electronic ***International Journal of Time Use Research***
2005, Vol.2, No 1, 33-53 ***dx.doi.org/10.13085/eIJTUR.2.1.33-53***

Non-response and population representation in studies of adolescent time use

Casey B. Mulligan, Barbara Schneider and Rustin Wolfe

Casey B. Mulligan
The University of Chicago
Social Science Research Building, Room 506, 1126 E. 59th Street
Chicago, IL 60637, USA
e-mail: c-mulligan@uchicago.edu

Barbara Schneider
The University of Chicago
1155 E. 60th Street, Room 369
Chicago, IL 60637, USA
e-mail: schneider-barb@norc.org

Rustin Wolfe
Department of Education, St. Mary's University of Minnesota
Twin Cities Campus, 2500 Park Avenue
Minneapolis, MN 55404, USA
e-mail: rwolfe@smumn.edu

Abstract

Researchers have debated which methods are most valid and reliable for studying time use. One key instrument for measuring time use is the time diary, which has unique analytic properties that, if not adjusted for, can bias estimates. To assess sampling and non-response bias and potential under- or overreports of various activities, we use three different datasets to compare adolescents' time use. Results of these comparisons are used to show how investigators can statistically adjust time use data to obtain more accurate estimates of time spent in various activities.

JEL-Codes: C80, C81

Keywords: Methodology for collecting, estimating, and organizing microeconomic data; correcting for sampling and non-response bias; comparing survey and ESM measures of time use

1 Non-response and population representation in studies of adolescent time use[1]

In January 2003 the Bureau of Labor Statistics (BLS) began the American Time Use Survey (ATUS), the largest time use survey ever conducted in the U.S. Approximately 3,200 households, a subsample of the Current Population Survey (CPS), were asked to recall their activities throughout a twenty-four hour period using a time diary. For each activity, sampled respondents identified where they were, whom they were with, and whether they were engaged in another activity (to capture multiple activities that occur simultaneously). It is expected that within a few years, analyses of the ATUS time diaries will be as extensive as those currently being used with the CPS questionnaires. Time diaries have unique analytic properties that, if not adjusted for, can bias estimates. As economists and sociologists tackle this new body of information, whether estimating labor supply, productivity, household behavior, or leisure time, questions of sampling bias, instrument burden, and non-response must be considered. This paper addresses some of the methodological problems associated with studies of time use and offers solutions for this type of analytic work.

One of the most frequently asked questions of time use studies, particularly those that rely on time diaries, is whether the demands of the instrument interfere too much with the lives of the subjects. Are those who agree to participate in such studies a biased sample of the population? It has been argued by some, such as Hochschild (1989) and Leete and Schor (1994), that time diary studies are burdensome, and that those who are willing to participate and who successfully complete the instrument represent a select sample of the general population. Determining whether and to what degree hypotheses such as these are valid, and advising analysts on the best use of time diary data, requires comparing time diary participants with other similar populations.

Juster and Stafford (1991), Robinson and Godbey (1997), and others have defended the time diary methodology, arguing that the errors associated with time diary estimates of time spent on certain activities appear to be random.[2] Supporters of time use studies have, for the most part, examined the responses of time use participants and compared them to single-item responses of individuals from other national studies. What has yet to be examined in detail are the characteristics of time use participants in contrast to nonparticipants. Specifically, are the characteristics of the nonparticipant population patterned in such a way that indicates a bias that could potentially alter the results of the responding population?

The purpose of this paper is to identify and quantify instances of sampling and non-response bias and potential under- or overreports of various activities among adolescents across three datasets: the Alfred P. Sloan Study of Youth and Social Development, the Current Population Survey (CPS) October 1992 and May 1993, and the National Education Longitudinal Study of 1988-

1 We appreciate the research assistance of Ananth Ramanarayanan and Jaeki Jeong with this paper.

2 Other studies on the reliability of time diary studies include Gershuny et al. (1986) who report some effects of the nature of the sample responding and the length of the study on response rates in studies of adult time use. Jacobs (1998) examined differences between self-reported work time from the 1992 National Survey of the Changing Workforce and the March 1997 CPS. He found that the standard self-reported measure of working time is a reasonably reliable indicator of time use.

1992 (NELS:88-92).[3] We chose the Sloan Study of Youth and Social Development because it is the largest study of adolescent time use based on the Experience Sampling Method (ESM), a form of time diary that relies on immediate reports of time use rather than reconstructed accounts of a day's or a week's activities. The other two datasets also include measures of adolescent employment based on survey responses. Because both NELS and the CPS are nationally representative samples, it is possible to compare estimates of time use, specifically adolescents' weekly work hours, across these datasets. Such comparisons of smaller with larger nationally representative datasets allow investigators to systematically address issues of non-response, reliability of measures, and representativeness. Although these results are based on adolescent samples, they have broader implications for studies estimating time use, specifically the development of procedures that correct for non-response bias.

2 Measuring time use

Sociologists interested in time use have directed their attention to understanding time use as a mechanism for viewing and analyzing social life and societal change, whereas those more interested in social psychology tend to consider time use as a concrete measure of human preferences, values, and behaviors. Regardless of how one frames the study of time use, one of the key concerns is finding the most efficient and accurate method for recording human behaviors. This task is challenging because it is known that people have a tendency to exaggerate the time spent on socially desirable activities and underreport time spent on activities that most view as socially undesirable, such as fighting and gambling.

How one collects time use information is subject to inherent sources of bias. Direct, unobtrusive, non-detectable observations of human lives, while certainly the most accurate method for measuring time use, are clearly logistically problematic as well as prohibitively expensive. If subjects know they are being observed, they may change their behaviors, thus producing non-typical estimates of time use, such as hours spent at work, with family, or recreating. The most common type of method is to ask subjects retrospective questions, such as those traditionally used by the Census Bureau, which ask a subject to report how many hours he/she worked the previous week. While this method has the advantage of brevity, it has been shown to produce somewhat unreliable time use estimates.

An alternative to retrospective survey questions is the time diary method, which requires individuals to record their activities and the times at which they occurred. There are several types of time diary. Perhaps the most well known is the full-day diary, used in a series of studies at the Universities of Maryland and Michigan. In this type of study, respondents are asked to name the activity or activities they were engaged in during each of 96 15-minute intervals over a 24-hour period. These full-day diaries are collected using a variety of procedures, including in-person interviews, telephone interviews, or return mail surveys. Some studies ask individuals to report their previous day's activities, while others introduce respondents to the diary and ask

[3] CPS and NELS use sampling procedures specifically designed to represent the general adolescent population, and their respective sampling parameter estimates are robust. We selected these two national datasets because they were fielded within several years of the Sloan study and contained comparable items related to demographic characteristics and time use.

them to begin using it when they wake up the next day. Another form of time diary, used primarily for estimating the number of hours worked, is the travel schedule diary. Individuals report what time they leave for work, how much time they spend commuting, and what time they return home.

