

Planning, designing, and executing the BLS American Time-Use Survey

From conception to implementation, the American Time Use Survey was 12 years in the making; its four developmental phases represented ever deeper levels of agency commitment and outside statistical support, as well as an evolution in thinking regarding survey estimation objectives, units of measurement, sampling plans, and data collection and coding protocols

Michael Horrigan
and
Diane Herz

This article describes the evolution of the American Time Use Survey (ATUS) from its inception as an issue of statistical policy interest in 1991 to its implementation in January 2003 as an ongoing monthly survey sponsored by the Bureau of Labor Statistics (BLS, the Bureau). This 12-year process included four developmental phases. Each successive phase represented a deeper level of agency commitment and outside statistical support. Reports referenced in the text reflect an evolution in thinking at both the Bureau of Labor Statistics and the Census Bureau regarding survey estimation objectives, units of measurement, the universe frame and sampling plan, and data collection and coding protocols.

First phase: policy environment

In 1991, a bill introduced into the 102nd Congress called for the Bureau of Labor Statistics to “conduct time-use surveys of unremunerated work performed in the United States and to calculate the monetary value of such work.” Although the bill, called the “Unremunerated Work Act of 1991,” did not make it out of committee, the existence of a bill naming the Bureau as a sponsoring agency spurred BLS management to begin studying the issue.

In April of the same year, the Bureau sent a representative to a conference sponsored by Statistics Canada on the measurement and valuation of

unpaid work. At the conference, it became clear that there was a strong sentiment in the international community that the lack of a time-use survey in the United States from which to measure the value of unpaid work was a significant data gap in the U.S. statistical system.

Following the conference, a BLS working group was convened to review the literature and summarize the conceptual issues related to measuring and valuing unpaid work. The initial focus of the group was on conceptual issues related to assigning a monetary value to time spent in unpaid work activities. For example,

- In assigning a wage value to time devoted to unpaid work, should one use the wage of a specialist (say, a gardener) or a generalist (say, an average worker) who may be hired to produce the good or perform the service?
- Should the quality of the goods produced or services performed in unpaid work be accounted for?
- How should one account for the marginal utility that may be experienced by the individual who engages in producing a nonmarket good or service?

In the context of the working group’s report, a time-use survey was viewed simply as the vehicle for collecting input data related to the conceptual

Michael Horrigan,
Assistant
Commissioner of the
Office of
Occupational
Statistics and
Employment
Projections, Bureau of
Labor Statistics,
formerly led the
development of the
ATUS; Diane Herz, a
supervisory economist
in the BLS Division of
Labor Force Statistics,
currently is managing
the project. E-mail:
herz.diane@bls.gov

issues raised. Very little effort was directed toward examining other applications of time-use data or toward the practical considerations of collecting such data. The initial working group issued its report in 1993.¹

Second phase: the pilot study

After issuing its report, the working group was disbanded, and the Bureau, while maintaining an interest in the valuation of unpaid work, was not actively engaged in further research on the issue. This period of inactivity, however, did not last long. In 1995, in Beijing, China, the United Nations held an international conference on the status of women. As it did at the Canadian conference, the issue of measuring and valuing unremunerated work emerged as a topic of substantial international interest. The Beijing conference's Platform for Action (item 206) stated, "national, regional and international statistical agencies should measure, in quantitative terms, unremunerated work that is outside national accounts and reflect its value in satellite or other official accounts that are separate from but consistent with core national accounts."²

The Beijing conference prompted the BLS Commissioner to ask the time-use working group to reconvene. Now, however, the group's focus shifted from investigating conceptual issues associated with unpaid work to examining the feasibility of collecting time-use data.

Between 1995 and 1997, the working group undertook two significant activities directly related to examining the latter task. First, the Bureau hired a survey contractor, Westat, to conduct a pilot study to test two alternative time-use questionnaires using telephone interviews. Second, the Bureau cosponsored a time-use conference with the MacArthur Network on Family and the Economy.

The BLS pilot study on time use was conducted in 1997. Drawing on other surveys (primarily one conducted by Statistics Canada), it provided a foundation for what would become the third phase of the working group's efforts. The pilot study discussed response rates, the collection of data on secondary activities, and how to probe for specific information. It also guided some subsequent research on when to call respondents.³ The first phase of the study included 21 cognitive interviews that focused on the ease or difficulty that respondents had in recalling activities from the previous day. The second phase was a random-digit-dialing sample of 1,000 households (1 person per household). The respondents were randomly divided into two groups. Members of the first group were asked what activities they were engaged in, when they performed each activity, and who was with them at the time. Members of the second group were asked the same questions, as well as whether they were doing anything else at the same time.

The results of the pilot study were presented in the fall of 1997 at a conference entitled "Time Use, Non-market Work, and

Family Well-Being," cosponsored by the Bureau and the MacArthur Network on Family and the Economy.⁴ The conference yielded many benefits. First, it introduced the Bureau to the international community of time-use researchers and survey practitioners. Second, it provided the Bureau and, in particular, the Commissioner, with substantial evidence to support the assertion that the lack of a time-use survey was "the biggest single gap in the Federal Statistical System."⁵ Third, it gave the BLS time-use working group critical feedback on its work to date and influenced the direction of work in the next developmental phase.

Two of the themes that emerged from the conference greatly influenced subsequent work on the survey.⁶ First, there was substantial debate over the desirability and the feasibility of measuring secondary activities. Although the theoretical value of such information was broadly supported, varying opinions were expressed about the ability to collect the data accurately and about how one might interpret results. Paper diary approaches that provide respondents the opportunity to list secondary activities, such as that utilized in the Australian time-use survey, are the best method; however, to be optimal, they must be combined with personal interviews, which permit the interviewer to probe diary entries in order to get accurate data. As a result, these approaches are quite costly. A computer-assisted telephone interview (CATI) allows for probes of secondary activities. However, the working group was concerned over the repetitive nature and associated respondent burden of asking, "What else were you doing?" after every activity reported. The discussion at the conference also pointed to the significance of childcare and, in particular, secondary childcare, as a key, if not *the* key, secondary activity. Some expressed the sentiment that capturing childcare well, even in the absence of data on other secondary activities, would be a significant accomplishment.

The second theme that emerged was the choice of the unit of observation in a time-use survey. Conference participants drew a sharp contrast between an approach in which all household members were interviewed and an approach in which only one individual per household was interviewed. The former is more consonant with household bargaining models, according to which choices made regarding time use are partly a function of how other members in one's economic unit are spending their time and the focus is on behavioral models of constrained choice. Ancillary information on the household also figures importantly, such as the ownership of capital (for example, household technology) that can influence the time spent doing unpaid work.

The participants noted that using the individual as the unit of observation would still allow reporting of many of the same concepts that multiple interviews would allow, although without the richness of detail that is particularly useful in testing household bargaining models. For example, it would be possible

to estimate the average time that married men with working wives spend doing housework.

Third phase: the Commissioner's charge

Following the BLS-MacArthur conference, the BLS Commissioner asked the working group to develop a more detailed plan for collecting time-use data. The resulting report became the blueprint for the ATUS. The Commissioner's request came against a backdrop of activities by the National Academy of Sciences (hereafter, National Academy).⁷ Having attended the BLS-MacArthur Network conference, members of the National Academy proposed holding a workshop on the value of time-use data for the U.S. statistical system. The Bureau was invited to present a report on how it would approach the collection of time-use data.

The report submitted by the Bureau to the National Academy's workshop was the working group's first full-fledged attempt to describe how the Bureau would conduct a time-use survey; as such, it stands in contrast to the eventual survey operation that was fielded.

Assumptions and constraints. Some key assumptions and constraints were imposed at the outset of the third phase of the development of the ATUS. These assumptions were the outgrowth of discussions that took place after the BLS-MacArthur conference and reflected the evolution of the thinking at the Bureau between the time of the conference and that of the Commissioner's charge to the working group:

- *Unit of observation:* One individual per sampled household.
- *Sample frame:* Households permanently leaving the Current Population Survey (CPS; "month-in-sample 8," or "MIS-8," households).
- *Collection mode:* CATI.
- *Reference period:* A 24-hour-day time diary (a listing of activities the respondent participated in on the day before the interview).

