

Chapter 7, "Twenty-Four Hours: An Overview of the Recall Diary Method and Data Quality in the American Time Use Survey,"

by Polly A. Phipps and Margaret K. Vernon

**from *Calendar and Time Diary: Methods in Life Course Research*,
edited by Robert F. Belli, Frank P. Stafford, and Duane F. Alwin
(SAGE Publications, Inc., 2009)**

Chapter 7, "Twenty-Four Hours: An Overview of the Recall Diary Method and Data Quality in the American Time Use Survey," by Polly A. Phipps and Margaret K. Vernon, appears in the book, *Calendar and Time Diary: Methods in Life Course Research*, which was published by SAGE Publications, Inc. Chapter 7 is in the public domain. You are free to use public domain material without specific permission. All other sections of this book are protected by copyright.

7

Twenty-Four Hours

An Overview of the Recall Diary
Method and Data Quality in
the American Time Use Survey

Polly A. Phipps and Margaret K. Vernon

Introduction

This chapter focuses on issues regarding the methodology and data quality inherent to the American Time Use Survey (ATUS). The ATUS is the first federally funded, ongoing, and nationally representative time use survey ever undertaken in the United States. The survey is sponsored by the U.S. Bureau of Labor Statistics (BLS), and data are collected by the U.S. Census Bureau. Since 2003, annual estimates of the amounts of time people spend in activities such as paid work, child care, volunteering, and socializing have been available and are posted for public consumption on the BLS Web site, along with public use microdata files.¹

We first describe the development process that led to the ATUS, outline analytical uses of the data, and provide an overview of the ATUS methodology. Second, we focus on current methodological issues that affect data quality, including measurement error associated with the diary instrument, computer-assisted telephone interviewing, and lack of secondary activities (activities

AUTHORS' NOTE: The views expressed here are those of the authors and do not necessarily represent the policies of the Bureau of Labor Statistics.

done while doing something else); and nonresponse error associated with the sampling design and response burden. While time diaries are considered a tried-and-true method, there is ample literature documenting survey error and trade-offs in different designs (National Academy of Sciences, 2000; Robinson, 1985, 1999). Also, there are methodological issues that are relatively unexplored, such as understanding respondents' cognitive processes when completing time diaries. We focus on both measurement and nonresponse error, describe current research that has assessed ATUS data quality, and suggest future research that could benefit and perhaps improve the ATUS.

ATUS Development Process

The United States has a long history of collecting data on time use, ranging from time-budget clocks collected by the Department of Agriculture in the 1920s and 1930s to measure the time use of homemakers (Stinson, 1999), to a rich history of time use diary surveys carried out at the University of Michigan beginning in the 1960s and at the University of Maryland beginning in the 1980s (Juster & Stafford, 1985; Robinson, 1999). In many countries, however, central statistical agencies conduct recurring time diary studies of their population: examples are Australia, Canada, Finland, Germany, Japan, Korea, New Zealand, the Netherlands, Norway, and Sweden, among others (Harvey & Pentland, 1999).

The BLS first began to consider the idea of carrying out a time use survey in 1991 when Congress introduced a bill that called for the Bureau to conduct surveys to measure the amount of time spent on "unremunerated" or unpaid work in the United States and to place a monetary value on such work.² While exploring the idea of conducting a time use survey, the BLS began to realize the extent of federally sponsored time use surveys in other countries, and that the international statistics community considered a U.S. time use survey an important statistical effort.

Launching a new federal survey is a major undertaking, requiring extensive design, testing, and funding approval (see Horrigan & Herz, 2004, for a detailed review of the ATUS development process). After initial design and development research, BLS undertook a time use feasibility test in 1997, presenting the results at a conference cosponsored by BLS and the MacArthur Network on Family and the Economy. In 1998, BLS put together a detailed plan for collecting time use data, forming the foundation for the ATUS methodology and funding proposal. The BLS simultaneously attended the 1999 National Academies workshop on the feasibility of collecting time use data, which helped to narrow and refine both methodological issues and analytical goals (National Academy of Sciences, 2000). In December of 2000 the survey received

approval and funding, and a BLS–Census Bureau team was formed to address management oversight, sampling and weighting, computerized instrument development, activity coding, and operations. Between December 2000 and January 2003, when data collection began, BLS undertook further pretesting, a field test, a dress rehearsal, and a prefielding period to resolve remaining problems.

Analytical Uses

While the ATUS was first considered as a way to measure and place a value on unremunerated work, many other analytical uses have since been brought to BLS's attention (Frazis & Stewart, 2004, 2007; Hamermesh, Frazis, & Stewart, 2005; Horrigan & Herz, 2004; Joyce & Stewart, 1999). Time diary data can be used across a wide range of disciplines to look at hours spent in paid work and productive nonmarket work, such as housework and child care. Work is just one of a number of major activities captured by time diaries; others include education, consumer purchases and services, eating and drinking, socializing and leisure, sports and exercise, personal care, religious participation, civic obligations, volunteering, and travel.

Time use data can be used to analyze the distribution and timing of activities over the course of a day, such as sleep, work, household, consumption, and leisure activities. Figure 7.1 shows that work activities claim most of the morning and afternoon hours of employed persons during the workweek, with household and consumer activities peaking in the early evening, and leisure and sports activities continuing into the late evening. Further detail on leisure time is displayed in Figure 7.2, showing that on average, a little over 5 hours a day is spent in leisure time, with television watching making up the greatest share of leisure hours.