The Experience Sampling Method (ESM) is yet another method for studying time use. Individuals are randomly notified by a beeper and, when signaled, record what they are doing and feeling. Participants typically respond to eight signals a day over the course of a week, which has the advantage of sampling over seven consecutive days rather than on a given day. Developed by Mihaly Csikszentmihalyi and colleagues (Csikszentmihalyi and Csikszentmihalyi, 1988; Csikszentmihalyi and Larson, 1984), the ESM was designed to examine how individuals spend their time, what they do, and how they feel during specific activities. The ESM has been used with diverse populations, including adults and adolescents, and has provided estimates on how much time adults watch television (Kubey and Csikszentmihalyi, 1990), and how much time adolescents spend on homework, socializing with friends, or being home alone (Csikszentmihalyi, Rathunde, and Whalen, 1993; Csikszentmihalyi and Schneider, 2000).[4]

Much like time diaries, the ESM has been criticized as being too burdensome; that is, the time and cognitive demands made on the respondent are more excessive than the demands typically made by surveys. There may also be an inherent selection bias with the method if people who agree to participate in the study differ systematically from those who do not agree to participate (Zuzanek, 1999). Critics have also suggested that individuals may underreport what they are doing simply because they do not wish to be interrupted. Such underreporting is often thought to occur more frequently during activities outside the home where respondents may be unwilling to answer the beep. However, these criticisms have yet to be explored systematically.

2.1 Specific problems in measuring time use

2.1.1 Retrospective duration

Although survey questions that measure time use avoid the problem of respondent burden, they are subject to a variety of inaccuracies. Horvath (1982), Bound et al. (1990), and others have shown that retrospective duration (RD) questions often suffer from recall biases, with some people forgetting episodes of time use altogether. Respondents may also erroneously estimate the time they spend on a given activity or activities during the period in question. When such retrospective items are compared with answers obtained through other methods, they have been shown to be somewhat unreliable, with individuals tending to overestimate the amount of time spent on certain activities such as paid work or housework (Robinson and Bostrum 1994; Lee and Waite, 2001).

Using the ongoing CPS, which interviews Americans throughout the year and asks them about their work situation in the prior week and in the prior year, Horvath (1982) constructed two estimates of unemployment during the calendar year. The first estimate aggregated the “last week” reports in all of the surveys for the calendar year, while the second used “last year”

[4] For a detailed description of the ESM and how it was used in the Sloan Study of Youth and Social Development, see Csikszentmihalyi and Schneider, 2000.

reports from the surveys of the following calendar year. He systematically found more unemployment with the first estimate, a discrepancy which is consistent with the hypothesis that more episodes of time use are forgotten when the time period to be recalled is longer and/or farther in the past. Bound et al. (1990) used the Panel Study of Income Dynamics Validation Study, which sampled hourly workers from a manufacturing firm for whom responses to RD questions could be cross-checked with their employer's records, to demonstrate how their retrospective reports of hours worked differed from those of the company. Juster and Stafford (1985) and the National Research Council (2000) have argued that RD questions encourage individuals to round off their responses, thus leading to exaggerated time use patterns. They also maintain that RD questions tend to evoke stereotypical or socially desirable responses rather than actual time allocations.[5]

Discrepancies in estimates of time spent on various activities have also been found between RD items and time diaries. Using several RD responses from the 1987 National Survey of Families and Households (NSFH), Marini and Shelton (1993) estimated that women spend a total of 31 hours per week on core housework, whereas men spend 16 hours. However, Juster and Stafford (1985), using 1981 time diary information, found that women spend 19 hours per week on core housework, and men spend 8 hours. Estimates consistent with those of Juster and Stafford were reported by Robinson (1985), who used 1985 time diary data for the same activities and found that women spend 19 hours per week on core housework, whereas men spend 10 hours. Based on their analyses and the results of other studies, Marini and Shelton (1993) conclude that time diaries may be the most valid and reliable method for estimating time use.

Another problem with RD questions relates to their wording, which takes on major significance since these items are asked only once and often not in conjunction with other items that could clarify specific time periods. For example, questions such as "Are you employed?" can be interpreted in many ways by the respondent. Does "being employed" mean working sometime during the month of the interview, during the week of the interview, or at the moment of the interview? Seemingly small differences in question wording can lead to very different estimates of time use.

One method for estimating the errors in RD questions is to ask the respondent at what times certain activities started and ended and then have the researcher calculate the actual time spent. This method, typically used to document travel schedules, has been successfully used to measure time worked. Duration of work is calculated based on arrival and departure times, taking into account travel times. Jacobs (1998) points out that many respondents might reliably report departure and arrival times because they must be aware of them in order to coordinate with employers, clients, and public transit. Furthermore, departure and arrival reports might be reliable even if the respondent does not accurately calculate durations of work time. The process of thinking about departure and arrival times may actually improve the accuracy of estimated

[5] To quantitatively illustrate these kinds of biases, we used the CPS web page to study employed men ages 25 to 54 who were surveyed in the March 2000 CPS. We found that more than 40 percent of these respondents indicated working exactly 40 hours in the week prior to the interview, while less than 1 percent reported working 39 or 41 hours (although Pencavel [1986] interprets the 40-hour week reported by the overwhelming percentage of employees as an accurate response showing how strict employers are regarding the time worked by their employees). This pattern suggests that people who work long hours tend to underreport their hours, and people who work less tend to overreport their hours.

time spent at work. While obtaining time estimates using a travel schedule may increase the precision of estimating hours worked over standard RD questions, there are several problems with these estimates, especially for people who combine child care or leisure activities with travel to and from work, and for those who work at home and do not commute.

2.1.2 Inference and burden in time use studies

One of the benefits of time diaries is that they account for an entire 24-hour period so that random errors occurring in one time use category are cancelled out by random errors occurring in another category. Further, time diaries obtain specific information about the day and time when an activity occurs. This is also the case for the ESM since respondents report their activities at the moment signaled. But while time diaries can provide very comprehensive information on the daily life of a person, they do not necessarily provide accurate estimates of time use over the course of a week or a month. For example, suppose an individual, using the single-day diary method, reports spending eight hours at work. Can we infer that person worked 40 hours in the week? Or 56 hours?[6] Gershuny et al. (1986) suggest that people tend to organize their lives by a weekly schedule. A study of one day may misrepresent a person's weekly schedule; for example, a person may typically work longer on Tuesdays but leave earlier on Fridays. The ATUS, which uses a full-day diary for a single day, may encounter such difficulties.