The choice of one individual per household (instead of multiple members of the household) as the unit of observation was a key point in the deliberations.

The group was sympathetic to the views of those advocating interviewing all members of a household. However, a number of countervailing views emerged. First, the perceived data gap in U.S. statistics entailed a clear interest in knowing how individuals spend their time (in addition to the obvious interest in household decisionmaking). Many of the potential uses of time-use data cited at the BLS-MacArthur conference did not require data to be collected from more than one individual in the household. These potential uses, or estimation objectives, included valuing

nonmarket work; providing an alternative measure of hours of work; and measuring time spent in various activities, including commuting, providing child care, time spent in training and education, and leisure time.⁸

Second, it was unclear why information on individuals' use of time, combined with ancillary information on household structure, could not be used to inform household bargaining models. For example, consider an activity such as grocery shopping. A time-use survey that interviews one individual per household permits reporting how the percentage of total grocery shopping time is conducted by husbands and wives in married-couple families. A time-use survey that includes all family members will provide the same statistic. What is missing from the survey of individuals is the conditional distribution of activities engaged in by a spouse while the other is doing the grocery shopping. Although such a distribution would provide richer behavioral data, no clear and compelling econometric argument was advanced that the gain in information resulting from interviewing every household member would be worth the large loss in sample yield (assuming a fixed budget).

For example, if one were to adopt the purist view that interviews with all household members are necessary to inform household bargaining models correctly, the possibility of low response rates from multiple family members (no matter what data collection methodology is used) would call into question the efficacy of such an approach. That is, at what point do missing data on some household members so dilute the quality of data needed to conduct research on household bargaining that it is not worth the expense and effort to obtain such data? Also, it can be argued that bargaining may occur over a longer period, such as a week, and that information about 1 day may not provide as rich a data source as is needed for some analyses. Finally, in surveys that attempt to interview all household members, systematic bias may be introduced in terms of who tends to be a respondent and who tends to be a nonrespondent. In particular, households which allocate their time so that someone is always home with the children will have a higher likelihood of missing an interview with the individual or individuals who are more likely to be away from home.

The choice to interview only one individual per household reinforced the decision to use CATI (which was tested early in the pilot study). Substantially lower costs per case with CATI than with personal interviews had already suggested that funding would be more easily obtained if a CATI approach were advanced. One advantage of a personal-visit protocol, were it selected, would probably have been higher response rates from multiple household members than would have been achievable with a CATI protocol. Once CATI was selected, however, Statistics Canada's report on low response rates that were experienced in attempting to interview a second household member in a CATI environment lent further support to the Bureau's decision to interview only one individual per household.⁹

The Bureau also considered a mail-out, mail-back protocol. While less expensive to administer than CATI, this protocol was deemed too risky because it might have generated unacceptable response rates and it would have eliminated the use of important probes (to ensure codable data) during the interview. The group also thought that ensuring that the correct respondent reported on the preassigned day (discussed later) would be more controllable in a CATI interview than with a mail-back diary or with a dropped-off diary and a field followup.

Selecting a stratified probability sample from the CPS was designed to enrich the demographic and economic information available on each individual, to reduce sampling costs, and to minimize respondent burden (because, in ATUS, many pieces of information would be imported rather than collected). The CPS also was compatible with the choice of CATI and was a relatively inexpensive sample frame, with recent phone numbers available for 95 percent of CPS households. Finally, by the end of their time in the CPS, many potential respondents are accustomed to answering interview questions by phone. (The ATUS sample person is the same as the CPS reference person in about 59 percent of cases.)

The choice of a 24-hour day as the frame of reference stands in contrast to longer (more retrospective) reporting protocols. A number of existing U.S. surveys already include reporting on the use of time over longer periods (such as “How many hours did you work last week?”). The choice in the ATUS was made to be consistent with most international practices on collecting time-use data and to minimize recall bias.¹⁰

The initial process. Given its charge, the working group concentrated on the following elements of survey design in preparation for the National Academy workshop:

- Primary and secondary estimation objectives of the survey.
- Sample size and the sampling plan.
- Data elements of the survey instrument.
- Operational considerations: systems development, training, field staff, and coding.
- Survey output.

The report delivered at the workshop can be viewed as a detailed first draft of BLS thinking about many of the elements of what has become the ATUS. After the workshop, the working group began anew on some of these elements, reconfirmed existing positions on others, and filled in gaps that had not been considered. For the purposes of this article, each element of the survey is considered in turn, and the development of the group’s thinking from the National Academy workshop to full production in January 2003 is traced. Exhibit 1 provides a concise summary of each of these elements. The choice of software for each system, the sampling stratification and weighting plan, the

variance estimation, the imputation and editing programs, and the training and certification procedures for coders and data collectors were not specified in the National Academy workshop, but were developed for production.

Discussion at the National Academy workshop. The presentation of the BLS proposal was met with strong overall support, despite serious misgivings by some on the choice of one individual as the sampling unit or the decision to interview each respondent only one time. Several attendees expressed the opinion that estimating the average time spent performing an activity in a week required multiple diaries from individuals, preferably two weekday and two weekend diaries. There was also support for repeating this approach for the same individuals several weeks during a year. In the end, the arguments were viewed as survey methodological preferences rather than absolute statistical requirements. Collecting 1-day diaries still would permit the production of all the desired estimates. The National Academy workshop report endorsed the BLS model, recognizing that the lack of a national time-use survey was an important gap to fill, regardless of disagreements over methodological issues.

Fourth phase: preparing for production

In December 2000, the survey received official approval and funding. A great deal of foundation work had already been completed, but each area would need to be revisited in more detail and become fully operationalized. Interagency work on the survey began in earnest, and joint BLS-Census Bureau teams were formed to cover management oversight, sampling and weighting, questionnaire design and development, activity coding, and operations. The survey was initially slated for full production in January 2002. A 4-month delay in budget approval that year had already set back the production schedule, and the systems requirements indicated that additional development time would be needed. New systems were required for the ATUS data collection instrument, the activity coding system, and call management at the call center. The starting date was rescheduled to January 2003.

Many activities occurred between funding in December 2000 and production in January 2003. The following were three primary ones:

- *Operations field test:* an extensive operations test in the summer of 2001.
- *Dress rehearsal:* a test of the questionnaire and operations with live cases in mid-2002.
- *Prefielding:* early fielding of the survey to resolve remaining problems (September–December 2002).

During the December 2000–January 2003 period, the Bureau of Labor Statistics and the Census Bureau continued to receive

Exhibit 1. American Time Use Survey elements: a comparison of analysis at the National Academy of Sciences conference and changes as of full production

Concept	Analysis at National Academy conference	Changes as of full production
Survey estimation objectives (uses)		
Primary sampling objectives	<p>Draw a sample size large enough to generate quarterly estimates of the proportion of the time spent in one- and two-digit activities for an average week, weekday, and weekend.</p> <p>These estimates would be presented for the entire U.S. population (16 years and older) and for selected demographic and labor force groups.</p>	<p>The scope of the sample was increased to include 15-year-olds because potential users expressed an interest in time-use statistics for teens. The CPS sample frame includes persons 15 and older.</p> <p>In addition, youths (various age cut-offs 15 years and under) were included in many other countries' time-use surveys.</p>
Secondary sampling objectives	<p>Generate annual estimates of the time spent in one-, two-, and three-digit activities for an average week, weekday, and weekend.</p> <p>Present these estimates for the entire U.S. population and separately for selected demographic groups.</p>	<p>Largely unchanged. The original team listed a series of demographic breaks. Actual publication detail depends on the sample that falls in each demographic and activity group cell.</p>
Periodicity	<p>Continuous for 14 months (2 months of prefielding, followed by 12 months of collection). Then repeat periodically.</p>	<p>Continuous indefinitely, with 4 months of prefielding before production.</p>
Reference period	<p>The day before the interview.</p>	<p>Unchanged</p>
The designated day	<p>Each respondent would be assigned a designated interview day for reporting about the previous day.</p>	<p>Unchanged</p>
How to handle noncontact on the designated day	<p>Call back exactly 1 week later, asking respondent to recall the previous 24-hour day. Call back again each week for up to 4 weeks.</p>	<p>Field period was extended to 8 weeks. This element was studied at length. As discussed subsequently, the possibility of substituting freely among Monday, Tuesday, Wednesday, and Thursday reference days was carefully examined.</p>
Sampling		
Choice of sampling frame	<p>Household addresses from Month-in-sample 8 of the Current Population Survey</p>	<p>Unchanged</p>
Unit of observation	<p>Randomly selected individual from each household</p>	<p>Unchanged</p>

