

**A STUDY IN THE PROCESS OF PLANNING, DESIGNING AND EXECUTING A SURVEY
PROGRAM: THE BLS AMERICAN TIME-USE SURVEY**

Michael Horrigan and Diane Herz

Bureau of Labor Statistics

In this paper, we describe 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 a new monthly survey sponsored by the Bureau of Labor Statistics. This twelve-year process included three separate survey methodological development phases prior to committing to a path of full operational development. Each phase resulted in one or more reports (see references) that documented our thinking at that point. And each consecutive phase represented a deeper level of agency commitment and outside statistical support and an evolution in our thinking on survey estimation objectives, units of measurement, the universe frame and sampling plan, and the data collection protocol.

The first three sections of this paper describe these three development phases, including details on the evolution in our thinking about the efficacy of a time-use survey as well as the transitions we experienced in securing both inside and outside agency support. The fourth section summarizes developments between funding and full production in January 2003.

1. The Policy Environment and Background: 1st Developmental phase

In 1991, a bill introduced into the 102nd Congress, the “Unremunerated Work Act of 1991” called for BLS 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 did not make it out of Committee, the

existence of a bill naming BLS as a sponsoring agency convinced BLS management to begin studying the issue.

In April of the same year, BLS sent a representative to a conference sponsored by Statistics Canada on the measurement and valuation of unpaid work. It was clear from this conference that there was 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.

As a result of these activities, the BLS Commissioner, Katharine Abraham, convened a working group to review the literature and summarize the conceptual issues related to measuring and valuing unpaid work. The focus of this initial work was mainly on conceptual issues related to assigning a monetary value to time spent in unpaid work activities. For example:

- Does one use the wage of a specialist (or a generalist) who may be hired to produce the good or perform the service?
- Should value be assigned based on the opportunity cost of one's time in unpaid work activities?
- Should the quality of goods produced or services performed be accounted for?
- How should one account for the positive marginal utility that may be experienced by the individual who chooses to engage in producing the non-market good or service?

In the context of this report, a time-use survey was viewed simply as the vehicle for collecting input data for this purpose. Very little effort was placed on examining the overall efficacy of collecting time-use data and its other applications. The working group issued its report in 1993 [Horrigan, et al, 1993].

2. The Pilot Study: 2nd Developmental phase

After issuing its report, the working group was disbanded, and the BLS, while certainly maintaining an interest in the issue of valuing unpaid work, was not actively engaged in further research on this issue. This state of affairs did not last long. In 1995, in Beijing, China, the United Nations held an

international conference on the status of women. Measuring and valuing unremunerated work emerged as a topic of substantial international interest. The conference's Platform for Action (item 206) stated that "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." (United Nations, 1995)

The BLS Commissioner asked the time-use working group to reconvene to examine the feasibility of collecting time-use data. While the impetus for this activity derived from the Beijing conference's call to measure the value of unpaid work, there was a decided shift in the working group's emphasis toward the efficacy of collecting time-use data and away from the conceptual issues surrounding valuing unpaid work that was the focus of the group's first report.

As a result, between 1995 and 1997, the working group undertook two significant activities directly related to examining the feasibility of conducting a time-use survey. First, BLS hired a survey contractor, Westat, to conduct a pilot study of two alternative versions of a time-use survey using a telephone methodology. Second, BLS co-sponsored a time-use conference with the MacArthur Network on Family and the Economy.

The BLS Pilot Study on time use was conducted in 1997. It drew on other surveys (primarily efforts by Statistics Canada) and provided a foundation for what would become the third phase of the working group's efforts. The Pilot Study provided information on response rates, the collection of simultaneous activities, and how to probe for specific information. It also guided some subsequent research on contact strategies (Stinson, et. al., 1998). The first phase of the study included 21 cognitive interviews centered on the ease and/or difficulty that respondents had in recalling activities from the

previous day. The second phase was a random-digit telephone interview with 1,000 respondents. The panel was randomly divided into two groups; the first group of respondents was asked what activities they were engaged in, when they performed the activity, and who was with them at the time. The second panel was asked the same questions and whether they were doing anything else at the same time.

The results of the BLS Pilot Study were presented in the fall of 1997 at the BLS and the MacArthur Network on Family and the Economy co-sponsored conference entitled, “Time Use, Nonmarket Work and Family Well-Being” (BLS, MacArthur, 1997). Co-sponsorship with the MacArthur Network yielded many benefits. First, it introduced BLS to the international community of time-use researchers and survey practitioners. Second, it provided the Bureau, and in particular, the BLS 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” (Nordhaus, 1997). Third, it provided the BLS time-use working group with critical feedback on their work to date, and influenced the direction of work in the third (as of yet unanticipated) developmental phase.

Two important themes emerged from the conference that influenced subsequent work.³ First, there was substantial debate over the desirability and the feasibility of measuring simultaneous activities. While there was broad support for the theoretical value of such information, varying opinions were expressed about the ability to collect such data accurately. Paper diary approaches that provide respondents the opportunity to list simultaneous activities, such as the Australian time-use survey, are the best method; however, to be optimal, they must be combined with personal interviews to permit interviewer probes of diary entries in order to get accurate data. As a result, this approach is quite costly.

³ There were numerous other important lines of inquiry explored at the conference in addition to the two reported here. The two themes we highlight figured prominently in our thinking in the next phase of survey methodological development. Other notable discussion points included a comparison of alternative coding systems, and research on reporting both activities and one’s emotional state at multiple times during a day, among others.

A Computer Assisted Telephone Interview (CATI), such as the Statistics Canada model, allows for probes of this type. 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 child care, and in particular, passive (or secondary) child care, as a key, if not *the key*, simultaneous activity. There was some sentiment expressed that capturing child care well, even in the absence of data on other simultaneous activities, would be a significant accomplishment.

The second theme that emerged was the debate over what should be the unit of observation in a time-use survey. Although the BLS was only reporting out the results of an initial feasibility pilot, the discussion at the conference was aimed at influencing BLS choices if a full-fledged survey were to become a reality. A sharp contrast was drawn by conference participants between an approach that interviewed all household members, and an approach that interviewed only one individual per household. The former approach is more consonant with household bargaining models. In these models, 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 (e.g. household technology) that can influence the efficiency of household production.