There are other inference problems using full-day time diaries that relate to reporting activities of short duration. Diaries may cause substantial under- or overestimates of activities of short duration, primarily because respondents may feel encouraged to report time use in regular blocks, such as 15 or 20 minute intervals, during which a short duration activity would either be ignored or grossly overstated. For this reason, it is important, regardless of the instrument, to include units of shorter and longer duration.

Another problem with time diaries relates to respondent burden. Assume time diaries could be obtained from an individual over the course of a week. Having to complete a diary every day may in fact be too burdensome, and individuals might over- or underreport activities, or simply stop participating. Now assume a group of individuals agreed to participate in such a project. One might suspect that they represent a biased group since it is likely that few respondents would tolerate the burden of reporting for seven consecutive days.[7] The ATUS may be an improvement over other methods since it makes fewer demands on respondents. That is an empirical question, however.

The ESM shares many of the problems of time diaries, including respondent burden, but has several advantages over both surveys and time diaries, including immediate response to signals, which minimizes recall bias, randomization of beeps, which eliminates stereotypical categorizations of time use, and data collection over several consecutive days rather than on a

[6] Some full-day diary studies have tried to address this problem by gathering diaries for a few selected days during the year and then constructing a synthetic week. Nevertheless, a synthetic week is not the same as a calendar week.

[7] The Economic and Social Research Council commissioned a seven-day diary study in Britain in 1983 and 1984. Gershuny et al. (1986) describe the study and explain how a similar one-day diary study had an 81 percent response rate, while the seven-day study had a 52 percent response rate.

single day. Although the ESM has been taken as the standard against which other time measurement methods should be evaluated (Robinson, 1985), there has been limited quantitative work comparing the ESM with other instruments. In this paper, we evaluate the burden of the ESM and its effects on response bias and compare it with other instruments. These comparisons show some of the ESM's limitations and strengths compared with other methods of studying time use.

3 Our approach

In our analyses we estimate whether Sloan adolescents' time use, which is constructed from repeated measures, is similar to estimates of time use obtained through single-point RD responses reported by adolescents in other studies. Attention is directed to those variables that (1) serve as proxies for time use, such as hours worked outside of school, or (2) are asked of the Sloan ESM sample and adolescents in other studies, such as gender, age, parents' employment status, and the days of the week and time of year the student works for pay.

Two dimensions of sample selection bias are examined: representativeness of the sample and non-response bias. The Sloan adolescent sample was not designed to randomly sample American teenagers nor was it seasonally representative, since more observations occurred in April, May, and October. However, the Sloan sample includes a significant number of observations for all nine months of the academic year and is therefore more seasonally representative than a number of other studies of adolescents. In our first analysis, we determine how representative the Sloan study is compared with two nationally representative samples of adolescents. To measure non-response to the ESM, ESM participants are compared to a larger sample of students who completed surveys but not the ESM. On the basis of both of these analyses, weights are constructed that can be used to estimate characteristics of the general adolescent population. The methods used here can be applied to other random and purposive samples to determine the generalizability of results from time use studies, such as the ATUS.

3.1 Sloan study design

The Alfred P. Sloan Study of Youth and Social Development is a national longitudinal study designed to examine adolescent's transitions into adulthood (see Csikszentmihalyi and Schneider, 2000). The study began in 1992 with 1,221 students in sixth, eighth, tenth, and twelfth grades who were followed over a five-year period at twelve sites across the U.S.[8] Localities were selected to satisfy the following criteria: variation in urbanicity, labor force composition, and race and ethnicity. Using 1990 U.S. census information, the sites were selected based on the degree to which their local economies were concentrated in manufacturing or service, as well as in their trend toward economic growth, stability, or decline over the past decade. The twelve sites were matched so that comparisons could be made among the school communities with respect to school size and the socioeconomic status and racial and ethnic

8 The twelve sites included 33 schools: 20 middle schools and 13 high schools. To provide variation in high school programs, two specialized schools were included in the sample: a mathematics and science academy and a magnet language academy. The remaining 11 high schools had more traditional comprehensive curricular programs.

diversity of the school populations. To ensure racial and ethnic diversity within schools, middle class African Americans and Hispanics were oversampled relative to their proportions in the national population.

For each school, two student samples were selected: ESM focal students, who were followed longitudinally, and Questionnaire-only (Q) students, who were not included in the longitudinal sample. The focal students were chosen from school-prepared enrollment lists of grades 6, 8, 10, and 12. Using a stratified design at each school, student selections at each grade level were made so that they were proportionately representative of gender, race, ethnicity, and level of academic performance. Based on student records, teachers rated each of these students as academically successful, working at grade level, or having academic problems. At each school, twenty-four students from each grade level were selected from lists prepared by the school using a random table of numbers.

The Q sample was selected using the same criteria as the focal sample. Specifically, the Q sample was designed to provide more information about the school and peer networks for each of the focal panel grades. Each year in the field, new Q samples were drawn from the grade the focal students were in. If a grade enrolled no more than 150 students, the Q sample consisted of the entire grade. Otherwise a random sample of 150 students was chosen from the grade enrollment lists. The Q students were administered the same questionnaires as the focal students, making it possible to aggregate information from both sample groups.[9]

Data were obtained using: (1) the Experience Sampling Method; (2) an in-depth interview; and (3) a battery of questionnaires, including a modification of instruments used in the National Education Longitudinal Study of 1988-1992. The Q students completed the questionnaires but were not interviewed and did not participate in the Experience Sampling Method; ESM students completed all instruments.

After completing a questionnaire pertaining to family characteristics, experiences in school, and plans for the future, the ESM students wore wristwatches programmed to beep randomly eight times daily in intervals between 7:30 am and 10:30 pm. The total schedule of beeps occurred over seven days and consisted of 56 signals.[10] Students completed a short questionnaire describing their activities and thoughts at the time of the beep. After the data were gathered, eight time slots were generated to categorize the various time slots across sites, schools, and cohorts: 7:29-9:17 am, 9:18-11:10 am, 11:11 am-1:03 pm, 1:04-2:57 pm, 2:58-4:49 pm, 4:50-6:42 pm, 6:43-8:35 pm, and 8:36-10:14 pm.

As explained below, subjects did not respond to all beeps, but the number of beeps to which a subject responded while engaging in a particular activity can be used to approximate a percentage of time engaged, and an absolute amount of time for the week. To do this, we first

[9] When the focal and Q samples are combined over the five years of the study, the total Sloan study sample is over 8,000 students. Analyses of the focal and Q samples revealed no differences in demographic characteristics, attitudes toward school, educational expectations, occupational aspirations, and other key variables used in this study.