Exhibit 1. Continued—American Time Use Survey elements: a comparison of analysis at the National Academy of Sciences conference and changes as of full production

Concept	Analysis at National Academy conference	Changes as of full production
Sample size	Sample required to achieve 2,000 completed cases per month at 70 percent.	Sample for 2003 was unchanged at about 3,270 per month. However, response rates that year averaged around 58 percent, yielding about 1,780 interviews per month. ¹
Periodicity of sample draw Questionnaire: Core time-use questionnaire	Monthly Same as in the original BLS pilot study, modeled after the Statistics Canada approach. Respondents are asked to report activity by activity, in sequence. For each activity reported, respondents are asked whom they were with, how long the activity lasted, and where they were.	Unchanged In cognitive testing, problems occurred with the “Who was with you?” question when people were away from home. The question was changed to “Who was in the room with you?” when the respondent was at home and “Who accompanied you?” when he or she was away from home.
Secondary activities	The implicit assumption in the NAS report was that secondary activities would be collected and coded.	Secondary activities are collected only when volunteered and will not be coded, except as needed for research purposes. The exception is childcare: a summary question measures secondary care. BLS is examining secondary activities volunteered by respondents in 2003 interviews and will continue to examine how to better collect these activities.
Summary questions	Ask respondents to identify each recorded activity for which they were paid.	Summary questions were significantly expanded. Four types of summary questions were included in production: questions on childcare, paid work, volunteering, and time spent away from home for 2 or more consecutive nights in the previous month.
Updates to CPS questions	Update the following CPS variables: household composition, total family income, labor force status of the respondent and his or her spouse or unmarried partner, and information on the respondent’s earnings and school enrollment.	Family income is not updated. Respondent’s labor force status is updated, except for the series on reasons for being out of the labor force.

Exhibit 1. Continued—American Time Use Survey elements: a comparison of analysis at the National Academy of Sciences conference and changes as of full production

Concept	Analysis at National Academy conference	Changes as of full production
Modules	Allocate 5 minutes of the questionnaire to special-topic modules. Do not specify the topics for these modules.	<p>Information on spouses' employment status (employed or not) and hours of work also are collected.</p> <p>Respondents are asked about whether they have children under 18 who do not live with them.</p> <p>Unchanged</p>
Activity coding	Adopt a variation of the coding system from the time-use survey of the Australian Bureau of Statistics.	The ATUS coding lexicon originally strongly resembled that of the Australian Bureau of Statistics.
Survey operations	Conduct a pretest and 3 months of pre-fielding before full production.	An extensive operations test was performed, as were a 2-month dress rehearsal (pretest) and 4 months of prefielding.
Target response rate	Adopt a 70-percent target response rate.	Unchanged
Staffing and training	Ensure that interviewers have experience coding.	Required that all interviewers also code. Considered and recommended a dedicated ATUS staff, but did not implement one, due to budget constraints.
Data dissemination and publication plans	Initial table shells were developed.	<p>Publication tables were developed for specific subject areas (for example, on unpaid work, leisure, and childcare), and a system was built to generate them.</p> <p>Public-use data files are being specified according to formats recommended by Andrew Harvey.²</p>

¹ These numbers reflect data from January through December 2003. A 35-percent sample reduction was implemented in January 2004 to keep survey costs in line with the survey budget.

² Andrew Harvey, "Guidelines for Time Use Data Collection and

Analysis," in Wendy Pentland and others, ed., *Time Use Research in the Social Sciences* (New York, Kluwer Academic/Plenum, 1999), pp. 19–46.

advice from outside groups, particularly the Federal Economic Statistical Advisory Council and the International Association of Time Use Researchers.

Survey estimation objectives. The primary and secondary objectives listed in exhibit 1 remained essentially unchanged, except for the expansion of the scope of the survey to include 15-year-olds. After generating table shells that summarized the time spent in a variety of activities, the working group started focusing on thematic tables that offered more detailed information on a variety of subjects, such as providing childcare, traveling, performing paid or unpaid work, volunteering, and participating in leisure activities. The table shells were developed on the basis of research areas brought to the group's attention in the conferences mentioned in this article, in other countries' time-use publications, and in meetings with future data users. A processing system was designed and deployed at the Bureau to generate the tables.

Periodicity of the survey. The National Academy report recommendation to draw the sample monthly did not change, but the survey administration plan was developed further. In the report, the working group suggested that the survey run for 14 months—2 months of prefielding and 12 months of production—and be followed with periodic time-use surveys. Budget process considerations had an impact on the final decision. It would have been very difficult, if not impossible, to secure funding for a time-use survey that would be conducted infrequently, because a continuous funding stream implies a continuous level of survey collection activity. Ultimately, instead of seeking funding for a time-use survey that would be collected, say, every 5 to 10 years, as is typical in most countries, a decision was made to seek funding for a continuous survey with sample building over time to permit more robust estimates and time-series analysis.

There also were discussions about whether the survey should be fielded evenly across the year and within months or whether the sample should be introduced in some months only or should be front loaded at the beginning of selected months. For ease of operation, and to represent all weeks and months equally across the year, a decision was made to introduce each month's sample evenly across 4 weeks. Each case would be called for up to 8 weeks. (See subsection titled "Survey operations" for a further rationale behind this decision.) Estimates, however, would be based on the date about which respondents were interviewed. (For example, first-quarter estimates represented interviews about January 1 to March 31, regardless of the sample month the cases were introduced.)

Reference period. The notion of asking someone to report, activity by activity, about the preceding day was the norm in international time-use data collection. This protocol was taken

as a given by the working group. Assigning a single designated interview day to each respondent in advance was a favored methodology because it would help control the distribution of interviews across the week. Following the recommendations of previous time-use researchers,¹¹ the working group initially recommended that individuals who could not be contacted on their assigned interview day would be called on the same day of the following week and interviewed about the day before the interview (so that the diary day would always be the same day of the week). There was concern, however, that this one-day-per-week schedule, also known as a *designated-day-with-postponement schedule*, would result in low response rates. Empirical work was conducted to examine the possibility of allowing some form of substitution. For example, if research showed that people spent time on all weekdays in a similar way, a decision could be made to allow individuals to be called on *any* weekday, rather than requiring a single day-of-week assignment.

A first step in assessing the feasibility of this approach was determining whether time-use profiles on weekdays were similar enough to one another to allow substitution. Research using data from the Environmental Protection Agency Time Diary Study conducted by the University of Maryland in 1992–94 showed that the Monday through Thursday profiles were similar to one another, that Friday's was only slightly different from those of other weekdays, and that the profiles of the 2 weekend days were different from weekday profiles and different from one another.¹² On the basis of these findings, the working group concluded that Monday-through-Friday substitution would be acceptable. However, because Saturday interviews (about Friday) are easier to obtain than other interviews, experiments were conducted with Monday-through-Thursday substitution only.