Many researchers are interested in comparing time use between different groups, such as time spent in child care between men and women. Time use information that is collected on an ongoing basis, as the ATUS data are, allows for across-time comparisons. For example, the data might be used to assess changes in commuting time over several years. Researchers are interested in using time use data as a measure of quality of life, for example, looking at free time or leisure time, as well as time spent with friends and family, as indicators of well-being. In addition, sleep researchers use the data to assess amount of time spent sleeping and when sleeping takes place. Time use data can be used to estimate the amount of time allocated to learning, such as time spent in class or in educational pursuits. There are also diverse users of time use data ranging from academia and government to business and legal communities. For example, both health and market researchers are interested in finding out how

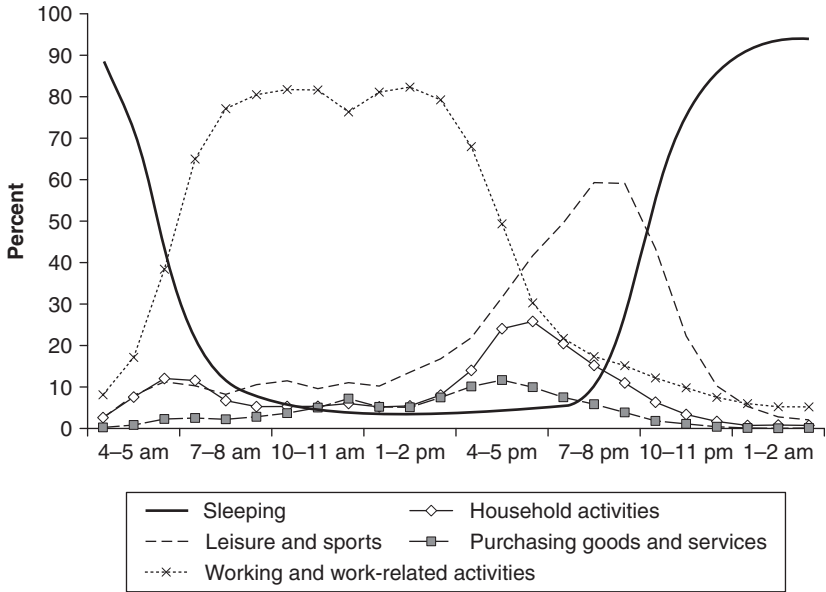


Figure 7.1 Percentage of Employed Persons Doing Selected Activities on Workdays by Hour of the Day

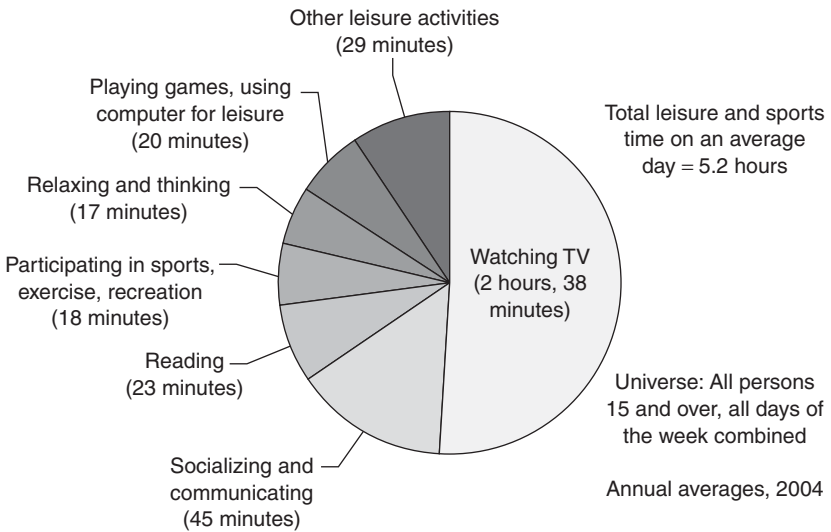


Figure 7.2 Average Hours per Day, Leisure and Sports Activities

much time people spend eating, and where eating takes place. The opportunity to analyze where activities take place in the latter example illustrates one of the unique features of time diary data, the collection of the activity location. Finally, lawyers are interested in using time use data to estimate the economic damages in personal injury and wrongful death cases. These are a few examples of how the ATUS data are being used, and there are almost certainly many more uses.³

Overview of ATUS Methodology

Data collection for the ATUS occurs continuously on nearly every day of the calendar year. The sample size per year is approximately 24,000. The sample is nationally representative, and includes an oversample of Hispanic and black households, as well as households with children, in order to improve the reliability of the estimates for those groups. One person is randomly selected from each sample household to participate in the ATUS; selected persons are also assigned a specific day of the week about which to report. Response rates averaged 57% in 2005,⁴ resulting in approximately 13,000 completed interviews. The ATUS interview is a computer-assisted telephone interview (CATI) that takes, on average, about 20 minutes to complete.⁵ During the interview, the preselected household member reports about his or her time use during the 24-hour period of the day before the interview. Interviewers use a set of scripted open-ended questions in tandem with conversational interviewing techniques⁶ to walk respondents chronologically through the 24-hour day, collecting information about time spent in activities beginning at 4:00 a.m. on the previous day up until 4:00 a.m. of the interview day. The interviewer uses precoded categories to record commonly reported activities, and for other activities, he or she types the verbatim responses in the activity lines of the diary grid, as seen in Figure 7.3. For each activity reported, the interviewer asks how long the activity took, recorded either as a duration or with start and stop times. Respondents are also asked questions about who was in the room with them (if at their own or another's home) or who accompanied them (if traveling or away from home) and where each activity took place, which are recorded using precoded categories.⁷ There are some exceptions to the "who" and "where" rules: "who" and "where" information is not collected for certain personal activities (such as sleeping or grooming), nor is "who" information collected when respondents report working or attending high school. After completing this sequence of questions on an activity, interviewers prompt respondents for the next activity by asking them what they did next.

