 14, 0 to 4, 5 to 14, or choose other groupings. The default for the use of KIDCNEED=1,2 captures all children ages 0 to 14.
 - Use STATEFIP to create state-level estimates; some estimates by region, metropolitan area, and county may be possible.

- Apply the person-level ASECVT weight when creating weighted estimates of the number of children in paid care.

Paid Child Care Expenditures Query. For extracting child care expenditures (SPMCHXPNS) by SPM family unit beginning in the 2010 survey year:

- Choose the distinct set of SPM family units using the SPMFAMUNIT unique identifier. Other definitions of households and families are available in the CPS but will produce inconsistencies when using the child care expenditure data.
- Select a single record from each SPM family unit for tabulating child care expenditures.²⁵ If using another identifier for households (e.g., HSEQ or SERIAL) to summarize child care expenditures, users should note that a small number of households have more than one SPM family unit. Another approach commonly used to identify one person per household is to use the first person record (PERNUM=1) in each household. This approach will not individually capture some of the additional families. Also, using the reference person (FAMREL=1) for each household will not capture all families because the reference person often answers on behalf of multiple families in the household.
- To identify those households with child care expenses, select only those households with child care expenditures (SPMCHXPNS >0). There are some very small dollar amounts reported for total annual child care expenditures that do not reflect regular use of paid care. To capture those using care on a regular basis, users may want to set a minimum annual child care payment.
- Other helpful parameters:
 - Use YEAR to restrict the years used in the analysis
 - Use YNGCH to specify the age of the youngest child
 - Use ELDC to specify the age of the eldest child
 - Use NCHLT5 to specify the number of children under the age of 5
 - Use NCHILD to specify the number of own children of any age
 - Use STATEFIP to create state-level estimates; some estimates by region, metropolitan area, and county may be possible
 - Use NFAMS to capture only those households with a certain number of families (e.g. only one family or more than one family)
 - Use SPMFTOTVAL to capture the total income of each SPM family unit.

Apply the SPMWT weight when creating weighted estimates by SPM family unit.

²⁵ A SQL subquery is used to identify the first person record in each SPM family unit. More specifically, it is not necessary to extract the first person but a single person from each SPM family unit. All members of the family unit have the same SPM weight and child care expenditure amount.

Child Age Groupings. Child care information in the ASEC is collected only for children ages 0 to 14. The first three sections of the CED child care report series commonly partition the overall data into two subgroups – younger children ages 0 to 4 and school-age children ages 5 to 14. The market for child care services is largely segmented along these two age groupings by both cost and market structure. Custom groupings of children using other age ranges are easily made using the AGE variable within the CPS.

Derived Child Care-Related Measures

Child care-related data reported in the CPS ASEC can be combined with many other CPS data series to create useful child care-related measures. Assembling the following group of summary measures produces a highly useful set of child care-related measures for a region:

1. number of children of child care age who are in paid care – count records where KIDCNEED=2;
2. total child care expenditures across all households or family units – sum all SPMCHXPNS>0 across the set of unique SPMFAMUNIT;
3. number of family units with children in paid child care – count the unique number of SPMFAMUNIT;
4. number of children of child care age who are not in paid care – count children KIDCNEED=1;
5. total number of children of child care age (under age 15) – derived from #1 and #4 above;
6. share of total children of child care age who are in paid care – derived from #1 and #5 above;
7. average child care expenditures per household or family unit in the region – derived from #2 and #3 above;
8. average child care expenditures per child in paid care in the region – derived from #1 and #2 above;
9. average income of all households or family units with children in paid care in the region – sum SPMFTOTAL for each unique SPM family unit with a child in paid care and divide by the number of SPM family units – derived using #3 above;
10. child care expenditures as a share of household income in the region – derived from #7 and #9 above; and
11. child care expenditures per child as a share of household income in the region – derived from #8 and #9 above.

The above derived data provide a multidimensional child care dataset at the national and state levels. These measures are used and discussed throughout the first three sections of the CED child care report series. Again, all derived child-care related estimates described above should use the appropriate sample weight.

Demographic and Regional Groupings. It is similarly straightforward to tabulate paid child care usage and expenditures along economic and demographic characteristics, including age, sex, race, income, education, marital status, and other measures. The breadth of data available in the basic CPS provides for rich analysis of paid child care usage by crossing it with demographic measures of households, families, and persons. The CED child care report series provides extensive analysis of child care related variables by income, education, sex, race, and other measures.

The child care data are also easily tabulated at both the national and state levels. State totals are easily aggregated to form custom regional totals.

Point Estimates vs. Multi-Year Averages. The decision to partition CPS data by geographic, economic, or demographic measures can quickly result in small sample sizes. This is especially the case when using a small subset of children in paid care, special characteristics of parents, or data for a small state.

Single-year point estimates formed using small samples may have very high standard errors (or uncertainty) relative to the true population estimate. The Technical Appendix describes the process for calculating standard errors and confidence intervals for summary estimates from the CPS ASEC. Many authors choose not to report measures when the standard error is high. An alternative is to use a multi-year average.



The Decision to Use CPS Child Care Data

The defining characteristic of child care data in the CPS survey is the coverage of paid forms of child care used to allow a parent to work. This aspect of the survey design may be either an advantage or disadvantage to the user but is the first consideration in the decision to use CPS data for child care-related research or analysis.

Why Use CPS Child Care Data?

If the underlying definition of paid care used in the survey is acceptable to the user, there are a number of advantages and disadvantages that must be considered when using CPS data.

Advantages of using child care data from the CPS include:

1. Consistent survey data available on both usage and expenditures.
2. The use of a highly consistent survey design and implementation process by the Census Bureau over the survey years of child care data collection.
3. Coverage availability at both the U.S. and state levels. National data are derived from the same dataset, making state totals additive to the national total.
4. Extended annual coverage now exceeding one decade for child care expenditures and two decades for child care usage.
5. The ability to partition child care data using detailed economic and demographic data on families and children available in the ASEC.
6. Coverage of children ages 0 to 14, which captures paid care usage of both preschool and school-age children.

There are also well-known limitations to the CPS child care data traced primarily to the survey's design:

1. No data are collected on the specific type of child care arrangement for each child.
2. The amount paid for care does not specifically capture the price of care but instead measures total annual household expenditures.
3. Expenditures are calculated per SPM family unit and cannot generally be identified for each child in care unless a family has only a single child in paid care.
4. No estimate is provided for the typical number of hours of paid care used each week, month, or year.
5. Some state samples in the CPS may be relatively small, particularly when slicing the data along multiple dimensions.

