

# **A Study in the Process of Planning and Designing a Survey Program: The case of time-use surveys at the BLS**

December 22, 1999

Michael Horrigan

Diane Herz

## **Abstract**

In this paper, we describe the process that a BLS working group used to develop recommendations for a U.S. survey on time use. We discuss the following elements, each of which are important in all survey design projects: the public policy environment and overall background surrounding the survey, starting assumptions and constraints, developing survey objectives, determining the appropriate unit(s) of measure and methods of data collection, questionnaire design, and designing a survey sample.

The task of the working group was twofold. First, the report needed to provide the basis for a decision on technical feasibility of a time-use survey. The recommendations needed to be specific enough to inform this decision. Second, it needed to serve as a blueprint: if a survey such as this is ever funded, the report can serve as the basis for a deeper and more considered examination of each element of the survey.

## **1. The Policy Environment and Background**

In 1998, BLS Commissioner Katharine Abraham set up a working group to examine the feasibility of conducting a survey of time-use. To set the stage for discussing how the working group approached this task, it is useful to provide background information on relevant BLS activities and other events related to time-use surveys that had preceded the establishment of this working group.

### BLS initial involvement.

BLS began exploring the feasibility of conducting a survey on unpaid work following the introduction of the Unremunerated Work Act into the 103rd Congress on February 18, 1993. The bill, which was introduced by Rep. Barbara-Rose Collins (D-MI), 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.” In April of the same year, BLS sent a representative to a conference sponsored by Statistics Canada entitled “The Measurement and Valuation of Unpaid Work.” Following the conference, a report was written for the BLS Commissioner that reviewed the literature and summarized some of the conceptual issues related to measuring and valuing unpaid work.

### U.N Conference on the Status of Women.

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 referred to actions that should be taken "...by national, regional and international statistical services and relevant governmental and United Nations agencies, in cooperation with research and documentation organizations, in their respective areas of responsibility." It included the following provisions.

"Develop an international classification of activities for time-use statistics that is sensitive to the differences between women and men in remunerated and unremunerated work, and collect data disaggregated by sex. At the national level, subject to national constraints:

- (i) Conduct regular time-use studies to measure, in quantitative terms, unremunerated work, including recording those activities that are performed simultaneously with remunerated or other unremunerated activities;
- (ii) Measure, in quantitative terms, unremunerated work that is outside national accounts and work to improve methods to assess and accurately reflect its value in satellite or other official accounts that are separate from but consistent with core national accounts." (United Nations, 1995)

In response to this interest, in 1997, BLS undertook two significant activities directly related to examining the feasibility of conducting a time-use survey: (1) BLS hired Westat to conduct a pilot study of two alternative versions of a time-use survey using a telephone methodology; and (2) BLS co-sponsored a time-use conference with the MacArthur Network on Family and the Economy.

#### BLS Pilot Study.

The BLS Time-use workgroup benefited a great deal from this research work. In the winter and spring of 1997, working under contract with Westat, the first phase of a feasibility test for the pilot survey was completed. This included conducting 21 cognitive interviews. These interviews revealed the kinds of difficulties that respondents have when confronted with the task of recalling and reporting all of the previous day's activities. They also provided insight into the types of activities that are most easily forgotten and left unreported by respondents. The information was used to create two alternative pilot survey instruments, including a version of the interview that included probes designed specifically to facilitate the reporting of nonmarket work.

Time-use interviews traditionally ask respondents to report (1) *what* they were doing, (2) *where* they were doing it, and (3) *with whom* they were doing it. In the BLS pilot study, this "standard" approach was used as a control condition. It was contrasted with an "enhanced" version. The enhanced version asked respondents (1) *what* they were doing and (2) *whether they were doing anything else at the same time*. This change of

approach was deemed important because cognitive tests showed that a great deal of housework and family care is done simultaneously with other activities.

During July of 1997, a field test was conducted to test the two approaches. BLS again worked with Westat. The goal was to complete 500 interviews using the standard time-use protocol and 500 interviews using the enhanced protocol. All interviews were random digit-dial calls (from the Westat telephone facilities in Rockville, Maryland). After the time reports were collected, the verbatim records of activities were delivered to coders who classified each of the reported activities into categories. Additional codes were used to identify whether each reported activity was a simultaneous or separate activity. Special codes were used for interviewer mistakes, missing information, refusals, and other errors.