Although the 24-hour recall diary is a fairly standard way to collect time use data, there are methodological challenges, including some that are unique

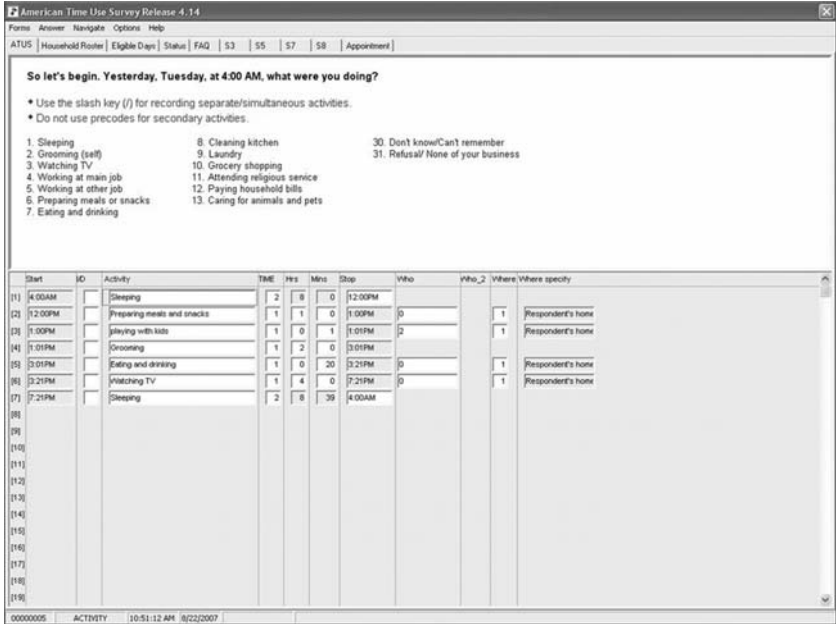


Figure 7.3 ATUS Computer-Assisted Telephone Interview, First Diary Screen

to the ATUS. We highlight some of these challenging issues, such as instrument design, computer-assisted telephone interviewing technology, and sample design, that impact the data quality of time use estimates.

Instrument Design

Survey cost and data quality considerations led BLS to decide upon a 24-hour recall diary to be used in tandem with computer-assisted telephone interviewing. In North America, time use information has been collected extensively through the use of the 24-hour recall diary method. The 24-hour recall instrument was used in 1975–1976 and 1981–1982 at the University of Michigan; in 1985, 1992–1994, 1995, and 1997–1998 at the University of Maryland; and in 1992 and 1998 at Statistics Canada. While there is some consensus that a personal interview combined with a leave-behind paper diary covering more than one day is the “Mercedes” of time diary survey administration (National Academy of Sciences, 2000), the 24-hour recall diary has been shown to

produce good data with much lower costs (approximately 3 to 1; Juster, 1986). Robinson's (1985) studies of reliability and validity of the 24-hour recall diary, conducted in the 1970s, found no systematic bias associated with the method when comparing it to paper diaries or activities respondents reported during a random hour of their diary day or when a pager was activated; nor did Juster (1986) in a comparison of husband and wife diaries. However, Robinson identified several problems, such as under- and overreporting, including the underreporting of activities with short time frames.

In the 1990s, the use of cognitive interviews became commonplace to assess respondents' comprehension of questions and ability to recall, judge, and communicate answers (Willis, 2004). While the 24-hour recall diary allows the respondent to describe his or her activities chronologically in an open-ended manner, little research had been conducted on the cognitive processes that respondents use to reconstruct their prior day. BLS and Westat researchers planning tests of the diary instrument were concerned about possible variation in data quality associated with differences in how respondents interpret "activities," the effects of highly salient activities, routine activities, and schemata (or everyday activities) on recall; and heuristic or rule-of-thumb strategies for estimating activity duration (Forsythe, 1997b).

Cognitive interviewing was carried out as part of the 1997 feasibility test to see how respondents interpreted "activities," and what strategies were used to recall events. Findings from the 33 cognitive interviews indicated that participants recalled activities and time durations more easily when their activities and times were atypical and/or when they had an appointment or time commitment. Some participants used a visualization technique as part of the recall process, while others used a decomposition strategy—recalling a larger block of time or major activity to set time boundaries, then recalling what happened during the time period. While participants who lived by the clock reported by the clock, others had difficulty identifying starting and stopping times, and reported time durations instead. To assist comprehension and recall, a checklist of common activities was tested, but was found less successful than general probes such as simply asking if participants did anything else during the time period. In the actual feasibility field test, a standard 24-hour recall instrument with no probes for additional activities was tested against an "enhanced" recall instrument. The enhanced instrument included probes asking respondents whether they did something else during the time period, and if so, whether they stopped their initial activity to do the additional activity. The average number of activities and time accounted for were higher under the enhanced version, particularly for nonmarket work (Forsythe, 1997a), as the enhanced recall instrument elicited secondary activities (activities done while doing something else). However, concern about respondent burden, and the complexity involved in programming the computer software to collect

additional activity time durations ultimately led to a decision not to attempt to collect secondary activities (Horrigan & Herz, 2004). Since the feasibility test, BLS has not conducted further research on respondent comprehension, recall, judgment, and communication associated with the questions used in the 24-hour recall diary. The computerized instrument, however, as well as interviewer instructions, has undergone numerous revisions since the ATUS has been in the field, based on the interviewer feedback about respondents' reactions and concerns that BLS obtains during regularly scheduled focus groups, debriefings, and refresher training sessions. Nonetheless, further research on understanding respondents' cognitive processes as they complete the 24-hour recall diary questions could aid in developing additional instrument and interviewer interventions to assist respondents in their task and further reduce measurement error.

Computer-Assisted Telephone Interviews

Compared to a paper diary, using a computer-assisted telephone instrument (CATI) as the mode of data collection and having interviewers verbally walk the respondent through his or her 24-hour time diary presents both data quality advantages and disadvantages. A major advantage of 24-hour telephone recall using a CATI system is consistency and control in how data are collected. CATI allows for interviewer monitoring to make certain that interviews are being conducted in a uniform manner. The survey task is the same, presenting the same challenges, for every respondent. When a paper diary is issued and respondents are asked to fill it out, there is more uncertainty about how and when respondents might choose to fill out the diary. Some respondents might fill out the diary as they go through their day while others might choose to fill it out at the end of the day, or even several days later, depending on the survey design.