[10] A seven-day schedule was used for all subjects, although the day of the week on which the schedule began varied by site, school, and cohort. While the beep-cycle starting day and ending days varied, the study was designed to begin between 11:11 am and 1:03 pm of the first day and to end between 9:18 am and 11:10 am of the seventh day—a total of seven complete days and 56 beeps.

calculated the ratio *r* of beep responses while engaged in the activity to the total beep responses. Since beeping occurred during approximately 15 “waking” hours (more precisely from 7:29 am to 10:14 pm, or 886 minutes) each day over a seven-day week, 105 waking hours (more precisely, 6,202 minutes) per week are represented. Each percentage point of *r* thus corresponds to 1.05 weekly hours (more precisely, 62.02 minutes). For example, 10.6 percent of beep responses occurred while subjects watched television (as a primary activity, see below), resulting in an estimate of 11 weekly waking hours (657 weekly waking minutes) of watching television.

3.2 Comparison groups

Comparison groups include the Census Bureau’s October 1992 and May 1993 Current Population Surveys (CPS) and the U.S. Department of Education’s 1988-1992 National Education Longitudinal Study (NELS:88-92). The two studies each share one advantage: they are designed exclusively to obtain national estimates of population demographics and labor force activity (U.S. Bureau of the Census, 2000). Unlike the decennial Census, which relies on a great many citizens completing and returning the questionnaire under no direct Bureau supervision, the CPS respondents are statistically sampled, and then located and questioned by trained interviewers (U.S. Bureau of the Census, 2000).[11] Teenage employment may be seasonal, so another advantage of the CPS is that its monthly surveys can be used to examine and, if necessary, correct for the effects of seasonality in our data. In case teenage employment seasonality is important, the May 1993 CPS is used for comparison with the first wave of the ESM (conducted in April and May of 1993). The May CPS does not measure grade in school, and does not measure employment for anyone less than 15 years old, so it does not allow for accurate measurement of the employment situation of students in sixth, eighth, or tenth grade. We therefore also used the October 1992 CPS to more accurately identify students by grade, and then to measure the October employment situation of twelfth graders.

Because the CPS provides limited information on adolescents, the NELS: 88-92 student survey was also used for purposes of comparison. NELS: 88-92 is a nationally representative study of adolescents that began in 1988 when 25,000 eighth graders in public and private high schools across the U.S. were surveyed. These students have been resurveyed four times: in 1990, 1992, 1994, and 2000. In our comparisons, we primarily use data from the NELS Second Follow-up (1992) since it was conducted at approximately the same time as the first wave of the Sloan study. The data collected include information from students, parents, teachers, and school administrators. In addition to basic demographic and family information, NELS: 88-92 includes variables measuring performance in school, educational aspirations, experiences in school, and experiences at work.

The purpose of these comparison groups is to isolate the two dimensions of sample selection bias: representativeness and non-response bias. With respect to determining the representativeness of the Sloan sample, we compare the Sloan “Q-only+ESM” sample of teenagers responding to a questionnaire—but not necessarily to the ESM—with CPS and NELS samples, under the assumption that CPS and/or NELS respondents adequately represent the

[11] See Hogan and Robinson (1993) for a discussion of how the Census Bureau uses statistical sampling to estimate undercounting in the decennial census, and how it appears that black and other minorities are undercounted in the census. Also note that the CPS serves as the sampling frame for the ATUS.

teenage population. With respect to potential non-response bias in the ESM we compare the Sloan full ESM sample with the sample of students who completed at least 15 of their 56 beeps (the ESM-15 sample). On average, the response rate was approximately 32 beeps per participant. As other studies have shown, this response rate is not unusual for adolescents (see e.g., Hoogstra, forthcoming). Given that adolescents are likely to respond to only half their scheduled beeps, it is important to examine patterns of non-response in the data.

4 Results - representativeness and non-response in the sloan sample

4.1 Age, gender, and parent employment comparisons

To examine the representativeness of the Sloan ESM samples, we compared the age and gender distributions in these samples with those in the CPS. Analyses of the ESM and CPS samples indicate that although age differences between the ESM and ESM-15 samples are statistically significant at the .95 confidence level, they are not substantively different. The age differences between the ESM-15 and CPS samples are statistically insignificant. The overall ESM sample is representative of the age distribution of those in school, and reflects the 11 percent high school dropout rate seen in the CPS and other surveys. The propensity of an ESM student to respond to at least 15 beeps declines with age. Seniors have a lower response rate to the ESM, and will be somewhat undercounted in an unweighted beep-level analysis (analyses not shown).

Table 1
Work Status of parents of CPS, NELS, and sloan study high school students

	Oct 1992 CPS	1992 NELS	1992-93 Sloan Q-only + ESM	1992-93 Sloan ESM students	1992-93 Sloan ESM students w/ ∃ 15 beeps
Neither parent works	6.8	2.9	2.1	2.0	2.0
Only father	27.7	10.6	13.6	15.1	15.7
Only mother	6.7	10.9	4.5	4.0	4.0
Both work	58.8	75.6	79.9	78.9	78.3
TOTAL	100	100	100	100	100
Respondents	5873	19379	2851	697	599

Note: NELS responses reported above are based on the student survey. When the parent survey is used to measure employment status of parents, then the numbers more closely resemble those of the CPS: neither parent works 8.8; only father 19.6; only mother 20.4; and both work 51.1.

Source: October 1992 Current Population Survey, National Education Longitudinal Study, Second Follow-up (1992) Student Survey; Sloan Study of Youth and Social Development Wave 1 (1992-93) Student Survey.

Analyses of the gender composition of the ESM and CPS samples show that girls are more likely to participate in the Sloan study, and are more likely to provide 15 or more responses during the week (overall, 55 percent of ESM participants are female; of those who responded to 15 or more beeps, 59 percent are female, a pattern that is consistent across grades). Tabulations for the CPS sample show more boys than girls (except among twelfth graders) in the school population, suggesting that ESM participation and response rates are greater among girls (analyses not shown).[12]

To further examine the representativeness of the Sloan ESM samples, we compared parent employment status across datasets. Table 1 shows that the distribution of work statuses of parents in the Sloan samples is similar to that reported in the NELS sample; both measures are based on student responses. The CPS numbers, however, are based on parent responses and are higher. Similarly, numbers based on the NELS parent survey resemble those in the CPS (see note in Table 1 below). As shown in the last three columns of the table, there is a slight, but statistically insignificant tendency for the ESM-15 sample to overrepresent students with father only working; it also slightly underrepresents students with both parents working relative to the ESM and Q-only+ESM samples.

4.2 ESM response rates by day of week

To examine patterns of non-response in the Sloan ESM samples, we compared ESM response rates across different times of the week. As described earlier, ESM beeps occurred at regular intervals during waking hours for a calendar week. Beeps were categorized according to four partitions of the week: “school time” (7:29am-2:57pm Monday-Friday), “after school weekday” (2:58pm-6:42pm Monday-Thursday), “school night” (6:43pm-10:14pm Sunday-Thursday), and “weekend” (2:58pm-10:14pm Friday, all day Saturday, and 7:29am-6:42pm Sunday). Analyses comparing response rates across these partitions indicate that the response rate differential is substantively and statistically significant, with response rates higher during school time than on school nights (analyses not shown). Response rates are related to gender, as noted earlier, but the time-of-week differential response is not. Although participation in the sample of students responding to at least 15 beeps declines with age, no systematic relationship between age and beep response rate was found, even when time of the week was taken into account.