The study found that the average completion time for all the interviews was 18.6 minutes. The enhanced interview averaged about 2 minutes longer than the standard approach. This increased time to complete the interview, however, was productive. With the enhanced interview, an average of 24.8 activities per respondent were reported; the standard interview averaged 21.9 activities per respondent. But more importantly, the specific reports of nonmarket work also increased. On average, the standard version collected 7.8 nonmarket work activities while the enhanced interview collected 8.8 nonmarket work activities per respondent

After examining the codes, about one hour per day of the total nonmarket work recorded by the enhanced version appeared to be work performed simultaneously with other activities. Because the standard time-use interview did not probe for simultaneous activities, respondents may not have been cued to the fact that small, secondary activities were also important and should be reported. Given the nature of child care and housework, as well as many kinds of leisure activities, there is reason to believe that the standard time-use survey may result in a slight undercount of certain behaviors (see the discussion of this bias in Szalai, 1972).

The BLS pilot survey, which had drawn on other surveys (primarily efforts by Statistics Canada) provided a strong foundation for the work of the time-use work group. From that survey, we had information on response rates, duration of interview, successful probes for information, and operational considerations. (Stinson, et. al., 1998)

### UN Expert Panel on Coding

One of our working group members attended a conference in New York on October 13-16, 1997, hosted by the United Nations' Secretariat statistical division for the purpose of developing an international coding scheme. The conference reinforced her appreciation (and, subsequently the work group's) for the need for international comparability of time-use statistics. It also brought her up to speed on the United Nations' coding scheme, one of several considered by the time-use survey and discussed briefly here and in more detail in Stinson (1999).

## 1997 Conference on Time Use

The November 1997 “Time-use, Nonmarket Work and Family Well-being,” conference also provided a great deal of information to the time-use work group when we convened in 1998. The conference was co-sponsored by the MacArthur Network on Family and the Economy and BLS. The conference, which took place at BLS, ran for two days and was well attended by researchers and survey methodologists from all over the world. Papers were presented on a wide range of topics. Economists, sociologists, and psychologists from academic institutions and governmental agencies contributed. About a third of the papers specifically addressed economic concerns that could be researched using time-use data. The other papers were fairly evenly distributed into the areas of family dynamics, childhood development, survey methodology, and public policy concerns, all of which related to time-use information. The session titles from the conference are listed below.

Economic Theory, Nonmarket Work, and Human Capital  
Economic Valuation of Nonmarket Work: International Perspectives  
Family Time: Quantities and Qualities  
Children and Time  
Measuring Nonmarket Labor Using a Time-Use Methodology  
Time-Use Surveys: Where should the BLS go from here?  
Time Use and Public Policy  
What do we do next?

Many important messages came out of the conference and served as a foundation for the work of the time-use working group. First, there is a great deal of policy and research interest in time-use data within the U.S. Second, the international community is eager for the U.S. Government to conduct a time-use survey. In the 1990s, governments in Australia, Canada, Denmark, Finland, France, Japan, New Zealand and Germany have conducted time-use studies (See Appendix 1 of the attached report for a more complete listing of the time-use studies that have been done in other countries). These data provide the U.S. with the potential for interesting cross-national comparisons of time use, as well as a wealth of survey experience from which BLS can draw when conducting a time-use survey. Lastly, it was evident from the conference that the lack of time-use data is regarded by some as a major statistical gap. Economist William Nordhaus, of Yale University, described time-use information as “the single most important gap in our federal statistics.”

## **2. Setting the Stage: The Charge to the Working Group**

Following the joint conference, BLS Commissioner Katharine Abraham asked that a working group be set up for the purpose of producing a report examining the feasibility of conducting a survey on how Americans spend their time. Specifically, the group was asked to consider elements of a time-use survey that would use the out-going rotation

groups of the CPS as the sampling frame. This survey would take advantage of the methodological and statistical expertise in the federal statistical system and would produce the first such information in our federal statistics.

### Assumptions and constraints

The following key assumptions and constraints, which were based on earlier research in the U.S. and other countries, provided guidance for our work:

- The purpose of the survey would be to estimate the time individuals spend in various activities;
- The sample for the survey would be drawn from the outgoing rotation groups of the monthly CPS;
- A 24-hour day time diary would be used (i.e. leading respondent through a 24-hour day);
- The data collection protocol would be a Computer Assisted Telephone Interview (CATI).