The discussion noted that using the individual as the unit of observation would still allow reporting of many of same concepts that multiple interviews would allow, although without the richness of detail that is particularly useful in developing household bargaining models. For example, the distribution of time-use activities by a single mother in a family with 3 children could still be tabulated from cross-sectional stratifications of the sample.

3. The Commissioner's Charge to the Working Group: 3rd Developmental phase

Following the International Conference on time use, the BLS Commissioner asked the working group to continue its work and to develop a more detailed report outlining a strategy for collecting time-use data that was to become the blueprint for the full operational form of the survey. This request came against a backdrop of activities by the National Academy of Sciences (BLS, 1998). Having attended the BLS/MacArthur Network conference, members of the National Academy of Sciences (NAS) developed a proposal for holding a workshop on the value of time-use data for the U.S. Statistical System. The BLS was invited to present a paper on how it would approach the collection of time-use data.

This section describes the report submitted by the BLS to the National Academy of Science's workshop on time-use surveys. This report was our first full-fledged attempt at describing a complete survey plan for conducting a time-use survey, and as such, stands in contrast, sometimes notably, to the eventual survey operation that has been placed in the field.

Assumptions and constraints

Some key assumptions and constraints were imposed at the outset of this third phase of the working group. These assumptions were the outgrowth of discussions that took place after the BLS/MacArthur Network International conference on time use data, and represented an evolution in our collective thinking between the conference and the timing of the Commissioner's charge to the working group.

- Collection mode: The data collection protocol would be a Computer Assisted Telephone Interview (CATI).
- Unit of observation: One individual per selected household.
- Sample frame: Outgoing rotation groups of the monthly CPS.

- Core concept: A 24-hour day time diary would be used (i.e. leading respondent through a 24-hour day) along with assigning respondents to designated interview days for reporting about the prior day.

Consider the choice of using CATI. The view had emerged that the lower costs of telephone interviews compared to personal-visit interviews suggested that the former would provide sufficient sample size for the overall budget parameters, while the latter would be too costly. The choice to interview only one individual from each household drawn from a stratified probability sample (discussed more fully below) served to reinforce the decision not to use a personal-visit protocol. Personal visits would probably provide better response rates from multiple household members than would a telephone protocol, but without this objective, a CATI approach would generate far more interviews for the same budget. Statistics Canada's report at the National Academy's workshop on the low response rates experienced in attempting to interview a second household member in a CATI environment lent further support to BLS's decision to interview only individual per household. In addition, a mail-out/mail-back protocol, while less expensive to administer than CATI, was felt to be too risky in terms of generating acceptable response rates and would eliminate the use of important probes during the interview. The group also thought that ensuring the correct respondent reported on the pre-assigned day would be more controllable in a CATI interview than with either a mail-back diary or a dropped-off diary with a field follow up.

The choice of one individual per household as the unit of observation (instead of multiple members of the household) was also a key point in our deliberations. Although we were certainly sympathetic to the views of those advocating interviewing all members of a household, there were a number of countervailing views that emerged. First, the perceived data gap in U.S. statistics included a clear interest in knowing how individuals spend their time (in addition to the obvious interest in household decision-making). Many of the potential uses of time-use data cited at the BLS/MacArthur

conference did not require collection of data from more than one individual in the household. These potential estimation objectives included valuing non-market work; verifying and interpreting data series from other extant surveys such as hours of work; measuring commuting time, time spent providing child care; time spent investing in training and education; and long-term shifts in our society toward (or away from) leisure time, among others.

Second, it was unclear why information on the time-use of individuals, 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. While this would provide richer direct behavioral data, there was no clear and compelling econometric argument advanced that the gain in information resulting from interviewing every household member would be worth the potential loss in sample yield under a fixed budget.

For example, if one were to adopt the purist view that interviews with all household members are necessary to correctly inform household bargaining models, the possibility of low response rates from multiple family members (no matter what data collection methodology is used) called into question the efficacy of such an approach. That is, at what point does 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? And finally, there is also a potential for systematic bias to be introduced in terms of who tends to be a respondent and who tends to be a non-respondent in surveys that attempt to interview all household members. In particular, the possibility exists that households that allocate their time so that

someone is always home with the children will have a higher likelihood of missing an interview with the individual(s) who is more likely to be not at home.

The choice of selecting a stratified probability sample from the Current Population Survey (CPS) was designed to enrich the demographic and economic information available on each individual, to maximize the efficiency of the sample design, and to minimize individual respondent burden (as many pieces of information would be imported rather than asked in ATUS). The CPS was also compatible with the choice of CATI and was a relatively inexpensive sample frame, as recent phone numbers are available for 95% of CPS households. And, by the end of their time in CPS, respondents are accustomed to answering interview questions by phone.

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

The initial process

Given its charge, the working group concentrated on the following elements of survey design in preparation for the NAS workshop. In particular, we examined:

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

The NAS workshop can be viewed as a detailed first draft of our thinking about many of the elements of what has become the American Time Use Survey. After the workshop, we began anew on

some of these elements, reconfirmed our thinking on others, and filled in gaps that were not considered as of the NAS workshop. For the purposes of this exposition we consider each element of the survey in turn, and trace the development of our thinking from the NAS workshop through to full production in January 2003. Table 1 provides a concise summary of each of these elements. A more detailed discussion of each element follows.