Along with no coverage of unpaid care, the primary limitation is that the CPS survey does not ask respondents about the specific types of care received or the quantity of care purchased. These limitations may be insurmountable if needing data on the type of care used, particularly the use of unpaid care. Because of the limitations of the survey design, CPS-based child care data are not well suited to addressing questions concerning the quality of care, types of care, frequency of care, or other specific characteristics of care arrangements available in each state. Other data sources may be better suited than the CPS for answering questions concerning the cost of care.

Despite the known limitations, CPS ASEC child care data does a good job of the basic tasks of counting the number of children in paid care and the amount spent by families on care. CPS data offer substantial benefits over many other sources of child care usage and cost data. These benefits include consistent survey methodology, annual continuity of data over an extended time frame, and state-level geographic coverage. The time series features and state-level coverage provided by CPS data are leveraged in the analysis throughout the CED child care report series.

Other Federal Sources of Child Care Data

Alternative sources on the usage and cost of child care services in the U.S. are not widely available. Two large-scale federal surveys – the Census Bureau's Survey on Income and Program Participation²⁶ (SIPP) and the U.S. Department of Education's Early Childhood Program Participation²⁷ (ECPP) survey administered as part of the National Household Education Surveys (NHES) program – provide much of the alternative data on U.S. child care usage and cost.

Much like the CPS survey, these survey-based sources simultaneously collect a wide range of economic and demographic data, which enables analysis of child care usage and expenditures by other household, family, and person characteristics. The primary drawback is that these surveys are generally performed only periodically and provide little continuous data for performing analysis of paid child care markets over time. Geographic coverage at the state level is also limited in both data sources.

²⁶ Details on the SIPP survey are available at: <https://www.census.gov/programs-surveys/sipp.html>. A detailed comparison of estimates from the SIPP and CPS ASEC are available in the following report: https://s3.amazonaws.com/sitesusa/wp-content/uploads/sites/242/2014/05/2007FCSM_Sae-ung-VIII-A.pdf.

²⁷ Details on the NHES-ECPP survey are available at: <https://nces.ed.gov/nhes/dataproducts.asp>.

SIPP. The Child Care Topical Module of the SIPP survey has been administered by the Census Bureau since 1985 and provides coverage of the number of children in care and the amount spent on child care. The survey has historically captured extensive data on all paid child care arrangements used by a family, regardless of cost, facility type, or licensing status. Total weekly child care expenditures are collected and can be extrapolated across children to form an estimate of annual spending. However, beginning with the 2014 panel, the survey no longer collects information by type of arrangement and is now based on all arrangements used for all children in the household, like the CPS survey.

The SIPP survey has other important limitations. SIPP uses a state-based survey design but is only suitable for state-level analysis of child care costs in certain states due to sample size concerns. Another limitation of SIPP child care spending data is the infrequent intervals over which the survey is conducted. The survey began collecting data in yearly overlapping panels in 2018 but has completed only three panels with this design (2018, 2019, and 2020). Prior panels were collected in wide intervals prior to 2018, occurring only in 1996, 2001, 2004, 2008, and 2014 over the prior 25 years.²⁸ This eliminates any opportunity to examine historical year-to-year changes in child care activity.

ECPP. The U.S. Department of Education collects data on the use and cost of a variety of types of child care arrangements through the NHES-ECPP survey program. Much like the SIPP survey, the NHES-ECPP program extends across several decades but is produced only at wide time intervals and does not provide a continuous annual dataset. NHES-ECPP surveys were conducted in 1991, 1995, 2001, 2005, 2012, 2016, and 2019, with a reduced set of questions asked in 1999.²⁹

A strength of the NHES-ECPP survey is the collection of extensive information on the cost and types of care used, including hourly measures of any cost burden. Coverage includes both paid and unpaid care, care both inside and outside the home, and extensive information on relative care (including paid relative care). In contrast, CPS provides no coverage of unpaid forms of care and no details on the types of care used.

An important limitation of the NHES-ECPP survey is that the sample size does not allow for state level analysis in all states despite fifty-state survey coverage. The survey also focuses only on children ages 6 and under who are not yet enrolled in kindergarten. This excludes more than half of all children under age 15 who are enrolled in paid child care.

Technical Issues for CPS Data Users

This section of the report addresses five technical issues faced by CPS data users. The issues are not specific to child care data but are instead traced to general use of the monthly CPS survey and the ASEC.

The first topic addresses concerns related to the use of data from the 2020 and 2021 survey years during the COVID-19 pandemic period. The second topic discusses the common question of how to calculate standard errors and create confidence intervals around estimates.

The third issue is faced by all users of CPS data from the 2014 survey year. Income questions in the survey were redesigned, and users must make choices concerning the sample used in forming summary estimates. The fourth issue surrounds concerns with using income measures in the CPS. The CPS is widely known to understate income because of the method used to record income in the survey.

Finally, we examine the ongoing issue faced by CPS users as annual population controls are applied to the data each year. Population controls can affect the historical comparability of CPS estimates over time.



²⁸ Child care questions were asked in SIPP panels in Apr. 1997-Jul. 1997, Apr. 1999-Jul. 1999, Oct. 2001-Jan. 2002, Feb. 2002-May 2002, Oct. 2002-Jan. 2003, Oct. 2003-Jan. 2004, Feb. 2005-May 2005, Jun. 2006-Sep. 2006, Jan. 2010-Apr. 2010, and Jan. 2011-Apr. 2011.

²⁹ For historical coverage of the NHES-ECPP survey, see: https://nces.ed.gov/nhes/surveytopics_early.asp.

COVID-19 Period

The events surrounding the Covid-19 pandemic resulted in substantial issues with the Census Bureau's implementation of the CPS in the 2020 and 2021 survey years. According to IPUMS reports:³⁰

During March of 2020, CPS data continued to be collected with some modifications in procedure for the safety of respondents and Census Bureau employees.

...

On March 20th, the Census Bureau suspended in-person data collection and closed two of the call centers that assisted in collecting CPS data. However, data collection continued by phone and efforts were made to reach households via phone that would normally have received an in-person visit. Response rates were lower than average by about 10 percentage points in March 2020.

Enumeration continued by phone exclusively in April through June of 2020. Response rates continued to be unusually low for incoming rotation groups that normally receive in-person interviews. Low response rates persisted during subsequent surveys for groups that entered the CPS during the pandemic.