While the 24-hour recall method utilizing CATI allows for more control, recalling activities that one did the previous day could be a more challenging task than filling out a diary as one goes through the day. Robinson's (1985) finding that leave-behind paper diaries elicited about 5% to 10% more activities than recall interviews could suggest that recalling activities that one did the day before and cataloging the activities orally is a more cognitively difficult task than recording activities in writing. Alternatively, it has been documented that within a survey, telephone interviews often take less time than other modes of administration, such as face-to-face interviews (McGrath, 2005). This could be because there is less chitchat on the phone compared with an in-person interview or because people simply don't want to stay on the phone as long as they would talk with an in-person interviewer or take to fill out a paper diary. Based on these findings, it might be expected that there would be fewer activities

reported in a telephone diary task compared with an alternative mode. To examine this assumption, the ATUS conducted an experiment to assess whether having respondents fill out a paper diary in advance of the telephone interview would improve data quality, such as the number of activities collected, variety of activities⁸ reported, and time estimates. The results of the test showed that there were no differences in the mean number of activities or types of activities reported or in the reporting of rounded time estimates between people who completed and did not complete an advance diary (National Opinion Research Center, 2002).

Another major advantage of using a telephone interview is that interviewers can probe for and record the level of detail of activities needed to code the data properly. ATUS has a set of 15 precoded commonly reported activities that interviewers select from (shown in Figure 7.3), and they record actual activities verbatim when the activity does not clearly fit into a precoded category. The ATUS uses a three-tiered activity classification system that allows for very detailed coding of diary activities (Shelley, 2005). ATUS interviewers are trained in both interviewing and coding, so they know both how to record activities to make the coding task more reliable and when to probe for more detail, such as when a respondent reports an ambiguous or vague activity. This helps data quality in several ways. First, if interviewers were not present to probe for the appropriate level of detail, there would be higher levels of incorrectly coded or uncodable activity data in the diary. For example, in the ATUS coding scheme, grocery shopping is differentiated from other types of shopping. If interviewers were not present to find out what type of shopping took place, respondents might provide the diary entry “shopping,” which would be coded as “shopping, not otherwise classified,” providing less precise data for analysis purposes. In addition to obtaining more accurate and codable data, using trained interviewers allows for more consistency and reliability in how data are both collected and coded. For example, trained interviewers prompt and probe for level of detail in a more consistent way than data would be gathered if each respondent personally wrote his or her own time diary. In addition to recording data, all interviewers code interviews. Interviewers do not code diary data they themselves collect, but code other interviewers’ diaries. In fact, each completed diary is coded separately by two different interviewers. If there is any inconsistency in how an activity is coded, the diary goes to an adjudicator who resolves the difference between the two codes. Thus, reliability checks are built in to make sure that the interviewers are collecting data consistently and with the appropriate level of detail. Having all interviewers trained and participating in both data collection and coding most probably leads to greater accuracy, consistency, and reliability in the data. Future research, such as an observational study of interviewer practices and coding reliability research, would help to verify the potential advantages of the ATUS operational methods.

In addition to interviewers being trained in recording and coding data, they are also trained in conversational interviewing techniques, which allows for checks to be built into the interview. Interviewers are trained to insert conversational data checks in several ways. First, though interviewers move forward in time through the instrument, they are trained to check back about timing of activities. For example, after a respondent says that he or she ate dinner for an hour, the interviewer might check back, "That brings us up to 7:00 p.m., does that sound about right?" This type of time check helps respondents to stay "on track" when remembering their day's activities and double checks with respondents on the sequence and timing of activities. The CATI instrument is flexible and accommodates changes, additions, and deletions of activities and times.

In addition to time checks, interviewers are trained to confirm an activity or activity length at any time in the interview if a respondent says something that doesn't make sense. There are also required checks built into the instrument asking about work breaks; activities other than work, school, or sleep lasting longer than 2 hours; and sleep lasting more than 10 hours. Often these checks reframe activities for respondents, helping them to catch mistakes in reporting and provide logical and coherent time data.

Finally, interviewers are trained to use one of several techniques to facilitate recall when respondents have difficulty remembering what activity they did next. One such technique is called working backwards. If a respondent can't remember what he or she did right after lunch, interviewers will ask, "Alright, what is the next thing you remember doing (and what time did that take place)?" Interviewers will then prompt respondents and try to jog their memories by saying, "You finished eating lunch at 1:00 and you can remember sitting down to watch Oprah at 4:00. Can you think of what you did in between?" This working backwards technique will often help respondents to fill in the gaps. Interviewers are also trained in visualization techniques. For example, they might ask respondents to picture where they were in order to facilitate recall. Following with the previous example, an interviewer might say, "You said that you went into the family room to watch Oprah at 4:00. Can you picture where you were right before you walked into the family room?" Although visualization is less commonly used by interviewers than working backwards, this technique does help respondents to remember forgotten activities. ATUS regularly conducts interviewer debriefings to determine how conversational interviewing and recall techniques are used and to develop refresher training and/or revise interviewer instructions accordingly. Further research to evaluate how conversational interviewing and specific recall techniques are used by interviewers and whether the techniques are successful in helping respondents reconstruct their day would provide important information on ATUS procedures and data quality.

Secondary Activities

Closely related to the telephone collection method, the automated instrument used to collect and synthesize data over the phone also plays a role in data quality. A drawback to using the current CATI instrument is that it cannot be configured to collect the duration of secondary or simultaneous activities (Horrigan & Herz, 2004). While the activity itself can be collected, time durations are programmed to correspond to a 24-hour time period, with no overlapping times. In contrast, time use studies that use leave-behind paper diaries and allow people to record primary and secondary activities find that people are often engaged in two or more activities at any given time (Robinson, 1985). If an ATUS respondent reports that he or she was working on the computer and listening to the radio, interviewers are trained to ask, "Which was your main activity?" Both activities are recorded (with the primary/main activity recorded first), but only the main activity is subsequently coded and output in the data, so the information about the secondary activity is lost. In other cases information on secondary activities is never captured at all, as the respondent is unlikely to offer the information, and there is no interviewer probe for additional activities. This may affect data quality for several reasons. First, without the systematic collection and coding of secondary or even tertiary activities, it is impossible to capture the full complexity and diversity of activities that humans engage in during a 24-hour day. When only primary activities are captured, 24-hours' worth of time is accounted for, but the *number* of activities that a person engages in is underestimated. Thus ATUS data do not provide a complete picture of the total number and different types of activities people engage in throughout the day, or of how people combine different activities at any given time. One such area of research that could be very interesting, for example, would be to learn more about how people engage themselves on their commutes to work. Some people might use this time to catch up on phone calls, to read or listen to music, or even sleep or eat breakfast.