4.3 Comparing teenage employment using the sloan, CPS, and NELS surveys

When it comes to using the ESM to measure time use, and work time in particular, there are three issues that must be addressed. First, how well do those Sloan students reporting at least 15 beeps represent the teenage population in terms of work histories or current work status?

[12] In addition to examining the representativeness of the Sloan sample with respect to age and gender through comparisons with the CPS sample, we also examined its representativeness with respect to school performance, using comparisons with the NELS: 88-92 sample. Both the NELS and Sloan questionnaires include questions posed for tenth graders about grades in four subjects (English, math, science, and social studies). Responses to these questions were aggregated for each respondent in both studies to compute a grade point average on a four-point scale. A comparison of the grade point averages for tenth grade students in the NELS sample and in the Sloan study ESM and questionnaire samples showed that averages were fairly similar for all of the samples, although there was a small and statistically significant gap between NELS and Sloan, with Sloan students having slightly higher GPAs. This difference appeared to be due to the schools sampled by the Sloan and NELS studies rather than non-response within the Sloan study.

Second, are Sloan work-related RD survey items comparable to work-related RD survey items from the NELS or CPS? Third, how do ESM estimates of time use compare with estimates derived from RD survey responses?

The Sloan study's work-related retrospective duration questions are different from those in the CPS. The Sloan study asks "Are you currently employed (have a paying job) or have you ever been employed?" to which valid responses are "never," "not employed now but was employed during this school year," "not employed this school year but was employed last summer," "was employed prior to last summer," or "currently employed." Note in particular that "currently" is rather open-ended. Does it refer to the day of the interview, the week of the interview, the month of the interview, or the semester of the interview? Also, does baby-sitting, yard work, or work at the family business count? These distinctions are expected to be more important for teenagers than for adults, since the former are less attached to the labor force, and time spent at school makes irregular, intermittent, and/or informal employment relatively more attractive.

Table 2
Percentage of high school seniors currently working -
a comparison of survey responses across three data sets

	10/92 CPS	5/93 CPS	1992 NELS	1992-93 Sloan Q-only +ESM	1992-93 Sloan ESM	1992-93 Sloan ESM w/ ∃ 15 beeps
Percent of currently working 12th graders (from surveys)	36.5 [1778]	39.5 [1927]	51.7 [16,070]	52.5 [708]	54.7 [170]	54.3 [138]
Average hours among currently working 12th graders (from surveys)	17.3 [666]	17.4 [778]	16.7 [7601]	19.0 [365]	18.8* [93]	17.7 [75]

Second Follow-up (1992) Student Survey; Sloan Study of Youth and Social Development Wave 1 (1992-93) Student Survey.

1 The following criteria were used in identifying 5/93 CPS 12th graders: those age 17, 18, or 19, enrolled in high school, who had completed the 11th grade, but did not have a high school diploma.
2 CPS observations are weighted using the household head's CPS weight.
3 Number of observations is reported in brackets.
4 In the Sloan study (Questionnaire-only and ESM samples), working is indicated by a "currently employed" response to the question "What is your job situation?"
5 In the Sloan study, hours worked at current job is reported in 0-10, 11-20, 21-30, 31-40, and 41+ hour intervals. We used the Sloan distribution of responses across these intervals, and CPS interval averages for seniors (7.0, 16.8, 25.8, 36.5, and 46, respectively), to compute Sloan average hours worked.
6 NELS statistics are weighted according to the NELS variable F2QWT, which weights the Second Follow-up sample to represent the 1992 U.S. population of twelfth graders.
* Significant at the p<.05 level in comparing ESM samples.

Source: October 1992 and May 1993 Current Population Survey; National Education Longitudinal Study

As shown in the last three columns of Table 2, which summarize Sloan survey responses, there are only minor differences across Sloan samples in the fraction of high school seniors "currently" employed. The NELS question about current employment, "What is your job situation?"is comparable to the question in the Sloan survey which is fairly open-ended. As might be expected, the percentages of those "currently working" are similar in both the NELS and Sloan samples (see Table 2).

In the CPS, RD questions about employment status are more specific, and we use the Census Bureau's concept of "currently employed and working" derived from those questions. In particular, "currently employed and working" refers to those who worked for pay some time during the survey week, plus those working 15 hours or more as unpaid family workers during the survey week. Table 2 shows that according to the various questionnaires, the CPS fractions of seniors currently working are substantially lower than in the NELS and Sloan samples. Given that NELS and CPS adequately represent the teenage population, this difference appears to be attributable to the survey question rather than the populations sampled (see also Committee on the Health and Safety Implications of Child Labor, 1998).

Sloan, NELS, and CPS asked about weekly hours usually worked on the current or most recent job, and we report the average for those "currently employed" in the second row of Table 2. The average hours reported by working twelfth grades is similar across datasets with most students working approximately 17 hours per week.[13] Note that the 17 percent of ESM participants who responded to fewer than 15 beeps tended to work longer hours if they were employed.[14] The average hours differences between the ESM and ESM-15 samples are statistically significant but the actual difference in terms of hours worked is small.[15] There are no statistically significant differences between either the ESM or the ESM-15 sample and the Q-only+ESM sample.

Table 3
High school seniors who ever worked: NELS and sloan surveys

	1992 NELS	1992-93 Sloan Q-only + ESM	1992-93 Sloan ESM Students	1992-93 Sloan ESM Students w/ ∃ 15 bee[s
Percent of 12th graders who ever worked (from surveys)	85.8 [16,070]	84.5 [708]	87.1 [170]	85.5 [138]

Note: Number of observations is reported in brackets.

Source: National Education Longitudinal Study of Second Follow-up (1992) Student Survey; Sloan Study of Youth and Social Development Wave 1 (1992-93) Student Survey.

As discussed above**,** the Sloan employment RD question can be used to measure whether a respondent has ever worked in his/her lifetime. NELS responses to the "What is your job

---

13 We further analyzed subgroup differences (e.g., gender) across these datasets to provide additional evidence of comparability. Analyses of the CPS data show that males are more likely to work than females; however, this difference is insignificant. Boys are more likely to work longer hours than girls and this difference is statistically significant. Similar trends were found for the Sloan survey sample (see Csikszentmihalyi and Schneider, 2000).