In July of 2020, in-person interviews began again in some areas of the country and the Census Bureau call centers that usually assist with CPS data collection also returned to interviewing in a limited capacity in this month. Response rates rose slightly compared to May and June, but remain lower than average.

...

In-person interviews were expanded to all areas of the country in September of 2020. In-person interviews were conducted only after attempting to reach all households by phone.

Additional research on the effects of the pandemic on the CPS is underway. Among the early releases, an NBER working paper³¹ (Montenovo, et. al.) finds evidence of non-random attrition or respondents in the CPS sample since the onset of the COVID-19 pandemic. Those who responded to the CPS in February 2020 but not in March are, on average, younger, more diverse, and less educated than those who responded in both February and March of 2020.

A recent Census Bureau working paper (Rothbaum and Bee) suggests that nonresponse rates in the CPS ASEC increased substantially in the 2020 survey year due to the COVID-19 pandemic.³² Nonresponse was also more strongly associated with income than in prior years, with high-income households more likely to respond than low-income households. Response patterns also reportedly changed by education and Hispanic origin when using alternative weights to reflect decreased nonresponse rates.

The evidence suggests a strong potential for systematic errors in the CPS ASEC results for the 2020 and 2021 survey years. Sample sizes are smaller than normal (see Figure 2); however the share of imputed values is not elevated in the period (see Figure 1).

Users must make their own determination of the usability of CPS ASEC data collected during the pandemic period. Users can choose to omit these estimates from the pandemic period in published works. If estimates are published, standard errors may be produced. Another alternative is to use multi-year averages across the period. IPUMS-CPS also recommends the use of entropy balanced weights (ASECWTCDV)³³ during the COVID period to adjust for nonresponse bias.

Confidence Intervals and Hypothesis Testing

Estimates formed using CPS data are sample-based and, as a result, entail both sampling and nonsampling error. Sampling error occurs when an estimate derived from a sample rather than the entire population differs from the true population value it represents.

The degree of sampling error is often measured using the standard error of the estimate, a widely used measure of variability. A standard error can be used to construct confidence intervals for the estimates if they are unbiased, typically at the 90 percent or 95 percent level. Standard errors can also be used in hypothesis testing to determine whether two estimates are significantly different.

Nonsampling errors are other errors not related to the sample selected. Typical sources of nonsampling error include poor sample design, data entry errors, nonresponse rates, missing entries, and survey implementation errors. The implications of nonsampling errors in the CPS are generally unknown.

30 See a detailed discussion on CPS issues during the pandemic at: <https://cps.ipums.org/cps/covid19.shtml>. BLS statements on the impact of COVID are available along with monthly response rates to the CPS in the 2019 to present period.

31 The paper is available online at NBER: https://www.nber.org/system/files/working_papers/w27132/w27132.pdf.

32 The report on nonresponse during the pandemic is available online at: <https://www.census.gov/content/dam/Census/library/working-papers/2020/demo/sehsd-wp2020-10.pdf>.

33 See details on the IPUMS-CPS website at: https://cps.ipums.org/cps-action/variables/asecwtcdv#description_section.

Standard Errors. There are two basic methods for calculating standard errors for estimates derived from CPS ASEC data.

1. **Use estimated measures of generalized variance produced by Census.**

The simplest approach to calculating standard errors uses formulas³⁴ provided by the Census Bureau along with a table of estimated factors and parameters.³⁵ The factors and parameters vary based on whether the data series represents a level, average, or change over time. Factors are provided for six measures: month-to-month changes, changes in monthly estimates one year apart, quarterly averages, changes in consecutive quarterly averages, yearly averages, and changes in consecutive yearly averages. The estimated standard errors can then be used to create confidence intervals around the estimate.

2. **Use a large set of replicate weights.**

An alternative method is to generate empirically derived estimates of the standard error. Instead of using a single weight to produce a single point estimate, this method uses a large number of potential (replicate) weights that are representative of the population. In the CPS, 160 separate weights at the household and person level are available for use in calculating standard errors.³⁶ This approach requires far more extensive programming ability. IPUMS-CPS offers this description of the basic process:



How Do I Obtain Replicate Standard Errors From IPUMS-CPS Data?

There are 3 main steps:

Run your analysis using the full-sample weights for ASEC (ASECWT and ASECWTH are the main CPS ASEC weights). Record the statistic you are interested in (e.g., the mean income of veterans, or the coefficient describing the relationship between income and whether one has health insurance coverage).

Run your analysis again using each set of replicate weights. First, run the analysis using REPWTP1, then again using REPWTP2, then again using REPWTP3, and so on up to the final set of replicate weights. After each set, record the statistic you are interested in. (N.B.: If you are analyzing a household-only file, be sure to use REPWT1, REPWT2, etc.)

Insert the above results into the following formula:

$$SE(X) = \sqrt{\frac{4}{160} \sum_{r=1}^{160} (X_r - X)^2}$$

where X is the result from the analysis using the full-sample weight and X_r is the result from the analysis using the r -th set of replicate weights.

Does Using Replicate Weights Make Any Substantive Difference?

In IPUMS testing of CPS data, replicate weights usually increase standard errors. This increase is generally not large enough to alter the significance level of coefficients, though marginally significant coefficients may become clearly nonsignificant. The more obvious effect of using replicate weights is on the width of confidence intervals, which can change substantially.

Source: Reproduced from IPUMS-CPS. Available online at: <https://cps.ipums.org/cps/repwt.shtml>.

IPUMS-CPS reports that using replicate weights generally produces larger standard errors than formula-based estimates based on Census-provided factors and parameters. Larger standard errors generally reduce the significance of estimates and widen estimated confidence intervals.

34 BLS provides a detailed discussion of sampling and nonsampling errors in the CPS: <https://www.bls.gov/cps/calculating-standard-errors-and-confidence-intervals.pdf>.

35 A table with the parameters and factors for calculating standard errors is available online at: <https://www.bls.gov/cps/parameters-and-factors-for-calculating-standard-errors.xlsx>.

36 Discussion of replicate weights in IPUMS-CPS is available online at: <https://cps.ipums.org/cps/repwt.shtml>.

2014 Change to Survey Panel Design

Special care must be taken with using CPS ASEC data from the 2014 survey year due to a shift in the underlying CPS survey methodology.³⁷ The 2014 survey utilized a split panel design to test a redesigned set of income questions. An enlarged sample was used and split into two smaller samples. A smaller sample of approximately 3/8ths of the total sample used new income questions while the other 5/8ths of the sample continued to use the traditional income questions. The size of the 5/8ths sample remains consistent with historical sample sizes.