One way to implement day-of-week substitution would be to use a *convenient-day schedule*—a schedule whereby respondents are called every day until they are interviewed or until an appointment is scheduled. There was concern, however, that this schedule could result in biased estimates because the probability of a day being selected as the diary day may be correlated with a respondent's time use. Generally, time-use researchers recommend using a designated-day over a convenient-day schedule, but there is very little empirical research to support that recommendation. A middle approach between a designated-day-with-postponement schedule and a convenient-day schedule is to use an every-other-day schedule, also called *designated day with postponement and substitution*.

To assess the potential bias associated with each of these contact schedules, Jay Stewart used computer simulations on mock time-use data to examine the impact on time-use estimates.¹³ He looked especially at the robustness of the various schedules to alternative assumptions about the patterns of activities across the week. The study indicated that the

convenient-day schedule introduced systematic bias into estimates of the time spent in various activities. In particular, time spent in activities engaged in at home was underestimated, while time spent in activities engaged in away from home was overestimated. The designated-day-with-postponement-and-substitution schedule generally did not introduce bias, but it was not as robust as the designated-day-with-postponement (no substitution) schedule.

The final decision about assigning designated days was made after the 2001 operations test (described later). In one test group, respondents were assigned to either a Tuesday/Thursday or a Monday/Wednesday group (that is, they could report on either of the two days assigned), doubling the number of eligible days per field period. The operations tests showed that the availability of a second eligible day during the same week increased response rates about 4 percentage points over an 8-week period. However, with the number of contact attempts held constant, there was no difference between the designated-day-with-postponement and designated-day-with-postponement-and-substitution schedules. After 8 weeks, the designated-day-with-postponement schedule yielded 59 percent, about the same as the 60 percent yielded in 4 weeks with the designated-day-with-postponement-and-substitution schedule and with the same number of contacts. Also, there was more day-of-week variation in responses when substitution was allowed. Because costs are based largely on the number of contact attempts, there was no advantage to allowing day-of-week substitution.

Sampling. The early basic framework for the sampling plan was developed and presented in the report to the National Academy workshop. The sample frame was identified as individuals leaving the CPS who had successfully completed their final (month-in-sample 8) interview. Using a subset of the CPS sample yielded several benefits, including the following:

- Advance selection of survey respondents by their demographic characteristics permitted an efficient sample to be drawn (certain demographic characteristics, such as race, did not require screening);
- Familiarity with the construction of the sample permitted the removal of some design features, to maximize ATUS sample efficiency.

Using unweighted CPS sample counts from month-in-sample 8 files and time-use distributions reported by F. Thomas Juster and Frank T. Stafford¹⁴ to develop parameters for estimating standard errors, Edwin Robison estimated the minimum sample size required to generate reliable estimates for the major subpopulations to be 12,000 per year.¹⁵ Robison assumed that this sample size would produce 9,000 completed interviews. He also estimated that an additional 12,000 samples (9,000 interviews) would be required to produce estimates for smaller

subpopulations specified in the survey's secondary objectives. In general, Robison estimated that 1,000 sample cases (750 interviews) in any particular cell would be needed to produce reliable estimates. To be conservative, the BLS-National Academy report recommended a slightly higher target for the sample: 21,000 completed interviews per year.¹⁶ On the basis of the experience of Statistics Canada, which achieved an 88-percent response rate,¹⁷ the Bureau set a conservative 70-percent target response rate. These sample size recommendations were used in conjunction with estimated Census production costs and BLS staff and research costs to estimate survey budget requirements.¹⁸

After funding was approved in late 2000, an interagency statistics team was formed to refine and finalize the sampling plan. Because the CPS was a household sample, the ATUS sample was stratified by means of household variables and was based on ensuring that reliable estimates could be made for minorities, labor force status groups (employed and not employed), and people with and without children. Labor market status and the presence of children are usually highly correlated with time use. The following stratification variables were chosen:

- Race and Hispanic origin of CPS householder (Hispanic; non-Hispanic, black; non-Hispanic, nonblack).
- Presence and age of children (under 6 years; 6 to 17 years).
- For households with no children, number of adults in the household (1; more than 1)

Sampling rates for each stratum differ in order to produce the desired sample size for various subgroups of the population and overall. The detailed reexamination of the sampling plan following the National Academy workshop led to the ultimate recommendation that the Bureau collect 2,000 completed interviews per month, or 24,000 completed interviews per year.

The questionnaire. The ATUS team members considered a number of issues in designing the ATUS questionnaire.

1. Core time-use questions. Many survey efforts require the development of a new survey instrument, a step that typically occurs early in the survey planning process. The time-use group had a draft partial questionnaire that had been developed from the 1997 pilot survey on the basis of some earlier surveys, particularly the Statistics Canada instrument.

ATUS questionnaire specifications were due to the Census Bureau shortly after the survey was funded, because software specifications, instrument programming, control systems development, and testing would take a long time to complete. The production plan included a summer 2002 dress rehearsal that required survey instruments to be ready well before full production. A draft questionnaire was submitted quickly, but work to refine the 2003 ATUS questionnaire continued until production began. The questionnaire needed to be refined for

several reasons: (1) the Census Bureau does not field untested questions; (2) the Bureau of Labor Statistics was committed to ensuring that questions were capturing the intended information; (3) results from tests and from the dress rehearsal led to many rewrites and retests of some questions; and (4) results from coding tests indicated that additional questions were needed to clarify some activities for coding (discussed later).

2. *The time-use diary.* The design of the time-use diary was fairly straightforward, because many paper diaries had been fielded in other countries and most used a grid with daily activities in rows and with associated information—who was with the respondent, where the respondent was, and how long the activity lasted—in columns. The Census Bureau software standard was Blaise, a package created by Statistics Netherlands. Blaise easily accommodated a grid structure, and the diary was programmed accordingly.¹⁹

3. *Secondary activities.* The enhanced version of the 1997 pilot questionnaire asked respondents not only what they were doing at a certain time, but also whether they were doing anything else at the same time. The pilot study showed that this version picked up more nonmarket work than the standard version, which did not directly address secondary activities.

At the time of the National Academy workshop, the Bureau recommended that a question about secondary activity be included in the survey, although there still was concern about the burdensome nature of asking this question after each recorded activity. Cognitive interviews indicated that many respondents wanted their secondary activities included in any measure of how they spent their time. However, many had difficulty specifying durations for these activities. In addition, it was not clear how to program the diary software to accept this additional information, and modifications to time computations in the software were extremely prone to errors. For the first year of production, it was decided that secondary activities would be collected (but not coded) only when respondents offered them. (The duration of an activity is collected for primary activities, but only the activity report is collected for secondary activities). Research on collecting data on simultaneous activities also was placed at the top of the research agenda during the first year of full production of the ATUS.

4. *Work summary questions.* Midway through the field period of the 1997 pilot test, researchers realized that the information that would be collected in the diary was insufficient for identifying and coding informal activities performed for income, such as making crafts for sale or babysitting. To supplement the existing information, a *summary* question—that is, a question that asked respondents to review in their minds the list of activities reported in the diary and identify one or more characteristics about each activity—was designed. The question asked re-

spondents to identify each recorded activity for which they were “paid or expect[ed] to be paid.” The additional information was used for coding. The pilot survey findings indicated that the question successfully identified income-generating events; the inclusion of a similar probe for coding purposes was thereupon recommended.

The questionnaire design team adopted this recommendation and considered additional questions as well, to better identify other types of work activities reported in the diary. The pilot survey question captured informal, income-generating activities, but did not clarify some activities that were done for one’s job, but were not reported as “work,” such as doing business paperwork or using the computer for work. In some cases, probes could be used during the interview to clarify activities. For example, interviewers were instructed to probe any reports of using the computer, asking “Was that done for work, school, or personal interest?” Still, customized probes could not be developed for all contingencies.