Concept		Analysis as of NAS	Changes as of full production
<u>a. Survey estimation objectives</u>			
	<i>Primary sampling objectives</i>	Draw a nationally representative sample of individuals 16 years of age or older to generate quarterly estimates of the proportion of the time spent in 1 and 2-digit activities for an average week, weekday, and weekend.	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 15-year olds, and youth were included in many other countries' time-use surveys
		These estimates would be presented for the entire U.S. population and separately for selected demographic and labor force groups.	Unchanged
	<i>Secondary sampling objectives</i>	Generate annual estimates of the time spent in 1-, 2- and 3-digit activities for an average week, weekday, and weekend.	Unchanged
		Present these estimates for the entire U.S. population and separately for selected demographic groups.	Largely unchanged. The original team listed a series of demographic breaks, but in production we acknowledge that publication detail depends on the sample that falls in each demographic and activity group cell.
<u>b. Periodicity</u>		Continuous for 14 months (2	Continuous indefinitely, with

		months of prefielding followed by 12 months of collection. Then repeat periodically.	4 months of prefielding before production.
<u>c. Frame of reference for reporting time-use activities</u>		Preceding day	Unchanged
	<i>The designated Day</i>	Each respondent would be assigned a designated interview day for reporting about the previous day	Unchanged
	<i>How to handle noncontact on the designated day</i>	Call back exactly one week later, asking respondent to recall the prior 24-hour day. Repeat call back each week for up to 4 weeks.	Reference period was changed to up to 8 weeks. This element was studied at length. As discussed below, we carefully examined the possibility of substituting freely among Mondays, Tuesdays, Wednesdays, and Thursdays.
<u>d. Sampling</u>			
	<i>Choice of sampling frame</i>	Household addresses from Month-in-sample 8 of the Current Population Survey	Unchanged
	<i>Unit of observation</i>	Randomly selected individual from each household	Unchanged
	<i>Sample size</i>	Sample required to achieve 2,000 completed cases per month at 70%.	Sample is unchanged at about 3,250/month. However response rates are averaging around 62%, yielding 1,750 interviews/month.
	<i>Periodicity of sample draw</i>	Monthly	Unchanged
<u>e. Questionnaire</u>			
	<i>Core time-use questionnaire</i>	Same as in the original BLS pilot, modeled after the Statistics Canada approach. Respondents are asked to report activity by activity in sequence. For each activity reported, respondents are asked who they	In cognitive testing, problems occurred with the “Who was with you?” question when people were away from home (e.g. at the shopping center. The question was changed to “Who was in the room with

		were with, how long the activity lasted, where they were	you?” at home and “Who accompanied you?” when away from home.
	<i>Simultaneous activities</i>	For each reported activity, respondents also asked if they were doing anything else at the same time	The implicit assumption in the NAS report was that simultaneous activities would be collected and coded. As of full production, simultaneous activities are being collected only when offered by respondents and will be not be coded, except as needed for research purposes. Cognitive work is being conducted to determine how to systematically collect simultaneous activities in the CATI interview.
	<i>Summary questions</i>	Ask respondents to identify each recorded activity for which they were paid.	Summary questions were significantly expanded. Four types of summary questions are included in full production: questions on child care, paid work, volunteering, and time spent away from home in the prior month (“missed days”).
<u>f. Updates to CPS questions</u>		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 for the respondent.	Family income is not updated. Labor force status is updated except for the series on reasons for being out of the labor force. Respondents are also asked about whether they have children under 18 who do not live with them.
<u>g. Modules</u>		Allocate 5 minutes of the questionnaire to special topic modules. Do not specify the topics for these modules, just use this as a placeholder.	Unchanged. Several Federal agencies and some private groups have expressed interest in purchasing 5 minutes of module time.

<u>h. Activity coding</u>		Adopt a variation of the coding system from the time-use survey of the Australian Bureau of Statistics	Significant changes have been made as a result of extensive testing on coding structures.
<u>i. Survey operations</u>		Recommendation of a pre-test and 3 months of pre-fielding before full production	An extensive operations test was done, as well as a 2-month dress rehearsal (pre-test) and 4 months of pre-fielding.
	<i>Target response rate</i>	Adopt a conservative 70% response rate, conservative as compared to the 88% response rate in the Canadian time use survey that uses a CATI methodology	Unchanged
<u>j. Data dissemination and publication plans</u>		Initial table shells were developed	Publication tables have been developed for specific subject areas and a system is being built to generate them (e.g. unpaid work, social contact, child care). Public-use data files are being specified according to formats recommended in Harvey (1999).

The following elements were not specified in the NAS workshop but have been or are still are being developed for full production: choice/style of software for each system; sampling plan; weighting plan; variance system; imputation and edit programs; and training and certification procedures for coders and data collectors.

Discussion at the National Academy Workshop

The presentation of the BLS proposal was met with strong overall support, despite serious misgivings by some on our choice of one individual as the sampling unit, and/or our decision to interview

each respondent only once. 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, we felt that the arguments could be viewed as survey methodological preferences as opposed to absolute statistical requirements. Collecting one-day diaries still would enable the production of all the desired estimates (albeit likely with higher variances than with a multiple diary and/or multiple household member approach). The NAS workshop report endorsed the BLS model, recognizing that the lack of any national time-use survey was an important gap to fill regardless of disagreements over methodological issues.

4. Getting ready for full production: the 4th developmental phase

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 fully operationalized. Interagency work on the survey began in earnest and joint BLS/Census 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 case management at the call center. BLS worked with the budget office to approve a January 2003 start date.

Many activities occurred between funding in December 2000 and production in January 2003.

Three primary ones were:

- an extensive operations test in the summer of 2001 – the “operations field test,”
- a test of the questionnaire and operations with live cases in mid-2002 –the “dress rehearsal,”

- and early fielding of the survey to work out remaining kinks from September-December 2002—“prefielding.”

These activities will be referred to in the subsequent section. During this period, BLS and Census continued to receive advice from outside groups, particularly the Federal Economic Statistical Advisory Council (FESAC) and the International Association of Time Use Researchers (IATUR).

a. Survey estimation objectives.

The primary and secondary survey estimation objectives listed in Table 1 remained essentially unchanged, except for the expansion of the survey scope to include 15-year olds. The focus on estimation changed from these overall objectives to classifying subject areas of interest and creating table shells. Shells were developed based on research areas brought to our attention in the conferences mentioned above, in other countries' time-use publications, and in meetings with future data users. For example, tables have been developed that will include data on time spent providing child care, traveling, doing paid or unpaid work, volunteering, and participating in leisure activities. Tables will also be developed on who people spent time with (e.g. “social contact”) and where they were (e.g. to examine travel patterns or influence emergency management planning). A processing system is being built at the Bureau of Labor Statistics to generate these tables.

b. Periodicity of the survey.

The NAS report recommendation to draw the sample monthly did not change, but the survey administration plan was further developed. In the NAS 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, as a continuous funding stream implies a continuous level of survey collection activity. Instead of advocating a time-use survey that would be collected, say every 5-10 years, as is typical in most countries, we recommended ultimately a continuous survey with sample building over time to enable more robust estimates and time-series analysis.