14 Not reported in the Table 2, but obtained by calculating the difference between the ESM and ESM-15 samples.

15 Although the ESM samples are relatively small in comparison to CPS and NELS, the fact that we are able to find statistically significant differences suggests that the size of the sample is large enough to detect differences across samples. However, it is important to note that while there is a statistical difference between the ESM and the ESM-15 samples, the actual difference in terms of hours worked amounts to only one hour per week.

situation?" question can also be used to determine whether a respondent ever had a job in his/her lifetime. There is almost no variation in the percentage of twelfth graders who have ever worked in either the NELS or Sloan datasets. Table 3 suggests that the sample of ESM students reporting at least 15 beeps is representative of the overall population in terms of propensity to work or have worked; the differences between the fraction working in that sample is not significantly different from that for the ESM and Q-only+ESM samples either substantively or statistically.

4.4 The ESM as a measure of work time

The ESM can be used to measure employment and hours, and in a way that is comparable to the CPS's "survey" week definition of "currently employed and working."[16] To measure employment in the sample of students responding to 15 or more beeps during the week, we took the fraction of those reporting at least one beep in the workplace, as shown in Table 4. We found an "employment rate" for seniors that was both similar to the CPS survey-based estimate and substantially different from the fraction of those responding affirmatively to the Sloan study's rather open-ended "currently" employed question.

Table 4
Working beeps of sloan ESM students compared with CPS survey responses

	CPS Questionnaire (1992-93)		Sloan ESM Responses (1992-93)	
			Unweighted	Weighted
	Oct 1992	May 1993	ESM w/ ∃15 beeps	ESM w/ ∃ 15 beeps
Percent working 12th graders	36.5	39.5	39.3	36.7
	[1778]	[1927]	[168]	[168]
Average hours among working 12th graders	17.3	17.4	15.2	15.9
	[666]	[778]	[67]	[67]

1 Number of observations is reported in brackets.
2 A "work beep" is one that occurred while the subject was at his or her workplace (e.g., even if during break time).
3 The sample of ESM students with 15 or more beeps consists of 168 students; 30 of those students failed to provide survey responses regarding their current employment. Therefore the ESM-15 sample size reported in Table 1 (n=138) is lower than the sample size reported above.

Source: October 1992 and May 1993 Current Population Survey; Sloan Study of Youth and Social Development Wave 1 (1992-93) Experience Sampling Method.

The reporting of *at least one* beep in the workplace is an obvious measure of employment, but beeps can be aggregated to obtain an estimate of the number of hours worked during the survey week. To do so, we first calculated the ratio r of beep responses at work to the total beep responses. The beeping schedule encompassed approximately 886 minutes per day over a seven-day week, which totaled 6,202 minutes per week. Each percentage point of r thus corresponds to

16 The CPS is conducted during the week of the 19th of a given month and asks about the previous week, so a more strict comparison of ESM and CPS would discard ESM responses provided during any week other than the week of the 12th. We have not found much CPS or Sloan study evidence of within-school-year seasonality of teen employment. Therefore, to maintain a larger sample size, we do not discard such ESM observations.

62.02 minutes. (For example, for all students who reported at least one beep at work, 11.0 percent of beep responses occurred while they were at work; we therefore estimated 11.3 weekly waking hours at work for those who worked at all.)

ESM response rates vary by time of week, and by gender, so ESM responses can be weighted by the inverse of the response rate for that gender/time of week to obtain a more accurate estimate of hours worked. A time-of-week-weighted estimate of work time is expected to be higher, since students are more likely to respond to ESM signals during the school day (when not engaged in paid work) and are less likely to respond at other times of the day.

The weighted and unweighted work hour estimates can be compared with hours estimates based on responses to questionnaires (retrospective duration questions). ESM estimates of the employment rate, and weighted ESM estimates of hours worked, are *very* similar to CPS RD questionnaire-based estimates. [17] To the extent that there are differences, they might be interpreted in three ways:

1) Reported beeps are imperfectly representative of adolescent time use (e.g., ESM non-response is especially high when the respondent is at work);

2) RD questionnaire estimates of the length of the workweek are imperfect; and

3) The CPS and Sloan study sample different populations.

The second interpretation has been made by authors of time diary studies of the workweek. For example, Robinson and Bostrum (1994) found that adults with shorter than average workweeks according to full-day diaries have RD workweek estimates that are longer than their diary estimates. While Robinson and Bostrum (1994) and Jacobs (1998) have different interpretations of those findings, both conclude that RD questionnaire estimates must be imperfect.[18] If this reporting bias carries over to teenagers, we would expect questionnaires to overestimate teenage work, since their workweeks are short relative to that of an average adult. However, there are a few reasons to suspect RD questionnaire biases to be different for teenagers than for adults. First, teenagers are typically “clock punching” hourly employees, and the process of punching the clock enables them to better estimate work hours than typical adult salaried employees. Second, teenage work schedules are much less regular (Committee, 1998), which makes it less likely that a teenager would accurately estimate his work hours for any given week.

[17] The calculations for the weights used in Table 4 are described in section 3.5 below.

[18] Mulligan (1998) used males from the 1985 full-day time diary study to compare diary with RD questionnaire time-worked estimates by age group (ages 25 to 64). He found RD estimates to exceed diary estimates the most for men in their 50’s and early 60’s—the group of men expected to have the shortest average workweek (see also Ruhm’s [1990] study of “bridge jobs and partial retirement”). His study is consistent with the finding of significant imperfections in RD workweek estimates, and perhaps also confirms Jacobs’ (1998) interpretation that RD questionnaires overestimate workweeks the most for adults with short workweeks.

Table 5
Work activities of sloan wave 1 students with at least 15 ESM responses

	Percent of beeps at work	
	Unweighted	Weighted
Working as primary activity	73.4	73.8
Working as secondary activity	6.4	5.9
Working neither as primary nor secondary activity:		
Homework	0.6	0.6
Talking with friends, in person	6.8	6.7
Talking with friends, other	1.4	1.3
Playing games	1.6	2.0
Watching television	3.9	4.0
Listening to music/radio	0.2	0.3
Doing a hobby	1.6	1.9
Personal care	3.9	4.0
Smoking	0.2	0.2

Note: The following activities were coded as "work" when done at the workplace: "thinking," "standing," "walking," "waiting," "driving," "nothing," "missing the beep," or "this study."

Source: The Sloan Study of Youth and Social Development Wave 1 (1992-93) Experience Sampling Method (ESM).

As shown in Table 4, when the survey question is restricted to a specific time period as in the CPS, then the responses regarding hours worked are similar to ESM responses which also are specific. [19] However, as shown in Table 5, teenage time at the workplace is the not the same as teenage time worked. The difference between teenage time at the workplace and teenage time worked can be seen by looking at what teenagers actually do at work. When beeped at work, teenagers reported working 80 percent of the time; the other 20 percent of the time they reported doing homework, talking with friends, playing games, watching television, listening to music/radio, doing a hobby, personal care, or smoking**.** These activities may be considered something other than "work"—even though done at work. The ESM shows that they are nontrivial and offers researchers some quantitative indicators of those activities.