Lisa Schwartz designed a second phase of paid-work tests to determine whether the ATUS could clearly identify work-related activities, particularly of individuals who worked in nonstandard work environments or had nonstandard work hours, because they were more likely to report work in vague terms.²⁰ The test included cognitive interviews and debriefings with salaried workers, self-employed persons, multiple jobholders, and freelancers. From the results of this testing, a second question was developed and administered right after the diary and before the question about income-generating activities: “You said that you were working from [insert start time(s) of work episode(s) reported] to [end time(s)]. Were there any other activities that were done as part of your job (or business)?” This question was followed by “Which ones?” asked twice to multiple jobholders—once each for the main job and for any other job(s).

After many debates about what constituted “work” and what activities might be “related to work,” but not actually work, coding rules were developed to direct how “yes” answers to the second question would be coded. Some activities would be coded as “work-related activities”; these included socializing, eating or drinking, and recreational activities (for example, taking a client to play golf) identified by respondents as done for their job or business. Others (for instance, grading papers) would be coded as work, because they were part of the respondent’s job, but were simply not reported as “work” in the activity description. Finally, some activities, such as smoking and sleeping, would be restricted from being coded as work.

5. *Summary questions on childcare.* Several rounds of testing were required in order to develop summary questions that would enable the Bureau to accurately measure the time respondents spent with children in their care. The 24-hour diary successfully captured “active childcare” activities, such as feeding or bathing children, but the “Who was in the room with you?” question did

not sufficiently identify secondary childcare. For example, a respondent may have been alone in the kitchen, but also responsible for the care of children elsewhere in the home or yard. Conversely, a respondent may have had a child in the room, but not have been responsible for the child at that time.

Statistics Canada had used a summary-question approach to identify activities during which respondents had provided secondary care to children. The ATUS questionnaire design team used that question as a starting point and expanded it to a series of questions to measure secondary care provided to one's household children, to one's own children residing elsewhere, and to others' children.

The team spent a great deal of time determining the appropriate concept to measure, eventually agreeing that secondary care of children was characterized by the respondent's being in the general vicinity of at least one child under 13 and specifically near enough to that child to provide assistance if necessary. To determine wording that would elicit responses in line with this definition of secondary care, BLS cognitive researchers conducted two focus groups.²¹ Participants were shown video clips of people providing this secondary care and were asked to describe it. From choices of terms presented to them, the group preferred "taking care of" and "looking after" one's children. They also offered a new term: "having children *in your care*." The design team thought that "taking care of" denoted more of a primary-care concept, and the diary had done a good job of capturing primary childcare activities (such as bathing or feeding children). The other two terms were tested in cognitive interviewing and the "in your care" wording was ultimately chosen.²²

6. Summary questions regarding absences from home. Some concerns also had arisen about a systematic bias that would occur in the survey: because respondents were asked about "yesterday," activities done during absences from home of 2 or more days would not be collected. Debates initially focused on how the Bureau might be able to get this information, such as by asking a subset or a new set of respondents to take a paper diary on a trip and record activities or by asking respondents in the telephone survey about activities engaged in during recent trips. However, carrying out a separate protocol to get these data would have been costly, and asking about detailed activities on recent trips would likely present recall problems.

Eventually, the questionnaire design team settled on obtaining enough useful information to begin to understand the extent of the bias: information on how many trips people had taken and the primary purpose of those trips. Questions were written to elicit this information, and interviews were conducted with a group of research subjects for whom business travel records were available. This approach enabled the researchers to evaluate the accuracy of reports about trips. Research showed that accuracy declined as the recall period increased and as the number of trips taken increased. Respondents had little or no

difficulty assigning primary purposes to their travel.²³ On the basis of the results of these tests, the Bureau decided to ask respondents only to report on trips taken during the month prior to their first eligible designated day.

Accordingly, the following questions were added to the survey:

Now I'd like to ask a few, very general, questions about times when you may have been away from home for business, vacation, or other sorts of trips. How many times were you away from home in the month of [month prior to first designated day]? Only report times when you were away from home for 2 nights or more. Let's start with the most recent trip you took in "month" (e.g., October). What was the purpose of that trip?...How many nights were you [insert purpose]?²⁴

7. Volunteering. During dress rehearsal and prefielding, Census Bureau coders reported difficulty in distinguishing between certain care and socializing activities, on the one hand, and volunteering activities, on the other. To clarify the distinctions involved, it became evident that a summary question on volunteering was required. Not long before, the Census Bureau had examined how to measure volunteering in a CPS supplement and had defined volunteering as unpaid work (except for expenses) done "for or through an organization." The CPS question on volunteering was adapted for the ATUS, with the reference period modified from the previous year to "yesterday": "Now I'd like to ask you a few questions about volunteer activities—that is, activities for which people are not paid, except perhaps [for] expenses. We only want you to include volunteer activities that you did for or through an organization. Yesterday, did you do any volunteer activities for or through an organization? Which [activities]?"

8. CPS updates. One of the most valuable advantages to using the CPS as the ATUS sampling frame is the wealth of information that is already known about respondents when they are selected for the survey. However, some pieces of information relevant to time-use analyses, such as a person's employment status, can change between the last CPS interview and the time-use interview. Accordingly, prior to the National Academy workshop, the questionnaire team discussed which elements of the CPS interview should be updated and examined whether other pieces of information should be collected during the time-use interview that would not be captured either by the basic time-use questionnaire or the update to the CPS elements. Respondent burden was considered in addressing these questions.

The team ultimately recommended that the survey should update the following CPS variables: household composition, total family income, labor force status of the respondent and

his or her spouse or unmarried partner, and earnings and school enrollment information about the respondent. After funding of the ATUS, the CPS questionnaire and skip patterns were examined in detail in order to understand the impact of these decisions on software development. All the questions and skips included in the ATUS would need to be reprogrammed for its data collection instrument, which was written in a language different from that of the CPS instrument. ATUS team members from the Census Bureau requested that unnecessary sections be excluded to reduce the programming burden. The team decided not to update the family income variable, because it is only a set of ranges and a decision had already been made to update the respondent's earnings. Questions on job search methods of the unemployed were retained, but the branch of the labor force status questions that distinguished reasons for being out of the labor force was not. A new question that would be asked in the ATUS would elicit information on the age and sex of all of the respondent's own children (under 18 years) who did not live with him or her.

9. Modules. Modules consisting of a series of questions on a specialized topic added at the end of the questionnaire hold the promise of allowing researchers to explore more fully social and economic issues related to time use. Modules also can be used to address data limitations resulting from some of the decisions made by the ATUS team members and described herein. For example, modules can be used to measure family allocation of time, asking the respondent questions about the time use of household members. In the National Academy report, the Bureau advocated the inclusion of 5-minute modules within the framework of an estimated 30-minute total interview. To avoid introducing added complexity during the first, stabilizing year of the survey, it was agreed that no modules would be implemented before January 2004 (1 year into production).

BLS thinking on ATUS modules remains basically unchanged since the National Academy workshop. Criteria for acceptable modules have been specified in a policy and include the following: only the designated person may be surveyed; the subject area must have some relevance to time use; and the module must run for at least 6 months.²⁵

Coding. The ultimate value of time-use data depends on the breadth, depth, and consistency of the activity classification system. Each activity is assigned a three-tiered activity code, using a detailed classification system and accompanying rules.

In describing the working group's early recommendations regarding activity codes, Linda Stinson noted that most of the existing activity classification systems evolved from the original structure developed by Alexander Szalai for the Multinational Time-Use Project of the 1960s.²⁶ The time-use group decided to select an existing classification system as a base in order to benefit from previous tests and code revisions, thereby saving time and money. The working group initially examined the

Eurostat classification system, which had been used by 18 countries at that time; the Australian system, which had modified some categories with the measurement of "unpaid work" in mind; and the United Nations' System, which had a basic framework compatible with the U.N. System of National Accounts. The National Academy report recommended a slightly modified version of the Australian system, which was quite detailed and best reflected the type of economy and activities engaged in in the United States.