There were discussions about whether the survey should be fielded evenly across the year and within months or whether sample should be introduced in some months only or front-loaded at the beginning of selected months. For operational ease, and to represent all weeks and months equally across the year, we decided to introduce each month's sample evenly across 4 weeks. Each case would be called for up to 8 weeks (see operations, below, for rationale). This created overlapping 11-week field periods. Estimates, however, would be based on the date about which respondents were interviewed (e.g. 1st quarter estimates represented interviews *about* Jan 1- March 31, regardless of the sample month the cases were introduced).

c. Frame of reference for reporting time-use activities.

The notion of asking someone to report, activity by activity, about the preceding day was a given at the beginning of our deliberations, as it was the norm in international time-use data collection. We also quickly settled on the idea of assigning a designated interview day to each respondent, since this would allow us to control the probability of selection. However, an ongoing concern was how to handle the situation of a respondent that missed his or her designated interview day. Under the theory that the probability of missing an interview day may be correlated with the use of time on the designated day (and the interview day)--and also because we wanted to control the representativeness of days across the week--we settled on a procedure that did not allow convenience substitution of interview days by respondents. Instead we specified a protocol to try again the following week (on the same day of the week), and repeat

this for some number of weeks. We initially recommended four weeks, consistent with Finland's time-use survey protocol; the NAS report also recommended further examination of the optimal period for the U.S. survey (Niemi, Kiiski, and Kiikkanen, 1986; BLS, 1998).

One justification for not allowing convenience day reporting was theoretical work done by a member of our design team. Jay Stewart (2002) performed a series of simulations designed to evaluate the potential biases associated with convenience day and designated day reporting. He specified a known set of activities and their distribution in terms of time, along with an assignment of two discrete probabilities of successful contact for any given day (easy and hard). In addition, he specified two types of postponement strategies to accompany designated day reporting; one that contacts a respondent a week later (for up to 4 weeks) and the other that assigns a preset pattern of substitution among Tuesdays through Fridays for interview attempts (assuming that Mondays to Thursdays as reference days are substitutable).

He found that any strategy involving convenience day sampling overestimated time spent in activities away from home and underestimates time spent in activities performed at home. There was no activity bias generated in designated day strategies, and the number of contact attempts required did not differ between the two alternative postponement strategies.

While Stewart's research did not provide theoretical evidence to support an advantage to allowing substitution, the Census Bureau was concerned about the very difficult potential survey requirement to attempt to contact a specific person on a specific day and then postpone subsequent interview attempts by a week each time. It was theorized that contact would be far more successful if respondents had multiple eligible interviewing days closer together. In particular, the Census Bureau was interested in the possibility of substituting across weekdays as designated interview days. We ruled out allowing consecutive day interviewing (e.g. if one was supposed to be interviewed on Tuesday about Monday and

was unavailable, they could not then be interviewed on Wednesday about Tuesday) because the probability of missing them on the first day was correlated with what they did on that day (e.g. they were engaged in activities primarily done away from home). If we had allowed consecutive day interviewing, again, we would overestimate activities done away from home. We also ruled out substitution of any days other than Monday through Thursday, as earlier time-use data sets showed that Fridays, Saturdays, and Sundays tended to have unique activity profiles (Stewart, 2000).

The final decision about substitution was made following our operations test (described later in this section). In one test group, respondents were assigned to either a Tuesday/Thursday or a Monday/Wednesday group (e.g. they could report on either of those days), doubling the number of their eligible interviewing days. As a preview of our findings, the operations test showed that the availability of a second eligible day during the same week increased response rates by about 4 percentage points. The technique was not implemented, however, because it introduced operational complexity, particularly to the case management software, and, more importantly, led to large differences in representation across days of the week.

d. Sampling

The early basic framework for the sampling plan was developed and presented in the report to the National Academy of Science's workshop. The sample frame was identified as households leaving the Current Population Survey who had successfully completed their eighth interview. The sampling unit, as discussed above, is one individual in the household. The statistician for the working group determined the minimum, maximum and optimal sample sizes for the survey. These estimates were used ultimately to estimate a survey budget (Robison, 1999).

Earlier decisions provided some boundaries for the discussion. The designated-day methodology forced us to think of a sample where individuals would be drawn for interview and then retained for several weeks if they were not available on their designated day of the week. Based on experience of Statistics Canada, which achieved an 88% response rate, a conservative 70% target response rate to the time-use survey was set.

Using a subset of the CPS provided many sampling benefits:

- We could identify and select by survey respondents' demographic characteristics in advance
- We knew the maximum potential monthly sample (72,000)
- We knew how the sample was constructed and could make revisions that would reduce costs but not substantially reduce standard errors
- We understood standard errors on CPS data as compared to a simple random sample

Using unweighted CPS sample counts from month-in-sample 8 files and time-use distributions reported in Juster and Stafford (1985) to develop parameters for estimating standard errors, Robison (1999) estimated that the minimum sample size for a time-use survey to generate reliable estimates for the major subpopulations in our primary estimation objectives to be 12,000/year (he assumed this would produce 9,000 completed interviews). He estimated that an additional 12,000 sample (9,000 interviews) would be required to produce estimates for a smaller subpopulations specified in our secondary objectives. In general, he estimated that 1,000 sample cases (750 interviews) in any particular cell would be needed to produce reliable estimates. To be conservative, the BLS NAS report recommended a slightly higher target for the sample: 21,000 completed interviews/year (BLS, 1998).

After funding, an interagency statistics team was formed to refine and finalize the sampling plan. Because the CPS was a household sample, stratification was done using household variables and was based on ensuring that reliable estimates could be made for minorities and for people with and without children, as the presence of children is usually highly correlated with time use. The following stratification variables and probabilities of selection were chosen:

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

The sampling rate for each stratum differs 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 NAS workshop led to the ultimate recommendation that we collect 2,000 completed interviews per month or 24,000 completed interviews per year.

e. The questionnaire.