The absence of secondary activities is less problematic for estimates produced for activities that are always done and coded as main or primary activities. For example, travel is always coded as a primary activity in the ATUS. Therefore, the travel that people engage in and the amounts of time in a 24-hour day that people spend traveling are well captured by the ATUS. The ATUS estimate for time spent in travel has been shown to be a valid measure. Bose and Sharp (2005) found that estimates for time spent in travel produced by the ATUS and the Bureau of Transportation Statistics are very similar. Work is another activity usually coded as a main activity. Frazis and Stewart (2004) compared estimates of respondents' hours worked from ATUS time diaries to Current Population Survey (CPS) estimates and found that they were quite

similar, when usual hours were controlled for, and the respondent's diary day falls inside the CPS reference week.

While certain types of activities, such as travel and work, that are done as primary activities are well captured by the ATUS, activities that are done frequently as secondary activities are not. One such activity we highlight that is of importance to the Bureau of Labor Statistics is child care. As was mentioned in the introduction, the ATUS was originally conceived as a way to place a value on nonmarket work. Accurately capturing time spent in child care is essential to this task. Yet much of the child care activity that caregivers engage in is often done as a secondary activity. Parents, guardians, and caregivers are likely to "keep an eye" on children and be responsible for them while engaging in primary activities such as traveling for errands, cleaning around the house, and even while doing personal care activities such as showering or dressing.

In order to improve estimates of the total amount of time spent in secondary child care, the ATUS cognitively tested questions measuring supervision of children and added a summary question that follows the main time diary section of the interview (Schwartz, 2001). The summary question asks, "During any part of the day yesterday was a child who is 12 years old or younger in your care?"⁹ If the response is yes, the interviewer checks a box next to the relevant primary activity to indicate that a child was being cared for, capturing passive secondary child care that occurred while the respondent was doing something else as a main activity. In 2003, this secondary child care question increased estimates of total time spent in child care by approximately 1.75 hours per day for the population.¹⁰

The secondary child care summary question was not designed to capture active secondary child care. Because the ATUS does not collect information on active secondary child care, the survey cannot provide information on specific active child care that takes place as a secondary activity. For example, reading and talking with children is of critical importance to children's academic and social development. If talking with one's child took place as a secondary activity (which one might assume would happen quite often), and was captured as secondary care, analysts would not be able to break apart how much time parents are spending doing this activity with children. Furthermore, Fisher (2005) suggested that time spent reading with children often occurs during travel; for example, reading in the car, on the train, or on the bus. Since travel is always coded as a primary activity in the ATUS, this reading activity would be captured as "care" and would not be separated from other types of more passive care such as "keeping an eye on" a child. These examples make the point that while the summary question method does increase time estimates of child care, the rough level of care captured does not allow for more refined analyses of the type of care being provided. Given the complexity of collecting child and parental activities and time, other studies have taken an alternative measurement

approach. For example, the Child Development Supplement of the Panel Study of Income Dynamics at the University of Michigan collected time diary data from children and/or their caregivers to produce estimates of time children spend interacting with caregivers (Stafford & Yeung, 2005).

Bianchi, Wight, and Raley (2005) compared secondary child care estimates from the ATUS (which uses the summary question measure) to the 2000 National Survey of Parents (NSP), which uses a probe asking, "What else were you doing?" to collect secondary in addition to primary activities. They found that there are large differences in time estimates of secondary care using these two different methods. For example, they report that the 2000 NSP picked up about 0.4 hours per day of secondary child care for fathers, while the 2003 ATUS picked up about 4.5 hours per day for fathers. As already noted, however, the ATUS question wording for child care is broader than typical definitions. The "in your care" language is designed to pick up on indirect or passive child care—or times when care or assistance could have been given—rather than more active types of care. It seems that when people report secondary activities using the "what else were you doing" probe they typically think of more active types of care. Recent work by Allard, Bianchi, Stewart, and Wight (2006) confirms this distinction. They compare the ATUS and NSP on three measures of child care and find that the surveys produce similar estimates of primary child care and time spent with children, the latter a broad measure calculated using the "who were you with" question asked of ATUS and NSP respondents for each primary activity. To measure secondary child care, the authors compare time estimates from the NSP secondary activity question "what else were you doing" to the ATUS "in your care" summary question, excluding primary child care, and times when the respondent and all children are asleep. Their results and conclusions indicate that the surveys are very different in regard to secondary child care. The ATUS is more consistent with a passive-care concept, capturing 5.8 hours per day of secondary child care compared to eight-tenths of one hour per day in the NSP, a measure more consistent with an activity-based concept.

While the ATUS has made substantial efforts to capture time spent in secondary child care, there may be other activities of interest that are being under-represented because they are often done as secondary activities. For example, the Economic Research Service sponsored new questions that were added to the ATUS in October of 2005 that capture secondary eating and drinking, or eating and drinking that takes place while the respondent is doing something else. Preliminary data suggest that the questions designed to capture secondary eating and drinking are increasing the number of episodes of eating and drinking reported, and increasing estimates for the amount of time spent eating and drinking. Other activities, such as watching TV, listening to the radio, talking on the telephone, talking with a spouse or other family members, and cuddling

family pets, also may be underrepresented. Thus, the operational limitation of using a CATI instrument that does not capture secondary activities impacts data quality in two ways. First, people often engage in more than one activity at any given time, so the instrument fails to capture the full complexity of human behavior. Second, certain *types* of activities that often take place as secondary activities are undercounted and underrepresented more than other types of activities, affecting the frequency counts and time estimates of such activities.