### Composition of the working group

Working group members brought a mixture of skill and experience to the project. Michael Horrigan, an economist and Director of the National Longitudinal Survey Program at the BLS led the group. Dr. Horrigan had experience in designing and conducting special surveys at the Bureau of Labor Statistics, including two surveys on employer-provided training. Linda Stinson, a research psychologist, brought broad experience in survey design and a thorough knowledge on the subject of measuring time-use. Three research economists, Diane Herz, Mary Joyce, and Jay Stewart, together contributed knowledge of economic theory, an understanding of the workings of the Current Population Survey, and experience designing special surveys. Midway through the process, a statistician, Ed Robison, joined the group to provide guidance and estimates of sample sizes sufficient to achieve the survey objectives.

The working group also had a wealth of information upon which to draw including reports from the BLS pilot study, prior research on time-use surveys by U.S. economists and sociologists, and the experiences of various countries that have conducted prior time-use surveys. A wealth of information exists on these surveys conducted by, among others, Canada, Australia, Finland, and more recently, Eurostat.

### **3. The initial process**

The intent of the first meeting was to lay out a process by which the group could systematically develop its recommendations. The Chair of the group developed a draft outline of the entire report prior to the meeting to use as a jumping-off point. This approach was based on the fundamental belief that the working group had to consider, in turn, each of the major milestones that one would encounter in actually running a survey.

A second, implicit belief was that just as these elements would follow a natural timeline progression, so too could the working group proceed in that fashion.

The original plan was to consider, in turn, the following major survey elements:

- (i). History of BLS involvement in time use
- (ii). Why are time-use data important?
- (iii). List the charge to the working group
- (iv). What is the primary estimation objective of the survey?
- (v). What is the secondary estimation objective of the survey?
- (vi). Sample size and the sampling plan
- (vii). Weighting, variances and imputation
- (viii). What data collection methodology should be employed to achieve these objectives?
- (ix). What are the data elements of the survey instrument?
- (x). Operational considerations: systems development, training, field staff, and coding
- (xi). Output from the survey

We immediately realized that much of the background work was already compiled in various documents related to the BLS Pilot, the time-use conference, and the survey documentation and analysis of other surveys from the U.S. (Robinson, 1997) and the survey efforts of other countries. As a result, question one was largely completed. We created a physical library to house all of the background information for reference throughout the development process. One group member created a notebook of time-use instruments from a range of countries.

We moved on to element (ii): Why are time-use data important? Here, we recognized that we all had a broad understanding of the political and economic debates on that subject already. Rather than devote meeting time to discuss justifications for time-use as the primary question, we chose to have a subgroup write a document answering that question (Joyce and Stewart, 1999). We also independently read up on literature. The authors of the overall justification document brought pieces of the discussion into our weekly meetings as it developed. They completed a draft in about two months, which was later incorporated into our report.

#### **4. Determining survey objectives: the fulcrum of the discussion**

As a group, we reviewed our charge and the known constraints for the project (e.g. using the CPS as a sampling frame and using CATI protocol). Our focus then moved to answering questions about our estimation objectives. We were pleased with the basic structure of the survey instrument used for the BLS pilot study, which was conducted in a CATI environment and worked with that survey instrument (or something very similar) in mind. (The questionnaire design is discussed in item 5 below.) Thus, we determined that

our efforts would concentrate on the data elements to be collected, and we would take as given the basic approach of leading a respondent through their day, activity by activity.

#### 4a. Our process: constraints, decisions, and deferrals

Much of the design work for the survey was done before we got started. The questions in the survey were fairly simple and already proven by others (see section 5 on questionnaire design, below.) Still, with questionnaire in hand, we needed to make decisions regarding the periodicity of the survey, sample stratification, reporting, sample design, and survey operations. We began by moving in a linear fashion through a set of questions, albeit recognizing that a feasibility study involves a lot of detours. Most of the questions were interconnected; some were entirely dependent on the answers to others. Over time, a general pattern to our process emerged. We would ask a high-level question. This would lead to more questions and ideas. Of those, we identified the ones we needed to answer immediately and move forward. Others were deferred to later meetings or to individuals or subgroups for further analysis.

As mentioned, we chose to determine the primary and secondary sampling objectives first. We opted for this question because we needed a precise statement of the estimation objectives needed in order to develop a sampling plan; that is, answering question (vi.) from the original work plan was dependent on answering questions (iv.) and (v.). Immediately, the question multiplied into more specific inquiries. At the same time, we decided that it would be inefficient to bring a survey statistician on board to the group until we had made these decisions.