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

ATUS questionnaire specifications were due to Census shortly after the survey was funded, as software specifications, programming, and testing would take a long time to complete. The production plan included a summer 2002 dress rehearsal that required instruments be ready well before full production. A draft questionnaire was submitted quickly, but work to refine the questionnaire continued through dress rehearsal, almost to full production. This occurred for several reasons: 1) a commitment was made that no questions would go on the ATUS that had not been successful in cognitive tests; and 2) results from tests and from the dress rehearsal led to many rewrites and retests of some questions; and 3) results from coding tests indicated that additional questions were needed to clarify some activities for coding. (See summary question section, below.)

Simultaneous activities. The enhanced version of the 1997 pilot questionnaire asked respondents not only what they were doing, but also whether they were doing anything else at the same time. The pilot study showed that this version picked up more non-market work than the standard version that did not directly address simultaneous activities.

At the time of the NAS workshop, we decided to recommend inclusion of this data element, although we were still concerned about the burdensome nature of asking this question after each recorded activity. Cognitive interviews indicated that many respondents wanted their simultaneous activities included. However, many had difficulty attaching time durations to their secondary (or tertiary) activities. In addition, it was not clear how to program the diary software to accept this additional information, and modifications to the diary section were always complicated as they involved time computations. A decision was made to collect but not code simultaneous activity reports (the duration is collected for primary activities, but only the activity report is collected for simultaneous activities) in the first year of production and to put simultaneous activities collection research at the top of the research agenda during the first year of full production. We are currently working with an outside vendor to develop the simultaneous activities research protocol.

Paid work summary questions. Midway through the field period of the 1997 pilot test of the time-use survey, researchers realized that the collected information was insufficient for accurately identifying and coding informal activities done for income, such as making crafts for sale or babysitting. In order to supplement the existing information, a *summary question*—that is, a question that asked respondents to review the list of activities reported in the diary in their mind and identify one or more characteristics about each activity--was designed, asking respondents 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 and recommended the inclusion of a similar probe for coding purposes.

The questionnaire design team adopted this recommendation, but 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 it did not clarify some activities that were done for one’s job but 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?” But customized probes could not be developed for all contingencies.

A second phase of paid work tests was designed to determine whether the ATUS could clearly identify work-related activities, particularly for individuals who worked in nonstandard work environments or had nonstandard work hours, as they were more likely to report work in vague terms. 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)?” followed by “Which ones?” The question is asked twice for multiple jobholders—once each for the main job and for 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 (e.g. taking a client to play golf) identified by respondents as

done for their job or business. Others would be coded as work, as they were part of the respondent's job, but were simply not reported as "work" in the activity description (e.g. grading papers). And some activities would be restricted from being coded as work, such as smoking and sleeping.

Child care summary questions. Several rounds of testing were required in order to develop child care summary questions that would enable BLS to accurately measure time respondents spent with children in their care. The 24-hour diary successfully captured "active child care" activities, such as feeding or bathing children, but the "Who was in the room with you?" question did not sufficiently identify "passive" or "secondary" child care. For example, a respondent may have been alone in the kitchen, but responsible for children elsewhere in the home or yard. Or, similarly, a respondent may have a child in the room but may not be responsible for them at that time.

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

The team spent a great deal of time discussing the concept we were trying to measure. We eventually agreed that passive care of children was characterized by the respondent being in the general vicinity of at least one child under 13, and specifically near enough to that child to provide assistance if necessary. We knew this language would not make a good survey question. To determine wording that would elicit responses in line with this definition of passive care, we conducted two focus groups. 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 child care activities. We tested the other two terms using cognitive interviewing and ultimately chose the “in your care” wording (Schwartz, 2001).

Dependent adult care. At the same time that the first round of questions were asked about child care, a set of cognitive interviews was done with caregivers in an attempt to determine ways to measure time spent in dependent adult care activities:

“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].”

In testing, the question had numerous problems, chief among which was the recognition that the terms “care”, “adults” and “disabilities” were unclear to and interpreted differently by respondents. Further clarification and testing of wording to convey the meaning of these terms was needed. In addition, some respondents did not like the “should not be left alone” language for adults, as 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 implying the adults were wrong/untrustworthy and the latter having too clinical a tone.

Recognizing the complexity of defining questions to measure disability—an ongoing multiyear process still taking place at BLS for a supplement to the Current Population Survey—a decision was made to defer the development of questions to measure care of disabled adults to a later date. Preliminary research work on measuring “eldercare,” a more restrictive concept, is currently being done by the BLS Office of Survey Methods Research for the ATUS. This work includes a literature search, expert consultation, and the development of questions for cognitive testing on ATUS.

Missed Days summary questions. Some concerns had also arisen about a systematic bias that would occur in the survey: since respondents were asked about “yesterday,” activities done on trips of two or more days would not be collected. Debates initially focused on how we might be able to get this information, such as by asking a subset or 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 done during recent trips. However, carrying on a separate protocol to get these data would be costly and asking about detailed activities on recent trips would present recall problems.

We eventually 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 enabled the researchers to evaluate accuracy of reports about trips. They learned 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 (Schwartz, 2001). Based on the results of these tests, BLS decided to ask that respondents only report on trips taken during the month prior to their first eligible designated day.

As a result, the following question(s) was(ere) 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 October. What was the purpose of that trip?...How many nights were you [insert purpose]?”*

Volunteering. During dress rehearsal and pre-fielding, Census Bureau coders reported difficulty in distinguishing between certain care and socializing activities and volunteering activities. To clarify the distinctions, it became clear that a summary question on volunteering was required. The CPS had recently 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 ATUS took the CPS question wording and modified the reference period from the prior year to “yesterday,” and the following question was added to the ATUS: *“Now I’d like to ask you a few questions about volunteer activities—that is, activities for which people are not paid, except perhaps 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?”* For those activities identified as volunteer activities, a follow-up question is asked about the name or type of organization. This will enable the coding of activities by type (e.g. food preparation volunteering) and by organization grouping (e.g. youth organizations).

f. Current Population Survey 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. Some pieces of information relevant for time-use analyses, such as employment status, could change between the last CPS interview and the time-use interview. Prior to the NAS workshop, we formed a subgroup to investigate which elements of the CPS interview should be updated. We also 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. And, we considered respondent burden when addressing these questions.