According to the Census Bureau:³⁸

Of the approximately 98,000 addresses in the 2014 CPS ASEC production sample, a subsample of about 30,000 addresses was randomly assigned to be eligible to receive the redesigned income questions. The remaining sample (approximately 68,000 addresses) was eligible to receive the set of ASEC income questions used in previous years, referred to here as the traditional income questions.

...

To examine the time series and make consistent comparisons, users may need to include two estimates for calendar year 2013. Estimates based on the portion of the sample that received the traditional income questions are most appropriate for comparing with 2012 and earlier years. Estimates based on the portion of the sample that received the redesigned income questions are the most appropriate for comparing 2013 with calendar year 2014 and beyond.

The practical issue with the enlarged sample is that count-based summary measures will be too high for 2014 unless only one of the subsamples is used. In IPUMS-CPS, both samples are weighted so that either sample is individually representative of the U.S. population.³⁹ The desired sample is selected using the HFLAG variable, where HFLAG=0 is the 3/8ths sample with the old income questions and HFLAG=1 is the 5/8ths sample with new income questions. If both samples are used, IPUMS-CPS makes the WTSUPP weight available with the resulting estimates equal to twice the U.S. population.

The Census Bureau provides little guidance on the use of the expanded sample to form estimates for 2014. The produc-

tion of two estimates as described in the Census quote above is used in many published estimates. If choosing only one subsample, the larger sample with the new income questions is likely preferred. Users can also use both subsamples in forming estimates. The approach taken throughout the CED child care report series is to use the full sample for all summary measures in 2014 and divide population-based estimates by two.

Measuring Income in the CPS

A well-known problem with the CPS is methodological issues with the tabulation of income. Some categories of income are "top-coded" at higher levels rather than reported using the full dollar amount. Top-coding is a traditional disclosure avoidance method used by the Census Bureau to prevent the identification of individuals. In this case, it protects the identity of extremely high-income individuals.

A recent report⁴⁰ by the Congressional Research Service provides a detailed view of the income top-coding problem.

"Some aspects of the Census Bureau CPS-ASEC data limit its usefulness in characterizing households at the top of the distribution. A key limitation derives from Census data recording and internal processing procedures, which effectively "top-code" individuals' four earnings categories at \$999,999 each, so that any individual's income above that limit is reduced to \$999,999 per category. In addition, Census data exclude capital gains income, which is an important source of income for certain top-income households because the distribution of wealth is also skewed (see the section below entitled, "Capital Income"). Census earnings data are top-coded at \$9,999,999 per earnings category at the time of data collection. Once collected, Census edits its income data to minimize the incidence of interviewer error or misreporting on the part of the individual interviewed. For the purposes of Census-published data tabulations (which are used in this CRS report) and public-use data, the internal processing limit is \$999,999 for each of the four individual earnings categories."

The practical result of top-coding is an understatement of true mean earnings in the U.S. when using CPS data. A report published by the Bureau of Labor Statistics provides a deeper evaluation of the problem and offers some insight into working within the limitations of the data.⁴¹

37 For more details, see: <https://www2.census.gov/library/publications/2014/demographics/p60-249description.pdf>.

38 For more details on the methodology change, see: <https://www.census.gov/topics/income-poverty/income/guidance/cps-methodology-changes.html>.

39 More information is available on the 2014 sample redesign at: https://cps.ipums.org/cps/three_eighths.shtml.

40 Congressional Research Service, The U.S. Income Distribution: Trends and Issues, January 13, 2021. Page 3. <https://crsreports.congress.gov/product/pdf/R/R44705>

41 See: Burkhauser, Richard V. and Jeff Larrimore. "Using internal CPS data to reevaluate trends in labor-earnings gaps." August 2009. Bureau of Labor Statistics. Monthly Labor Review. pp. 3-18. Available online at: <https://www.bls.gov/opub/mlr/2009/08/artifull.pdf>.

TOP-CODING IS A WELL-DOCUMENTED PROBLEM for the CPS, but until recently, the only available strategy for mitigating the problem has been to place further restrictions on the data, either by using consistent top-coding or by discarding the cell means provided by the Census Bureau from 1995 onward. As a result, calculations have tended to understate true mean earnings in the United States. When comparing earnings across two groups within the population that are top-coded at different rates, all previously available top-code correction schemes may lead to a misstatement of the earnings gap between the groups.

IPUMS-CPS provides a detailed evaluation of the changes in top-coded income values in historical CPS data along with suggested methods for using income derived from IPUMS-CPS datasets.⁴²

Annual Population Controls

The Census Bureau makes annual population control adjustments⁴³ to CPS data to include the latest information available about population change and to incorporate improvements in the estimation methodology. BLS incorporates the annual population control adjustments into the CPS estimates each January. The adjustments are based on the latest decennial census population count, supplemented with birth and death data and estimates of net international migration. Following a decennial census, a new population base is introduced along with the adjustments. The adjustments influence the underlying population totals as well as population estimates by sex and race. After the population controls are determined, CPS weights used to create summary measures are then updated.

The January 2022 population control adjustments introduce several changes to the CPS population estimates:⁴⁴

The population adjustments incorporated with household survey data for January 2022 reflect the introduction of a blended 2020 population base, which combines population totals from the 2020 Census and demographic characteristics from other sources. The January 2022 adjustments reflect updated birth and death statistics, which include the effects of the coronavirus (COVID-19) pandemic; new estimates of net international migration, which reflect reduced international migration due to the pandemic; and methodological improvements to the estimation process. The vast majority of the population change, however, is due to the change in the base population from Census 2010 to the blended Census 2020 base.

The adjustments may increase or decrease the population level, depending on whether the latest information indicates the population estimates had trended high or low. Conceptually, the effects of the annual population control adjustments represent cumulative over- or under-estimation of the population since the last decennial census point.

IPUMS-CPS provides a description of the population control process from their viewpoint:

Estimates on the entire population are prepared by projecting forward the resident population from the last available census. These projections are derived by updating the demographic census data from a number of other data sources that account for death, births and net migration. About 3 years after every census (i.e., 2003 for the 2000 Census and 2013 for the 2010 Census), the Census Bureau updates its independent population control and provides a new weight for the relevant years.