Specifying the estimation objectives became a fulcrum around which we systematically addressed other issues. Some of the questions, all of which required further research and discussion, included:

- Is the concept of time-use best measured as a proportion or a level? The answer to this question would have a large impact on sample size requirement for achieving goals for the precision of estimates.
- What periodicity should be used for estimating the concept? Should it be monthly, quarterly, annually, every 3 years, etc?
- How should the sample be stratified by sub-population group? From the literature, what patterns have emerged in the U.S. and in other countries that we need to measure?

One way to describe the process and sequence of analytical events in the development of the survey objectives is to work backwards from our decisions. Our ultimate choices of estimation objectives follow.

#### *Primary sampling objective*

The main sampling objective is to draw a nationally representative sample of individuals 16 years of age or older that would be used to generate quarterly estimates of the proportion of time spent in various 1- and 2-digit activity categories for an average week, weekday, and weekend. These activity categories are discussed more fully below. These estimates would be presented for the entire U.S. population 16 years of age or older as well as for a limited number of comparison groups such as men and women, and groups defined by age categories.

#### *Secondary sampling objective*

A secondary sampling objective is to generate annual estimates of the proportion of time spent in various 1-, 2- and 3-digit activity categories for an average week, weekday, and weekend. These estimates would be presented for the entire U.S. population 16 years of age or older as well as for a larger number of subgroup populations defined by gender, age, race, ethnicity, family structure, presence of children, education, employment status, and location based on a rural/urban stratification.

#### 4b. Specific questions

Some of the detours we took along the way to developing the survey objectives are described below:

##### *How often, for how long, and at what level of detail?*

One of the first issues we addressed was that of the periodicity of reporting. We knew that a time-use survey would have to collect data throughout the entire year to overcome seasonal effects related to how individuals spend their time. We were not sure about whether to produce estimates for quarters or months. We were also not clear about the level of detail on activities that we needed to report and how often.

We knew there were various systems for coding activities in existence, and although we did not need to decide immediately on which exact coding system we would recommend adopting, we did need to decide on the relationship between the periodicity of reporting estimates and the level of detail. Would we report 1- or 2-digit detail on a quarterly basis and 3-digit detail on an annual? Would we instead recommend a three-year or a five-year cycle for 3-digit detail?

This question provides a good example of how the process worked. First, we had a constraint: All coding systems are designed with 3 levels of detail. We needed to plan to report at similar levels of detail. Second, we had some questions that needed to be addressed immediately, including: What level of detail would we report? How often? Third, we identified some topics to be discussed later including:

- What coding system would we recommend for the U.S. time-use survey? This

required a survey of existing codes.

- What do we know about coding accuracy and its implications for training, computer systems design, and staffing requirements, etc? What recommendations did we need to make related to this knowledge?

Originally, we were envisioning a rotation scheme such as in the Statistics Canada survey, in which time-use is asked every 7 years. However, we had little basis for justifying such a long rotation scheme. Although it may be true that time-use patterns do not change much over time in Canada, we had no evidence to suggest that this is the case in the United States. In addition, the Canadian model also requires building a survey system in which particular topics are investigated each year, with time-use being rotated in every 7 years. The fact that modules to the Current Population Survey already accomplish the objective of providing a flexible survey vehicle to investigate various survey topics meant that this would not be an appropriate objective for this survey vehicle. Finally, the challenge of determining the feasibility of just a time-use survey is difficult enough without adding to it the burden of specifying topics that would be investigated in the 'off-years'.

We settled on the recommendation for a continuous, core time-use survey. The continuous survey would give us time-use estimates that contained information for all seasons. One great benefit of this periodicity is that we can supplement the core questionnaire with modules on topics specifically related to time use. In this process, we identified yet another element of the report to be examined at a later date. (Modules are discussed later in this document.)

The continuous collection decision meant that we would construct 3-digit estimates annually and 1- to 2-digit estimates quarterly. We could have reported on 3-digit estimates over a longer period than one year. However, we thought that annual estimates would be in greater concordance with other major U.S. statistical series.

Clearly, this decision was pivotal in our deliberations. Other elements of the sampling objectives that we specified included the sampling stratification variables and the choice of reporting period concepts of an average weekday, weekend, and week. These decisions reflected a number of side excursions taken by the group. When the group needed to identify the sample stratification variables, for example, we spent a couple of weeks looking at the kinds of information and the sample sizes of subgroups from MIS-8 sample from the (March 1997) CPS. We also looked at how patterns of time-use were reported for subgroups by other countries.