We 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 for the respondent. After funding, we examined the CPS questionnaire and skip patterns in detail to understand the impact of these decisions on software development. All the questions and skips we included in ATUS would need to be reprogrammed for the ATUS data collection instrument. ATUS team members from the Census Bureau requested that we exclude unnecessary sections to minimize programming burden. We decided not to update the family income variable as it is only a set of ranges and we had decided already to update the respondent's earnings. We also excluded the branch of labor force status questions that distinguished reasons for being out of the labor force. We kept jobsearch methods of the unemployed, however. And, we asked for new information to be added to the household roster—information on the age and sex of any children of the respondent who did not live with him or her.

g. Modules

Modules—a series of questions on a specialized topic added at the end of the questionnaire—hold the promise of allowing researchers to more fully explore social and economic issues related to time use. They could also fill in limitations of some of our decisions. For example, module questions could be used to measure family allocation of time, asking the respondent questions about the time use of household members. In the NAS report, we 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, we agreed that no modules would be implemented before January 2004 (one year into production).

Our thinking on modules basically remains unchanged since the NAS workshop. We have refined the criteria in a module policy, which includes requirements such as: 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 (BLS, 2003).

h. Activity coding

The ultimate value of time-use data depends on the breadth, depth, and consistency of the activity classification system. Restrictions on disclosure of potentially identifying respondent information prevent the Census Bureau from including activity verbatim reports provided by respondents (e.g. "I washed my hair") on public-use data files. Rather, each activity is assigned a 3-tiered activity code using a detailed classification system and accompanying rules.

Stinson (1999) described the working group's early recommendations regarding activity codes at some length. She 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 1960's (Szalai, 1972). 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 NAS report recommended a slightly modified version of the Australian system, as the system was quite detailed and best reflected the type of economy and activities done in the United States.

After funding, an interagency coding team further customized the system to include activities unique to the U.S., to remove some unique to Australia, and ensure that activities were consistent with a 4-fold typology of time developed by Dagfinn Ås (BLS, 1998). This lexicon was then submitted to IATUR members and to a team of BLS cognitive psychologists for review. During the review process, many took issue with the 4-fold typology of time that grouped activities into "necessary," "committed," "contracted," and "free" time. Most thought the overall framework was appealing, but 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 can easily combine them to represent these concepts.

The coding team found that international comparability across systems was not as simple as we had expected. Even such seemingly straightforward activities as eating were handled differently in different countries. For example, some countries coded eating in restaurants as socializing ("free" time) while others coded such eating as eating ("necessary" time). Andrew Harvey, president of IATUR, confirmed that international comparability at detailed levels did not exist, as no complete crosswalk had been built across systems. Two systems that harmonize time-use data at aggregate levels across countries have been developed, however, and we are working with the designers to be sure that U.S. data are included⁴

The removal of the comparability constraint freed up the coding team to change the focus of lexicon revisions to coder usability (ability to locate the right code) and enhancing analytical value. To address the latter, the proposed 3-tier, 3-digit system--which allowed 9 categories within each tier--was

⁴ The Multinational Time Use Study was launched in the 1980s by Professor Jonathan Gershuny of the University of Essex. The idea was to create a cross-national and historical archive of time use datasets. All the datasets in the archive have been harmonized into a common set of variables including time-use activities and various demographic and socio-economic characteristics of respondents. The harmonized file currently contains 35 different surveys, from more than 20 countries. It also contains 41-time use activity codes. The development of these common activity codes was itself constrained by the richness (or lack of) activity codes in the original surveys. In recent years, an alternative harmonized time-use activity schema has been

replaced with a 3-tier, 6-digit system, that allowed up to 99 subcategories (most categories still have 10 or fewer).

To implement recommendations from the team of research psychologists, the ATUS team conducted several usability tests of the coding system with Census Bureau coders. We developed a training package on the proposed system and worked with the Census Bureau to identify 6 individuals in their Jeffersonville, Indiana facility that had coded on other surveys. Ultimately, 3 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 (BLS, 2002). These tests were very productive and led to both small and large changes in the classification system. A few of them included:

- 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 4 largely purchasing categories, such as "consumer purchases" and "professional services" [See box with lexicon categories].
- Removal of the "media use" category as many did not think of TV, radio, etc. together. Separate categories were developed for each type of media use.
- Removing and revising ambiguous examples under various categories
- Agreeing on the best location / conceptual definition for many difficult-to-code activities like "looking for things" and "waiting for [x]"

In the fall of 2001, BLS worked with Westat, a private research firm, to conduct an additional, larger-scale test. The test was designed to measure coding accuracy and efficiency across time (returns to experience) and to evaluate the BLS coding training. The test involved 9 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

developed as part of the Harmonised European Time Use Study (HETUS), based on some 20 surveys having been carried out since 1999.

⁶ 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.

recommendations to improve the classification system, the coding software, and the training. Most have been adopted (Westat, 2001).

The coding tests, as well as work at BLS to specify coding rules and analytical tables, routinely pointed out difficult-to-code activities. The list is too long for this chapter, but some of the most difficult categories included travel, waiting, packing and unpacking things, work-related activities, and computer use. After the tests, much work was done during 2002 to address these issues, including revising coding categories, subcategories, and particularly rules and job aids for training. A review of the proposed coding system, including how it compared to several other countries' systems, was completed by Anne Gauthier (2001) that also led to a number of important changes in the classification system. As mentioned earlier, difficulties distinguishing between care, socializing, and volunteering (e.g. for "I read to my blind neighbor") 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 a verification and adjudication process that required a second coder assign a code to each activity and an adjudicator to rule on the correct code. Feedback from the dress rehearsal and prefielding activities was also used to modify the system, right up to production.