4.5 Constructing weights

Although unweighted ESM estimates of time use are fairly close to estimates from other studies, these results suggest that ESM non-response is substantively significant in two dimensions: time of week and gender.[20] Estimates that better characterize the wider adolescent population and more closely match estimates from other studies might therefore be obtained by weighting beeps

[19] Gershuny et al. (1986) report a similar finding in their British study of adult time use. They found average hours worked to be similar in a time diary sample and in a more standard employment-questionnaire sample. Perhaps surprisingly, their point estimates suggest that those who work long hours are *more* likely to respond in a diary study than in an employment-questionnaire study.

[20] As Kalton (1985) notes, day of week is typically not randomly sampled in time diary studies and therefore has to be adjusted for. Since the ESM is typically conducted over a period of a week, with participants beginning the beeping schedule on different days, variation in response rates over days of the week can be computed and response rates can be adjusted accordingly (see Jeong, forthcoming).

according to their time of week and the gender of the respondent. Table 6 reports the weights used in this analysis. There are six weights: two in the gender dimension (w^m and w^f, with $w^m+w^f = 1$) and four in the time-of-week dimension (w_1, w_2, w_3, w_4, which sum to one). The gender weights are reported in the last row of the table, and the time-of-week weights in the last column. The weight for each type of beep (e.g., a school-time beep, or a beep from a male subject) is proportional to the inverse of the frequency of that type of beep in the sample of beeps obtained from those providing at least 15 responses to the ESM. The weights for subtypes of beeps (e.g., male school-time beeps) were calculated as the product of the corresponding beep-type weights and are reported in the interior of the table.

Table 6
Time-of-week and gender weights for sloan ESM wave 1 beeps
(in sample of those reporting at least 15 beeps)

	Gender		
Time of week	Male	Female	Either
School time	0.122	0.082	0.204
After school weekday	0.135	0.092	0.227
School night	0.160	0.109	0.269
Weekend	0.179	0.121	0.300
Any time	0.596	0.404	1

Note: Each of the eight weights in the interior is the product of its time-of-week weight (from the last column) with its gender weight (from the last row).

Source: Sloan Study of Youth and Social Development Wave 1 (1992-93)
Student Questionnaire and Experience Sampling Method (ESM).

The probability that a male (or female) respondent would be included in the ESM-15 sample is inferred by comparing October 1992 CPS and ESM-15 participation rates by gender; this comparison is used to calculate the column sums w^m and w^f in Table 6. Algebraically,

CPS_g (ESM_g) is the number of CPS (ESM-15) respondents of gender g, respectively.

$$w^g \equiv \frac{CPS_g / ESM_g}{\sum_{i=mf} CPS_i / ESM_i} \tag{1}$$

Note that this equation has the two properties required of weights: (1) summing to one across groups (gender in this case); and (2) proportional to the inverse sampling probability (in this case, gender g is ESM-sampled with probability ESM_g/CPS_g).

The probability that a beep at a particular time of week would be included in the sample is inferred from the response rates by time of week, and used to calculate the row sums.[21] Algebraically, the weight w_s for time-of-week s is the inverse of the probability of inclusion:

[21] For example, the school-time beep response rate was 0.686. The inverse of 0.686 is 1.458. Summing inverse response rates for all four time slots (1.458+1.618+1.923+2.146), we get 7.145. The school-time beep weight of 20.4 percent is then calculated as 1.458/7.145. In this way, the weights sum to one across time slots, and each slot's weight is inversely proportional to its response rate.

$$w_s \equiv \frac{1/r_s}{\sum_{i=1}^{4} 1/r_i} \tag{2}$$

where r_s is the response rate for time slot s. The eight subtype weights shown in the interior of Table 6 were calculated as the product of the corresponding column and row sums. For example, $w_1^m = 0.122 = 0.596 \cdot 0.204 = w^m \cdot w_1$.

5 Conclusions

5.1 Tradeoffs between ESM and retrospective duration questions

In selecting a particular method to measure time use, a researcher is confronted with several decisions. How can the study be economically administered? Can responses be compiled from a sufficiently representative sample? Can responses be expected to be accurate, and interpreted by subjects as they are interpreted by study designers? Retrospective duration surveys such as the CPS and NELS:88-92 are practically useful in that they can be administered in a single session. In contrast, the ESM is certainly more difficult to administer since participants are required to fill out response forms several times a day over an extended period of time. We suspected that selecting a population of subjects willing to complete the ESM would introduce some respondent selection bias. And, in fact, girls are overrepresented in the Sloan ESM sample and among those who filled out the ESM. Older students were less likely to respond to beeps, although this difference does not appear to be significant. However, with respect to specific characteristics of parents' employment, the work status of parents of ESM students appears to be representative of the overall population of households with adolescent children.

But as some have suggested, people who agree to complete the ESM may be more organized and diligent. Students who completed the ESM had slightly higher grade point averages than students in the NELS sample, with the highest grades being reported by those students who completed 15 or more beeps. These higher rates may be confounded by the fact that there are more females in the ESM sample, and girls tend to have higher grades than boys in elementary and high school.

The other more problematic issue regarding the ESM is response rates by activity and time of week. We found that after-school and weekend beeps were underreported. This problem can be handled through weighting procedures, and we have shown how it is possible to weight the sample, adjusting for non-response by time of week and for the overrepresentation of females. What is perhaps most surprising is that even though the ESM tends to have lower response rates after school and on weekends, when estimating the percent of adolescents who have worked, the results from the ESM are nearly identical with national samples. These results suggest that ESM responses reporting on activities outside the household and outside of school are not as spurious as some have assumed that they may be (Jacobs, 1998). This comparability is also achieved when comparing CPS and ESM estimates of the average hours worked by high school seniors. However, if we weight the ESM sample by differential response patterns, the percent who are working remains consistent with national samples, but the average hours worked by seniors is slightly lower than the CPS.

5.2 Implications for other studies of time use

Overall, there is little tendency for a potentially burdensome instrument like the ESM to undercount the activities performed by "busy" people, like teenagers with jobs or students with good grades. While "busy-ness" may not be an important factor in discouraging responses, "compliance" may be an important factor in encouraging it, and there are systematic differences across people in this regard. Response rates were significantly different between the genders, and somewhat different according to grades received in school. These two results are likely to generalize to other potentially burdensome time use instruments, and it seems that a demographic-weighting strategy, as shown above, is one strategy for adjusting for some of this non-response.[22] When it comes to analysis of the ATUS, it may also be advisable to add variables that predict compliance, like GPA, to the CPS questionnaire items.