After funding of the ATUS, an interagency coding team customized the system further, to include activities unique to the United States, to remove some unique to Australia, and to ensure that the activities mentioned were consistent with a fourfold typology of time developed by Dagfinn Ås.²⁷ This lexicon was then submitted to members of the International Association of Time Use Researchers and to a team of BLS cognitive psychologists for review. During the review process, many took issue with the new fourfold typology that grouped activities into "necessary," "committed," "contracted," and "free." Most thought that the overall framework was appealing, but they noted numerous exceptions that could be made to the rules. As a result, the use of these rules as a guiding principle was dropped. However, categories were still maintained in a way that users easily can combine them to represent the four concepts involved.

The coding team found that international comparability across systems was not as simple as had been expected. Even such seemingly straightforward activities as eating were coded differently in different countries. For example, some countries coded eating in restaurants as socializing (and therefore occupying "free" time), while others coded such eating as just that—eating (occupying "necessary" time). Andrew Harvey, president of the International Association of Time Use Researchers, confirmed that international comparability at detailed levels did not exist. Still, two systems that harmonize time-use data at aggregate levels across countries have been developed, and those working on the ATUS are collaborating with the designers to be sure that U.S. data are included.²⁸

The removal of the comparability constraint freed up the coding team to change its focus from revisions of the lexicon to usability by the coder (ability to locate the right code) and enhancement of the analytical value of the survey. Toward the last of these ends, the proposed three-tiered, three-digit system, which allowed 9 categories within each tier, was replaced with a three-tiered, six-digit system that allowed up to 99 categories in each tier. This system is more flexible than a three-digit system, because it can accommodate the introduction of new codes.

To implement recommendations from the team of research psychologists, the ATUS team conducted several usability tests of the coding system with Census Bureau coders. Ultimately, three separate coding tests were conducted at the Census Bureau, each one introducing a revised lexicon and more

extensive training than the first, and the last test introducing coding software.²⁹ These tests were highly productive and led to both small and large changes in the classification system, including the following:

- Removal of the “activities related to goods and services” category present in many time-use classification systems. Coders did not associate this title with the elements it included, such as grocery shopping and purchasing medical services. The category eventually expanded into four largely purchasing categories, such as “consumer purchases” and “professional services.”
- Removal of the “media use” category, because many did not think of television, radio, and other media together in one category. Separate categories were developed for each type of media use.
- Removing and revising ambiguous examples under various categories.
- Agreeing on the best location or conceptual definition for questions on many difficult-to-code activities, such as “looking for things” and “waiting for [something or someone].”

In the fall of 2001, the Bureau worked with Westat, a private research firm, to conduct an additional, larger scale test designed to measure coding accuracy and efficiency over time (returns to experience) and to evaluate BLS training in coding. The test involved nine coders with a wide range of experience coding other survey data. The test showed that coding speed increased rapidly with experience, and coding accuracy increased as well, though not as quickly. Westat made a number of recommendations to improve the classification system, the coding software, and the training. Most were adopted.³⁰

The coding tests, as well as work at the Bureau to specify coding rules and analytical tables, routinely pointed out difficult-to-code activities. Some of the most difficult categories were travel, waiting, packing and unpacking things, work-related activities, helping others, and computer use. After the tests, much work was done during 2002 to address these issues; among aspects of the coding system that were revised were coding categories, coding subcategories, and, particularly, rules and job aids for training. A review of the proposed system, including how it compared with several other countries’ systems,³¹ led to a number of important changes in the classification system. As mentioned earlier, difficulties distinguishing between care, socializing, and volunteering made it clear that a summary question on volunteering needed to be added to the ATUS. Systems were specified and developed to run coding operations, including verification and adjudication processes that required a second coder to assign a code to each activity and an adjudicator to rule on the correct code. Feedback from the dress rehearsal and prefielding activities also was used to modify the

system, right up to production. Ultimately, a system with the following 17 top-tier categories was developed:

- Personal care
- Household activities
- Caring for and helping household members
- Caring for and helping non-household members
- Work and work-related activities
- Education
- Consumer purchases
- Purchasing professional and personal care services
- Purchasing household services
- Obtaining government services and civic obligations
- Eating and drinking
- Socializing, relaxing, and leisure
- Sports, exercise, and recreation
- Religious and spiritual activities
- Volunteer activities
- Telephone calls
- Traveling

Survey operations. A few specifics of the survey operations were discussed in the development of the sampling objectives and were suggested in the National Academy report. Telephone interviewing was a starting assumption. A subsequent recommendation was made to use a designated-day methodology with 4 weeks of callbacks, doubling the length of the 1997 pilot study reference period. The various recommendations provided a beginning set of operational assumptions, but a great deal of work remained.

A BLS-Census field-test team was chartered to recommend detailed ATUS operations. The team was particularly concerned about how to fulfill the difficult and unprecedented requirement that the Census Bureau contact a specific household member on a predesignated day. While the 1997 pilot study provided guidance on extending the calling period, it was not clear how many calls should be made over how many weeks to achieve the 70-percent response rate target and how different mail-out or day-of-week substitution techniques might affect the survey results.

Using 3,200 cases from the outgoing rotation groups of the CPS, the field-test team designed three experiments that were run concurrently by the Census Bureau in April through June of 2001. The 1997 pilot results indicated that efforts would need to be made to increase both contact rates (reaching the designated person) and response rates (gaining cooperation). Thus, the studies focused on maximizing these two objectives, rather than on collecting codable activity data. A paper questionnaire was developed that included both a short diary (from 4 A.M. to noon) administered by phone and debriefing questions designed to provide insight into contact- and response-related issues.³² The effects of the following methods on contact, response, and costs were studied:

- *Priority mail*: All respondents received an advance letter and brochure from the Census Bureau. Some received the materials by priority mail, while others received them by regular mail.
- *Substitution*: Some respondents had the option of substituting between 2 eligible days per week (either Monday/Wednesday or Tuesday/Thursday), while others were eligible to report on a specific weekday only.
- *Proactive appointment setting*: Some respondents were called in advance of their interview day and were asked to set an appointment. Others were called only on their interview day.
- *Field visits as opposed to calling*: Some respondents were called for up to 8 weeks; others were called for up to 4 weeks and then were visited for up to 4 more.
- *Incentives*: Respondents were divided into three incentive groups—those who received no incentive, those who received a debit card for \$20, and those who received a debit card for \$40. Those who received debit cards were given the PIN number if they completed the survey.
- *Incentives would be used only for households whose telephone numbers the Census Bureau lacked*. When costs came in, it was clear that the Bureau would not be able to afford incentives for each case. The shorter field period required to get to 70 percent did not reduce costs enough to offset the cost for incentives, even with a reduced face value. As a result, an 8-week field period and a no-incentive protocol were chosen for households for which the Census Bureau had a recent phone number. Incentives were chosen for households for which the Census Bureau lacked such a number. A \$60 incentive induced 41 percent to complete the survey. Because this group included underrepresented demographic groups, and because survey advisors (including the Federal Economic Statistical Advisory Council) felt strongly that those groups should be included in the sample, a \$40 incentive was implemented for full production.³³
- *Substitution would not be used*. It was somewhat surprising that the availability of a second eligible day during the week increased response rates by only about 4 percentage points over an 8-week period. Substitution was not implemented because it did not reduce costs and it increased the variability of responses across days of the week.

Two tests covered all of the variables just listed and were carried out with the 95 percent of the sample for whom the Census Bureau had telephone numbers. An additional study was developed for the 5 percent of the sample for whom the Census Bureau had no phone number. (Most had responded to the CPS in personal-visit interviews.) This group received a \$60 debit card, and their letter asked them to call the toll-free number of the Census Bureau to respond. They had up to 4 weeks to call in and complete an interview; those who had not done so were visited on their designated interview days for up to 4 more weeks in an attempt to secure an interview.