ATUS coding lexicon: First-tier categories	
01	Personal Care
02	Household Activities
03	Caring for and Helping Household Members
04	Caring for and Helping Non-household Members
05	Work and Work-Related Activities
06	Education
07	Consumer Purchases
08	Professional and Personal Care Services
09	Household Services
10	Government Services and Civic Obligations
11	Eating and Drinking
12	Socializing, Relaxing, and Leisure
13	Sports, Exercise, and Recreation
14	Religious and Spiritual Activities
15	Volunteer Activities
16	Telephone Calls
17	Traveling

i. Survey operations

A few of our survey operations were discussed in the development of the sampling objectives and were suggested in the NAS report. Telephone interviewing was a starting assumption. We subsequently recommended a designated-day methodology with 4 weeks of callbacks, doubling the length of the 1997 Pilot study reference period. These recommendations provided a beginning of operations, 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 pre-designated day. While the 1997 Pilot study provided guidance to extend the calling period, it was not clear how many calls should be made over how many weeks to achieve the 70% response rate target, and how different mail-out or day-of-week substitution techniques might affect results.

The team designed 3 experiments that were run concurrently by the Census Bureau in April-June 2001 using 3,200 cases from the outgoing rotation groups of the CPS. 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 codeable activity data. A paper questionnaire was developed that included a short diary (from 4am to noon) and debriefing questions designed to provide insight into contact and response. Detailed test designs are described in Piskurich et. al. (2001). The effect 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 among 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 to set an appointment. Others were called only on their interview day only.

Field visits versus calling: Some respondents were called for up to 8 weeks; others were called for up to 4 and then visited for up to 4 more.

Incentives: Respondents were divided into 3 incentive groups. They received debit cards for \$0, \$20, or \$40 and were given the PIN number if they completed the survey.

The first of the 2 tests covered all of the variables above and were done with the 95% of the sample for whom the Census Bureau had telephone numbers. An additional study was developed for the 5% 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 were called for up to 4 weeks, and visited for up to 4 more to secure an interview.

Response rates varied a great deal across the treatment groups. The highest response rates were achieved with larger incentives and with field visits, both very expensive operations. Incentives also sped response; for example, a 70% response rate was achieved in only 2 weeks with a \$40 debit card; with no incentive, a 69% response rate was reached after 8 weeks. To analyze the data and make operational choices, contact and 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 very effective in reaching respondents, and the costs were relatively small.

No field visits. Field visits increased response by about 4 percentage points as compared with a full-CATI operation. However, they were prohibitively expensive and would require training interviewers across the country rather than in one location.

No proactive appointment setting. We were surprised to find that calling in advance to set an appointment did not increase response. It did, however, increase costs.

Incentives to nonphone households only. When costs came in, it was clear that BLS would not be able to afford incentives for each case. The shorter field period to get to 70% 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 telephone households. Incentives were chosen for nonphone households, however. A \$60 incentive induced 41% to complete the survey. Because this group included underrepresented demographic groups and survey advisors including the FESAC felt strongly that they should be included in the sample, a \$40 incentive was implemented for full production.⁶

No substitution. It was also surprising that the availability of a second eligible day during the week only increased response by only about 4 percentage points. The substitution technique was not implemented because it introduced operational complexity and led to large differences in representation across days of the week.

j. Staffing and training

In the NAS report, we asserted that it would be desirable for interviewers to have experience with the coding system. As we prepared for full production, we became more convinced of the necessity of this approach. As the survey questionnaire and coding lexicons were developed, the number of probes that were needed in the interview to allow correct coding of activities expanded significantly. A good

interview required strong listening skills to remember these probes in conducting the back-and-forth “conversational interview” required by ATUS. It also required an interviewer who understands the probes and their relationship to coding.

Most surveys have separate interviewing and coding staffs, and we were concerned that it might be difficult to find people who were inclined to do both tasks. This was tested for the first time in the dress rehearsal; debriefings with Census Bureau interviewers indicated that they thought activity coding experience was not only important (and very enjoyable), but was critical to conducting good time-use interviews. For similar reasons, the oversight team also thought that ATUS should have a dedicated staff, but the prohibitive cost of maintaining a dedicated staff full time necessitated allowing interviewers to work on other surveys as well as ATUS. However, we continue to require that interviewers also code, and extensive training requirements for both activities are mandatory for anyone working on the survey.

The difficulty of the interview and coding influenced our approaches to training. Based on our training experiences in the dress rehearsal and prefielding, we learned that training worked best if training on the coding rules preceded the interviewing training. We contracted with a vendor to develop a training package that the Census Bureau facility could deliver without assistance from headquarters. The developer included extensive audio in the computer-based training, as the CATI interview required skills in listening and immediate, customized probing.

j. Data dissemination and publication plans

During the development of the NAS report, the working group examined reports from other time-use surveys and reports of CPS data to and developed some illustrative table specifications at that time. Work on tables continued after funding and is still ongoing. Meetings with advisory groups and with

users as part of our systems requirements gathering meetings provided information on the types of analyses users planned to do with ATUS data.

The ATUS team is continuing to develop table shells in subject area groupings, such as child care, unpaid work, travel, and social contact, and will soon have data to populate them. This tables we actually publish will depend on cells sizes and variances, and will be released to the public in mid-2004, probably through a news release. Some tables will be made available to users on line at the BLS web site (www.bls.gov/tus). Data will be presented using a variety of concepts, such as time spent doing an activity during an average day, weekday, or week, or the proportion of the population that engaged in a particular activity. Most tables will include tabulations by some demographic characteristics or by labor force status.

Many time-use users will be interested in microdata files. The data file formats are also still in development, but will most likely adhere to recommendations by Harvey (1999) that call for the following three types of files:

File type	Description
Respondent summary file	Contains summary information for each respondent (such as age, sex, cumulative time spent in various activities, at various locations).
Episode file	Contains episode-level information for each activity (such as activity code, who was present, where the activity occurred.)
Time Points file	Contains the activity codes for activities performed at pre-specified intervals during the day (e.g. at 5pm, the respondent was eating; at 5:15pm, the respondent was doing the dishes, etc.)

The current plan is to produce SAS or ASCII microdata files for distribution via CD-ROM or through the BLS web site.

Concluding remarks

This paper traces the development of the American Time Use Survey from its inception as an issue of statistical policy interest in 1991 to its implementation in January 2003 as a new monthly survey sponsored by the Bureau of Labor Statistics. The development of the American Time Use Survey represents a coalescence of work including academic inquiry and debate, survey methodological design, operational testing, production management, and by no means least important, a strong and growing consensus among government agencies as to the need for these data.