This inquiry led to a series of must-answer questions:

- What stratification variables would we choose?
- What information would we need to update in the time-use interview? Example, employment status.
- How quickly could the Census Bureau pick a sample, and turn over the cases (and information sheets from the MIS8 interview)?

- What new information would we need from the respondent?

Deciding that we wanted to come up with representative estimates for an average weekday, weekend, and week was probably the easiest and least debated element of the estimation objectives. We debated whether producing estimates for Saturday and Sunday separately or whether estimating the whole weekend should be our sampling objective. We settled on the reporting for the weekend as a primary objective. However, the sample will be stratified by Saturdays and Sundays, so separate estimates will also be possible.

Can you remember what you did yesterday?

Another major issue we deliberated was the "designated day." The discussion around how to collect survey responses referring to a pre-chosen date both broadened and deepened our thinking. Our decisions in this area significantly influenced the operational recommendations made in the report.

The notion of asking someone to report, activity by activity, about the preceding day was a given at the beginning of our deliberations. We also quite quickly settled on idea of assigning a designated-day to each respondent, since otherwise the probability of selection would not be known in advance. What we did not know immediately was how we were going to handle respondents who we could not interview on their designated day. This gave rise to critical thinking on a variety of fronts:

- If a respondent's designated interview day is a Tuesday (they are reporting about Monday), and he misses the interview, should we try to get him on Wednesday (to report about Monday)? Or should we call back the next Tuesday? Ultimately, we decided on the latter. We also specified that the data collection protocol would include 4 callbacks, using Finland's protocol of 3 callbacks as a point of reference.

If a respondent is unavailable for three Tuesdays and is ultimately interviewed on the 4<sup>th</sup> try, should we find out if something unusual happened on those days that is related to their time use? Our answer was to build this information into the questionnaire.

- What are the sampling and weighting implications of a designated-day strategy?
- What are the computer systems implications for developing an automated call scheduler that will keep track of these designated days and callbacks?

Do hours of time use add up to 24 per day?

As we continued to grapple with our sampling objectives and the various related issues, we decided to use a visual technique to support our work. We designed table shells as soon as we had enough basic information to specify them. Writing up these

specifications early had a number of advantages that would likely carry over to any survey design effort.

- We learned what elements of the survey questionnaire might be problematic when we would later need to translate question/answer categories into meaningful rows or stubs in a table
- We gained insight into variables that we were missing or had not clearly specified by visualizing them in an analytical context
- We learned which elements we might be able to do without and which would likely have very small samples (such as 3-digit activity codes crossed by multiple demographic variables)

This technique was especially useful when we further developed recommendations on reporting the 24-hour day. We knew from the BLS pilot survey that the non-market work that individuals do is often done at the same time as other things. However, we were not clear on how to report simultaneous activities, as they brought the daily total to more than 24 hours.

The November 1997 Time Use Conference emphasized the need for us to restore the 24-hour clock in our reports of time-use data. We also knew that we had to carefully examine simultaneous activities. We could not move forward until we resolved these issues. Table shells helped us answer these questions by constructing the consequences of each option visually. In the end we developed two distinct types of tables (see Appendix 4 of attached document):

- *24-hour tables* in which all solo activities are listed (at the 1- or at the 2- digit level of detail), and all simultaneous activities are grouped into one category. A supplementary table would be constructed to provide detail on the simultaneous pairs, triples, etc.
- *Thematic tables* in which solo activities related to the theme (example: time spent providing child care) are listed, as are all pairs, triples, etc of simultaneous activities that include that activity as part of the simultaneous set. The total hours add up to total spent in that type of activity per weekday, weekend, or week in a quarter or annually.

The table shell process also had a number of side benefits:

- Specifying tables concepts, row stubs, and column wafers helped greatly to refine our thinking on the level of coding detail we wanted to present (and the kinds of coding systems we wanted to use). Thus, we addressed one of our items deferred earlier.
- Reacting to visual, analytical output helped to refine our thinking on our choices of

sample stratification variables, periodicity, and our choice of weekend versus weekday

#### 4c. Summary on sampling objectives

In determining the primary and secondary sampling objectives of the survey--and particularly in using the table shell technique--we made significant progress or resolved a number of questions that, in our initial planning, were expected to be resolved at a later point in the process. These included determining the data collection method, the data elements, and the output from the survey. We also identified some new ideas, such as the development of special-topic modules. We also left on the table some of the must-answer questions such as how fast the information pre-loads from the MIS-8 interview could be turned around for the time-use interview, which items should we update from the basic CPS interview, and what new items should we collect.