Response rates varied a great deal across the treatment groups. The highest were achieved with larger incentives and with field visits, both expensive operations. Incentives also sped response; for example, a 70-percent response rate was achieved in only 2 weeks with a \$40 debit card; with no incentive, a 69-percent response rate was reached after 8 weeks. To analyze the data and make operational choices, contact rates, response rates, and costs were examined for each methodology. The following operational choices were made:

- *Priority mail would be used*. Priority mail appeared to be highly effective in reaching respondents, and the costs were relatively small.
- *Field visits would not be made*. Field visits increased response by about 4 percentage points, compared with a full-CATI operation. However, they were prohibitively expensive and would require training interviewers across the country rather than in one location.
- *Proactive appointment setting would not be used*. Calling in advance to set an appointment did not increase response. It did, however, increase costs.

Staffing and training. The National Academy report suggested that it would be desirable for interviewers to have experience with the coding system. As preparation for full production continued, the necessity of this approach was confirmed. It became clear that the number of probes that were needed in the interview to allow correct coding of activities expanded significantly with the development of the coding lexicon. To collect high-quality time-use data, an interviewer must listen to subtle cues from the respondent and remember to probe when necessary to obtain enough detail to code activities. In most surveys, interviewers read questions verbatim. In the diary portion of the ATUS, they must instead conduct a “conversational interview,” taking in information as it is provided by the respondent and probing when necessary.

There was concern that it might be difficult to find people who were inclined to do both tasks. This hypothesis was tested for the first time in the dress rehearsal; debriefings with Census Bureau interviewers indicated that they thought that activity-coding experience not only was important and enjoyable, but was critical to conducting good time-use interviews. For similar reasons, the BLS-Census Bureau oversight team also thought that the ATUS should have a dedicated staff. However, the prohibitive cost of maintaining a dedicated staff necessitated allowing interviewers to work on other surveys as well as the ATUS. In production, interviewers are required to code, and extensive training and certification is mandatory for everyone working on the survey.

The difficulty of the interview and of coding influenced approaches to training. Dress rehearsal and prefielding experiences indicated that training in the rules of coding should be provided prior to training in interviewing techniques. The Bureau contracted with a vendor to develop a training package that the call center staff could deliver easily without assistance from headquarters. The developer included an extensive audio feature in the computer-based training, given that the CATI interview required skills in listening and immediate, customized probing.

Data dissemination and publication plans. To develop its initial table specifications, the working group examined publications based on other time-use surveys, as well as BLS reports that used CPS demographic and labor force data. Work on tables has continued since then. Meetings with advisory groups and with outside users provided information on the types of analyses planned. Tables were developed by subject area groupings, including childcare, unpaid work, travel, and leisure. The Bureau published a subset of these tables in September 2004, through a news release. The data are available to users online at the BLS website.³⁴ Data were presented through several concepts, including time spent doing an activity and the proportion of the population that engaged in a particular activity on an average day. Most tables included tabulations by demographic characteristics or labor force status.

Many time-use users will be interested in microdata files. The data file formats are still in development, but most likely will adhere to recommendations by Andrew Harvey that call for the following three types of files:³⁵

- *Episode file.* Contains episode-level information for each activity (such as the activity code, who was present when the respondent engaged in the activity, and where the activity occurred.)
- *Respondent summary file.* Contains summary information on each respondent (such as age, sex, and cumulative time spent in various activities at various locations).
- *Time-points file.* Contains the activity codes for activities performed at prespecified intervals during the day (for example, at 5 P.M., the respondent was eating; and at 5:15 P.M., the respondent was doing the dishes.)

The current plan is to produce SAS and ASCII microdata files for distribution through the BLS website.

Update since January 2003

A number of important changes were made to the ATUS after several months of production and continued research.

Response rates. Response rates for cases in the panels for 2003 came in substantially below the 70-percent targeted rate.

The response rate for households for which the Census Bureau has a telephone number was 58 percent, while the rate for households for which the Census Bureau has no telephone number was 33 percent. (These households called in to complete the interview.) A number of parallel efforts, including the establishment of a response rate investigation team at the Census Bureau, were put in place to investigate the cause of the low response rates. The team examined calling and response patterns by interviewer, time of day, and respondent characteristics to influence targeted methods for improving response. In January 2004, a response analysis survey was conducted with about 50 respondents and 40 nonrespondents to determine their reasons for participating in the ATUS or declining to participate in the survey. The study also solicited suggestions for improving the survey materials, the interview, or the contact protocol. In addition, the Bureau of Labor Statistics is conducting an examination into whether, and to what degree, nonresponse bias exists in ATUS estimates.

Secondary activities. Proposals were solicited from outside vendors as to how to identify the best way to measure secondary activities. Among the proposals were tests that required substantial software changes if the results were successful. For example, new time computation fields would need to be added to the ATUS diary for the duration of the secondary activities, and new screens would need to be added for cases in which the respondent could not determine the length of the activity, but could provide a range for its duration. Before embarking on this extensive research into collecting information on all secondary activities, the Bureau decided to examine the reports of volunteered secondary activities to determine whether a more targeted approach to measure *some* secondary activities might be more effective than an approach that would seek to measure *all* secondary activities. This research will begin in 2005.

Reduction in the sample size. Once the survey was in full production, actual costs could be measured. They showed that the cost of maintaining the initial sample size (about 3,375 cases per month) exceeded the ATUS budget. In the first year, savings from development years paid the full survey costs. However, in order to bring costs in line with the annual survey budget for future years, the monthly sample had to be reduced by about 35 percent, to 2,200 cases per month. The Census Bureau computed variances under several sample reduction strategies. The Bureau of Labor Statistics implemented an evenly distributed reduction across all stratification groups, rather than a reduction in targeted groups only, because the first approach minimized increases in variances for overall time-use estimates. The reduction was implemented in January 2004.

THIS ARTICLE HAS TRACED THE DEVELOPMENT of the American Time Use Survey from its inception in 1991 as an

issue of statistical policy interest to its implementation in January 2003 as a new monthly survey sponsored by the Bureau of Labor Statistics. Along the way were healthy debates over the choice of one individual per household as the survey respondent, the cognitive research that led to the inclusion of various summary questions, the transition to full production, and a description of the future work that remains.

The development of the ATUS represents a coalescence of work that includes academic inquiry and debate, survey methodological design, operational testing, production management, and a strong and growing consensus among government agencies as to the need for the kinds of data the survey captures. Many individuals were involved in this process. Some contributed through advisement at the Mac-

Arthur or National Academy conferences or at the Federal Economic Statistical Advisory Council meeting or other advisory meetings, through contractual relationships, or through projects and conferences sponsored by the International Association of Time Use Researchers. Others worked at the Bureau of Labor Statistics or the Census Bureau designing and running tests, securing funding for the project, building and testing software, providing training, conducting interviews, and coding activities. Former BLS Commissioner Katharine Abraham, under whose leadership much of the early work and the securing of funding were completed, was critical to the endeavor, as was current BLS Commissioner Kathleen Utgoff, who continued to support the project. Finally, the ongoing interest and support of the National Academy played a key role as well. □

Notes

¹ Michael Horrigan, Maury Gittleman, Mary Joyce, and others, *The Measurement and Valuation of Unpaid Work*, Report of the BLS Working Group (Bureau of Labor Statistics, 1993).

² *Platform for Action*, Report from the Fourth World Conference on Women (Beijing, United Nations, 1995).

³ Linda Stinson, Angie Becher, Barbara Forsyth, and Kerry Levin, *Using a time-use approach to measure the frequency and duration of non-market work*, BLS internal report (Bureau of Labor Statistics, 1998).

⁴ *Agenda for the Conference on Time Use, Non-market Work, and Family Well-Being* (Washington, DC, Bureau of Labor Statistics and the MacArthur Network on the Family and the Economy, 1997).