Sample Design Methodology

The ATUS uses a household population-based sample to provide estimates on how the U.S. population uses its time. Respondents are age 15 and older. The ATUS sample is selected from households that have completed their eighth and final Current Population Survey (CPS) interview. Households become eligible for selection into the ATUS sample 2 months after their final CPS interview. The CPS is a BLS-sponsored monthly panel survey that collects information about employment, unemployment, and household composition. Drawing the sample from CPS households is a very efficient way to select a scientific sample because the demographic, employment, and contact information of respondents is already known. It also decreases respondent burden because the ATUS can update household composition and selected demographic and employment information rather than collect it for the first time. In addition, the use of the CPS as a sampling frame reduces sampling costs. While drawing the ATUS sample from CPS households does decrease respondent burden and allows for an efficiently selected sample, this method is not without its liabilities.

RESPONSE RATE AND BURDEN

First, to be selected into the ATUS sample, the household must have completed the final CPS interview (wave 8). Wave 8 on average has the highest response rate of the eight CPS panels: 93.2% in 2004. However, the approximately 6.8% of sample members who refuse the CPS interview or are not able to be contacted have no chance of being selected for the ATUS sample. This could be a source of error, including bias. For example, Dixon and Tucker (2000) found multiunit housing structures; some demographic groups, including males and persons of black race; people enrolled in school; and those with higher than usual hours worked were associated with higher CPS nonresponse. How this might affect the ATUS sample has not yet been examined.

Second, the ATUS has a lower than desired response rate, ranging from an average of 56% to 58% per year over the first 3 years of production.¹¹ Of the 44% nonresponse in 2003,¹² the types of nonresponse outcomes, from highest to lowest are: refusals (19.3%); unknown eligibility due to an incorrect or

unconfirmed telephone number (9.9%); inability to contact respondents on the designated day (7.3%); and other noninterviews, including language barriers and health issues (7.2%; O'Neill & Dixon, 2005).¹³ In 2004, the ATUS recontacted a subset of respondents and nonrespondents to discuss why they did or did not choose to participate in the ATUS (O'Neill & Sincavage, 2004). Of this subset, approximately 32% of nonrespondents and 81% of respondents participated in the recontact survey. The main reason nonrespondents gave for not participating in the ATUS was their previous CPS participation. Both nonrespondents and respondents reported that they felt that they were doing more than their "fair share" and that they were tired of the Census Bureau calling them. ATUS nonrespondents also stated several other reasons for not participating, however; survey fatigue was by far the most often stated reason. Thus, drawing the sample from CPS households increases sample efficiency and decreases survey costs, including a reduced interview length, but has a tendency to overburden and fatigue some persons who are selected to participate in both the CPS and ATUS.

Other features of the sampling design also contribute to nonresponse. In particular, the ATUS randomly selects one person from a household and presselects a designated day of the week for which that person is to report.¹⁴ This design ensures the sample is representative and all days of the week are equally represented.¹⁵ However, designating one person means that if the person is unavailable, an alternative person in the household cannot complete the survey as a proxy respondent, as is the practice in the CPS. In addition, when the designated person is busy on the assigned reporting day, he or she either might be more difficult to contact or may refuse to participate. For example, if the designated reporting day is a Thursday, and the respondent is out of the house working all day and then goes to a class in the evening, he or she would be very difficult to contact. The ATUS sample stays in rotation for 8 weeks, so a person assigned to report on a Thursday would be contacted every Thursday for 8 weeks in a row.¹⁶ However, if this person is out of the house every Thursday, he or she would be difficult to contact on any Thursday. Selecting a designated person and assigning a day of the week helps to ensure a representative sample, but contributes to both noncontact and refusal to participate in the ATUS.

Potential problems in contacting the selected respondent and obtaining a high response rate were identified as early as the 1997 feasibility test (Levin, 1997). In response, BLS put a great deal of effort into field testing alternative advance contact methods, number of eligible interviewing days, advance interview scheduling, 4- versus 8-week data collection duration, telephone versus in-person interviews, and incentives. The procedures with the highest return on investment were adopted (within budget constraints), including use of priority mail for advance contact, an 8-week data collection period, and incentives for households without a telephone number (Piskurich, Nelson, & Herz, 2001). Based on informational gaps and reasons for participation identified in the 2004 survey of respondents and nonrespondents, BLS added an "Information for

Respondents” link to its Web site and introduced revised advance materials. Further research on survey sample design—including the role that increasing cell phone use might be playing with regard to noncontact—and field procedures is now under way to identify interventions that could reduce nonresponse.

EFFECTS OF NONRESPONSE

Relatively low response rates of approximately 56%–58% have raised concern that there might be bias in the time use estimates. Two studies have been undertaken to assess nonresponse bias. O’Neill and Dixon (2005) found that several demographic characteristics including race, age, marital status, and presence of relatives in the household were related to the probability of being unable to contact a sample member, or for the sample member to directly refuse to participate. But their findings showed little evidence of bias across time use categories. Using CPS data and propensity scoring to model who among the actual respondents looked most like nonrespondents, they found that the modeled nonrespondents spent less time in household activities and more time in personal care and religious activities. The category of personal care (sleeping, showering, grooming, and personal/private activities) showed the biggest evidence of bias; however, the difference was still relatively small, approximately 12 minutes out of a total of 12 hours. The second study, by Abraham, Maitland, and Bianchi (2005), found a number of similar factors were related to the probability of participating in the survey, such as age, race, education, marital status, hours worked by the respondent and his or her spouse, ages of children in the household, presence of other adults in the household, home ownership, and urbanicity. However, when Abraham and colleagues constructed new weights based on response propensities and compared the time use estimates to those produced using BLS standard weights, there were only minimal differences. Both of these studies have congruent findings that suggest that despite nonrandom nonresponse, the estimates show little evidence of bias due to nonresponse. Abraham and colleagues are planning further research using ATUS telephone call history data to compare time use estimates of difficult and easy to contact respondents.