Two important points should be noted here. First, the lag between when the Census is conducted and when the CPS weights are updated is about 3 years. While the Census data are being processed, the CPS files are made available using the weighting scheme from the US Census prior to the latest Census. Second, once the files are updated, the old weights become obsolete and are replaced in the IPUMS data extract system. Published estimates from the lag years that use the old weights are not always updated. For example, 2010 poverty estimates were released in ASEC using the 2000 population controls. Once the 2010 population controls were made available, IPUMS-CPS replaced the ASEC 2010, 2011, and 2012 weights that are based on the 2000 population control with weights that are based on the 2010 population controls.

IPUMS-CPS provides only the most up-to-date weights with online data extracts, but older weights remain available to users for download.⁴⁵

The key implication of annual population control adjustments for users is that the process can render historical comparability of estimates from the CPS ASEC invalid. The annual population controls introduce the greatest uncertainty into monthly data from the core CPS. The use of recent monthly CPS data that stretches across years may not be directly comparable as controls are integrated in January data each year.

⁴² Income top-coding information from IPUMS-CPS is available at: https://cps.ipums.org/cps/topcodes_tables.shtml.

⁴³ Additional information about household survey population controls is available at: <https://www.bls.gov/cps/documentation.htm#pop>.

⁴⁴ Discussion of the January 2022 population control adjustments is available online at: <https://www.bls.gov/web/empsit/cps-pop-control-adjustments.pdf>.

⁴⁵ For the treatment of population controls in IPUMS-CPS, see: https://cps.ipums.org/cps/spm_weights_update_2014.shtml.

Summary

The CPS provides both policymakers and child care researchers with a robust historical dataset on paid child use and expenditures. The data have several advantages, including a consistent surveying methodology, an extended historical time frame, and state-level coverage.

All novice users of the CPS should study the introductory section of the report on the design and structure of the dataset and seek supplementary materials if needed. We also recommend that users closely evaluate the questions and methodology used in the survey interview process. In addition, users should evaluate both the sample sizes underlying the child care data and the prevalence of imputed values.

The discussion in the report focuses on IPUMS-CPS as a helpful source for access to raw CPS child care data. After extracting data in tabular form, several methodological tips are provided to assist users in assembling a broad set of child care-related measures. Care must be taken in choosing the appropriate sample weight when forming summary measures of the data.

Users of CPS child care data will eventually encounter the five technical issues discussed in the report. These issues include survey issues with the CPS during the COVID-19 epidemic, the calculation of standard errors and confidence intervals, steps required when using data from the 2014 survey, issues with the measurement of income in the CPS, and the effects of ongoing population controls on the data.

Finally, we hope this information and guidance will help support further research on child care using the CPS datasets.

Appendix - CPS Interviewer Questions - Child Care

The survey questions and instructions for the child care questions in the CPS ASEC are reproduced below. Items in blue reflect interviewer instructions for use in eliciting appropriate survey responses. The full technical manual for the 2021 CPS ASEC survey is available online at: <https://www2.census.gov/programs-surveys/cps/techdocs/cpsmar21.pdf>.

10.2 CHILD CARE

Q95

Now we want to ask about some of your expenses for children.

Did (you/ anyone in this household) PAY for the care of (your/their) (child/children) while (you/they) worked in 2020?

- Include: All child care expenses including preschool and nursery school expenses, before and after school care, and summer care.
- Do not include: cost of kindergarten or grade/elementary school.

- 1 Yes
2 No

Q95A

Which children needed care while their parents worked?

- Enter all that apply, separate using the space bar or a comma.
- Probe: Anyone else?
- Enter 96 for All persons
- Enter 0 if none

CCFREQ

What is the easiest way for you to tell us how much was paid for child care while (you/they) worked in 2020: weekly, every other week, twice a month, monthly, or yearly?

- 1 Weekly
2 Every other week
3 Twice a month
4 Monthly
7 Yearly

CCAMT

How much was paid (weekly/every other week/ twice a month/monthly) for child care?

- Include child care payments made for all children in the household.
- For example, if there are two adults in the household with childcare expenses use the total paid by both adults. Do not try to separate the payments. Record one total for the entire household.

CCNUMPAY

How many (weekly/every other week/twice a month/monthly) payments did (you/they) make during 2020?

(1-52), (1-26), (1-24), (1-12) _____

CCTOT

Then (you/they) paid \$(amount) altogether in child care while (you/they) worked during 2020. Does that sound about right?

- 1 Yes
2 No

CCEST

What is your best estimate of the correct amount (you/they) paid for child care while (you/they) worked in 2020?

_____ Best-Fit Line

Glossary

Best-Fit Line

A best fit line through a scatter plot of data best expresses the linear relationship between those points. The straight line provides the best approximation of the relationship between the data points. The slope of the line provides a quantitative estimate of both the direction and magnitude of the relationship. Best fit lines are also commonly referred to as trendlines or linear regression lines.

Birth Rate

Measures of the birth rate reflect the number of births in a population over time (typically one year). Commonly used measures of the birth rate include the crude birth rate and fertility rate.

Capital Investment

Capital investment is the purchase of tangible and intangible assets by firms, governments, and individuals for the purpose of pursuing their business and operating goals and objectives. Capital investment is measured in the child care report using net private fixed investment as defined by BEA. The estimated measure of capital is net of depreciation and includes the broad asset categories of equipment, structures, and intellectual property. Public sector assets are excluded from the analysis. State-level estimates are formed by partitioning national data on net private fixed assets at the industry level based on a region's share of national household earnings at the industry level.

Child and Dependent Care Tax Credit (CDCTC)

The Child and Dependent Care Tax Credit is a federal tax credit available to pay for the care of eligible children and adult dependents (qualifying persons) to enable taxpayers to work, look for work, or attend school. The credit is calculated based on income and covers a percentage of expenses incurred for the care of qualifying persons. For tax year 2021, the American Rescue Plan Act of 2021 extended the credit up to \$4,000 for one qualifying person and \$8,000 for two or more qualifying persons and made the credit temporarily refundable.

Children of Child Care Age

Children ages 14 and under are considered most likely to participate in formal or informal child care. The group of children of child care age is divided into two groups in the report: younger children ages 0 to 4 and older children ages 5 to 14. This definition follows the Current Population survey which measures paid child care usage for children ages 14 and under.

Child Care Cost Burden

The child care cost burden of a household reflects expenditures on paid child care as a percentage of household income. This measure reflects the notion that the cost burden of child care is best measured relative to ability to pay. The cost burden is calculated as child care expenditures divided by total household income. Burden can be measured on a per child basis or for all children in a household.