⁵ William Nordhaus, remarks made in "Session VI: Time-Use Surveys: Where Should the BLS Go from Here?" in Lynn Hatch, ed., *Summary of the Conference on Time Use, Non-market Work, and Family Well-Being* (Washington, DC, Bureau of Labor Statistics and the MacArthur Network on the Family and the Economy, 1997).

⁶ Numerous other important lines of inquiry were explored at the conference in addition to the two reported here, although those two themes figured prominently in the group's thinking in the next phase of the methodological development of the survey. Other notable points for discussion included a comparison of alternative coding systems and research on reporting both activities and one's emotional state at multiple times during the day.

⁷ Bureau of Labor Statistics Time-Use Survey Working Group, *A Report on the Feasibility of Conducting a Time-use Survey*, paper presented at the National Academy of Sciences workshop on Time-Use Measurement and Research, Washington, DC, 1998.

⁸ See the following *Monthly Labor Review* articles: Mary Joyce and Jay Stewart, "What can we learn from time-use data?" August 1999, pp. 3–6; Anne E. Winkler, "Measuring time use in households with more than one person," February 2002, pp. 45–52; and Lisa K. Schwartz, Diane Herz, and Harley Frazis, "Measuring intrahousehold allocation of time: response to Anne E. Winkler," February 2002, pp. 53–59.

⁹ Statistics Canada's report was presented at the National Academy's workshop. (See Lorna Bailie, "Remarks made in 'Session VI: Time-Use Surveys: Where should the BLS go from here?'" in Hatch (ed.), *Summary of the Conference*.)

¹⁰ Linda Stinson, *Measuring How People Spend Their Time*, paper presented at the American Statistical Association meetings, Washington, DC, August 1999.

¹¹ Brian L. Kinsley and Terry O'Donnell, "Marking Time: Methodology Report of the Canadian Time Use Pilot Study—1981," in *Explorations in Time Use*, vol. 1 (Ottawa, Department of Communications, Employment and Immigration, 1983); Graham Kalton, "Sample Design Issues in Time Diary Studies," in F. Thomas Juster and Frank T. Stafford, eds., *Time, Goods, and Well-Being* (Ann Arbor, MI, University of Michigan, Institute of Social Research, 1985), pp. 333–51; Ingrid Lyberg, "Sampling, Nonresponse, and Measurement Issues in the 1984–85 Swedish Time Budget Survey," in *Proceedings of the Fifth Annual Research Conference* (Bureau of the Census, 1989), pp. 210–38; and Andrew Harvey, "Guidelines for Time Use Data Collection and Analysis," in Wendy Pentland and others, eds., *Time Use Research in the Social Sciences* (New York, Kluwer Academic/Plenum, 1999), pp. 19–46.

¹² Jay Stewart, "Alternative Indexes for Comparing Activity Profiles," paper presented at the 2000 conference of the International Association for Time Use Research, Belo Horizonte, Brazil, 2000.

¹³ *Ibid.*

¹⁴ Juster and Stafford, eds., *Time, Goods, and Well-Being*.

¹⁵ Edwin Robison, *Sampling and Reporting in Time-Use Surveys*, paper presented at the American Statistical Association meetings, Washington, DC, August 1999.

¹⁶ Bureau of Labor Statistics Time-Use Survey Working Group, *A Report on the Feasibility*.

¹⁷ Statistics Canada's time-use surveys were conducted as add-ons to that nation's General Social Survey in 1992 and 1998 and were not independently fielded.

¹⁸ Robison, *Sampling and Reporting*.

¹⁹ The questionnaire design team deliberated the order of the questions for some time. CPS updates that were essential to the administration of the diary were placed before the diary questions began. Other CPS updates were placed after the diary and summary questions, because they were deemed less important to collect than the diary information. The most sensitive questions—on earnings—were placed at the end of the questionnaire in case they triggered a refusal to answer.

²⁰ Lisa K. Schwartz, "The American Time Use Survey: cognitive pretesting," *Monthly Labor Review*, February 2002, pp. 34–44.

²¹ Linda Stinson, *Report on Cognitive Testing Phase 1: The American Time Use Survey Summary Questions*, BLS internal report, 2000.

²² Lisa Schwartz, "Minding the Children: Understanding How Recall and Conceptual Interpretations Influence Responses to a Time-Use Summary Question," unpublished internal paper, BLS Working Paper Series (Bureau of Labor Statistics, 2001). At the same time that the first round of questions was asked about childcare, a set of cognitive interviews was conducted with caregivers in an attempt to determine ways to measure time spent providing dependent adult care activities as a secondary activity. (See Stinson, *Report on Cognitive Testing Phase I*.) The wording of the question adopted was similar to that used in the childcare interviews:

In addition to the activities you just told me about, we are interested in finding out about the time you spent looking after adults and children 13 and older who cannot take care of themselves because of a physical or psychological problem. Yesterday, did you spend any time looking after anyone living in the household 13 or older who cannot or should not be left alone because of a physical or psychological problem? Please tell me when you were looking after [name].

Testing indicated that the question had numerous problems. Chief among them was the recognition that the terms "care," "adults," and "disabilities" were unclear to, and interpreted differently by, respondents. In addition, some respondents did not like the "should not be left alone" language for adults, because many adults needed care, but could also be left alone. Some found "looking after" pejorative. The alternative terms "keeping tabs on" and "monitoring" were also rejected, the former possibly implying that the adults were wrong or untrustworthy, the latter having too clinical a tone.

Recognizing the complexity of defining questions to measure disability—a multiyear process is taking place at the Bureau to develop a series of questions to do just that—a decision was made to defer the development of questions to measure care of disabled adults to a later date. The current focus has shifted to measuring care of adults as a primary activity and to developing a series of questions to identify overall time spent providing "eldercare," a more restrictive concept.

²³ Schwartz, "The American Time Use Survey."

²⁴ Because of programming difficulties, these data will not be available until 2005.

²⁵ "ATUS Module Policy," BLS internal document.

²⁶ Alexander Szalai, *The Use of Time: Daily Activities in Urban and*

Suburban Populations in Twelve Countries (The Hague, Mouton, 1972); cited in Stinson, *Measuring How People Spend Their Time*.

²⁷ Bureau of Labor Statistics Time-Use Survey Working Group, *A Report on the Feasibility*. Dagfinn Ås (1920–77) was a leading time-use researcher who was one of the planners of a multinational time-use budget study conducted in the 1960s. He was Norwegian.

²⁸ The Multinational Time Use Study was launched in the 1980s by Professor Jonathan Gershuny of the University of Essex in the United Kingdom. The idea was to create a cross-national and historical archive of time-use data sets. All the data sets in the archive have been harmonized into a common set of variables, including time-use activities and various demographic and socioeconomic characteristics of respondents. The harmonized file currently contains 35 different surveys from more than 20 countries, as well as 41 time-use activity codes. The development of these common activity codes was itself constrained by the richness (or sparseness) of activity codes in the original surveys. In recent years, on the basis of some 20 surveys having been carried out since 1999, an alternative harmonized time-use activity schema has been developed as part of the Harmonised European Time Use Study. (Visit <http://www.iser.essex.ac.uk/mtus/index.php> on the Internet.)

²⁹ Kristina Shelley, "Activity Coding in the American Time Use Survey: A Report of 3 Tests," BLS working paper, 2004.

³⁰ *Final Report: Research Services for Usability Testing and Lexicon Evaluation: The American Time Use Survey* (Rockville, MD, Westat, 2001).

³¹ Anne Gauthier, *BLS 2003 Codes of Activities: Comparisons over Time and across Nations*, draft report, 2002.

³² Detailed test designs are described in Karen Piskurich, Dawn Nelson, and Diane Herz, "Maximizing Respondent Contact in the American Time Use Survey," in *Proceedings of the 2001 American Association of Public Opinion Research Conference* (2001).

³³ The \$40 amount was chosen because respondents indicated in the debriefing section of the test that \$20 was too low and \$50 was too high.

³⁴ www.bls.gov/tus.

³⁵ Harvey, "Guidelines for Time Use."