We also used weights to adjust for time-of-week differences in response rates. A precise analogue for this kind of bias may not exist for time diary studies, but there is the question of the level of detail a time diary respondent might provide, and whether this level of detail might vary with the time of week. Recall that a completed time diary consists of a sequence of activities and their time intervals, with the partitioning of time largely an object of choice by the respondent. For example, a time diary respondent is free to respond "at work from 9 am to 5 pm," or to offer an additional level of detail for the activities he performed at work (or elsewhere) between 9 am and 5 pm. It seems likely that the level of detail provided will be a function of time of week (perhaps with less detail during work time), so that short duration activities will be undercounted to a degree determined by the time of week in which they tend to occur. One strategy for dealing with this problem is to weight time diary measures of short duration activities according to their time of week, as we did with all activities in the ESM, to obtain noticeably more accurate estimates of time use.

Ideally, comparison of time use measures should examine different methods for obtaining time use data, such as time diaries versus ESM and survey responses. Since there were no time diary studies of adolescents collected during the early 1990s, when other studies used in these analyses were conducted, it was not possible to compare time diary with data obtained using surveys and ESM. It would be beneficial if future studies included multiple methods for obtaining time use data so that estimates based on these different methods could be compared. When different measures of time use are available for the same sample, issues of non-response are more obvious. If the sample is not representative, however, weighting procedures will still need to be employed, and comparisons with nationally representative samples will help identify how the data should be weighted.

[22] In fact, since the CPS forms the sampling frame for the ATUS, ATUS users should have an easier time creating demographic weights because they do not have to account for a difference in sampling frames, as we did with the Sloan study.

References

Bound, J. et al. (1990), Measurement error in cross-sectional and longitudinal labor market surveys: Validation study evidence, in: Hartog, J., Ridder, G. and J. Theeuwes (eds.), *Panel Data and Labor Market Studies*, New York, North Holland, 1-19.

Committee on the Health and Safety Implications of Child Labor (1998), *Protecting youth at work*, Washington, DC, National Academy Press.

Csikszentmihalyi, M. and I.S. Csikszentmihalyi (1988), *Optimal experience: Studies of flow in consciousness*, Cambridge, Cambridge University Press.

Csikszentmihalyi, M. and R. Larson (1984), *Being adolescent: Conflict and growth in the teenage years,* New York, Basic Books.

Csikszentmihalyi, M. and B. Schneider (2000), *Becoming adult: How teenagers prepare for the world of work*, New York, Basic Books, 2000.

Csikszentmihalyi, M., Rathunde, K. and S. Whalen (1993), *Talented teenagers: The roots of success and failure,* Cambridge, Cambridge University Press.

Gershuny, J. et al. (1986), Time budgets: Preliminary analyses of a national survey, in: *The Quarterly Journal of Social Affairs*, Vol. 2, No. 1, 13-39.

Hochschild, A. (1989), *The second shift: Working parents and the revolution at home*, New York, Viking.

Hogan, H. and G. Robinson (1993), What the Census Bureau's coverage evaluation programs tell us about differential undercount, Washington, DC, Census Bureau.

Hoogstra, L. (forthcoming), The design of the 500 Family Study, in: Schneider, B. and L. Waite (eds.), *Being Together, Working Apart: Dual-Career Families and the Work-Life Balance*, Cambridge, Cambridge University Press.

Horvath, F.W. (1982), Forgotten unemployment: Recall bias in retrospective data, in: *Monthly Labor Review*, Vol. 105, No. 3, 40-43.

Jacobs, J.A. (1998), Measuring time at work: Are self-reports accurate?, in: *Monthly Labor Review*, Vol. 121, No. 12, 42-53.

Jeong, J. (forthcoming), Obtaining accurate measures of time use from the ESM, in: Schneider, B. and L. Waite (eds.), *Being Together, Working Apart: Dual-Career Families and the Work-Life Balance*, Cambridge, Cambridge University Press.

Juster, F.T. and F.P. Stafford (eds.) (1985), *Time, goods, and Well-Being*, Ann Arbor, MI, Survey Research Center, The University of Michigan.

Juster, F.T. and F.P. Stafford (1991), The allocation of time: Empirical findings, behavioral models, and problems of measurement, in: *Journal of Economic Literature*, Vol. 29, No. 2, 471-522.

Kalton, G. (1985), Sample design issues in time-diary studies, in: Juster, F.T. and F.P. Stafford (eds.), *Time, Goods and Well-Being,* Ann Arbor, MI, Survey Research Center, The University of Michigan, 93-112.

Kubey, R. and M. Csikszentmihalyi (1990), *Television and the quality of life: How viewing shapes everyday experience,* Hillsdale, NJ, Erlbaum.

Lee, Y.-S. and L.J. Waite (2001), *Measuring housework in the family economy: Survey and ESM measures of housework time of mom, dad, and teen in working families*, presented at the conference Dutiful Occasions: Working Families, Everyday Lives, Ann Arbor, MI, University of Michigan

Leete, L. and J.B. Schor (1994), Assessing the time-squeeze hypothesis: Hours worked in the United States, 1969-89, in: *Industrial Relations*, Vol. 33, No. 1, 25-43.

Marini, M.M. and B.A. Shelton (1993), Measuring household work: Recent experience in the United States, in:

Social Science Research, Vol. 22, No. 4, 361-82.

Mulligan, C.B. (1998), Substitution over time: Another look at labor supply over the life cycle, in: *NBER Macroeconomics Annual 1998*, Vol. 13, 75-152.

National Research Council, Committee on National Statistics (2000), *Time-use measurement and research*, Washington, DC, National Academy Press.

Pencavel, J. (1986), Labor supply of men: A survey, in: Ashenfelter, O. and R. Layard (eds.), *Handbook of Labor Economics*, Amsterdam, North Holland, 103-124.

Robinson, J.P. (1985), The validity and reliability of diaries versus alternative time use measures, in: Juster, F.T. and F.P. Stafford (eds.), *Time, Goods, and Well-Being*, Ann Arbor, MI, Survey Research Center, The University of Michigan, 33-62.

Robinson, J.P. and A. Bostrum (1994), The overestimated workweek: What time diary measures suggest, in: *Monthly Labor Review*, Vol. 117, No. 8, 11-23.

Robinson, J.P. and G. Godbey (1997), *Time for life*, University Park, PA, Pennsylvania State University Press.

Ruhm, C.J. (1990), Bridge jobs and partial retirement, in: *Journal of Labor Economics*, Vol. 8, No. 4, 482-501.

U.S. Bureau of the Census (2000), *Current population survey - Design and methodology*, Census Bureau Technical Paper 63, Washington, DC, Census Bureau.

Zuzanek, J. (1999), Experience sampling method: Current and potential research applications, paper prepared for the Time-Use Measurement and Research Workshop of the Committee on National Statistics organized by the National Research Council, Washington, DC, May 27-29.