### **5. Questionnaire design**

Our discussion of the survey instrument ran parallel to our discussion of the survey estimation objectives, and ultimately occupied much of our thinking in the latter stages of working on this report. There are three elements to the questionnaire: the basic time-use questionnaire, the CPS elements, and the modules.

#### 5a. The basic time-use questionnaire.

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, however, had a draft questionnaire from the BLS pilot at the beginning of the group's effort. We again worked from constraints, to questions, to deferrals. We knew that in all time-use surveys, "a verbatim description of the day's activities is collected along with an assignment of the approximate starting and stopping times for each activity, recorded either in free format or in fixed 5-10 minute intervals" (Stinson, 1999). We were also already committed to carrying out the survey using Computer Assisted Telephone Interviewing (CATI) protocol.

The BLS pilot survey questionnaire had been developed based on some earlier surveys, particularly the Statistics Canada instrument. The enhanced version of the 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 nonmarket work than the standard version that did not directly address simultaneous activities. The group inquired about other useful findings from research that might improve the pilot questionnaire. (The questionnaire design is discussed in more detail in the attached report and in Stinson, 1999).

Midway through the field period of the 1997 pilot test of the time-use survey, researchers had realized that the collected information was insufficient for accurately

identifying and coding each instance of “paid work.” In order to supplement the existing information, an additional summary question was designed, asking respondents to identify each recorded activity for which they were paid. The additional information was used for coding purposes only and did not provide variables for the datafile.

The time-use working group took advantage of the Pilot survey and recommended the inclusion of a similar probe for coding purposes in the time-use survey. The process of examining the questionnaire led us to more detailed questions about interviewer skills and training as well as activity coding that we addressed later.

#### 5b. CPS elements

The great advantage of using the CPS as a sampling frame is the wealth of information that is already known about the eligible respondents. However, we realized that some pieces of information particularly relevant for time use, such as their employment status, could change between their last CPS interview and the time-use interview. We formed a subgroup to investigate which elements of the CPS interview should be updated during the time-use survey. We also took this opportunity to ask if there were other basic pieces of information that 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. We also needed to consider respondent burden when addressing this issue. In the end, we settled on the following:

- Update household composition. The respondent burden should be minimal because the information will be carried over from the MIS-8 interview and we expect that in the vast majority of cases there will be no changes in household composition.
- Update total family income. This refers to a question on the control card that asked respondents to indicate where the family’s income falls in a list of 7 income categories. Although there could be maybe some sensitivity to this question, the response burden would be minimal.
- Collect labor force status using the basic CPS instrument. Again respondent burden would be minimal because much of the information will be carried over from the MIS-8 interview.
- Collect employment status of the spouse or unmarried partner. Again respondent burden would be minimal because much of the information would be carried over from the MIS-8 interview. To minimize respondent burden, we recommended that no industry and occupation information be collected for employed spouses.
- Update earnings information for all respondents. These data would be collected using dependent interviewing unless the respondent indicated that he or she had changed jobs. If this imposed too much of a burden on respondents, we could collect information on earnings only for those respondents who started a new job since their

MIS-8 interview. This would include people who changed jobs, as well as employed people who were not employed at the time of the MIS-8 interview.

- School enrollment. School enrollment questions are currently asked of 15-24 year-olds. Because time-use can vary by enrollment status, we recommended also collecting this information for respondents who are 25 years of age and older.

It should be noted that because respondents will be drawn from the CPS, it will be possible to match respondents' time-use data with their responses to various CPS supplements such as the March Income Supplement, the October Education Supplement, the February Displaced Worker/Tenure and Contingent Work Supplements. This would broaden tremendously the number of research possibilities and enrich our understanding and interpretation of the data. However, there may be additional background information not collected in the CPS that also would improve time-use analyses. We did not attempt to list all the possible variables of interest but recognized that such an exercise would need to be done before arriving at a final questionnaire.

### 5c. The modules

During the revisions of the questionnaire and the discussions of the periodicity of our survey, we discussed the idea of adding questions on special topics to the regular time-use survey. These questions may enable researchers to more fully explore social and economic issues related to time use. They may also fill in limitations of some of our decisions: for example, we had decided that the individual, rather than the family, would be our unit of observation in the time-use interview. In a module, it would be possible to examine family allocation of time in greater detail. We briefly listed topics in our first set of recommendations. While the entire report was in peer review, we returned to this subject, examining possible modules in more depth. (See box.)