Civilian Non-Institutional Population

The civilian non-institutional population measures those persons ages 16 and older and their children not on active duty in the Armed Forces or residing in institutions (e.g., correctional institutions or long-term care facilities for the aged).

Cointegration

Cointegration is a statistical concept that refers to the long-run co-movement of two or more data series over time. If variables are found to be cointegrated, estimates can then be made of the long-run elasticity between two cointegrated factors over time. The concept of cointegration is closely tied to the notion of Granger causality but focuses on the long-run dimension of the relationship among a group of variables over time. A set of cointegrated variables maintain a long-run equilibrium relationship over time, with any short-run deviations from the long-run relationship corrected over time through an error-correction process. In fact, two cointegrated variables are expected to have short-run Granger causality present in at least one direction.

Confidence Interval

A confidence interval reflects a margin of error or degree of uncertainty surrounding a statistical estimate. Using the standard error and a predetermined confidence level (typically 90% or 95%), a numerical estimate of the interval containing the true value of the parameter is formed. Confidence intervals are also used in hypothesis testing.

Correlation

Correlation is a statistical measure of the degree of linear dependence between two series over a specified period. Correlated series tend to move in coordination with one another over time. Positively correlated variables tend to move in the same direction; negatively correlated series tend to move in the opposite direction. If X and Y are correlated, the calculated correlation is the same for both series in each direction in the period.

Cost-of-Living

Cost of living reflects differing prices across geographic areas for a range of typical living expenses including housing, food, energy, and other items. Measures of the cost of living are often used to compare how costly it is to live in one geographic area versus another. Cost of living adjustments are made in the report using state-level regional price parity (RPP) indexes produced by the Bureau of Economic Analysis (BEA) along with the national implicit price deflator to adjust for national price changes over time.

Crude Birth Rate

The crude birth rate is the number of births per 1,000 population in a geographic area.

Current Population Survey (CPS)

The Current Population Survey, also commonly referred to as the household survey, is a sample-based monthly survey of about 60,000 eligible households. It provides a comprehensive body of data on the U.S. labor force by demographic and labor force characteristics.

Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC)

A widely used supplement to the Current Population Survey is the Annual Social and Economic Supplement (ASEC) conducted by the Census Bureau every February, March, and April. The supplement collects data on health insurance coverage, work experience, income from all sources, receipt of noncash benefits, poverty, migration, geographic mobility, and other special topics. The CPS ASEC also collects data on the number of children in paid child care and the expenditures of households and families using paid care. Use of the ASEC requires a tradeoff from monthly to annual data but provides a broader sample and larger universe than the basic CPS.

Educational Attainment

Educational attainment refers to the highest level of education that an individual has completed. Attainment is often measured using the number of years of education completed, especially when used to describe the average attainment across the population of a geographic region. Attainment is distinct from the level of schooling that an individual is attending currently.

Elasticity

Elasticity is an economic concept used to measure the percentage change of one economic variable in response to a change in another. The response is deemed elastic (or highly responsive) if the resulting change in a variable is more than proportional to the initial change and inelastic (or not highly responsive) if less than proportional.

Employment-Population Ratio

The employment-population ratio (or employment ratio) is a measure of labor force attachment that measures the share of the population activity employed. The ratio is calculated as the number of employed workers divided by the civilian noninstitutional population. The employment ratio does not consider unemployed workers as attached to the labor force. As a result, the employment ratio is far more volatile than the labor force participation rate across the economic cycle.

Family

A family is defined in the Current Population Survey (CPS) as a group of related individuals who are all members of the same household. Multiple families can be domiciled within the same household.

Female Labor Force Participation Rate

The female labor force participation rate measures the rate of participation of women in the labor force.

Fertility Rate

The fertility rate is the number of births per 1,000 women ages 15 to 44 in a geographic area.

Goods-Producing

Goods-producing sectors of the economy are those that produce products rather than services. These typically include NAICS sectors covering farming; forestry, fishing, and related activities; mining; and manufacturing.

Granger Causality

Granger causality is a statistical test of the usefulness of one variable in forecasting future values of another. Granger causality is present between two variables if future forecasts of variable X are improved by using variable Y in its prediction, above the level present when using only information about the history of X. Granger causality can be present in a single direction from either X to Y or Y to X, in both directions (bi-directional), or may not be present at all. If there is no Granger causal relation found from Y to X, Y is deemed strictly exogenous to X in providing useful forecasting information. Granger causality also differs greatly from measuring the correlation between two time series. Correlation simply measures the linear dependence between two series over a specified period. If X and Y are correlated, the calculated correlation is the same for both series in each direction in the period. Granger causality, however, measures statistical predictability in both directions and in the time dimension.

Great Recession

The Great Recession refers to the steep decline in economic activity associated with the U.S. recession lasting from December 2007 to June 2009, as well as downturns in national economies globally. It is the longest recession in the post-World War II period and generally considered the most significant economic downturn since the Great Depression.

Household

Survey data from the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC) are organized using samples of households or dwellings. A household is defined as all persons who occupy a single dwelling unit. A dwelling unit is a room or group of rooms intended for occupation as separate living quarters and having either a separate entrance or complete cooking facilities for the exclusive use of the occupants. In a small percentage of cases, multiple family units occupy a household.

Household Income

Household income represents all combined forms of gross income, both earned and unearned, for all members of a household ages 15 and over.

Hypothesis Testing

Hypothesis testing is the general process of testing whether statistical results are meaningful. Hypothesis tests are performed using standard errors to determine whether two estimates are significantly different. Hypothesis tests also provide evidence on whether your results happened by chance or will be repeatable.

Imputed Values

Imputed values are Census Bureau estimates of missing responses traced to two sources – nonresponding households and item nonresponse. Some households provide no responses at all while others do not respond to certain items in the survey. The Census Bureau notes that most of the imputations in the CPS are due to nonresponding households.

IPUMS-CPS

IPUMS provides census and survey data from around the world integrated across time and space. IPUMS integration and documentation makes it easy to study change, conduct comparative research, merge information across data types, and analyze individuals within family and community context. Data and services available free of charge. IPUMS CPS harmonizes microdata from the monthly U.S. labor force survey, the Current Population Survey (CPS), covering the period 1962 to the present. Data include demographic information, rich employment data, program participation and supplemental data on topics such as fertility, tobacco use, volunteer activities, voter registration, computer and internet use, food security, and more. IPUMS-CPS, University of Minnesota, www.ipums.org.