This process involved the work of a great many individuals, too numerous to name here. Some contributed through advisement at the MacArthur or NAS conferences, FESAC meetings, through contracting relationships or through IATUR projects. Others worked at BLS or the Census Bureau to run tests, build and test software, and conduct interviews. Former BLS Commissioner Abraham, under whose leadership much of the early work and the securing of funding was completed, was absolutely critical to this endeavor, as was the continued interest and support of the National Academy of Sciences.

We attempted to trace the most important elements that influenced the development of what has become the American Time Use Survey. This included a retelling of the 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. We hope that we have been true to the historical record, and any mistakes in emphasis or fact, while unintentional, remain the province of the authors. The American Time Use Survey has the potential to fill many gaps in our knowledge of our society and economic processes, and we are proud to have been part of its development.

Appendix A: The Software

Both Microsoft's Visual Basic and Statistics Netherlands' Blaise were considered as possible software packages for the CATI instrument, as they both satisfied the most sophisticated software requirement: development of an activity grid that computed times. The selection of the Census Bureau as the agency to implement the survey drove this decision; the Census Bureau was moving to Blaise as their standard for CATI instruments.

There was some concern that the diary portion of the instrument would be difficult to program in Blaise, and Census was relatively new to the software. BLS brought in an experienced Blaise programmer from Westat to prototype the diary portion of the questionnaire while specifications for the rest of the instrument were written at the Census Bureau. While there were many gains to this prototyping process, one of the more difficult management lessons was learned here. The rapid prototyping process used with Westat produced a good instrument, but it was very difficult to pass this "third party code" from Westat to the Census Bureau. In hindsight, a team including representatives from both agencies might have been a better approach. The specification writing and programming process was also instructive, the main lesson being that programmers should be "at the table" during the requirements development phase, to provide suggestions for efficient solutions and to buy into the questionnaire design. Later in the process, Capability Maturity Model (CMM) techniques were employed that dramatically improved the questionnaire implementation process.

REFERENCES

Baile, Lorna (1997), Remarks made in “Session VI: Time-Use Surveys: Where should the BLS go from here?” in the written summary for the “Conference on Time Use, Non-Market Work, and Family Well-Being,” co-sponsored by the Bureau of Labor Statistics and the MacArthur Network on the Family and the Economy, November 20-21, 1997, edited by Lynn Hatch.

Bureau of Labor Statistics, Time-Use Survey (TUS) Working Group. (1998, October 23). A Report on the Feasibility of Conducting a Time-use Survey, presented at the National Academy of Sciences workshop on Time-Use Measurement and Research.

Bureau of Labor Statistics and the MacArthur Foundation (1997), “Agenda for the Conference on Time Use, Non-Market Work, and Family Well-Being,” November 20-21, 1997, Washington D.C.

Bureau of Labor Statistics. (2002). Activity Coding in the American Time Use Survey: a report of 3 tests. BLS internal report.

Gauthier, Anne (2002, August 21). BLS 2003 Codes of Activities: Comparisons Over Time and Across Nations. Draft report

Harvey, Andrew. (1999). Guidelines for Time Use Data Collection and Analysis. Chapter 2 in *Time Use Research in the Social Sciences*, edited by Pentland et. Al. New York: Plenum Publishers.

Horrigan, Michael et al (1993), “The Measurement and Valuation of Unpaid Work”, BLS Working Group Report, Washington D.C.

Joyce, M. and Stewart, J. (1999) Why Collect Time-Use Data? Paper presented at American Statistical Association meetings, Washington DC, August 1999.

Juster, F. T. and Stafford, F. P., [Eds.] (1985). *Time, Goods, and Well-Being*. Ann Arbor, MI: Institute for Social Research, the University of Michigan.

National Academy of Sciences, “Time-Use Measurement and Research: Report of a Workshop,” Committee on National Statistics, National Research Council, 2000.

Niemi, I., Kiiski, S. and Kiikkanen, M. 1986. *The Use of Time in Finland*. Helsinki: Central Statistical Office of Finland.

Nordhaus, William (1997), Remarks made in “Session VI: Time-Use Surveys: Where should the BLS go from here?” in the written summary for the “Conference on Time Use, Non-Market Work, and Family Well-Being,” co-sponsored by the Bureau of Labor Statistics and the MacArthur Network on the Family and the Economy, November 20-21, 1997, edited by Lynn Hatch.

Piskurich, K., Nelson, D. and Herz, D. (2001). *Maximizing Respondent Contact in the American Time Use Survey*. Proceedings of the 2001 American Association of Public Opinion Research conference.

Robinson, J. and Godbey, G. (1997). *Time for life: The surprising ways Americans use their time*. Pennsylvania State University Press.

Robison, E. (1999). Sampling and Reporting in Time-Use Surveys. Paper presented at American Statistical Association meetings, Washington DC, August 1999.

Schwartz, Lisa. (2002, February). “The American Time Use Survey: cognitive pretesting” *Monthly Labor Review*, pp. 34-44.

Schwartz, L., Herz, D., and Frazis, H. (2002, February). “Measuring intrahousehold allocation of time: response to Anne E. Winkler.”

Stinson, L., Becher, A., Forsyth, F., Levin, K. (1998). Using a time-use approach to measure the frequency and duration of non-market work. Final report on the 1997 BLS Pilot Test.

Stinson, Linda (1999) Measuring how people spend their time. Paper presented at American Statistical Association meetings, Washington DC, August 1999.

Stinson, Linda. (2000). “Report on Cognitive Testing Phase 1: The American Time Use Survey Summary Questions.” (Bureau of Labor Statistics internal report)

Stewart, J. (2002, August 21). Assessing the Bias Associated with Alternative Contact Strategies in Telephone Time-use Surveys. Internal BLS working paper.

Stewart, J. (2000, May 26). Alternative Indexes for Comparing Activity Profiles. Internal BLS working paper.

Szalai, A. (1972). *The use of time: Daily activities in urban and suburban populations in twelve countries*. The Hague: Mouton.

United Nations (1995, September). *Platform for action*. Report from the Fourth World Conference on Women. Beijing, China.

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

Winkler, Anne E. (2002, February) "Measuring time use in households with more than one person," *Monthly Labor Review*, pp. 45-52.