## **6. Activity codes**

In her paper for this ASA conference, Linda Stinson describes our decision regarding activity codes at some length. She notes that most of the currently-used activity classification systems have evolved from the original structure developed by Alexander Szalai for the Multinational Time-Use Project of the 1960's (Szalai, 1972).

As in other areas of this work, the time-use group decided to select an existing classification system (at least as a base) in order to benefit from previous tests and code revisions, thereby saving time and money. We selected three systems to examine in depth because they each had strengths.

The time-use group examined the Eurostat classification system, which has been used by 18 countries to date and provide direct international comparability. We also examined the Australian system, which has redefined some of the primary categories (from

Szalai) in ways that allow "unpaid work activities" to be more accurately identified. (Stinson, 1999). Finally, we considered the United Nations' International Trial Classification System, which has a basic framework that is compatible with the U.N. System of National Accounts. Ultimately, we chose to recommend a slightly modified version of the Australian system.

Here again we followed a similar process. Our constraint was international comparability. Our guide was analytical relevance. We seriously considered three existing surveys. External questions spurred on by this exercises included what level of detail to report in working tables and published tables and how to best train interviewers and coders to collect information specific enough to fit the coding scheme.

## **7. Sampling and survey operations**

After this long journey through the definition of primary and secondary sampling objectives and the related tasks of questionnaire selection/design and coding selection, we were ready to tackle our remaining questions from our initial work plan. These included:

- (vi). Sample size and the sampling plan.
- (vii). Weighting, variances and imputation
- (x). Operational considerations: systems development, training, field staff, and coding.

### 7a. Sample size and the sampling plan

In the development of our sampling objectives, we answered many important questions for designing a sample for the survey. We knew in advance that the CPS MIS-8 group would be the universe from which a sample would be drawn. We had also identified the periodicity (monthly), reporting periods (average week, weekend, and weekday both quarterly and annually), and demographic and other relevant variables to be included in analysis. And, we had identified the sampling unit as one individual in the household. It was at this point that we asked a statistician to join the working group to lead us through calculations to determine the minimum, maximum and optimal sample sizes for the survey. These samples would be used to estimate a survey budget. This process is described in detail in Robison (1999).

Our earlier decisions provided some boundaries for the discussion. The designated-day methodology, for example, forced us to think of a sample where individuals would be drawn for interview and then maintained if they were not available on their designated day. Prior research also played a role. We assumed a 70% response rate to the time-use survey when designing the survey sample. This was a conservative estimate based on the relatively high response rate of 88% in Statistics Canada using the same basic data collection methodology. We chose to err on the side of aiming too low rather than choosing an initial sample size that would result in too few respondents to

meet our sampling objectives.

Using a subset of the CPS provided many benefits:

- We could identify and select by survey respondents' demographic characteristics in advance
- We knew our maximum potential sample--72,000 adults
- 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
- We would have a substantial database of information about the respondent that could be used for further analysis

The statistician (Ed Robison) whittled down the potential sample (from the initial 72,000). He took into account estimated response rates, adhered to our methodological choice of one person per household, and thinned out the sample in the less-populous states (that are overrepresented in the CPS in order to produce good State-level estimates). He then took advantage of earlier time-use research. Using evidence on the dispersion of sample estimates provided by Juster and Stafford (1985), he was able to derive recommended sample sizes for both national and subgroup estimates.

A spreadsheet was developed that used unweighted CPS counts from MIS-8 samples as a foundation. He computed an entire distribution of time-use proportions for various subpopulations and sample sizes (10 characteristics with proportions summing to 1.0. From this, he determined that a sample of 35,000 with 24,000 completed cases per year (or 2,000 completed cases per month) was needed to satisfy sampling objectives.

#### 7b. Weighting, variances, and imputation

The sampling statistician also developed some initial recommendations regarding the weighting plan. It is clear that a more considered analysis of weighting, variance estimation, and imputation requirements will be needed if the survey ever becomes operational.

### **8. Survey operations**

Many of our survey operations were discussed in the development of primary and secondary sampling objectives. At the outset of the project, we knew we would be using telephone interviewing. We subsequently decided on a designated-day methodology with 4 weeks of callbacks. We knew the type of coding that would be required of interviewers.

One of the aspects of the interview that influenced our recommendations on staffing and training was the recognition of the fact that the time-use survey requires a very flexible interview, with interviewers using probes to get at simultaneous activities and to define the activities to the level of specificity required for accurate coding.