Labor Force

The labor force includes all persons in the civilian non-institutional population classified as either employed or unemployed. The labor force does not change as individuals move from employment to unemployment, and vice versa. The labor force changes only when new entrants enter the labor force or existing participants exit.

Labor Force Status

Labor force status measures the degree of labor force attachment for persons ages 15 years and older. Persons are generally classified as either in the labor force or not in the labor force. Those in the labor force are further classified as either employed or unemployed. Many persons are not in the labor force due to school, retirement, health, personal choice, and other factors. Members of the Armed Forces are excluded from most measures of work status.

Labor Force Attachment

Labor force attachment is a general economic term referring to a person's status as a participant in the labor force. Persons attached to the labor force include those either employed or unemployed. Those who are unattached do not participate in the labor force. The two most widely used measures of the degree of labor force attachment for the population of a geographic area are the labor force participation rate and the employment-population ratio.

Labor Force Participation Rate

The labor force participation rate is the most widely cited measure of labor force attachment and is calculated as persons in the labor force (either employed or unemployed) divided by population (civilian noninstitutional) ages 16 and over. In other words, it captures the percentage of the population of a geographic area that is either employed or unemployed and looking for work. The inclusion of the unemployed is the key characteristic of the participation rate versus the employment ratio, another popular measure of labor force attachment.

Maternal Labor Force Participation

Maternal labor force participation refers to the labor force participation of women with children.

Mean Reversion

Mean reversion refers to the tendency of observations in a data series to gradually move towards the long-term mean of the series. Large deviations away from the mean are expected to be followed by a return toward the mean in a mean reverting series.

Median Household Income

For households, the median income represents the level of household income where half the households in a geographic region (including those with no income) earn more and half earn less. Median household income is also referred to as the midpoint of the income distribution or the 50th percentile of household income.

Outlier

An outlier is an observation or data point that differs significantly from others in the same sample. Outliers can be due to measurement error or may simply reflect unusual and unexpected behavior among the observations in the sample. Outliers are sometimes excluded from the data set to gauge the sensitivity of any statistical findings to the presence of the outlier(s).

Paid Child Care

Paid child care is defined in the report as any form of child care arrangement for a child ages 0 to 14 for which a parent makes a direct expenditure on care to enable them to work. This follows the definition of paid child care used in the Current Population Survey (CPS). Paid options can include both formal and informal care arrangements such as neighbors or friends, which may or may not be regulated by states.

Panel Data

Panel data refers to data observations categorized for a given entity or data measure observed across time. Panel data is also known as longitudinal or cross-sectional time series data. Panel data used throughout the child care report is defined for multiple states (cross sections) and multiple time periods.

Panel Model

Panel model techniques are statistical tools and methods that simultaneously utilize the information contained in the economic behavior of an entity or data measure across time. Unit root, Granger causality, and cointegration tests used throughout the child care report are all panel modeling techniques that use a 50-state panel dataset. The use of a panel of states rather than national data can provide for more robust estimates of the fundamental factors driving paid care usage.

Personal Income

Personal income includes all forms of income that persons receive in return for their provision of labor, land, and capital used in current production and the net current transfer payments that they receive from business and from government.

Per Capita Income

Per capita income measures the amount of income earned per person in a geographic region. Per capita income is commonly used as a measure of standard of living of the population in a region.

Population Control Adjustments

The Census Bureau makes annual changes to historical CPS population data and sample weights to include the latest information available about population change and to incorporate improvements in the estimation methodology. BLS incorporates these annual population control adjustments into the CPS estimates each January. The adjustments are based on the latest decennial census population count, supplemented with birth and death data and estimates of net international migration. The key implication of annual population control adjustments for users is that the process can render historical comparability of estimates from the CPS ASEC invalid.

Prime Working Age Women (ages 25-54)

Women of prime working age are those ages 25 to 54 who actively participate in the labor force. These women are more likely to participate in the labor force than younger and older women and have likely completed pre-career education and training.

Probability Value (p-value)

A probability value, or p-value, is a statistical parameter used within hypothesis testing that determines the probability of obtaining the observed results assuming a given probability distribution of the test statistic and that the null hypothesis is true. In other words, it is the predetermined level of probability at which statistical significance is found. A p-value of 0.05 (5 percent) or lower is typically considered the threshold of statistical significance.

Public Preschool Education

Public preschool includes a range of publicly funded early childhood education programs accessed by children before they begin compulsory education at the primary school level. Public pre-kindergarten (or Pre-K) programs are commonly available to children ages 4 to 5 in many states (i.e., 5-year-old children not yet enrolled in public kindergarten). In some states, public preschool also serves 3-year-old children. Publicly funded preschool could be located in a school or in a mixed delivery setting such as child care centers and family child care homes depending upon state or local school district decisions.

Quartile

A quartile is a statistical tool used for summarizing data by dividing the observations into four groups that are more-or-less of equal size. Data is often ranked along some measure of the value of the underlying data and then assigned to quartiles. As with other forms of quantiles (e.g., terciles, quintiles, deciles, etc.), quartiles provide a convenient means of comparing data across grouped intervals.

Real Personal Income

Personal income calculated at its nominal, or current, value and then adjusted for the effects of inflation over time is deemed real personal income. At the state level, an additional adjustment is made to nominal personal income to reflect state-level differences in cost-of-living when calculating real personal income. The cost-of-living adjustments are made using Regional Price Parity (RPP) indexes developed by the Bureau of Economic Analysis.

Replicate Weights

A common method used to create standard errors for use with statistical estimates from the CPS is to generate empirically derived estimates of the standard error. Instead of using a single weight to produce a single point estimate, a large number of potential (replicate) weights that are representative of the population are used. In the CPS, 160 separate weights at the household and person level are available for use in calculating standard errors. The process of producing the estimate is replicated using each of the 160 weights and the results are used to calculate a statistical estimate of the standard error.

Sample Size

Sample size refers to the number of individual observations in a sample of data.

Sampling (Nonsampling) Error

Sampling error occurs when an estimate derived from a sample rather than the entire population differs from the true population value it represents. Nonsampling errors are other errors not related to the sample selected. Typical sources of nonsampling error include poor sample design, data entry errors, nonresponse rates, missing entries, and survey implementation errors. The implications of nonsampling errors in the CPS are generally unknown.

Service-Providing

The service-providing sectors of the economy produce intangible services instead of goods. A range of services are produced by both private and public sector entities under the NAICS classification system.