Conclusion

The BLS has accomplished much in bringing the ATUS to implementation, building on the hard work and successes of the time use research community, and filling a major void in U.S. national statistics. The ATUS is now in the 7th year of data collection. Yearly time use estimates are published by BLS, and there is an active community of analysts using the data for both substantive and methodological research.

The current ATUS methodology impacts data quality in a variety of ways. The 24-hour diary is considered a tried-and-true method, allowing respondents to report their daily activities in a relatively open-ended way. Yet relatively little research has been conducted on how respondents understand and structure the 24-hour recall diary task, the strategies they use to recall activities and times, how they judge the adequacy of a response and communicate it to the interviewer, and how interviewers shape the interaction. Computer-assisted telephone interviewing provides consistency in how activities are reported compared with a paper diary that individuals fill out at their leisure. It is also much less expensive than an in-person interview. However, the current CATI program doesn't allow for collection of time durations for secondary activities, reducing the accuracy of time estimates for activity categories often done as secondary activities.

Sample design methodology also impacts data quality in different ways. A sample drawn from the CPS helps keep costs down and lessens burden in the form of a shorter interview, but increases nonresponse due to respondent fatigue. Randomly selecting one person from a household and interviewing him or her on a designated day ensures representation of the sample and days of the week, but also contributes to survey nonresponse. Yet, even with lower than desired response rates, studies conducted so far have not found bias associated with nonresponse.

As with most surveys, many methodological decisions and trade-offs were made when designing the ATUS. Options were carefully considered: Studies and tests were conducted, experts and users were consulted, and quality and cost issues were weighed. With the survey solidly in place, evaluation of current methodology with an eye to quality improvement is under way at BLS. The ATUS presents numerous opportunities for further assessment of data quality, including additional research on respondent cognitive processes, respondent-interviewer interaction, coding reliability, measurement of secondary activities, sample design and nonresponse, respondent burden, and nonresponse bias.

Notes

1. Available at <http://www.bls.gov/tus>.
2. The bill was not enacted. National economic accounts have not included non-market production (e.g., domestic activity, child care, shopping, and education), a component of total productive activity. A major conference was held in 1994 on the Measurement and Valuation of Unpaid Work (see Statistics Canada, 1994). In addition, the 1995 Beijing Women's Conference called for the collection of time use data to measure nonmarket production (Harvey & Pentland, 1999).
3. For many examples of how time use data are used, see presentations and posters from the ATUS Early Use Conference at the University of Maryland, <http://www.atususers.umd.edu/papers/atusconference/index.shtml>.

4. Based on American Association for Public Opinion Research Response Rate 2 calculations (AAPOR, 2006).

5. This includes the time to complete the ATUS as well as a temporary module on secondary eating and drinking added to the survey in 2005; the module adds an average of 5 minutes.

6. Conversationally flexible interviewing techniques allow interviewers to engage in more ordinary conversation during the interview and deviate from the structured script to assure respondent understanding of questions (Schober & Conrad, 1997). See section on Computer Assisted Telephone Interviewing for detail on specific techniques.

7. There are 22 “Who” categories (e.g., alone, spouse, household child, friends). There are 26 “Where” categories that are associated with either a place (e.g., home, workplace) or mode of transportation (e.g., bus). See the ATUS User’s Guide (Bureau of Labor Statistics, 2006) for further detail.

8. The types of activities were tested using the first tier or aggregate level coding scheme. In addition, there was no difference between the conditions in the rounding of time periods.

9. This wording slightly varies if there are children under 12 living in the household, but the “in your care” concept is the same.

10. The population includes persons aged 15 and up. A small percentage of secondary child care time is imputed.

11. The expected response rate for the ATUS was approximately 70% (Horrihan & Herz, 2004).

12. O’Neill and Dixon calculate nonresponse outcomes using data that have undergone final editing and processing. During the processing, completed cases with poor quality data are moved into the nonresponse category, lowering final annual average response rates by 1% to 3%.

13. BLS considers CPS households that have moved as ineligible for the ATUS, and thus, are not considered nonrespondents. Abraham, Maitland, and Bianchi (2005) suggest that these households should be considered as “non-contacts.”

14. See Stewart (2002) for an assessment of contact strategies, including designated and convenient day contacts and field periods, potential activity bias, and recommendations for the ATUS procedures.

15. The ATUS sample is split between weekdays and weekends. Approximately 50% of the sample is selected to report for a specific weekday (10% for each day), and 50% for a weekend day (25% each on Saturday and Sunday).

16. A field experiment tested whether substitution of days increased response rates (Piskurich, Nelson, & Herz, 2001). The findings indicated that response rates are equivalent after eight attempts whether or not substitution was used. Since substitution tended to increase reporting for Wednesdays, and reduce reporting on Fridays, ATUS adopted the eight attempts rather than a substitution procedure.

References

- Abraham, K., Maitland, A., & Bianchi, S. (2005). *Nonresponse in the American Time Use Survey: Who is missing from the data and how much does it matter?* University of Maryland, ATUS Early Results Conference. Retrieved from <http://www.atusers.umd.edu/papers/atusconference/index.shtml>