As a result, we concluded that interviewers must be highly trained and able to probe with unscripted questions to collect the appropriate level of detail for coding verbatim responses. We included a recommendation that the field interviewers be given training on coding as a way of improving their skills in probing. We also recommended that the time-use survey be staffed by an interview team dedicated to the survey. We recommended that the usual pre-testing done in a survey be supplemented by field operations that commence at least three months early. This would enable the interview staff to ‘hit the ground running’ when real data collection begins.

The development of software for coding was recognized as a very important operational element that would effect the quality of the data. Although we did not go into great depth on this element, we discussed various minimum capabilities that a software system should possess. A computer-aided coding system should provide coders with information on how prior similar verbatims were coded. It should also provide coders with user-friendly access to all levels of codes as they review verbatims. The coding software should also include training modules to keep the coders' skills at high levels.

## **9. What about doing two things at once?**

In the development of our initial recommendations, the group chose not to ask respondents to divide up the time they spent in simultaneous activities. The group’s collective thinking was that asking such a question would be such a burdensome strategy that it would be impossible and would likely yield data of dubious quality. We knew that the problem remained to be solved, but we implicitly assumed that the solution would be applied as a post-data collection imputation rather than as a part of the questionnaire design.

We had not found a solution, however. If satisfactory post-data collection adjustment were not developed, asking the respondent to self-report would have been the second-best alternative. We thought that such reporting was problematic from the point of view of interrupting the flow of the survey and providing high quality data.

While this question might have been answered in the development of sampling objectives, it actually was addressed much later. A deeper group consideration about the subject was spurred on by comments received in independent peer review after the completion of our recommendations. It was not until other individuals outside the working group, and in this case, outside the BLS, read the report that we recognized the issue for ourselves.

### *What was the solution?*

Based on comments from the Bureau of Economic Analysis, we considered several alternatives for splitting up simultaneous activities: 1) dividing the time spent by the number of activities; 2) assigning weights arbitrarily to activities reported as primary and

secondary by respondents; and 3) assigning weights to activities based on patterns of time-use by similar respondents. After much discussion, we recommended the third option, as it was grounded in economic theory. These options are described in more detail in the attached report.

## **10. Communication and Peer Review**

Communicating with internal and outside experts throughout the process of developing recommendations was important for keeping tabs on both the policy environment and methodological developments. The time-use working group kept in contact with experts in a number of other countries. This kept us up to speed on revisions in other survey instruments (from which we were learning). We also maintained contact with a number of advocacy groups interested in the eventual output. This helped us understand how our users might approach the data once the survey was completed.

Peer review was an important ingredient in this process. We needed to know when we were ready to get critical review. As this is true in the academic review process, it is also true in survey design. This paper has suggested that the evolution of our thinking on time-use surveys followed a somewhat linear path, albeit with significant detours and a few roadblocks. In fact, a more precise graphical analogy may be a series of concentric circles, as each element overlapped in certain areas with other elements.

Each time we wrote out the survey plan--including all of the elements mentioned in this paper--we achieved a particular level of specificity. Our intention was to get the major elements correct in the first few attempts. With the exception of providing recommendations on how to divide time spent in simultaneous activities, our report to the Commissioner achieved these objectives. It provided a lot of detail on each of the elements and was a good initial basis for developing potential budget estimates. It was later expanded to include recommendations for dividing simultaneous activities. Ideas and recommendations for potential time-use modules were also added.

## **11. What's next?**

Our recommendations on the content of the survey are noted throughout this paper and are written in detail in the attached report. Regarding the process of development of these recommendations, it might best be described as nonlinear, possibly as concentric circles. We began with a list of development areas and a great deal of material on which to draw. As we resolved issues, we influenced others and asked brand new questions. There was a great deal of interdependency between sections of our process. Our initial project plan provided a rough guide, but much of what we developed relied on a combination of others' research, prior experience, reading, and "going with the flow" of our thinking.

There is work left to do. In fact, as of this writing we are in the middle of examining our report in light of comments we have received on our report from the May 27-28, 1999 Workshop on Time-Use Surveys sponsored by the National Academy of Sciences. We fully expect to have an addendum to this paper that summarizes these further deliberations by the time of the ASA meetings in August 1999.

## REFERENCES

Joyce, M. and Stewart, J. (1999). What can we learn from 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.

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.

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.

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. [Online] [www.undp.org/fwcw/dawoff.htm](http://www.undp.org/fwcw/dawoff.htm)