Standard Error

The degree of sampling error in a statistical estimate is often measured using the standard error of the estimate, a widely used measure of variability. A standard error can be used to construct confidence intervals for the estimates if they are unbiased, typically at the 90% or 95% level. Standard errors can also be used in hypothesis testing to determine whether two estimates are significantly different.

Stationarity

A stationary data series will have a mean, variance, and autocorrelation structure that is stable over time. Visually, stationary series tend to be mean reverting and do not trend strongly upward or downward. They also do not have periodic patterns such as seasonality. A non-stationary data series, or one with a unit root, may have to be differenced one or more times to achieve stationarity. The level of integration, denoted as $I(i)$, is used to describe the number of times (i) a data series must be differenced to achieve stationarity.

Statistical Causality

The statistical notion of causality tests for the increased predictability of the future path of one variable, X, using another variable, Y. While not addressing the issue of economic causality in the traditional sense, tests of statistical causality provide an empirical measure of the historical responses and timing embodied in the relationships among data series. Granger causality is a common approach to testing for statistical causality. Because economic causality also operates in the time dimension, economic causal relations are often informed using Granger-type methods, particularly in forecasting applications.

Statistical Significance

In statistical hypothesis testing, a result is statistically significant if it is deemed unlikely to have occurred due to chance given the stated hypothesis tested. Statistical significance is usually determined by rejection of the null hypothesis.

Subsidies and Cost Offsets

Several federal and state subsidies, tax credits, and other forms of cost offsets are available to assist families in meeting the cost of paid child care. Subsidies and offsets examined in the report include those provided through the Child Care and Development Fund (CCDF), the Temporary Assistance for Needy Families (TANF) block grant, and the Child and Dependent Care Tax Credit (CDCTC).

Supplemental Poverty Measure (SPM) Program

The Supplemental Poverty Measure program is an effort by the Census Bureau to extend the official U.S. poverty measure by taking account of many of the government programs designed to assist low-income families and individuals that are not included in the official poverty measure. The new SPM poverty measure also considers work-related expenses, one of which is the cost of paid child care to enable a parent to work.

Time Series Analysis

Time series analysis describes a group of statistical techniques and methods for analyzing time series data to extract meaningful characteristics of the data. These techniques are used most often to examine relationships present between variables over different points in time. The Granger causality and cointegration tests used in the reports are widely used methods of time series analysis.

Time Series Data

Time series data is a collection or sequence of data observations collected over time intervals. Time series data is commonly collected on an hourly, daily, weekly, monthly, quarterly, or annual time interval and indexed in time order.

Todo-Yamamoto Method (TY Method)

All Granger causality tests in the child care report are implemented within a VAR model framework using the method of Todo and Yamamoto (1995). The TY method allows for causality testing among a group of data series within a system framework. The system includes an equation for each data series with the series as the dependent variable and the remaining variables as explanatory (independent or right-hand side or) variables. The TY method is noteworthy in that it is robust to the presence of unit roots, or the order of integration of the time series. The base VAR used in the tests is augmented, or overfit, by including an additional lag of the level of each variable as an additional exogenous variable in each equation of the VAR.

Top-Coding

Top-coding is a traditional disclosure avoidance method used by the Census Bureau to prevent the identification of individuals. Top-coding is a well-known problem with the tabulation of income in the CPS. Some categories of income are 'top-coded' at higher levels rather than reported using the full dollar amount. The practical result of top-coding is an understatement of true mean earnings in the U.S. when using CPS data.

Traded Activity (or Openness)

Traded activity is defined as production for trade outside a region, or a region's degree of openness. The concept traces its origins to the notion of enhancing the 'basic' industries located within a region. Basic industries produce goods and services that are exported for sale outside the local market. This includes trade with other states as well as internationally. States with large manufacturing, mining, and Federal government sectors (including military) tend to have the most traded activity with outside regions. Traded activity captures spending from outside the region which in turn helps support the development of the region's 'non-basic' sectors. Non-basic industries are believed to merely recirculate existing purchasing power, which exerts less influence on overall regional growth than an equivalent injection of spending from outside the region.

Unit Root Test

Unit root tests are used to test the stationarity of a data series and establish its degree of integration. $I(0)$ variables are stationary in levels (no unit root) and require no differencing, while $I(1)$ variables have a unit root and must be differenced once to achieve stationarity. An $I(2)$ series is one that must be differenced twice to achieve stationarity. Most nonstationary series are $I(1)$ and become stationary after differencing once. Few data series require differencing twice (or more) to achieve stationarity.

Unpaid Child Care

Some families may use unpaid child care, which reflects time children spend out-of-the-home. However, for purposes of this report, only the use of paid child care was reviewed. The report compares the average income of families with children age 14 and younger that use paid child care compared to families with children of the same age that do not use paid child care. The same analysis is also included for families with children under age 5 that use paid care compared to families with children under age 5 that do not use paid care.

Vector Autoregressive (VAR) Model

Vector autoregressive models are a time series technique used to investigate the relationships among a group of time series variables. The estimated model includes an equation for each data series with the series as the dependent variable and the remaining variables as explanatory (independent or right-hand side or) variables. Each equation in the VAR model includes only each variable's lagged (or past) values, lagged values of the other variables in the model, and an error term. The VAR model imposes no structural assumptions on the data but instead treats all data in the model as endogenous to the system.

Women of Working and Childbearing Age (ages 18-54)

The population of women ages 18 to 54 are of both working age and childbearing age. These women are the most likely to use paid child care services for children ages 0 to 14. This measure captures a broader group of women than prime working age women (ages 25-54) by including younger women ages 18-24 who are typically of childbearing age.

Abbreviations and Acronyms

ACS	American Community Survey
BLS	Bureau of Labor Statistics
BEA	Bureau of Economic Analysis
CCDF	Child Care and Development Fund
CDCTC	Child and Dependent Care Tax Credit
CED	Committee for Economic Development of The Conference Board
CPS	Current Population Survey
CPS ASEC	Current Population Survey, Annual Social and Economic Supplement
DBMS	Database Management System
ECPP	Education Early Childhood Program Participation
HHS	U.S. Department of Health and Human Services
IPUMS-CPS	IPUMS-Current Population Survey
IRS	Internal Revenue Service
NAICS	North American Industrial Classification System
NHES	National Household Education Surveys
NIEER	National Institute for Early Education Research
RPP	Regional Price Parity
SIPP	Survey on Income and Program Participation
SPM	Supplemental Poverty Measure
SQL	Structured Query Language
TANF	Temporary Assistance for Needy Families
VAR	Vector Autoregressive Model