- Allard, M. D., Bianchi, S., Stewart, J., & Wight, V. (2006). *Measuring time spent in child-care: Can the American Time Use Survey be compared to earlier U.S. time-use surveys?* Washington, DC: Bureau of Labor Statistics.
- American Association for Public Opinion Research. (2006). *Standard definitions—Final dispositions of case codes and outcome rates for surveys*. Retrieved from http://www.aapor.org/pdfs/standarddefs_4.pdf
- Bianchi, S., Wight, V., & Raley, S. (2005). *Maternal employment and family caregiving: Rethinking time with children in the ATUS*. Retrieved May 21, 2008, from University of Maryland, ATUS Early Results Conference Web site: <http://www.atususers.umd.edu/papers/atusconference/index.shtml>
- Bose, J., & Sharp, J. (2005). *Measurement of travel behavior in a trip-based survey versus a time use survey*. Retrieved May 21, 2008, from University of Maryland, ATUS Early Results Conference Web site: <http://www.atususers.umd.edu/papers/atusconference/index.shtml>
- Bureau of Labor Statistics. (2006). *American Time Use Survey user's guide*. Retrieved from <http://www.bls.gov/tus/atususersguide.pdf>
- Dixon, J., & Tucker, C. (2000, May). *Modeling household and interviewer nonresponse rates from household and regional characteristics*. Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Portland, OR.
- Fisher, K. (2005, November). *Examining the dynamics of childcare using the American Time-Use Survey and USA heritage time use data sets*. Paper presented at the 2005 International Association for Time Use Research Conference, Halifax, Canada.
- Forsythe, B. (1997a, November). *Assessing data quality: How well did our approach work?* Paper presented at the Conference on Time Use, Non-Market Work and Family Well-Being. Washington, DC: Bureau of Labor Statistics and MacArthur Network on the Family and the Economy.
- Forsythe, B. (1997b). *Cognitive laboratory research plan for collecting time use data*. Washington, DC: Bureau of Labor Statistics, Office of Survey Methods Research.
- Frazis, H., & Stewart, J. (2004). What can time-use data tell us about hours of work? *Monthly Labor Review*, 127, 3–9.
- Frazis, H., & Stewart, J. (2007). Where does time go? Concepts and measurement in the American Time Use Survey. In E. R. Berndt & C. R. Hulten (Eds.), *Hard-to-measure goods and services: Essays in honor of Zvi Griliches* (National Bureau of Economic Research Studies in Income and Wealth, Vol. 67). Chicago: University of Chicago Press.
- Hamermesh, D. S., Frazis, H., & Stewart, J. (2005). Data watch: The American Time Use Survey. *Journal of Economic Perspectives*, 19, 221–232.
- Harvey, A. S. (2003). Guidelines for Time Use Data Collection. *Social Indicators Research*, 30, 197–228.
- Harvey, A. S., & Pentland, W. E. (1999). Time use research. In W. E. Pentland, A. S. Harvey, M. P. Lawton, & M. A. McColl (Eds.), *Time use research in the social sciences* (pp. 3–14). New York: Kluwer Academic/Plenum.
- Horrigan, M., & Herz, D. (2004). Planning, designing, and executing the BLS American Time-Use Survey. *Monthly Labor Review*, 127, 3–19.
- Joyce, M., & Stewart, J. (1999). What can we learn from time use data? *Monthly Labor Review*, 122, 3–6.
- Juster, F. T. (1986). Response errors in the measurement of time use. *Journal of the American Statistical Association*, 81, 390–402.
- Juster, F. T., & Stafford, F. P. (Eds.). (1985). *Time, goods, and well-being*. Ann Arbor: University of Michigan, Institute for Social Research.

- Levin, K. (1997, November). *Administrative details and costs: An assessment of feasibility*. Paper presented at the Conference on Time Use, Non-market Work and Family Well-Being. Washington, DC: Bureau of Labor Statistics and MacArthur Network on the Family and the Economy.
- McGrath, D. E. (2005). Comparison of data obtained by telephone versus personal visit response in the U.S. Consumer Expenditures Survey. *2005 Proceedings of the American Statistical Association*, Survey Research Methods Section. Alexandria, VA: American Statistical Association.
- National Academy of Sciences. (2000). *Time-use measurement and research: Report of a workshop*. Washington, DC: National Academy Press.
- National Opinion Research Center. (2002). *Pretesting of the Advance Diary for the American Time Use Survey*. Washington, DC: Bureau of Labor Statistics, Office of Employment and Unemployment Statistics.
- O'Neill, G., & Sincavage, J. (2004). *Response Analysis Survey: A qualitative look at response and nonresponse in the American Time Use Survey*. <http://www.bls.gov/tus/home.htm>
- O'Neill, G., & Dixon, J. (2005). Nonresponse bias in the American Time Use Survey. *Proceedings of the American Statistical Association*, Survey Research Methods Section. Alexandria, VA: American Statistical Association.
- Piskurich, K., Nelson, D. V., & Herz, D. E. (2001 May). *Maximizing respondent contact in the American Time Use Survey*. Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Montreal, Canada.
- Robinson, J. P. (1985). The validity and reliability of diaries versus alternative time use measures. In F. T. Juster & F. P. Stafford (Eds.), *Time, goods, and wellbeing* (pp. 33–62). Ann Arbor: University of Michigan, Institute for Social Research.
- Robinson, J. P. (1999). The time-diary method: Structures and uses. In W. E. Pentland, A. S. Harvey, M. P. Lawton, & M. A. McColl (Eds.), *Time use research in the social sciences* (pp. 47–87). New York: Kluwer Academic/Plenum.
- Schober, M. F., & Conrad, F. G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly*, 61, 576–602.
- Schwartz, L. K. (2001). *Minding the children: Understanding how recall and conceptual interpretations influence responses to a time-use summary question*. Retrieved from <http://www.bls.gov/tus/home.htm>
- Schwartz, L. K. (2002). The American Time Use Survey: Cognitive pretesting. *Monthly Labor Review*, 125, 34–44.
- Shelley, K. (2005). Developing the American Time Use Survey activity classification system. *Monthly Labor Review*, 128, 3–15.
- Stafford, F., & Yeung, J. (2005). The distribution of children's developmental resources. In D. S. Hamermesh & G. A. Pfann (Eds.), *The economics of time use* (pp. 289–313). Amsterdam: Elsevier.
- Statistics Canada. (1994). *International Conference on the Measurement and Valuation of Unpaid Work: Proceedings*. Ottawa, Ontario, Canada: Author.
- Stewart, J. (2002). Assessing the bias associated with alternative contact strategies in telephone time-use surveys. *Survey Methodology*, 28, 157–168.
- Stinson, L. (1999). Measuring how people spend their time: A time-use survey design. *Monthly Labor Review*, 122, 12–19.
- Willis, G. (2004). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage.