American Time Use Survey Substitution of Days

Final Report

March 29, 2013

Prepared for:
Bureau of Labor Statistics
Washington, DC

Prepared by:
Westat
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

Westat®

Table of Contents

Table of Contents

Part II

Table of Contents

Part III

Part III (continued)

Table of Contents

Part IV

General Introduction

Time use surveys, such as the American Time Use Survey (ATUS), present unique methodological challenges to survey designers. In addition to obtaining probability samples of persons, time use surveys must also obtain probability samples of days of the week in order to adequately describe the level of activity in the population. This essentially means assigning members of the sample to participate in the survey on a specific day of the week. However, the random assignment of a person to a specific day increases the amount of effort required to recruit members of the sample to participate in the survey, because the sample person may not be available (or willing) to participate on the day to which she was assigned.

Any design feature that decreases sample persons' ability or willingness to take part in a survey may also lead to lower response rates and the potential for nonresponse bias. Indeed, ATUS response rates have been under 60 percent since the survey's inception in 2003. Research into the effect of nonresponse on the time use estimates has shown that there are systematic relationships between demographic or background characteristics of the designated ATUS respondent and his or her probability of responding (Abraham, Maitland, and Bianchi, 2006; Fricker and Tourangeau, 2010). Some studies have also found evidence of potential bias in the data (Abraham, Helms, and Presser, 2009).

The ATUS sample is drawn from households that have completed the final round of the Current Population Survey (CPS). A specific designated person (DP) from these CPS households is then assigned a specific day of the week to report on; they are asked about the activities that they took part in from 4 a.m. the previous day to 4 a.m. on the day of the interview. Although the designated person is called up to eight weeks on the designated day of the week for the ATUS interview, this design feature still represents a significant challenge to achieving a higher a response rate.

Westat has been investigating the possibility of relaxing some of these restrictive design features that might lower the response rate to the ATUS. The pages that follow present the results of Westat's findings. Parts I-III of this report are specific to substitution of the day of week. Part I presents a literature review of the different methods used in time use studies. Part II presents the results of analysis of the existing ATUS microdata and other relevant data to understand the potential impact of allowing day of the week substitution. Part III presents an experimental design to implement a day of the week experiment. Part IV presents the results of an analysis of the existing ATUS microdata to understand the potential impact of substituting the designated person.

Literature Review on Day of Week Substitution in Time Use Surveys

Part I

1 Background and Introduction

Time use surveys, such as the American Time Use Survey (ATUS), present unique methodological challenges to survey designers. In addition to obtaining probability samples of persons, time use surveys must also obtain probability samples of days of the week in order to adequately describe the level of activity in the population. This essentially means assigning members of the sample to participate in the survey on a specific day of the week. However, the random assignment of a person to a specific day increases the amount of effort required to recruit members of the sample to participate in the survey, because the sample person may not be available (or willing) to participate on the day to which she was assigned.

Any design feature that decreases sample persons' ability or willingness to take part in a survey may also lead to lower response rates and the potential for nonresponse bias. Indeed, ATUS response rates have been under 60 percent since the survey's inception in 2003. Research into the effect of nonresponse on the time use estimates has shown that there are systematic relationships between demographic or background characteristics of the designated ATUS respondent and his or her probability of responding (Abraham, Maitland, and Bianchi, 2006; Fricker and Tourangeau, 2010). Some studies have also found evidence of potential bias in the data (Abraham, Helms, and Presser, 2009).

The ATUS sample is drawn from households that have completed the final round of the Current Population Survey (CPS). A specific designated person (DP) from these CPS households is then assigned a specific day of the week to report on; they are asked about the activities that they took part in from 4 a.m. the previous day to 4 a.m. on the day of the interview. Although the designated person is called up to eight weeks on the designated day of the week for the ATUS interview, this design feature still represents a significant challenge to achieving a higher a response rate.

The goal of this literature review is to understand different methods for allocating respondents to days of the week. This paper begins with a review of the empirical literature investigating the different methods. Next, we review the methodologies of current and past time use studies. Finally, we discuss a few surveys that use designs that are similar to those in time use studies and encounter similar difficulties.

2 Review of Empirical Literature

The literature discusses several issues related to the allocation of respondents across days of the week. This section of the literature review covers the potential advantages and disadvantages of the different approaches and, where possible, discusses some of the empirical findings regarding the effects of each approach.

Convenient Day Versus Designated Day. The most basic decision to be made with respect to day of week assignment is whether one allows the sample person to complete the time diary on a day that is convenient for the interviewer or respondent or whether the research design designates the day of the week on which the sample person is to complete the diary. We refer to a "diary" here even though ATUS does not use a physical diary, but instead collects diary-like information over the telephone.

There are advantages and disadvantages to the convenient day approach. The main advantage is that this approach allows greater flexibility in call scheduling. The convenient day approach allows the interviewer to make multiple contacts within a short period of time rather than waiting an entire week to contact sample persons who were not available on their assigned day. A disadvantage is that there is no probability selection mechanism to accurately represent days of the week. Many of the earliest time use surveys conducted in the 1960s and 1970s completed the interviews on what amounted to a convenient diary day. The interviewers may have been instructed to collect roughly equal numbers of diaries across days of the week; however, no probability mechanism was used to select the days (Kalton, 1985). Convenient day assignment also occurs by default when the interviewer is allowed to make repeated contact attempts regardless of the day on which the initial attempt was made.

The convenient day approach can be biasing when the level or type of activity performed varies systematically by day of the week and when respondents hold a preference for particular days. Kinsley and O'Donnell (1983) highlight at least three reasons why a sample person might want to delay the diary day. First, he or she might be too busy to participate and refuse to complete the interview on the assigned day. Second, the sample person might be engaged in socially unacceptable behavior and want to avoid participation. Third, the sample person might be unavailable to complete

the interview. Recent research demonstrates that noncontact can be a significant contributor to nonresponse in time use surveys like the ATUS (Abraham, Maitland, and Bianchi, 2006).

Unlike the convenient day approach, the designated day approach provides permits the selection of a probability sample of days. The primary difficulty with this approach is that only a certain percentage of interviews will be completed on the first attempt. Kalton (1985) suggested two procedures that might be applied when the interview is not completed on the first attempt. First, one could extend the recall period, allowing the data to be collected two or three days after the designated day. This would give field staff more time to contact the sample person. Kalton suggested that the recall could only be extended two to three days for a weekday diary date since memory decay occurs quite rapidly for weekdays beyond that point. However, decay may take somewhat longer for weekend days since these weekend days are more distinctive than weekdays. Second, if it is not possible to complete the interview in a reasonable amount of time from the designated date, the survey could complete the interview on the next closest date that falls on the same day of the week. Kalton also suggested that advanced warning about the interview date would be particularly helpful for time use studies.

There is a clear tradeoff here. On the one hand, the designated day approach has the advantage of providing a probability mechanism to represent all days of the week. On the other hand, the design is more difficult to implement. Interviewers may push more reluctant respondents to complete the diary on the designated day to avoid having to call back one week later; this may lead to lower quality reports of time spent in activities (Kinsley and O'Donnell, 1983). For example, reluctant respondents may provide less complete diaries, reporting fewer activities in the interview. Recent research by Fricker and Tourangeau (2010) found some evidence for lower quality data from reluctant respondents, depending on the data quality indicator being used. Kinsley and O'Donnell also point out that it may be more difficult to convince the respondent to make a commitment if the interview must be completed on a designated day.

Experience from the Swedish Time Budget Survey highlights some of the difficulties with completing diaries on the designated day. Respondents to that survey were asked to complete a telephone recall diary four days before they filled out a self-administered diary on a designated day. Lyberg (1989) reports that only 47 percent of the sample agreed to complete the self-administered time diary on the designated date. And, even though the protocol called for completing the telephone recall diary four days prior to the self-administered diary, this only happened 43 percent of

the time. Somewhat encouragingly, Lyberg found that the propensity to respond to the diary was not associated with day of the week.

There are actually very few studies that make direct comparisons between the convenient day and designated day approaches. One was an experiment done as part of a large scale pilot study for the Canadian Time Use Survey (Kinsley and O'Donnell, 1983). Kinsley and O'Donnell reached three conclusions based on this study. First, the estimated time spent away from home is higher under the convenient day approach. This is because respondents may not be reached on an initial contact attempt when they are not at home, but may be reached on the following day when the respondent is at home. Second, the designated day approach is somewhat more difficult to implement. Although there were no differences in response rates under the two approaches, the designated day approach required more telephone calls to complete the interview. Third, there was a tendency for respondents to report fewer activities under the designated day approach. On three out of the seven days, respondents reported more activities under the convenient day approach.

A study by Stewart (2002) used simulations of time use data to look specifically at the possibility of substituting similar weekdays from Monday to Thursday. The simulation actually examined several different contact strategies: convenient day, designated day, designated day with postponement, and designated day with postponement and substitution. In addition, he examined two different types of potential biases. First, he looked at activity bias, which occurs when the probability of interviewing a potential respondent is correlated with the respondent's activities on that particular diary day. Second, he looked at noncontact bias, which occurs when differences in the probability of contacting individuals are caused by differences in the activities across individuals. The simulation also assumed different patterns of "hard to contact" and "easy to contact." Like Kinsley and O'Donnell's study, Stewart's simulations showed that the convenient day approach produced higher estimates of time spent in activities away from home. He also found that extending the field period from four weeks to eight weeks virtually eliminated any difference in noncontact bias between the designated day with postponement and designated day with postponement and substitution strategies.

The potential for bias with the convenient day approach has led many authors to prefer the designated day approach (Harvey, 1993; Harvey, 1999; Kalton, 1985; Kinsley and O'Donnell; Stewart, 1985).

Extending the Recall Period Versus Postponing the Interview. Another design option that may interact with how sample persons are allocated to a day of the week is the length of the recall period. Most time use studies such as the ATUS ask the respondent to report the activities that they were engaged in over the previous 24 hours from the day of the interview. The advantages of reporting about the previous day are obvious from the perspective of memory. The task of recalling all activities in the previous 24 hours can be difficult. Hence, the recall should take place as close to the reference day as possible. In addition, memories across different days may begin to blend together when the reference period is too distant from the designated day.

Even though there are reasons for asking about the prior 24 hours from a measurement perspective, this limits the ability of field staff to make contact within the allotted period of time and increases the potential for nonresponse. Hence some surveys, such as the Canadian General Social Survey, relax this requirement and allow interviewers to call up to two days after the reference period.

There has been relatively little research on the length of the recall period in time use surveys. General guidance from the literature and from experienced time use researchers is that recall is difficult more than two days after the diary day (Pentland et al., 1999). Harvey (1990; 1993) cites work by Klevmarken comparing 24 and 48 hour diaries; that work found "under-reporting of some activities, a decreased number of activities reported and instances of memory lapse during some substantial time period." However, the same study by Klevmarken found small differences in the average duration of activities between yesterday and day before yesterday diaries.

Other guidance suggests that memory becomes more difficult with a longer recall period, but that the effect of recall period length may depend on whether the diary day is a weekday or weekend day (Michelson, 2005). Juster (1985) examined data from the 1975-76 and 1981-82 Surveys of Time Use in the United States. The original design of these studies was for 24 recall of the previous day from the interview. However, because the research design called for interviews with spouses to be done on the same diary day, it sometimes turned out that one or the other spouse was not available for an interview on the assigned date. Hence, the absent spouse was asked to complete the diary on a later day, which extended the recall period. This led to a relatively small proportion of the sample having a recall period of more than 24 hours. Juster found that the mean number of activities reported declined as the recall period got longer; however, this effect was limited primarily to weekdays. The general explanation for this result is that the effect of memories blending together might be greater for weekdays than weekends since weekdays tend to be more similar to each other than weekend days.

Day of Week Variability in Time Use. The extent to which similar activities are performed on different days is an important consideration in determining which days could be substituted for one another. Fortunately, this is an area where there has been a great deal of research. This section will focus on how some of the major findings in the literature regarding the participation in activities across days of the week.

The major distinction made between days is, not surprisingly, between weekdays and weekend days. Market-work-oriented activity is concentrated on weekdays, whereas entertainment and leisure activities are concentrated on the weekends (Hill, 1985). Research by Kalton (1985) addressed the topic of how many days of the week time use studies should sample. In order to draw a conclusion on this topic he examined data from the 1975-76 Surveys of Time Use in the United States. He found small differences in activity during individual weekdays with the exception of Friday. On Friday, people worked more, cooked less, did more child care, and spent less time in leisure activity like watching television than on Monday through Thursday. Saturday and Sunday were different from the rest of the week and different from each other. This was true across nearly all categories activity; however, Sunday included by far the most time in passive leisure activity such as watching television.

Some research suggests that there may actually be more variability across weekdays than one might expect. This is particularly true when one looks at differences in different demographic groups. Research by Zuzanek and Smale (1999) looked at individuals at different stages of the lifecycle and how they use their time. For example, they examined how gender, age, marital status, and presence of children influence how a person spends their time across days of the week. Using this approach, they found that the typical weekday – weekend dichotomy might not fully explain all of the meaningful differences between days of the week. They used data from the Canadian General Social Survey and reached a slightly different conclusion from Kalton (1985). In addition to weekends being different from weekdays and Fridays being different from other weekdays, Zuzanek and Smale also found that Mondays are different from Tuesday, Wednesday, and Thursday. However, these differences assume different forms depending on the demographic group. For example, during the week, working women spent the fewest hours working on Monday. Mothers at home spend the most time on housework and child care on Monday compared to other days of the week.

Summary indices of time spent in different activities can also be used to understand differences in time use by day of the week. Stinson (1999) calculated the Szalai T similarity formula using time use data from the National Human Activity Pattern Survey to compare days of the week. Here is the formula for Szalai T:

$$T_S = \sqrt{\frac{\sum_{i=1}^{k}\left(\frac{a_i - b_i}{a_i + b_i}\right)^2}{k}},$$

in which a_i is the time spent on activity i on one day and b_i is the time spent on the same activity another day; k is the total number of activities across the two days.

These global measures compare the total time spent in multiple activities in a day and range in value from 0 meaning that the days are totally similar to 1 meaning totally dissimilar. Stinson found that Friday, Saturday, and Sunday are different from the rest of the week. Monday through Thursday show the most similarity. There was also some variability by demographic group. She found that those with children and those employed part-time show less similarity across days. Stewart (2006) took a similar approach using the same data as Stinson, but calculated several different types of indices in order to test the sensitivity of the findings to the calculation of the index. His conclusions concur with the primary conclusions of Stinson. He also found that Friday, Saturday, and Sunday are different from the rest of the week.

Optimal Calling Times and at Home Patterns. The literature on optimal calling times is also relevant when deciding the extent to which substituting days of the week may affect the time use estimates. This literature also provides insight into whether the likelihood of contacting someone varies by day of the week. It also provides insight into whether day of week substitution might lead to overestimates of time spent in activities away from home, as suggested by Kinsley and O'Donnell (1983) and Stewart (2002). Weeks et al. (1980) reports the results of initial contact attempts from an in person screening of households. They found that there was little variation in the proportion of households contacted during weeknights (Monday-Friday). The best overall day to make contact was Saturday, when more than half of the screened households had an eligible person at home during most of the day. Weeks, Kulka, and Pierson (1987) performed the same type of analysis on telephone call data and found that roughly 75 percent of first calls were answered on weekday nights and 70 percent of calls were answered on Saturday mornings, Saturday evenings, and Sunday

evenings. Hence, this suggests that the time of day that one makes calls may vary by day, but the overall ability to make contact may not differ much across days.

One can also use time use data to understand at home patterns. Groves (1989) reports on a study by Hill (1978) that used time use data to estimate the time of day that people are at home. Maitland (2006) examined at home patterns in the 2003 and 2004 ATUS. He found that similar proportions of respondents were at found at home throughout the daytime hours Monday-Friday. However, the proportion of people at home during day time hours was slightly higher on Monday. In addition, the proportion of people at home in the evening was lower on Friday than on other weekdays. The highest proportion of people being at home during the daytime hours occurred on the weekend with Sunday having the highest percentage of people at home during these hours. Saturday evenings were similar to Friday evenings and Sunday evenings were similar to weekday evenings.

In summary, much of the literature on suggests that there is an important distinction between weekdays and weekends. However, there may also be differences between days of the week such as Monday or Friday and other weekdays. Saturdays also appear to be different from Sunday. Important differences in time use across days of the week may also depend on which demographic groups are of primary interest in a research program.

Review of Past and Current Surveys 3

This section reviews the methodology of current and past time use studies. We performed a selective review of some of the more recent time use surveys involving more general populations. The studies reviewed were some of the most commonly cited studies in the literature. In addition, the selected studies highlight some of the main differences between methodologies used in time use surveys. However, as can be seen in Table 1, many of the studies share the main characteristics of the ATUS. Many of the studies reviewed here can be found through links on the Centre for Time Use Research home page. Links to the specific studies are shown in the references to this paper.

Relatively few of the studies we examined allow day of week substitution. There were three studies that did allow some degree of substitution. Substitution often arises due to particular challenges with the study design, such as the need to interview two people about the same day, taking a convenient day approach, or making an effort to increase the response rate near the end of a survey. The Disability and Use of Time survey is a supplement to the Panel Survey of Income Dynamics (PSID). Married couples age 60 or over from the PSID are selected for this survey. At the beginning of the study, each couple is assigned a designated day as well as a backup day in case they cannot be reached on the designated day. The designers assigned the couples randomly to all possible combinations of the weekday and weekend diary days. In contrast, the 1975-76 Americans' Use of Time Survey included day of week substitution, since respondents were not randomly assigned to days of the week initially due to cost constraints (Kalton, 1985). The Princeton Affect and Time Survey made a similar decision; however, this survey began with the sample being randomly assigned to days of the week. Data collection shifted to a convenient day approach towards the end of the field period to increase the response rate to the survey.

Although most studies collect time use data for one day, some studies do ask about more than one day to measure within person variability in time use. Table 3-1 shows some variability with respect to how many days of the week that are sampled by each study. The latest guidelines for the harmonised European time use surveys recommend sampling two days of the week—a weekday and a weekend day (Eurostat, 2009). However, each participating country makes its own decision on how many days to sample (Eurostat, 2005). The Disability and Time Use study in the United States also samples one weekday and one weekend. The 1975-76 Americans' Use of Time survey sampled four

Table 3-1. Time use survey methodologies

Survey title	Year	Substitution of days	Substitution of persons	Number of diary days	Recall period	Mode	Response rate	Universe	Stand-alone survey
American Time Use Survey	2003-2011	No	No	1	24 hours	CATI	54.6% (RR2)	U.S. 15+	No
Americans' Use of Time	1975-1976	Yes	No	4 (2 weekdays, Saturday, Sunday)	24 hours (up to 7 days)	PAPI, CATI	72% initial wave, 44% all 4 waves	U.S. Adults	Yes - Panel
Americans' Use of Time	1985	No	No	1	24 hours	Mail, CATI, PAPI	51% Mail, 67% CATI	U.S. 12+	Yes
Canadian GSS	1998	No	No	1	24 (up to 48 hours)	CATI	77.6%	Canadians 15+	Yes
Disability and Use of Time	2009	Yes	No	1 weekday, 1 weekend	24 hours	CATI	73% (at least one partner)	U.S. Married, 60+	No
Dutch Time Use Survey	1975- 2005	No	No	7	24 hours	In person/ Self admin.	35%	Dutch 12+	Yes
Family Interaction, Social Capital, and Trends in Time Use	1998-1999	No	No	1	24 hours	CATI	56%	U.S. Adults	Yes
Harmonized European Time Use Surveys	2008	No	No	2 (1 weekday, 1 weekend)	24 hours	In person	NA	European residents Age 10+	Yes
National Human Activity Pattern Survey	1992-1994	No	No	1	24 hours	CATI	63%	U.S. households	Yes
National Time Diary Study	1994-1995	No	No	1	24 hours	CATI	65%	U.S. Adults	Yes
Princeton Affect and Time Survey	2006	Yes	No	1	24 hours	CATI	37% (RR3)	U.S. 16+	Yes

days including two weekdays, a Saturday, and a Sunday. The Dutch Time Use Survey has respondents complete a diary for each day of the week. This survey also has one of the lowest response rates of the time use studies we reviewed (see Van Ingen, Stoop, and Breedveld, 2011).

The 24 hour recall period is also by far the most common among the studies reviewed. However, there are occasions where recall periods of longer than 24 hours are allowed. Similar to day of week substitution, extension of the recall period often arises in reaction to challenges with a specific feature of a research design or to increase overall response rates. For example, as discussed in the previous section, the Americans' Use of Time study required couples to be interviewed about the same diary day. When one of the members of the couple was not available, the recall period was extended to as many as seven days for the absent respondent. In contrast, the research design for the Canadian General Social Survey allows recall periods of up to 48 hours to improve response.

Review of Studies With Similar Methodologies

4

Other studies face similar methodological issues related to assigning sample persons to days of the week. We first review briefly a couple of these studies conducted by Westat in the areas of food intake and travel. Next we review two studies from the literature: one study of school instruction and one study of health and aging. Finally, we end with a general discussion of substitution of sample units.

The National Household Travel Survey. The National Household Travel Survey (NHTS) provides information to assist transportation planners and others who need comprehensive data on travel and transportation patterns in the United States. The NHTS is a survey of the civilian, non-institutionalized population of the United States age 16 and over. The survey details the daily travel patterns of Americans by having respondents report about the daily trips that they make. Research shows that they travel patterns of the Americans varies significantly by day of the week (e.g., Yun and O'Kelly, 1997). Hence, respondents are asked to provide reports of trips using a designated 24-hour travel day that starts at 4:00 AM of the day assigned and continues until 3:59 AM of the following day.

The NHTS is a CATI survey that includes a household recruitment interview, travel diary, and call-back interview to collect travel detail from the respondent. The purpose of the travel diary was to aid in recall during the call-back interview. This allowed a somewhat extended recall period for the interview. Reminder calls were also made to the respondents to remind them to complete the diary on their designated day. One-seventh of the sample telephone numbers were assigned to each day of the week. Respondents to the recruitment interview were assigned a date 10 to 14 days in the future to allow time for diary mailings to reach the household. Travel detail was collected from the respondents during the call-back interview. Respondents were typically called on the day following their assigned diary day and calling continued for up to seven days until a completed interview was obtained.

Continuing Survey of Food Intakes by Individuals. The Continuing Survey of Food Intakes by Individuals is intended to provide information on the dietary status of individuals in the U.S. population. The survey involved three years of the continuous data collection beginning in 1994. A nationally representative sample of individuals of all ages was asked to report food intake for two

nonconsecutive days through the administration of 24-hour dietary recalls. Proxy interviews were only allowed for sample persons under six years of age and those who were unable to complete the interview due to physical or mental limitations. Prior to the survey, it was believed that making an appointment might influence a person's eating behavior if they knew that they were going to be asked what they had eaten. Hence, appointments were only made the first time an intake interview was to be completed.

The data collection procedures were designed to ensure that at least ten percent of the day 1 food intake diaries were completed on each day of the week. Interviewers were given a three day period within which they could interview a sample person for the day 1 intake part of the survey. Repeated visits were made within this period to complete the interview. However, interviewers were permitted to change the day of week after making repeated visits on different scheduled days at different times. The day 2 intake interview was to take place 3-10 days after the day 1 intake interview. The procedures called for the day 2 interview to take place on a different day of the week than the first interview. Only about two percent of the day 2 intake interviews took place on the same day of the week as the day 1 intake interview.

Other Examples From the Literature. We reviewed two other studies from the literature that illustrate different approaches to sampling days.

Research on classroom instruction has primarily relied on classroom observations and end of the year surveys to measure the effects of different teacher practices on student behavior. Teaching behavior is highly variable over the course of the school year. Hence, the choice of days on which one sends observers to a classroom can have a major impact on the findings of a study. Rowan (2009) reports on efforts to have teachers log data at the end of the day and over extended periods throughout the school year. This increases the burden on respondents, but reduces the impact of sampling time. Rowan reported on some comparisons between teacher logs and observational data and also between teacher logs and end of year surveys. Camburn and Barnes (2004) found that in comparison to observer data, teacher logs are most accurate for grosser levels of detail and for more frequent events. The logs seemed to help prevent teachers from overlooking quick and routine aspects of teaching. In addition, Camburn and Han (2006) found that teachers report higher frequencies of teaching practices with end of year surveys compared to logs.

Hurd and Rohwedder (2009) report on the Consumption and Activities Mail Survey (CAMS) that is a subsample of respondents in the Health and Retirement Study (HRS). The CAMS measures time

use, but has a different set of objectives compared to a survey like the ATUS. For example, it is important to capture a sufficiently representative set of activities that can be related to characteristics of the respondents measured in the HRS. In addition, the CAMS attempts to capture social interactions that are difficult to measure with a one-day diary. Given these constraints, the CAMS asks how many hours the respondents spent over the last week for activities that occur more frequently such as sleeping and housecleaning. The survey asks how many hours over the last month for less frequent activities such as going to movies. Hurd and Rohwedder compared various estimates of time use between the CAMS and ATUS and despite many methodological differences between the two surveys the time use estimates were quite similar.

General Treatment of Substitution in the Literature. Last, we reviewed sampling texts on the general treatment of substitution as a method to reduce nonresponse. Texts such as Kish (1965) and Lohr (1999) discuss substitution of sampling units and warn that substitution should be reported in the results. Lohr indicates two reasons why substitution is not a preferred method for dealing with nonresponse. First, if the reason a sample unit did not respond is related to the characteristic of interest, there is still nonresponse bias. Second, substitution causes the sample to no longer have known probabilities of selection. Lohr also provides an example from the National Longitudinal Study in which two schools were selected from each strata to participate and two other schools were selected as substitutes from each strata. Followup studies found a consistent bias due to substitution and nonresponse in this study.

Conclusion 5

This paper reviews the literature related to day-of-week substitution in time use surveys. More specifically, we have reviewed the empirical literature related to day-of-week substitution, methodologies related to past and current time use surveys, and methodologies of surveys that face similar issues in allocating days of the week to sample members.

A few recurring themes emerged throughout the review. First, there is general consensus that a designated day-of-week approach is needed to minimize bias in the time use estimates. Second, the literature makes a major distinction between weekdays and weekends. However, days like Monday at the beginning of the work week and Friday at the end of the work week sometimes appear to be slightly different from Tuesday – Thursday. Saturday and Sunday have their own unique patterns. Finally, some of the conclusions from the literature are based on relatively little published research. Although much of the literature is suggestive, there are very few studies directly addressing important issues such as whether a convenient day approach leads to significantly different estimates from the designated day approach. We only found one study that addressed this issue experimentally. We were unable to find any studies that experimentally compared the different approaches to day-of-week substitution that Stewart (2002) compared in his simulations. As recognized by Harvey (1990; 1993), there also needs to be more research on the acceptable length of the recall period in time use surveys. Additionally, does the acceptable length of the recall period differ for weekdays and weekends? These gaps in the literature leave open the possibility for some much needed methodological research that could help inform the design of the ATUS and future time use surveys.

References

Abraham, K.G., Helms, S., and Presser, S. (2009). How Social Processes Distort Measurement: The Impact of Survey Nonresponse on Estimates of Volunteer Work in the United States. *American Journal of Sociology* 114(4): 1129-1165.

Abraham, K.G., Maitland, A., and Bianchi, S. (2006). "Nonresponse in the American Time Use Survey: Who is Missing from the Data and How Much Does It Matter?" *Public Opinion Quarterly* 70(5):676-703.

Camburn, E., and Barnes, C.A. (2004). "Assessing the Validity of a Language Arts Instruction Log Through Triangulation." *Elementary School Journal* 105(1):49-74.

Camburn, E., and Han, S.W. (2006). *Factors Affecting the Validity of Teachers' Reports of Instructional Practices on Annual Surveys*. Madison: Wisconsin Center for Education Research.

Fricker, S., and Tourangeau, R. (2010). "Examining the Relationship Between Nonresponse Propensity and Data Quality in Two National Household Surveys." *Public Opinion Quarterly* 74(5): 934-955.

Groves, R.M. (1989). *Survey Errors and Survey Costs*. New York: Wiley.

Harvey, A.S. (1990). "Time Use Studies for Leisure Analysis." *Social Indicators Research* 23(4): 309-336.

Harvey, A.S. (1993). "Guidelines for Time Use Data Collection." Social Indicators Research 30(2/3): 197-228.

Eurostat. (2005). Comparable time use statistics: National tables from 10 European countries. Luxembourg: Office for Official Publications of the European Communities.

Eurostat. (2009). Harmonized European Time Use Surveys. Luxembourg: Office for Official Publications of the European Communities.

Hill, M.S. (1985). "Patterns of Time Use." In *Time, Goods, and Well Being*, ed. F. Juster and F. Stafford. Ann Arbor: University of Michigan, Institute of Social Research, 133-176.

Hill, D.H. (1978). "Home Production and the Residential Electric Load Curve." *Resources and Energy* 1: 339-358.

Hurd, M.D., and Rohwedder, S. (2009). "Time Use in the Older Population: Variations by Socioeconomic Status and Health." In *Calendar and Time Diary Methods in Life Course Research*, ed. R.F. Belli, F.P. Stafford, and D.F. Alwin. Los Angeles: Sage.

Juster, F.T. (1985). "The Validity and Quality of Time Use Estimates Obtained from Recall Diaries." In *Time, Goods, and Well Being*, ed. F. Juster and F. Stafford. Ann Arbor: University of Michigan, Institute of Social Research, 63-91.

Kalton, G. (1985). "Sample Design Issues in Time Diary Studies." In *Time, Goods, and Well Being*, ed. F. Juster and F. Stafford. Ann Arbor: University of Michigan, Institute of Social Research, 333-351.

Kinsley, B., and O'Donnell, T. (1983). "Marking Time: Methodology Report of the Canadian Time Use Pilot Study – 1981." In *Explorations in Time Use* (vol. 1), Ottawa: Department of Communications, Employment and Immigration.

Kish, L. (1965). *Survey Sampling*. New York: Wiley.

Lohr, S. (1999). *Sampling Design and Analysis*. New York: Duxbury Press.

Lyberg, I. (1989). "Sampling, Nonresponse, and Measurement Issues in the 1984-85 Swedish Time Budget Survey." Proceedings of the Fifth Annual Research Conference. Washington DC: Bureau of the Census, 210-238.

Maitland, A. (2006). "Contacting Sample Households: Lessons from Time Use Data." Presentation at the 61st Annual Conference of the American Association for Public Opinion Research. Montreal, Canada.

Michelson, W. (2005). *Time Use: Expanding the Explanatory Power of the Social Sciences*. London: Paradigm Publishers.

Pentland, W.E., Andrew, S., Harvey, M. Lawton, P., and McColl. M. (1999). *Time Use Research in the Social Sciences*. New York: Kluwer Academic/Plenum Publishsers.

Rowan, B., Camburn, E.M., and Correnti, R. (2009). "Using Time Diaries to Study Instruction in Schools." In *Calendar and Time Diary Methods in Life Course Research*, ed. R.F. Belli, F.P. Stafford, and D.F. Alwin. Los Angeles: Sage.

Stewart, J. (2002). "Assessing the Bias Associated with Alternative Contact Strategies in Telephone Time-Use Surveys." Washington DC: Bureau of Labor Statistics.

Stewart, J. (2006). "Assessing Alternative Dissimilarity Indexes for Comparing Activity Profiles." *International Journal of Time Use Research*. 3(1): 49-59.

Stinson, L. (1999). "Day-of-Week Differences and Implications for Time Use Research." Proceedings of the American Statistical Association, Survey Research Methods Section. Alexandria, VA: American Statistical Association, 998-1003.

Van Ingen, E., Stoop, I., and Breedveld, K. (2011). "Nonresponse in the Dutch Time Use Survey: Strategies for Response Enhancement and Bias Reduction" *Field Methods* 21(1):69-90.

Weeks, M.F., Jones, B.L., Folsom, R.E., and Benrud, C.H. (1980). "Optimal Times to Contact Sample Households." *Public Opinion Quarterly* 44(1): 101-114.

Weeks, M.F., Kulka, R.A., and Pierson, S.A. (1987). "Optimal Call Scheduling for a Telephone Survey." *Public Opinion Quarterly* 51(4): 540-549.

Yun, D., and O'Kelly, M.E. (1997). Modeling the Day-of-the-week Shopping Activity and Travel Patterns. *Socio-Economic Planning Science* 31(4): 307-319.

Zuzanek, J., and Smale, B.J.A. (1999). "Life-Cycle and Across-the-Week Allocation of Time to Daily Activities." In *Time Use Research in the Social Sciences*, ed. W. Pentland, A. Harvey, M.P. Lawton, and M.A. McColl. New York: Kluwer Academic/Plenum Publishsers.

Links to time use studies reviewed in Table 1.

American Time Use Survey

http://www.bls.gov/tus/atususersguide.pdf

Americans' Use of Time (1975-76)

http://www-2009.timeuse.org/information/studies/data/downloads/usa/1975/1975-76-time-use-codebook.pdf

Americans' Use of Time (1985)

http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/9875/detail#access-and-availability

Canadian GSS

http://www.statcan.gc.ca/pub/12f0080x/4194543-eng.pdf

Disability and Use of Time Survey

http://psidonline.isr.umich.edu/DUST/dust09_UserGuide.pdf

Dutch Time Use Survey

See Van Ingen, Erik, Ineke Stoop, and Koen Breedveld. 2011. "Nonresponse in the Dutch Time Use Survey: Strategies for Response Enhancement and Bias Reduction" Field Methods 21(1):69-90.

Family, Interaction, Social Capital and Trends in Time Use

http://www-2009.timeuse.org/information/studies/data/downloads/usa/1998-1999/US98-99.pdf

Harmonized European Time Use Surveys

http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-08-014/EN/KS-RA-08-014-EN.PDF

National Human Activity Pattern Survey

http://exposurescience.org/heR.doc/library/heR.ActivityData/html/survey.methods.html\

National Time Diary Study

http://www-2009.timeuse.org/information/studies/data/downloads/usa/1994-1995/USA-1994-95-sample-design.pdf

Princeton Affect and Time Diary Survey

http://krueger.princeton.edu/akrueger/pages/atus-subjective-well-being-module

Links to other studies reviewed

National Household Travel Survey

http://nhts.ornl.gov/

Continuing Survey of Food Intakes by Individuals

http://www.ars.usda.gov/Services/docs.htm?docid=7764

American Time Use Survey Microdata Analysis for Day of Week Substitution

Part II

1 Background and Introduction

Time use surveys, such as the American Time Use Survey (ATUS), present unique methodological challenges to survey designers. In addition to obtaining probability samples of persons, time use surveys must also obtain probability samples of days of the week in order to adequately describe the level of activity in the population. This essentially means assigning members of the sample to participate in the survey on a specific day of the week. However, the random assignment of a person to a specific day increases the amount of effort required to recruit members of the sample to participate in the survey, because the sample person may not be available (or willing) to participate on the day to which she was assigned.

Any design feature that decreases sample persons' ability or willingness to take part in a survey may also lead to lower response rates and the potential for nonresponse bias. Indeed, ATUS response rates have been under 60 percent since the survey's inception in 2003. Research into the effect of nonresponse on the time use estimates has shown that there are systematic relationships between demographic or background characteristics of the designated ATUS respondent and his or her probability of responding (Abraham, Maitland, and Bianchi, 2006; Fricker and Tourangeau, 2010). Some studies have also found evidence of potential bias in the data (Abraham, Helms, and Presser, 2009).

The ATUS sample is drawn from households that have completed the final round of the Current Population Survey (CPS). A specific designated person (DP) from these CPS households is then assigned a specific day of the week to report on; they are asked about the activities that they took part in from 4 a.m. the previous day to 4 a.m. on the day of the interview. Although the designated person is called up to eight weeks on the designated day of the week for the ATUS interview, this design feature still represents a significant challenge to achieving a higher a response rate.

This paper reports the results of a microdata analysis of the ATUS data concerning day of week substitution. In preparation for this analysis, Westat reviewed the literature related to day-of-week substitution in time use surveys in general. More specifically, the empirical literature related to day-of-week substitution, methodologies related to past and current time use surveys, and methodologies of surveys that face similar issues in allocating days of the week to sample members.

A few recurring themes emerged throughout the review. First, there is general consensus that a designated day-of-week approach is needed to minimize bias in the time use estimates. Second, the literature makes a major distinction between weekdays and weekends. However, days like Monday at the beginning of the work week and Friday at the end of the work week sometimes appear to be slightly different from Tuesday – Thursday. Saturday and Sunday have their own unique patterns.

The goal of this analysis is to use the ATUS microdata to determine the feasibility of day-of-week substitution, to estimate potential bias associated with substitution, and inform an experimental design to test a procedure for day-of-week substitution.

Methods 2

There are two important empirical questions that have to be considered in substituting days of the week in the ATUS design. The most obvious question is whether time use estimates are similar across different days of the week. Substitution becomes more feasible the more similar the estimates are across different days of the week.

A couple of different approaches were taken to address the similarity of time use estimates across days of the week. First, we compare key estimates of time use across different days of the week. For example, we compare estimates of the average time spent in each of the major activity codes from the ATUS activity file across days of the week. This gives an indication of which activities vary across days of the week and which days are the most similar or different.

Next, we take a more global approach to comparing time us across days of the week by creating summary measures of time use for each day of the week. For example, one can create an activity profile for each day of the week that summarizes time use on that day across all different types of activities. This approach is similar to that of Stewart (2006) and Stinson (1999) who compared different dissimilarity indexes that summarize differences in time use between days of the week. Stewart (2006) found that the weighted absolute difference dissimilarity index was the best summary measure since it is robust to the level of aggregation and is easily interpreted as the average proportional difference in the time spent in all activities. The weighted absolute difference can be computed by using the following formula.

$$T_{WAD} = \sum_{i=1}^{k} \frac{|a_i - b_i|}{a_i + b_i}\left(\frac{a_i + b_i}{2880}\right)$$

In this formula a_i is the time spent in activity i by group A and b_i is the time spent in activity i by group B. The groups in our analysis would be the different days of the week.

We also calculate the Szalai T index shown below:

$$T_S = \sqrt{\frac{\sum_{i=1}^{k}\left(\frac{a_i - b_i}{a_i + b_i}\right)^2}{k}},$$

in which a_i is the time spent on activity i on one day and b_i is the time spent on the same activity another day; k is the total number of activities across the two days.

The second issue is how response propensities vary by day of the week. Substitution is more feasible the smaller the differences in respondent availability and cooperation across different days of the week.

There are also different approaches that can be taken to understand the availability and cooperation of respondents across days of the week. The ATUS includes a rather extensive set of data that can be used to understand the availability and cooperation of respondents by day. First, we utilize the information on the ATUS case history and call history file to explore whether there are different days of the week on which respondents are more likely to be contacted or cooperative.

Response propensity models are common tools for understanding the factors that influence the probability of response. These are logistic regression models predicting whether or not a sample case responded to the survey. Our analysis begins by predicting the probability of responding to the survey using the designated day of week from the ATUS case history or call history file. If there is no relationship between time use and response propensity for a given day, substitution by day is less likely to distort ATUS estimates.

Since the designated ATUS respondent is selected from the outgoing rotation of the CPS there is other information available about ATUS respondents that are helpful in the response propensity models (see, for example, Fricker and Tourangeau, 2010). The ATUS-CPS file contains a vast amount of demographic information about the designated person and the household that they live in. Variables from the CPS were added to the logistic regression model in order to understand if any differences due to assigned day can be explained by differences in the demographic characteristics of the DPs.

Finally, the ideal data set for this analysis would consist of individuals being interviewed on multiple days across the week. This would allow one to measure the association between levels of activity across different days of the week, but for the same person. Unfortunately, the current design of the ATUS is for one person to be interviewed on a single day of the week. Our final set of analyses simulates a data structure where respondents are interviewed on more than one day of the week by grouping similar types of respondents together. This data structure was achieved through the following steps.

1. Respondents from each weekday were grouped into classes by sex (0 = Male, 1= Female), presence of a spouse (0 = No, 1 = Yes), presence of children (0 = No, 1 = Yes), age (1 = 15-30, 2 = 31-60, 3 = 61+), and employment status (0 = Unemployed or Not in the Labor Force, 1 = Employed). Each class for each specific day was output to a separate data file.

2. The cases within each data file were assigned a random number using the SAS function RANUNI and the cases were sorted by this random number.

3. Within each class, cases from different pairs of days were merged together.

4. For each pair of days, all classes were concatenated to produce a complete data file that particular pair of days.

5. Cases that did not have a match were deleted from the data file. This occurred when one particular day had more members of a class than the day that it was being paired with.

This procedure resulted in one data set to analyze for each pair of days. We then calculated the mean level of the main activities and net difference for the main categories of time use for each pair of days.

Results 3

The analysis begins by examining time use estimates by day of the week. Next we show how response propensity and outcome rates vary by day of the week. Finally, we present the results of an analysis that groups together similar respondents.

Time Spent in Major Activities. Our first set of analyses examines time use by day of the week. Recall that the time use literature generally includes three main findings with respect to differences in time use across day of the week. First, weekends are generally different from weekdays. Saturday and Sunday are different from each other. Monday and Friday can be slightly different from Tuesday to Thursday.

The ATUS microdata confirm many, if not all, of these findings from the literature. Table 3-1 illustrates the amount of time (in minutes) spent by respondents in each of the major activity codes in the ATUS. First, weekends are definitely different from weekdays. Respondents spend more time with personal care activities, household activities, consumer purchases, and leisure activities on the weekends. They spend less time with work and education activities on the weekend. Some differences also emerge for the weekdays. Respondents spend slightly less time in work activities on Monday and Friday compared to Tuesday, Wednesday and Thursday. On Friday, respondents tend to spend less time in work and more time in leisure activities and eating and drinking compared to Tuesday, Wednesday and Thursday. On Monday, respondents also tend to spend less time in work and slightly more time in leisure activities and household activities.

It is easier to substitute days if respondents engage in similar levels of the activities across days. Overall, the numbers in Table 3-1 suggest that Tuesday, Wednesday, and Thursday are the most similar days. Both Monday and Friday seem to have some idiosyncrasies that might make them less substitutable. The differences between Friday and other weekdays may be more obvious than the differences seen between Monday and the other weekdays.

Table 3-1. Mean time spent in major activities by diary day (in minutes)

Time use category	Monday (n=9,089)	Tuesday (n=9,037)	Wednesday (n=9,020)	Thursday (n=8,822)	Friday (n=8,793)	Saturday (n=22,177)	Sunday (n=22,886)
Personal care activities	556.11	552.87	555.15	549.70	537.27	584.42	622.51
Household activities	108.25	103.35	106.59	102.71	101.15	134.35	125.33
Caring for and helping household members	32.60	35.43	34.10	34.09	30.70	26.92	26.22
Caring for and helping nonhousehold members	11.25	11.41	12.45	12.27	12.80	17.14	12.73
Work and work related activities	267.05	282.31	277.20	281.4	263.14	96.21	67.19
Education	36.20	36.66	38.72	33.82	27.10	9.53	12.00
Consumer purchases	30.91	30.76	29.44	31.42	41.52	56.99	40.42
Professional and personal care services	7.50	8.63	8.34	8.70	9.57	5.06	1.83
Household services	1.59	1.26	1.25	1.17	1.91	1.45	0.34
Government services and civic obligations	0.70	0.91	0.83	0.83	0.87	0.24	0.11
Eating and drinking	69.45	68.88	70.48	70.18	78.29	81.79	80.00
Socializing, relaxing, and leisure	262.47	253.48	247.71	258.77	278.42	353.24	357.50
Sports, exercise, and recreation	20.98	20.12	19.89	20.08	21.55	31.28	23.68
Religious and spiritual activities	3.76	4.07	6.71	4.48	4.46	7.18	38.00
Volunteer activities	8.28	8.99	9.82	9.85	9.60	12.27	11.28
Telephone calls	7.21	7.05	7.24	6.07	7.01	5.95	7.12
Travel	1.43	1.42	1.73	1.44	1.50	2.08	1.61

Note. All means are weighted using the ATUS final weight. Includes all ATUS data from 2005-2011.

Analysis of Activity Profiles. We also made a more global comparison of activity profiles across days of the week. This was done using two dissimilarity indexes. These indexes compare time use across days and over the entire vector of major activity codes. Following Stinson (1999) and Stewart (2006) we summarize the data using Szalai's T dissimilarity index and the weighted absolute-deviation index as summary measures. Scores closer to zero indicate that two days are more similar and scores closer to one indicate that two days are more dissimilar for each of these indexes.

Table 3-2 shows these two dissimilarity indexes for the ATUS data. Looking at Szalai's T we see that the indexes are the highest between weekdays and weekends. As expected this indicates the greatest dissimilarity is between weekdays and weekends. The index values comparing different weekdays are much lower. Generally, the highest index values for weekdays occur for comparisons with Friday.

This is true for Tuesday, Wednesday, and Thursday. Monday is an exception to this pattern. It appears that Monday is actually more similar to Friday than another weekday like Wednesday.

Table 3-2. Dissimilarity indexes comparing activity profiles by diary day

Comparison of	to	Weighted absolute difference	Szalai's T
Monday	Tuesday	.014	.050
	Wednesday	.015	.086
	Thursday	.015	.061
	Friday	.030	.078
	Saturday	.144	.270
	Sunday	.166	.398
Tuesday	Wednesday	.010	.068
	Thursday	.007	.061
	Friday	.035	.080
	Saturday	.157	.286
	Sunday	.179	.403
Wednesday	Thursday	.013	.061
	Friday	.038	.099
	Saturday	.154	.270
	Sunday	.176	.387
Thursday	Friday	.030	.079
	Saturday	.153	.272
	Sunday	.176	.393
Friday	Saturday	.135	.255
	Sunday	.158	.399
Saturday	Sunday	.053	.281

Source. ATUS data 2005-2011.

Szalai's T has some disadvantages that Stewart (2006) covers in a review of various dissimilarity indexes. Mainly, the formula weights each activity equally so that proportional differences in short duration activities have the same effect on the index as proportional differences in longer duration activities. In addition, activities that are significantly different in percentage terms across days have a large effect on the index. Stewart (2006) found that the weighted absolute-difference performs better than Szalai's T, because it weights the differences by the fraction of time spent on the activity and uses the absolute differences rather than squaring the summation of the differences.

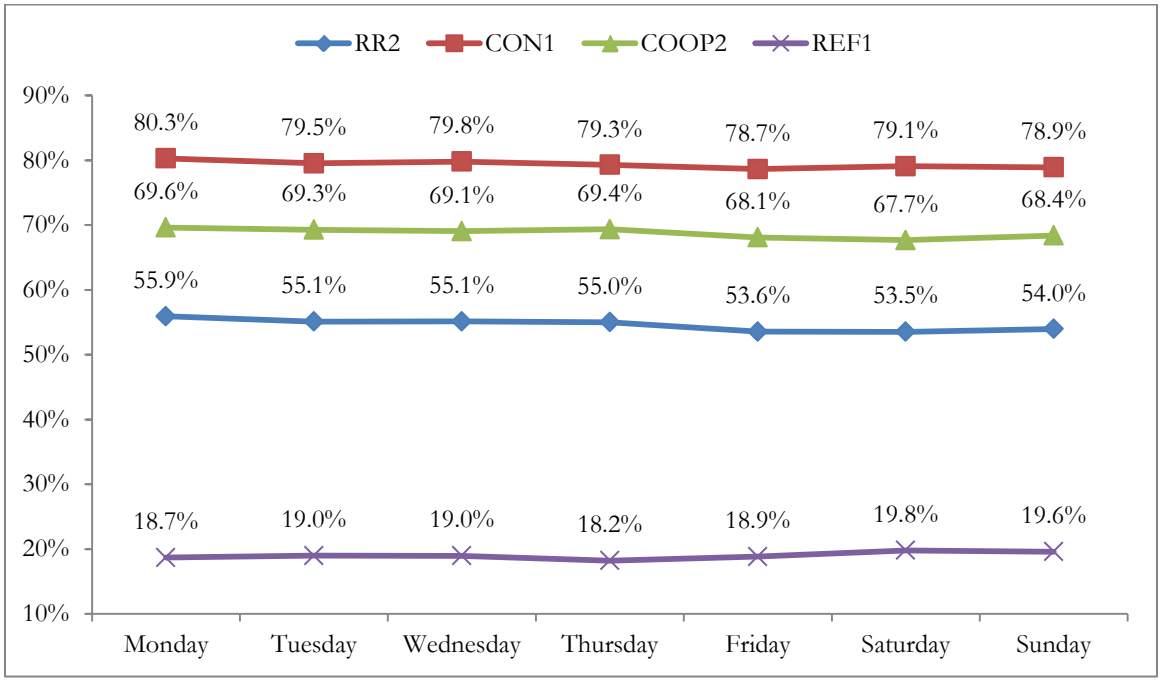
The picture is slightly different using the weighted absolute difference index. It is still clear that weekdays and weekends are the most dissimilar days. However, the picture is now clearer with

weekdays. Among weekdays, the highest index values consistently occur with Fridays. Comparisons among Monday – Thursday are typically half the size of the index when these days are compared to Friday.

In summary, the results from the dissimilarity indexes differ slightly depending upon which index is used. However, the results may suggest that among weekdays, differences between Friday and the other weekdays are more important than differences between Monday and other weekdays.

Analysis of Outcome Rates. We next examined whether it was more difficult to contact sample households and obtain completed interviews on different days of the week. Figure 3-1, illustrates different final outcome rates by day of the week in the ATUS data from 2005-2011. The final outcome rates were calculated based on information from the ATUS case history file.

Figure 3-1. Final outcome rates by day of the week (ATUS data 2005-2011)

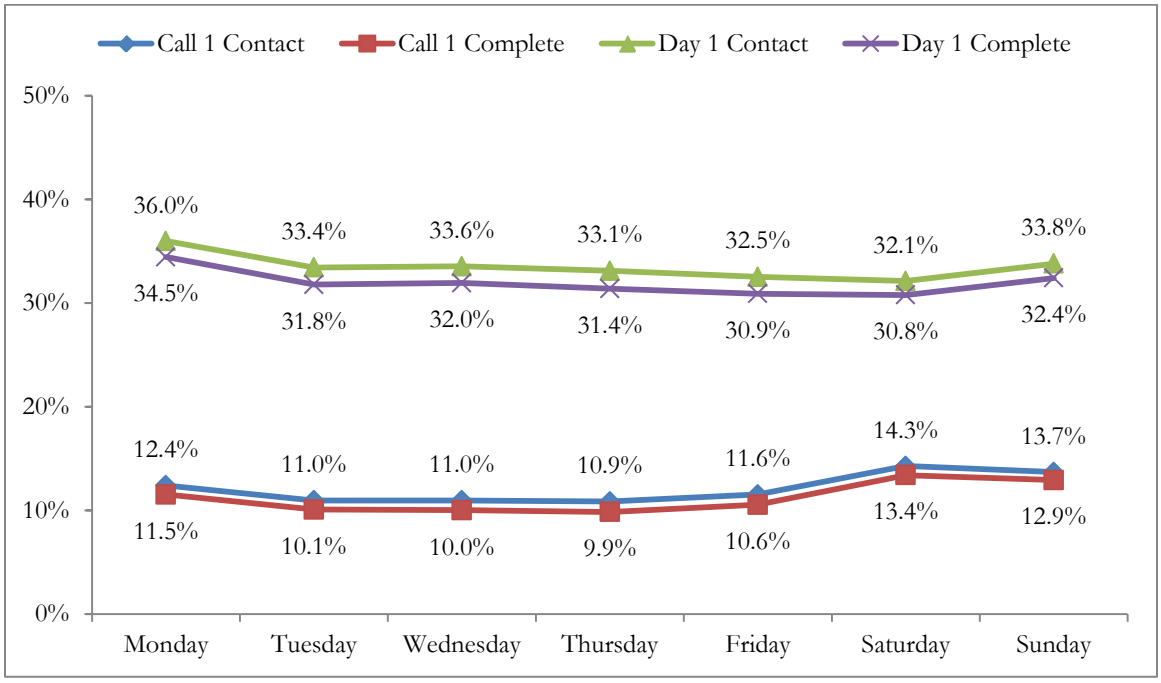


Overall the differences in outcome rates are relatively small across days of the week. Response rates are highest at the beginning of the week and generally decline over course of the week. The response rate for Monday is 55.9 percent over this time period and falls to 53.5 percent for those scheduled to be interviewed on Saturday. Contact, cooperation, and refusal rates show a similar pattern across the week.

The overall outcome rates in Figure 3-1 indicate how difficult it may be to complete an interview given the current ATUS protocol. In considering whether to substitute days, it is more relevant to understand more interim call outcomes. For example, one may consider substituting a day for someone after a complete is not obtained after the first call or after the first day on which they were called. Using the ATUS call history and case history files we were able to identify when cases were contacted and when they were completed.

Figure 3-2 illustrates how the probability of obtaining contact or completed interviews varies by day of the week. The bottom two lines show the probability of contacting a case and the probability of completing an interview on the first call is highest on the weekends. The top two lines show the probability of contacting a case and completing an interview after the first day of calling. The top two lines are more similar to the final outcome rates in Figure 3-1. That is, the probability of contacting a case and completing an interview are highest at the beginning of the week and generally decline throughout the week.

Figure 3-2. Percentage of cases contacted and completed on the first call and first day by day of the week, ATUS 2005-2011 (not eligible excluded)



Modeling Response Propensity. We next ran some logistic regression models predicting the probability of response. Once again, we use an indicator for response status and day of week from the ATUS case history file. Table 3-3 shows the results of three models.

The first model only includes indicators for day of the week. Monday is the reference category in this model. As expected from Figure 3-1, the odds of responding are significantly lower on Friday, Saturday, and Sunday compared to Monday.

Model 2 includes predictors for age, sex, presence of a spouse, employment status, presence of children, and whether the sample unit was in school. Since we need information on both respondents and nonrespondents, these predictors are taken from the ATUS-CPS file. We chose these predictors because they are indicators of where a respondent is in the lifecycle and thus serve as relatively good indicators of how different respondents might spend their time. All of the demographic variables except the presence of children were predictive of response. The differences in response propensity between Monday and Friday, Saturday, and Sunday remain even after controlling for the demographics.

Table 3-3. Adjusted odds ratios logistic regression predicting probability of response

	Model 1 (n=167,132)	Model 2 (n=167,132)	Model 3 (n=160,177)
Sunday	.93*	.94*	.97
Tuesday	.97	.97	1.03
Wednesday	.97	.97	1.03
Thursday	.96	.96	1.01
Friday	.91*	.91*	.97
Saturday	.91*	.91*	.98
Age 24-55		1.23*	1.08
Age 56-61		1.78*	1.42*
Age 62 and over		1.91*	1.30*
Female		1.17*	1.14*
Partner present		1.37*	1.38*
Employed full time		1.24*	1.49*
Employed part time		1.37*	1.51*
Unemployed		1.17*	1.34*
Have children		.98	1.04
In school		1.20*	1.31*
Contacted after week 1			.07*

Source: ATUS data 2005-2011.

Notes: All estimates are weighted by the ATUS base weight. Standard errors are shown in parentheses.

* $p < .0001$.

Model 3 adds an indicator for whether the respondent needed to be contacted after the first scheduled interview day. For example, someone might not have been available on Monday, October 22, so they were contacted on October 29 or subsequent weeks. Not surprisingly the odds of obtaining response decline significantly if the interview is not completed during the first

scheduled interview day. In addition, the differences in response by day of week disappear when we control for whether the cases are contacted after the first scheduled interview day. In other words, given that a sample unit needs to be contacted after the first scheduled interview, there is no difference in the probability of obtaining a response by day of the week.

The Effect of Substitution on Time Use. We next looked at whether those who would need to substitute a day are likely to have different time use patterns than those who would not. For this analysis week looked at differences in time use between those who did not complete the interview on the first scheduled diary day and those who did. Table 3-4 presents the results of three different OLS regression models predicting the time spent in work activity, household activity, and social activity. We have once again included demographic characteristics that are predictive of time and an indicator for whether the interview was completed after the first scheduled diary day.

All three models show that those who need complete the interview after the first scheduled diary day show significantly different time use patterns than those who complete the interview on the first scheduled diary day. For example, respondents who do not complete the interview on the first scheduled diary day work significantly more hours in work activity and spend significantly fewer hours in household activity and social activity. Hence it does appear that those who would need to substitute a day are likely to have different time use patterns than those who would not.

Table 3-4. OLS regression predicting number of minutes spent in selected major activities

	Work activity (n=89,824)	HH activity (n=89,824)	Social activity (n=167,132)
Intercept	33.84* (4.73)	75.81* (2.73)	452.73* (4.73)
Completed after week 1	13.05* (2.01)	-4.85* (1.13)	-8.60* (1.67)
Age 24-55	41.58* (4.56)	31.52* (2.05)	-19.21* (3.62)
Age 56-61	19.72* (4.66)	44.79* (3.19)	1.98 (4.64)
Age 62 and over	-14.51 (4.99)	42.28* (2.91)	38.30* (4.49)
Female	-44.05* (1.68)	46.63* (1.15)	-46.49* (1.91)
Partner present	-2.39 (2.23)	30.63* (1.25)	-37.64* (1.97)
Employed	301.97* (1.67)	-46.97* (1.61)	-138.12* (2.51)
Unemployed	19.54* (2.32)	9.76* (2.89)	-10.41 (4.80)
Have children	-14.83* (2.35)	1.52 (1.24)	-39.30* (2.10)
In school	-49.69* (3.97)	-33.97* (1.91)	-61.35* (3.35)

Source: ATUS data 2005-2011.

Notes: All estimates are weighted by the ATUS final weight. Standard errors are shown in parentheses.

* $p < .0001$.

Analysis of Matched Cases. The ideal data set for our analyses would include time use data for each respondent on multiple days of the week. This way we could examine the correlations in time use for the same person on different days of the week. However, the primarily goal of the ATUS is to produce national estimates of time use and not to look at intra-person variability in time use, Hence, the ATUS data only measures time use for one respondent on one diary day.

Our final analysis simulates the ideal data set by matching similar cases and comparing time use estimates. We first matched cases based on a combination of sex, presence of a spouse, presence of children, age, and employment status. In total, the combination of these variables created 48 cells on which to match respondents. For this analysis we look at net difference rate. For example, we calculated the average amount of time that the matched cases spent in each main activity on each pair of diary days. We then subtracted the mean for day 2 from the mean for day 1. The results are shown in Table 3-5.

We now highlight some of the significant net differences in Table 3-5. Respondents spend less time in personal care activities on Friday compared with Monday, Tuesday, Wednesday, and Thursday. Respondents spend less time at work on Monday compared with Tuesday, Wednesday, and Thursday. Respondents also spend less time at work on Friday compared with Tuesday, Wednesday, and Thursday. There is not much difference in the amount of time spent at work on Monday and Friday. Respondents spend somewhat more time in household activities on Monday compared with Tuesday, Wednesday, Thursday and Friday. Respondents spend more time in leisure activity in Monday compared with Tuesday and Wednesday. Finally, respondents spend more time in leisure activity on Friday compared with Monday, Tuesday, Wednesday, and Thursday.

Overall the results in Table 3-5 are consistent with those shown in Table 3-1. Some significant differences occur between Monday and Friday and the rest of the weekdays.

Table 3-5. Net differences (in minutes) across pairs of days for matched cases

Time use category	Mon-Tue (n=8,758)	Mon-Wed (n=8,743)	Mon-Thu (n=8,591)	Mon-Fri (n=8,587)	Tue-Wed (n=8,750)	Tue-Thu (n=8,611)	Tue-Fri (n=8,614)	Wed Thu (n=8,658)	Wed-Fri (n=8,567)	Thu - Fri (n=8,474)
Personal care activities	2.22	2.70	4.88	18.91	0.77	2.89	16.50	1.71	14.98	13.54
Household activities	7.06	5.15	8.78	12.04	-2.04	1.76	5.43	4.20	7.34	3.40
Caring for and helping household members	0.28	1.16	1.11	5.58	0.87	0.88	5.29	-0.15	3.81	4.15
Caring for and helping no household members	-0.01	-0.16	-0.47	-1.12	-0.03	-0.24	-0.83	-0.17	-0.87	-0.69
Work and work related activities	-19.13	-21.81	-19.04	-4.08	-2.33	-0.07	15.18	1.86	17.97	15.46
Education	-0.43	-0.61	1.26	6.88	0.11	1.91	6.95	1.91	7.60	5.28
Consumer purchases	1.03	1.08	-1.14	-10.17	-0.19	-2.14	-11.03	-2.01	-11.09	-9.21
Professional and personal care services	-1.24	-0.83	-1.42	-2.07	0.40	-0.31	-0.79	-0.52	-1.05	-0.59
Household services	0.26	0.48	0.41	-0.28	0.20	0.14	-0.54	-0.10	-0.80	-0.70
Government services and civic obligations	-0.30	-0.11	-0.12	-0.17	0.11	0.10	0.15	0.05	0.04	0.01
Eating and drinking	0.11	-0.66	-1.14	-7.69	-0.79	-1.39	-7.79	-0.61	-6.82	-6.58
Socializing, relaxing, and leisure	12.05	18.38	7.69	-15.63	5.90	-4.60	-27.71	-9.84	-33.38	-23.12
Sports, exercise, and recreation	-0.04	1.14	0.21	-0.55	1.09	0.32	-0.65	-0.89	-2.01	-1.05
Religious and spiritual activities	-0.25	-3.41	-0.52	-0.44	-3.16	-0.43	-0.16	2.71	2.94	0.34
Volunteer activities	-1.00	-1.46	-1.10	-0.56	-0.44	-0.07	0.35	0.38	0.78	0.53
Telephone calls	0.12	0.54	1.29	0.85	0.42	1.25	0.78	0.54	0.12	-0.37
Travel	0.00	-0.22	0.10	-0.01	-0.33	0.08	-0.13	0.31	0.45	-0.05

Conclusion 4

This paper summarizes an analysis related to day-of-week substitution in time use surveys. Overall, the findings of this analysis are consistent with the existing time use literature that we reviewed on day-of-week substitution.

Overall we found some differences in time use across days of the week. As expected, weekdays differ from weekends. Respondents appear to have slightly different time use patterns on Monday and Friday compared to Tuesday, Wednesday, and Thursday. However, not all of our analyses agreed with this conclusion. For example, the dissimilarity indexes indicated that any differences between weekdays occurs between Friday and the remaining weekdays. In contrast, our analysis of time spent in main activities showed that Monday also has some differences with other weekdays.

We also examined differences in response propensities by day of the week. Response propensity is generally highest at the beginning of the week and then declines the later in the week that someone is scheduled to be interviewed. We did find that cases that require substitution (i.e., those interviewed after the first scheduled interview day) engage in somewhat different levels of activity than those that do not require substitution. For example, one might expect that those who are most likely to substitute days are those who engage in more activities such as work and other related activity outside the home.

References

Abraham, K.G., Helms, S., and Presser, S. (2009). How Social Processes Distort Measurement: The Impact of Survey Nonresponse on Estimates of Volunteer Work in the United States. *American Journal of Sociology* 114(4): 1129-1165.

Abraham, K.G., Maitland, A., and Bianchi, S. (2006). "Nonresponse in the American Time Use Survey: Who is Missing from the Data and How Much Does It Matter?" *Public Opinion Quarterly* 70(5):676-703.

Fricker, S., and Tourangeau, R. (2010). "Examining the Relationship Between Nonresponse Propensity and Data Quality in Two National Household Surveys." *Public Opinion Quarterly* 74(5): 934-955.

Kalton, G. (1985). "Sample Design Issues in Time Diary Studies." In *Time, Goods, and Well Being*, ed. F. Juster and F. Stafford. Ann Arbor: University of Michigan, Institute of Social Research, 333-351.

Stewart, J. (2002). "Assessing the Bias Associated with Alternative Contact Strategies in Telephone Time-Use Surveys." Washington DC: Bureau of Labor Statistics.

Stewart, J. (2006). "Assessing Alternative Dissimilarity Indexes for Comparing Activity Profiles." *International Journal of Time Use Research*. 3(1): 49-59.

Stinson, L. (1999). "Day-of-Week Differences and Implications for Time Use Research." Proceedings of the American Statistical Association, Survey Research Methods Section. Alexandria, VA: American Statistical Association, 998-1003.

Appendix A
Analysis of Dutch Time Use Survey Data

As mentioned earlier in this report, the ideal data set for this day of week analysis would consist of individuals being interviewed on multiple days across the week. This would allow one to measure the association between levels of activity across different days of the week, but for the same person. Although the ATUS is not designed for such an analysis, there are other data sets that do fit this description.

The 2005 Dutch Time Use Survey measured time use on 2,204 individuals across all seven days of the week. This survey had the respondents fill out a time diary for each day of the week. Table A-1 shows the results of the analysis of time use patterns in this survey by day of the week. Column three of the table shows the differences in the activity profiles for the respondents. We used the weighted absolute difference (WAD) at the case level to examine the differences in the activity profiles. For example for one individual we compared the activity profile of this person on Monday with the same person's activity profile on Tuesday. The number in column three represents the mean WAD. The mean WAD is around .2 for comparisons between weekdays and around .3 when weekdays are compared with Saturday or Sunday. The mean WAD is also around .3 when Saturday is compared to Sunday. One may also note that there is a consistent tendency for the mean WAD to be slightly higher for Friday compared to other weekdays.

The remaining columns of Table A-1 show the correlations in main activities for the same individual across days of the week. First, the table shows the relative size of the correlations between days varies by type of activity. Correlations in work activity across weekdays are in the range of .7-.8, whereas correlations in consumer purchases are much lower in the range of .2 or less. Second, the table confirms what we have seen so far about the time use patterns across days of the week. The table shows that within activities and among days the weekdays are correlated the strongest. Once again, there is consistently a weaker correlation between Friday and the other weekdays. The Correlations are the weakest between weekdays and weekend days.

Table A-1. Association between time use within same individual across days of the week – 2005 DTUS (n = 2,204)

Comparison of	to	Activity profile	Personal care	Household activities	Care for household members	Work	Education	Consumer purchases	Professional services	Civic obligations	Eating and drinking	Leisure	Recreation	Religion	Other
Monday	Tuesday	.19	.57	.69	.59	.80	.69	.24	.33	.32	.59	.57	.34	.35	.35
	Wednesday	.21	.48	.63	.58	.71	.63	.17	.36	.22	.53	.52	.38	.41	.32
	Thursday	.20	.48	.67	.55	.73	.67	.18	.30	.28	.49	.50	.31	.37	.29
	Friday	.23	.43	.64	.52	.66	.64	.17	.29	.23	.47	.37	.33	.24	.20
	Saturday	.34	.27	.42	.47	.11	.42	.09	.32	.16	.35	.24	.22	.28	.09
	Sunday	.34	.33	.41	.52	.10	.41	.06	.35	.20	.38	.27	.14	.30	.14
Tuesday	Wednesday	.20	.55	.67	.56	.76	.67	.16	.36	.23	.59	.58	.29	.42	.33
	Thursday	.19	.53	.69	.52	.78	.69	.20	.36	.27	.52	.55	.41	.59	.33
	Friday	.23	.45	.65	.50	.69	.65	.16	.20	.24	.47	.40	.35	.34	.24
	Saturday	.34	.26	.42	.45	.14	.42	.04	.29	.13	.38	.26	.16	.36	.08
	Sunday	.35	.27	.36	.55	.11	.36	.04	.18	.16	.36	.25	.16	.29	.09
Wednesday	Thursday	.20	.60	.66	.58	.75	.66	.19	.36	.37	.53	.52	.28	.50	.32
	Friday	.23	.46	.62	.53	.66	.62	.16	.31	.28	.47	.40	.36	.38	.26
	Saturday	.33	.27	.44	.47	.11	.44	.09	.32	.12	.37	.25	.20	.41	.13
	Sunday	.34	.27	.39	.55	.08	.39	.04	.22	.18	.36	.23	.14	.31	.09
Thursday	Friday	.21	.54	.67	.54	.73	.67	.15	.27	.23	.56	.45	.33	.41	.21
	Saturday	.33	.36	.43	.43	.14	.43	.07	.38	.18	.45	.24	.22	.42	.11
	Sunday	.35	.28	.39	.48	.09	.39	.07	.31	.21	.31	.23	.19	.32	.07
Friday	Saturday	.31	.38	.47	.49	.23	.47	.13	.28	.36	.46	.36	.23	.39	.30
	Sunday	.34	.20	.40	.51	.14	.40	.03	.28	.20	.31	.23	.17	.19	.07
Saturday	Sunday	.28	.38	.47	.49	.23	.47	.13	.28	.36	.46	.36	.23	.39	.30

American Time Use Survey Experimental Design

Part III

1 Introduction

Time use surveys, such as the American Time Use Survey (ATUS), present unique methodological challenges to survey designers. In addition to obtaining probability samples of persons, time use surveys must also obtain probability samples of days of the week in order to adequately describe the level of activity in the population. This essentially means assigning members of the sample to report in the survey on a specific day of the week. However, the random assignment of a person to a specific day increases the amount of effort required to recruit members of the sample to participate in the survey, if the respondent is required to report on the previous day. This is because the sample person may not be available (or willing) to participate on the day to which she was assigned.

Any design feature that decreases sample persons' ability or willingness to take part in a survey may also lead to lower response rates and the potential for nonresponse bias. Indeed, ATUS response rates have been under 60 percent since the survey's inception in 2003. Research into the effect of nonresponse on the time use estimates has shown that there are systematic relationships between demographic or background characteristics of the designated ATUS respondent and his or her probability of responding (Abraham, Maitland, and Bianchi, 2006; Fricker and Tourangeau, 2010). Some studies have also found evidence of potential bias in the data (Abraham, Helms, and Presser, 2009).

The ATUS sample is drawn from households that have completed the final round of the Current Population Survey (CPS). A specific designated person (DP) from these CPS households is then assigned a specific day of the week to report on; they are asked about the activities that they took part in from 4 a.m. the previous day to 4 a.m. on the day of the interview. Although the designated person is called up to eight weeks on the designated day of the week for the ATUS interview, this design feature still represents a significant challenge to achieving a higher a response rate.

This report details an experiment to test a procedure for allowing day of week substitution in the ATUS. In designing this experiment, Westat reviewed the literature related to day-of-week substitution in time use surveys in general. Westat also conducted an analysis of the existing ATUS microdata to determine if day of week substitution is feasible. We begin with a brief review of the findings from the literature review and microdata analysis. Then we present specifications for an experiment to test a procedure for allowing day of week substitution.

Background 2

The proposed experimental design was developed based on the background information from the literature review and our analyses of the existing ATUS microdata. This section of the report includes some background information on our experimental design, which is described in Section 3.

Selection of Days to Allow Substitution

Our earlier report analyzing the ATUS data about day-of-week substitution showed some clear patterns between days of the week. Overall, our analysis showed that Monday-Thursday were the most similar days in terms of time use. We came to this conclusion by looking at differences in specific time use estimates across days of the week. We also looked at more general activity profiles by day of the week.

Here, we will briefly summarize the results of the differences between specific time use estimates by day and refer the reader to the earlier report for more detail. We found that respondents spend more time with personal care activities, household activities, consumer purchases, and leisure activities on the weekends. They also spend less time with work and education activities on the weekend. Some differences also emerge for the weekdays. Respondents spend slightly less time in work activities on Monday and Friday compared to Tuesday, Wednesday and Thursday. On Friday, respondents tend to spend less time in work and more time in leisure activities and eating and drinking compared to Tuesday, Wednesday and Thursday. On Monday, respondents also tend to spend less time in work and slightly more time in leisure activities and household activities. We concluded from these specific estimates that Monday and Friday seem to have some idiosyncrasies; however, the differences between Friday and other weekdays may be more obvious than the differences seen between Monday and the other weekdays.

Table 2-1 shows how the activity profiles compare by day of the week. Both the weighted absolute difference and the index called Szalai's T concur that Monday-Thursday are the most similar days. Friday is slightly different from the other weekdays. Saturday and Sunday are different from each other and the other weekdays.

Table 2-1. Dissimilarity indexes comparing activity profiles by diary day

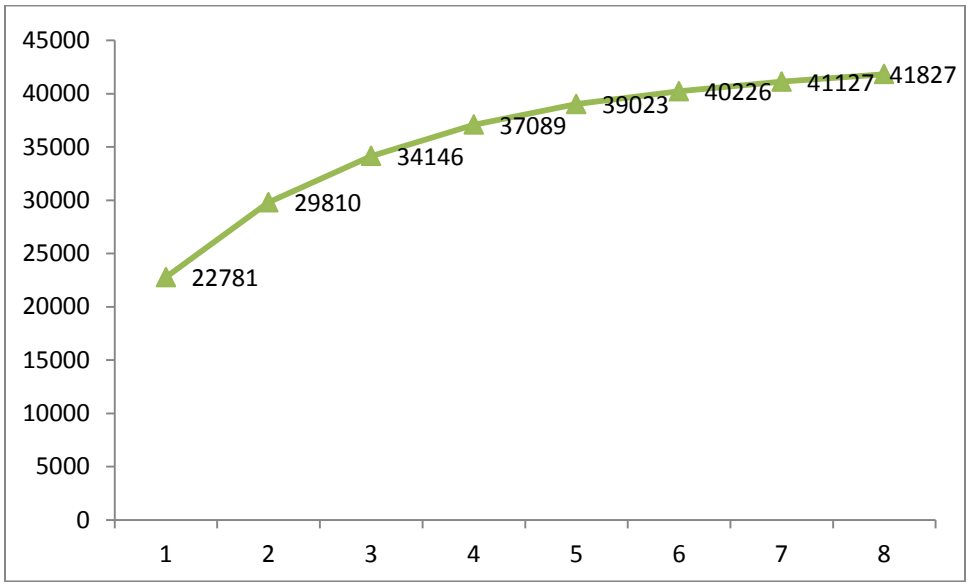
Comparison of	to	Weighted absolute difference	Szalai's T
Monday	Tuesday	.014	.050
	Wednesday	.015	.086
	Thursday	.015	.061
	Friday	.030	.078
	Saturday	.144	.270
	Sunday	.166	.398
Tuesday	Wednesday	.010	.068
	Thursday	.007	.061
	Friday	.035	.080
	Saturday	.157	.286
	Sunday	.179	.403
Wednesday	Thursday	.013	.061
	Friday	.038	.099
	Saturday	.154	.270
	Sunday	.176	.387
Thursday	Friday	.030	.079
	Saturday	.153	.272
	Sunday	.176	.393
Friday	Saturday	.135	.255
	Sunday	.158	.399
Saturday	Sunday	.053	.281

Source. ATUS data 2005-2011.

Identifying Cases for Substitute Days of the Week

An important question to address in substituting days of the week is what event would trigger day-of-week substitution. Analyses of the ATUS call history data show that more than one half of the completed interviews occur during the first week that a case is introduced. As shown in Figure 2-1, seven in ten (71%) completes are achieved within two weeks after a case is introduced.

Figure 2-1. Cumulative number of completes by weeks, Monday-Thursday diary day cases, ATUS 2004-2011



In addition, cases completed within the first two weeks require significantly less effort than those cases that remain after the first two weeks. Figure 2-2, shows the probability of obtaining a completed interview among the active cases by week. The first week, nearly thirty percent of the sample completes the interview on the first day of interviewing. Approximately fourteen percent of the remaining sample completes the interview on week two. The probability continues to decline until about week 6 or 7 towards the end of the field period when only about five percent of the remaining cases will end up completing the interview.

One can also see that the level of effort increases over the field period by looking at the average number of call attempts by week. Figure 2-3 shows the average number of call attempts by the week that cases are retired from the ATUS sample. The red line shows the total number of call attempts and the green line shows the number of initial call attempts until the first contact (among those whom contact was made.) On average, fewer than three call attempts are made on cases that are retired from the sample on week 1. A little more than six call attempts on average are made on cases that are retired on week 2. The average number of call attempts increases to more than eight by week 3 and rises by at least four call attempts on average after that point in the field period. Hence, Figure 3 demonstrates that the amount of effort expended on cases using the existing ATUS data collection protocol increases around week three or four. The green line showing the number of contact attempts until initial contact shows a similar trend.

Figure 2-2. Proportion of active cases each week completing the ATUS interview, Monday-Thursday diary day cases, ATUS 2004-2011

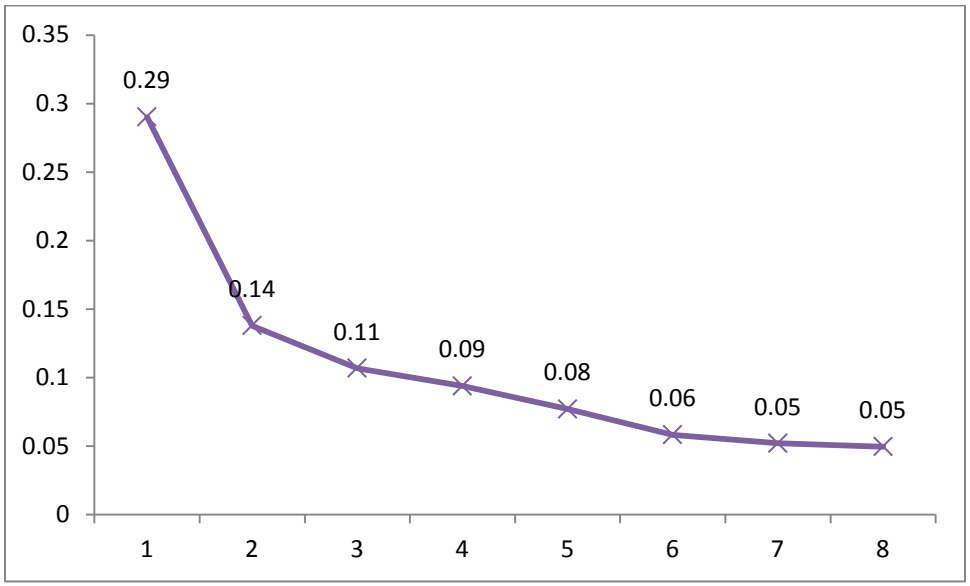
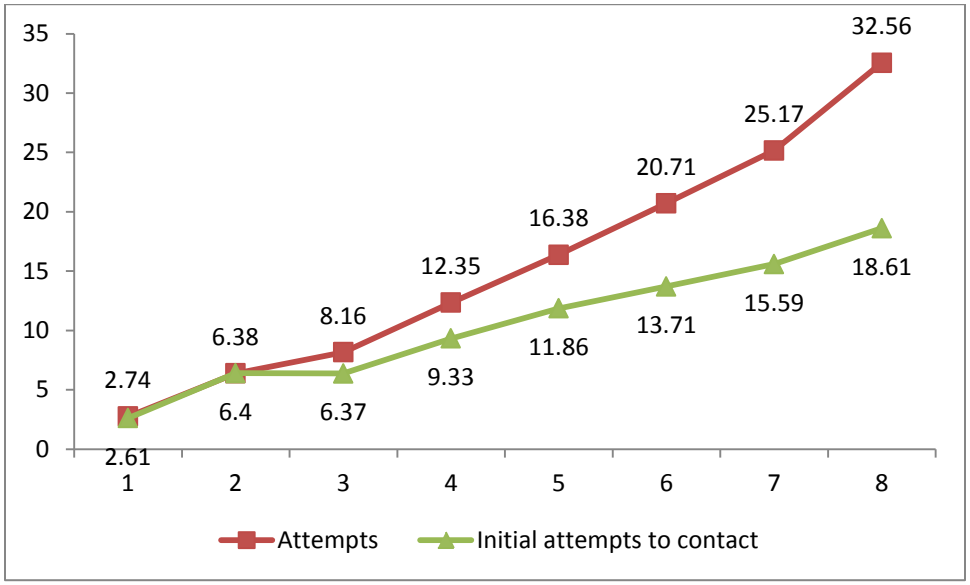


Figure 2-3. Total attempts and attempts to initial contact by week that cases are retired from the sample, Monday-Thursday diary day, ATUS 2004-2011



Extending the Recall Period

The advantages of reporting about the previous day are obvious from the perspective of memory. The task of recalling all activities in the previous 24 hours can be difficult. Hence, the recall should take place as close to the reference day as possible. In addition, memories across different days may begin to blend together when the reference period is too distant from the designated day. Some surveys, such as the Canadian General Social Survey, relax this requirement and allow interviewers to call up to two days after the reference period. There has been relatively little research on the length of the recall period in time use surveys to guide researchers about the extent to which the recall can be extended.

General guidance from the literature and from experienced time use researchers is that recall is difficult more than two days after the diary day (Pentland et al., 1999). Harvey (1990; 1993) cites work by Klevmarken comparing 24 and 48 hour diaries; that work found "under-reporting of some activities, a decreased number of activities reported and instances of memory lapse during some substantial time period." However, the same study by Klevmarken found small differences in the average duration of activities between yesterday and day before yesterday diaries.

Other guidance suggests that memory becomes more difficult with a longer recall period, but that the effect of recall period length may depend on whether the diary day is a weekday or weekend day (Michelson, 2005). Juster (1985) examined data from the 1975-76 and 1981-82 Surveys of Time Use in the United States. The original design of these studies was for 24 hour recall of the previous day from the interview. However, because the research design called for interviews with both spouses to be done on the same diary day, it sometimes turned out that one or the other spouse was not available for an interview on the assigned date. Hence, the absent spouse was asked to complete the diary on a later day, which extended the recall period. This led to a relatively small proportion of the sample having a recall period of more than 24 hours. Juster found that the mean number of activities reported declined as the recall period got longer; however, this effect was limited primarily to weekdays. The general explanation for this result is that the effect of memories blending together might be greater for weekdays than weekends since weekdays tend to be more similar to each other than weekend days.

Experimental Design 3

The proposed experimental design was developed based on the background information in the previous sections. The design includes the following features:

1. A mechanism that identifies cases eligible for substituting day of the week.
2. A Monday-Thursday diary day substitution procedure.
3. A Monday-Thursday call day substitution procedure.

We propose a design that includes procedures for substituting both the call day and the designated diary day. The advantages and disadvantages of each approach are shown below in Table 3-1.

Table 3-1. Advantages and disadvantages of substituting the call day and designated diary day

Substitute diary day		Substitute call day	
Advantage	**Disadvantage**	**Advantage**	**Disadvantage**
Increased flexibility with call scheduling	Contact bias	Increased flexibility with call scheduling	Memory decay
Less nonresponse	Activity bias	Less nonresponse	

The two approaches share some common advantages compared to the current ATUS methodology. Both approaches ease the restriction of contacting a household on a specific day of the week. This has the advantage of increasing the flexibility of call scheduling. The increased flexibility with call scheduling should lead to more efficiency, less effort, and higher success rate in contacting households. For example, households that are consistently not contactable on a Tuesday may be easier to reach on another day of the week.

The two approaches have different potential disadvantages that could increase the error in time use estimates. There are two closely related sources of potential error associated with substituting the diary day. There is a potentially biasing effect of being able to contact individuals on days when they are more likely to be at home. This in turn could result in the respondent being more likely to report on activities they do when they are not at home if they are reporting about a previous day when contact attempts were unsuccessful because the respondent was not at home on that day.

Another potential source of error with substitution of the diary day depends on how the substitution is made. The ATUS sample is a probability sample of days where 10 percent of the sample is initially allocated to each weekday and 25 percent are allocated to each weekend day. Substituting the designated day will potentially alter this allocation. If a convenient diary day approach is taken after some designated point in the field period this could potentially alter the distribution of the sample allocated across the different days of the week. A probabilistic reassignment of the diary day (e.g., from one weekday to another randomly selected weekday) would minimize the effects of reallocation. In addition, weighting could be used to adjust the final data to reflect the desired distribution of days of the week. However, overall, the data is still likely to demonstrate an upward bias in the time spent in activities done away from home.

Substitution of the call day eliminates the risk of contact and activity bias. However, extending the call day further away from the designated diary day increases the effect of memory decay on the time use estimates. The scant research available on the extent of memory decay suggests that the number of activities reported declines by about 10 percent per day (e.g., Juster, 1986). However, this research tends to be more suggestive than conclusive due to the nonexperimental designs used in the previous studies. There appear to be certain factors that affect the rate of decay. For example, memory decay is less substantial for those who work and have more regular schedules. In addition, there is less memory decay for weekends than weekdays.

Research Design

We next present two options for a proposed preliminary study to fill a gap in the literature on the effect of memory decay on time use estimates. Then, we describe a larger experimental design to test the effects of the substitution of the diary day and call day.

Preliminary Study on Length of the Recall Period

We propose two different designs for a small scale study to investigate the effect of extending the recall period on time use estimates. We recommend that either study design be implemented over the telephone to maintain consistency with the current ATUS design. The questionnaire for this study is shown in Appendix A.

The key dependent variable in this preliminary study is the average number of activities reported. Table 3-2 illustrates the number of activities reported across days of the week. In general, the number of activities reported is relatively similar for Monday through Friday diary days with respondents reporting roughly 21 activities each day. However, respondents tend to report fewer activities on Saturday and Sunday diary days with respondents reporting around 19 activities on average.

Table 3-2. Number of activities reported by day of week

Day	Average number of activities reported
Monday	20.7
Tuesday	20.9
Wednesday	20.9
Thursday	20.9
Friday	20.9
Saturday	18.9
Sunday	18.6
Overall	19.8

The first design is a between subjects design where the respondents are called and asked to report about only one time period. In other words, the respondent is called and asked to report about yesterday, the day before yesterday, or two days before yesterday. A summary of this design is shown below in Table 3-3.

Table 3-3. Number of subjects per treatment for a between subjects experimental design

	Report on yesterday	Report on day before yesterday	Report on 2 days before yesterday
Diary day/type of subject	Number of respondents		
Monday – Friday (Employed)	67	67	67
Monday – Friday (Not employed)	67	67	67
Saturday or Sunday	67	67	67
Total	201	201	201

Key parameters in power analysis. power=.8, type I error (one-tailed) =.05, effect size = 2 activities.

We conducted a power analysis to determine the number of cases that would be needed to detect an effect of lengthening the recall period. Power is the ability to detect an alternative hypothesis. In general, there are two types of errors of interest in statistical testing. The first type of error is called Type I error (or alpha) and occurs when we incorrectly conclude that a treatment has an effect, but

in truth the treatment does not have an effect in the population (i.e., there is a null effect in the population). Type II errors (or Beta) occur when we incorrectly conclude that a treatment does not have an effect, but in truth the treatment does have an effect in the population. The power of a test is equal to 1 – Beta. Power is the conditional probability that we will correctly conclude that a treatment has an effect given that there is a true effect in the population.

Power analyses found that this design requires at least 199 cases in each condition to detect an average difference of two activities between the conditions. This effect size was arrived at based on the analyses of Juster (1986), who found that memory decays roughly 10 percent per day. Given that the average respondent reports approximately 20 activities, an effect size of 2 is appropriate for these analyses. We also assume that the statistical test is using the traditional .05 level of significance for a one-tailed test, which is appropriate since previous research suggests that respondents tend to forget activities over time. Last, we assume power of .8 which means that we have an 80 percent chance of detecting a difference of two events between the treatments. We have balanced the design on factors from the literature known to affect the number of activities that respondents are able to report such as whether the respondent works and whether the respondent reports about a weekday or weekend.

Implementation of the between subjects design is relatively straightforward. Respondents reporting about yesterday for Monday to Friday would be called Tuesday to Saturday similar to the current ATUS design. Those reporting about the day before yesterday for Monday to Friday would be called Wednesday to Sunday. Finally, those reporting about 2 days before yesterday would be called Thursday to Monday. Saturday and Sunday cases would follow a similar pattern.

The second option is to conduct a within subjects design as shown in Table 3-4. For this design, a respondent would be called and asked to report on either yesterday and the day before yesterday or yesterday and 2 days before yesterday. This design is somewhat more complicated to implement; however, the within subject design increases the power of the experiment. The increased power arises due to the correlation between reports by the same individual. For example, power will be increased if respondents who report a higher number of activities yesterday are also likely to report a higher number of activities the day before yesterday.

Power analyses found that this design requires at least 121 cases in each condition to detect an average difference of two activities between the yesterday report and the day before yesterday report or the 2 days before yesterday report. Once again, we assume that the statistical test is using the traditional .05 level of significance for a one-tailed test and power of .8. We assume a mild

correlation of .4 between the respondent reports about yesterday and either the day before yesterday or 2 days before yesterday. We have balanced the design on one additional factor relative to the between subject experiment. We now have approximately one half of the respondents report on yesterday first and the other half report on yesterday last to minimize the effect of the order that we ask about these two time periods.

Table 3-4. Number of subjects per treatment for a within subjects experimental design

Diary day/type of subject	Report on yesterday and day before yesterday	Report on yesterday and 2 days before yesterday
	Number of respondents	
Ask about yesterday first		
Monday – Friday (Employed)	16	16
Monday – Friday (Not employed)	16	16
Saturday or Sunday	32	32
Ask about yesterday last		
Monday – Friday (Employed)	16	16
Monday – Friday (Not employed)	16	16
Saturday or Sunday	32	32
Total	128	128

Key parameters. power=.8, type I error (one-tailed)=.05, correlation=.4, effect size=2 activities.

This design would be slightly more complicated to implement than the between subjects design. The calls would need to be scheduled so that respondents are reporting about either a weekday or weekend day. This strategy avoids measuring differences in the number of activities reported due to differences between the two types of days. For example, the weekday interviews for those recalling activities about yesterday and the day before yesterday would need to take place between Wednesday and Saturday. The weekday interviews for those recalling activities about yesterday and 2 days before yesterday would need to take place between Thursday and Saturday. A similar strategy would be undertaken for weekend activities.

The preliminary study would inform the decision about whether one should include experimental conditions varying the length of the recall period in the larger experimental design. If the preliminary study does not provide evidence that memory declines significantly as one extends the recall period then it would be preferable to include extended recall conditions in the experimental design. The study may also provide evidence that extended recall conditions are more feasible under some conditions than others. For example, it may be that memory decay is significant for weekdays, but

not weekends. In addition, the study may find that memory decay is significant only when asking about two days prior to yesterday.

Another possible indicator of memory decay that we consider in the analysis is the extent to which time use reported by a respondent approximates his or her usual time use pattern. For example, it is possible that as the recall period is extended, the respondent will rely on what they usually do on a typical day. Hence, the questionnaire in Appendix A includes a few brief questions on the respondents' usual time use patterns.

The choice between the two study designs may be driven by both cost and assumptions about the measurement process. The second design or within subjects design would be cheaper to implement since it requires fewer cases in each condition. However, there may be some reactivity in the responses since respondents are asked to report about more than one time period. In addition, comparing differences in the reports about two different time periods requires the assumption that there actually is no true difference between the two time periods. For example, the difference in the reported number of activities between yesterday and the day before yesterday is due to memory decay and not because there were actual differences in the number of activities in which the respondent engaged.

ATUS Experimental Design

The experimental design includes conditions for both weekend and weekday cases. The design also includes conditions that substitute either the call day or the diary day. The sample is drawn from 13,800 cases supplemental cases completing their eighth and final interview from the Current Population Survey. Table 3-5 summarizes the conditions in the experiment for the weekday cases. The cases from the current ATUS design serve as a control group in the experimental design. The first four conditions in the experiment would be applied to weekday cases. Conditions five and six apply to weekend cases.

In condition 1, we would substitute the diary day for someone who has not completed the ATUS interview after week 2 of interviewing. This substitution would only occur for respondents who are completing the interview between Tuesday and Friday regarding diary days between Monday and Thursday. Those who have not responded after two weeks of calling would be randomly assigned to one of the three remaining weekdays for the remainder of the field period. For example, one third of the Monday cases who do not complete the interview by week two will have their diary day

reassigned to Tuesday, another third will have their diary day reassigned to Wednesday and the remaining cases will be reassigned to a Thursday diary day. We would expect that changing the call strategy to another day will increase the response rate. The effect could be somewhat limited since the call strategy is limited to only one other randomly selected day. However, this condition minimizes the risk of selection bias by randomly assigning the sample person to another diary day. In addition, condition 1 allows us to test whether it makes a difference which day that the cases are assigned. It may be better to assign the respondent to a day that is further away from the next day. For example, if the reason for nonresponse on the originally assigned day is that the designated person was not home, assigning cases to the next day will result in over reporting of activities performed away from home.

Table 3-5. Summary of experimental design for weekday (Monday – Thursday) cases

		Weekday cases = 9,200			
Condition	Control	Condition 1	Condition 2	Condition 3	Condition 4
Description	Current ATUS Design	Substitute diary day randomly between Monday and Thursday	Substitute any diary day between Monday and Thursday	Substitute call day – ask about day before yesterday	Substitute call day – ask about 2 days before yesterday
Example		Tuesday switched to Wednesday	Tuesday switched to any Tuesday-Friday	Wednesday asking about Monday	Thursday asking about Monday
Details		After week 2 – sample person is assigned a new call day between Tuesday and Friday	After week 2 – sample person is switched to a protocol with contact attempts made between Tuesday and Friday	After week 2 – sample person keeps the same diary day, but is switched to a protocol with reporting beginning 4 AM on day before yesterday	After week 2 – sample person keeps the same diary day, but is switched to a protocol with reporting beginning 4 AM 2 days before yesterday
Sample	10,400	2,300	2,300	2,300	2,300
Eligible	9,360	2,070	2,070	2,070	2,070
n	5,200	1,155	1,155	1,155	1,155

Condition 2 is similar to condition 1 except that rather than being randomly assigned to a diary day, the sample person is called every weekday from Tuesday through Friday until an interview is completed. Condition 2 has the greatest potential for maximizing response rates, but also has higher risk of selection bias than condition 1.

Conditions 3 and 4 substitute the call day and extend the length of the recall period rather than substitute the diary day. Respondents would be randomly assigned to report on time use on the day before yesterday in condition 3. In condition 4, the respondents would be asked to report about 2 days before yesterday.

Table 3-6 summarizes the conditions for the weekend cases. In condition 5, the call day is substituted so that the respondent is asked about the day before yesterday. The call day is substituted so that the respondent is asked about two days before yesterday in condition 6. Once again, sample individuals are switched to this protocol after call attempts in the first two weeks are unsuccessful.

Table 3-6. Summary of experimental design for weekend (Saturday and Sunday) cases

		Weekend cases = 4,600	
Condition	**Control**	**Condition 5**	**Condition 6**
Description	Current ATUS Design	Substitute call day – ask about day before yesterday	Substitute call day – ask about 2 days before yesterday
Example		Monday asking about Saturday	Tuesday asking about Saturday
Details		After week 2 – sample person keeps the same diary day, but is switched to a protocol with reporting beginning 4 AM on day before yesterday	After week 2 – sample person keeps the same diary day, but is switched to a protocol with reporting beginning 4 AM 2 days before yesterday
Sample	13,000	2,300	2,300
Eligible	11,170	2,070	2,070
n	6,400	1,155	1,155

Power Analysis

The sample sizes in Figures 2-3 and Figure 2-4 were arrived at by performing power analysis. Before we present the results of the power analysis, we will describe the different types of analyses that we anticipate using the data from the experiment. The primary comparisons for the power analysis are between the current ATUS design and each of the experimental conditions.

One set of analyses will examine the differences in response rates across the conditions. We expect condition 2 to increase response rates the greatest. Conditions 1, 3, and 4 should also increase the

response rate by changing the call strategy, but we expect this effect to be weaker than in condition 2.

We will also analyze call history data to understand the level of effort associated with each condition. We expect greatest reduction in the level of effort with condition 2. For example, the greater flexibility with call scheduling should maximize the opportunity to make contact with the respondents at a time that is convenient for them. Conditions 1, 3, and 4 should also provide some reduction in effort compared to the normal ATUS protocol. However, we would still expect slightly higher levels of effort compared to condition 2.

A third set of analyses will examine differences in the time use estimates. We would expect time use estimates to be more likely to be affected as the protocol shifts towards a more convenient day approach. Therefore, the estimates in condition 2 might differ from the other conditions with stricter diary day assignment procedures. For example, the respondents in condition 2 might be interviewed on a day when they are more likely to be at home; however, they would be more likely to report about activities that happened away from home since they are reporting about the previous 24 hours.

A final set of analyses are specific to conditions 3-6. These analyses will help determine the effect of the length of the recall period. The primary dependent variable in this analysis will be the number of activities reported. We expect the number of events reported to decline as the recall period becomes longer. However, based on research by Juster (1985), we expect this effect to be greater for weekdays than weekends.

The power analyses in Table 3-7 and Table 3-8 were performed with these analytic goals in consideration. Table 3-7 shows the number of respondents needed to detect differences between mean estimates of time use activity for different levels of power and different sizes of effects. The power analysis in Table 3-7 is based on comparisons between weekday (Monday-Thursday) cases in the experimental condition and weekday cases from the ATUS regular data file. A key feature of the experimental design is the use of the regular ATUS cases as the control group. This has the effect of increasing the power of the comparisons due to the large number of cases in the ATUS. Comparisons between weekend cases from the experimental conditions and weekend cases from the regular ATUS design will require slightly fewer cases due to the larger number of cases on the weekend.

Table 3-7. Number of respondents required for differences in mean time use estimates, by size of effect and desired power

	Small effect			Medium effect			Large effect		
Desired Power	.7	.8	.9	.7	.8	.9	.7	.8	.9
Personal care activity (M = 573, SD = 143)	746	987	1412	169	216	294	71	90	121
Household activity (M = 124, SD = 143)	737	974	1393	167	214	290	70	89	120
Care for HH members (M = 39, SD = 88.8)	680	897	1274	155	199	270	68	87	117
Care for non HH members (M = 12, SD = 56.2)	918	1226	1784	203	261	355	89	113	152
Work (M = 174, SD = 259)	760	1006	1441	172	220	299	75	96	129
Education (M = 18, SD = 85.7)	821	1090	1571	184	236	321	80	103	138
Consumer purchases (M = 42, SD = 75.6)	836	1112	1605	187	240	326	82	104	140
Professional services (M = 7, SD = 32.9)	867	1154	1671	193	248	337	84	108	145
Household services (M = 1, SD = 13.3)	1383	1895	2894	289	373	511	125	160	216
Civic (M = .5, SD = 8.7)	514	671	940	120	153	207	53	67	90
Eating (M = 76, SD = 64.1)	816	1083	1560	183	235	319	80	102	137
Socializing (M = 303, SD = 214)	728	962	1374	165	212	287	72	92	124
Recreation (M = 22, SD = 70.0)	701	925	1318	160	204	277	70	89	120
Religious (M = 15, SD = 52.0)	767	1015	1455	173	222	301	76	97	130
Volunteer (M = 11.0, SD = 54.9)	869	1158	1676	193	248	338	85	108	145

Table 3-8. Sample size required for differences in mean contact attempts, by size of difference and desired power

Power	Difference in mean contact attempts		
	1	2	3
.7	1160	268	118
.8	1521	343	150
.9	2142	464	202

The effect sizes in Table 3-7 are the expected mean differences divided by the pooled standard deviation. That is, we assume that the standard deviations between the control cases from the original ATUS design and the experimental conditions are equal. A small effect is approximately 10 percent of the standard deviation, a medium effect is approximately 20 percent of the standard deviation, and a large effect is approximately 30 percent of the standard deviation. The standard deviations were computed using all of the ATUS data from 2003-2011. The power values in Table 3-7 represent levels of power that are traditionally viewed as acceptable in the literature. Power of .8 is a commonly used standard in the literature. SAS PROC POWER was used to arrive at the estimates of the sample size needed to detect a given size of an effect for a given level of power. Sample code used in the power analysis is shown in Appendix B. As an example, the value of 746 in

the upper left cell of the table means that 70 percent of the samples of this size would detect a difference of 14 minutes in personal care activity between the ATUS original cases and the experimental condition given that the true difference is at least 14 minutes. A difference of 14 minutes represents a small effect size (that is, the 14/143 is .10, which is a small effect size.) The value in the far right of row one means that 90 percent of the samples of size 121 would detect a difference of 14 minutes in personal care activity between the ATUS original cases and those in the relevant experimental condition.

Table 3-8 demonstrates the samples sizes needed to detect differences in mean contact attempts between the experimental conditions and the original cases from the ATUS design for a given level of power. A total sample of approximately 1521 cases in an experimental condition would be needed to reliably detect a mean difference of 1 contact attempt.

Table 3-9 shows the power to detect differences in response rates for samples of a given size. A difference of five percent between the original ATUS cases and the experimental conditions could be detected quite reliably with a total sample of 750 cases or more in a condition.

Table 3-9. Power to detect differences in response rate, by difference and sample size

Sample	Difference in response rate		
	5%	10%	15%
750	.75	.99	.99
1500	.95	.99	.99
2000	.98	.99	.99

These analyses suggest that sample sizes of 1000 to 1500 per experimental group would provide adequate power for the planned analyses.

References

Abraham, K.G., Helms, S., and Presser, S. (2009). How Social Processes Distort Measurement: The Impact of Survey Nonresponse on Estimates of Volunteer Work in the United States. *American Journal of Sociology* 114(4): 1129-1165.

Abraham, K.G., Maitland, A., and Bianchi, S. (2006). "Nonresponse in the American Time Use Survey: Who is Missing from the Data and How Much Does It Matter?" *Public Opinion Quarterly* 70(5):676-703.

Fricker, S., and Tourangeau, R. (2010). "Examining the Relationship Between Nonresponse Propensity and Data Quality in Two National Household Surveys." *Public Opinion Quarterly* 74(5): 934-955.

Harvey, A.S. (1990). "Time Use Studies for Leisure Analysis." *Social Indicators Research* 23(4): 309-336.

Harvey, A.S. (1993). "Guidelines for Time Use Data Collection." Social Indicators Research 30(2/3): 197-228.

Pentland, W.E., Harvey, A.S., Lawton, M.P., and McColl, M. (1999). *Time Use Research in the Social Sciences*. New York: Kluwer Academic/Plenum Publishers.

Juster, F.T. (1985). "The Validity and Quality of Time Use Estimates Obtained from Recall Diaries." In *Time, Goods, and Well Being*, ed. F. Juster and F. Stafford. Ann Arbor: University of Michigan, Institute of Social Research, 63-91.

Michelson, W. (2005). *Time Use: Expanding the Explanatory Power of the Social Sciences*. London: Paradigm Publishers.

Appendix A
Questionnaire for Preliminary Study

[ONE HALF OF RESPONDENTS IN THE WITHIN SUBJECTS DESIGN ARE ASKED ABOUT YESTERDAY FIRST AND ONE HALF ARE ASKED ABOUT YESTERDAY LAST.]

Ask about previous yesterday.

Now I'd like to find out how you spent your time yesterday, [yesterday's day & date], from 4:00 in the morning until 4:00 AM this morning. I'll need to know where you were and who else was with you. If an activity is too personal, there's no need to mention it.

1. So let's begin. Yesterday, [previous weekday] at 4:00 AM, what were you doing? /What did you do next?
2. How long did you spend [ACTIVITY]?
3. Who was with you? / Who accompanied you?
4. Where were you while you were [ACTIVITY]?
5. You did not report any eating or drinking yesterday. Did you do any eating or drinking yesterday as your main activity?

Ask about previous 48 or 72 hours

Now I'd like to find out how you spent your time [TARGET DAY], [TARGET DAY AND DATE], from 4:00 in the morning on [TARGET DAY] until 4:00 AM [DAY AFTER TARGET DAY]. I'll need to know where you were and who else was with you. If an activity is too personal, there's no need to mention it.

1. So let's begin. [weekday], at 4:00 AM, what were you doing? /What did you do next?
2. How long did you spend [ACTIVITY]?
3. Who was with you? / Who accompanied you?
4. Where were you while you were [ACTIVITY]?
5. You did not report any eating or drinking yesterday. Did you do any eating or drinking yesterday as your main activity?

Ask about usual hours in different activities

Now I would like to ask you about how much time you usually spend in different activities on a typical [TARGET DAY].

1. How many hours do you usually spend sleeping on a typical [WEEKDAY/WEEKEND]?

2. How many hours do you usually spend watching television on a typical [WEEKDAY/WEEKEND]?

3. How many hours do you usually spend working on a typical [WEEKDAY/WEEKEND]?

4. How many hours do you usually spend preparing meals or snacks on a typical [WEEKDAY/WEEKEND]?

5. How many hours do you usually spend eating or drinking on a typical [WEEKDAY/WEEKEND]?

6. How many hours do you usually spend in educational activities such as taking class or doing homework on a typical [WEEKDAY/WEEKEND]?

7. How many hours do you usually spend in exercise and recreation activities on a typical [WEEKDAY/WEEKEND]?

8. How many hours do you usually spend volunteering on a typical [WEEKDAY/WEEKEND]?

Debriefing question

How difficult was it for you to recall the amount of time that you spent [SPECIFIC ACTIVITY (E.G., WORKING, SLEEPING, ETC.)] [YESTERDAY, THE DAY BEFORE YESTERDAY, 3 DAYS AGO]?

Not at all
Slightly
Somewhat
Very

How confident are you that you recalled all of the activities that you did [YESTERDAY, THE DAY BEFORE YESTERDAY, 3 DAYS AGO]?

Not at all
Slightly
Somewhat
Very

Appendix B
Sample SAS Code for Power Analysis

```
/*power analysis for preliminary study between subjects design*/
proc power;
twosamplemeans test=diff
meandiff = 2 4 6
stddev = 8
npergroup = .
power = .8
alpha = .05
sides = 1;
PLOT X=POWER Min=.6 Max=.9 Key=OnCurves;
run;

/*power analysis for preliminary study within subjects design*/
proc power;
pairedmeans test=diff
meandiff = 2 4 6
stddev = 8
corr = .4 .5 .6
npairs = .
power = .8
alpha = .05
sides = 1;
PLOT X=power Min=.6 Max=.9 Key=OnCurves VarY(Panel BY Corr);
run;

/*power to detect differences in personal care activity*/
proc power;
twosamplemeans test=diff
meandiff = 14 28 43
stddev = 143.9
groupns = 5200 | .
power = .7 .8 .9
alpha = .05
sides = 2;
run;
```

```
/*power to detect differences in mean contact attempts*/
proc power;
twosamplemeans test=diff
meandiff = 1 2 3
stddev = 13
groupns = 10400 | .
power = .7 .8 .9
alpha = .05
sides = 2;
run;

/*power to detect differences in response rate*/
proc power;
twosamplefreq test=pchi
proportiondiff = .05 .10 .15
refproportion = .50
groupns = 10400 | 2000
power = .;
run;
```

American Time Use Survey Microdata Analysis for Substitution of the Designated Person

Part IV

Background and Introduction 1

The American Time Use Survey (ATUS) is a unique source of information about the amount of time that Americans spend on a diverse range of activities, including "non-market activities," such as housework. The ATUS contributes valuable information used by researchers from a variety of social science disciplines. For example, it has increased our understanding of the amount of time families spend on housework and child rearing activities. It has also helped economists understand how Americans value unpaid activities such as volunteering.

Surveys are increasingly facing difficulty with obtaining responses from sample households (for a recent review, see Brick and Williams, 2013). The ATUS is no exception. Since its beginning in 2003, the ATUS has obtained response rates under 60 percent. Although this response rate compares favorably with many other telephone surveys, there is room for improvements in the response rates and such improvements might reduce any nonresponse bias in the current data. Research into the effect of nonresponse on the time use estimates has shown that there are systematic relationships between demographic or background characteristics of the designated ATUS respondent and his or her probability of responding (Abraham, Maitland, and Bianchi, 2006; Fricker and Tourangeau, 2009). Some studies have also found evidence of potential bias in the time use estimates (Abraham, Helms, and Presser, 2009).

The ATUS has several design features that make it difficult to obtain a higher response rate. The sample is drawn from households that have completed the final round of the Current Population Survey (CPS). A specific designated person (DP) from these CPS households is then assigned a specific day of the week to report about the activities that they took part in from 4 a.m. the previous day to 4 a.m. on the day of the interview. Although the designated person is called up to eight weeks on the designated day of the week in order to complete the ATUS interview, this design feature still represents a significant challenge to achieving a higher a response rate.

These constraints could be relaxed in either of two ways to improve response rates. First, the DP could report about a different day from the one that was originally assigned. This design change was investigated in a previous report submitted as part of this task. The analyses described there showed that weekdays, specifically Monday-Thursday, are similar in terms of the patterns of time use that respondents engage in on those days. However, Friday, Saturday, and Sunday each demonstrate

different patterns of time use that would make it difficult to substitute those days for any other day without potentially biasing the data. Hence, a systematic design change that reassigned sample persons from one of the four interchangeable weekdays to another of the interchangeable weekdays would be feasible.

Another possibility is to replace one DP with another DP from within the same household (at least in households with multiple adults). This report explores this possibility, using the available ATUS microdata. First, we use matching to explore potential differences in time use activity between persons within a household. Second, we conduct a simulation to investigate the effects of allowing DP substitution at different points in the field period.

Matching 2

The current design of the ATUS does not allow us to directly estimate the effect of substituting another individual for the designated person (DP). For example, the ideal data set for such an analysis would include multiple persons from the same household asked about time use during the same reference period. Therefore, we can only simulate this arrangement with the current ATUS microdata.

Matching

The primary goal of the analysis in this chapter was to select a random individual within the designated person's household to replace the DP person and then find a match for that replacement person to serve as a donor for the randomly selected individual's time use. We undertook the following steps to prepare for the analysis.

1. First, we began with ATUS roster file for years 2003-2011. This data file lists all members of the designated persons' household for DPs that responded to the ATUS in these years.

2. We deleted everyone from the roster file under the age of 15 since they are not eligible to participate in the ATUS. We also deleted single person households.

3. We divided the roster file so that we had a "donor data file" consisting of respondents from odd numbered years of data collection (2005, 2007, 2009, and 2011). Respondents from even numbered year data years of data collection (2004, 2006, 2008, and 2010) were used as the "designated person data file". We also merged estimates of time use from the main activity categories for the designated respondent onto the designated respondent data file. This information would be used in analysis to compare the designated respondents' time use with the selected substitute's time use.

4. We assigned everyone on the designated person data file a random number and sorted the data file in ascending order by TUCASEID and the random number.

5. We then selected a person from each household with a DP to be a substitute for that DP. We did this by creating a new data file that took the case with the lowest random number within each household on the designated person data file. For two-person households, the other member of the household was selected. For three person

households, a randomly selected member of the household was selected as the person to substitute.

6. We merged information about the substitute person onto the new data file with those who were selected to be substitutes. We used this information to find a donor to match the person who was selected to be replaced. We merged the following information onto the data file for each selected substitute.

   a. The day about which the substitute would report (1=Monday-Thursday, 2=Friday, 3=Saturday, 4=Sunday). We combined Monday to Thursday based on results from our review and analysis of day of week substitution.

   b. The age of the substitute (1=15-30, 2=31-60, 3=61 or older)

   c. The substitute's sex (0=Male, 1=Female)

   d. Whether or not the substitute had a spouse (0=No, 1=Yes)

   e. Whether or not the substitute was enrolled in high school or college (0=No, 1=Yes)

   f. The substitute's employment status (0=Not employed, 1=Employed part-time, 2=Employed full-time)

   g. Whether or not the substitute's spouse was employed (0=No/No spouse, 1=Yes)

   h. The age of children in the household (0=No children, 1=Youngest child age 0-5, 2=Youngest child age 13-17)

   i. The number of eligible respondents age 15+ in the household (2=2, 3=3 or more)

7. We attached the same information from step 6 to the cases in the donor data file. We also attached estimates of time use from the main activity categories to each case in the donor data file. This information would be used in analysis to compare the designated respondents' time use with the selected substitutes time use.

8. We then used SAS PROC SQL to match cases with similar characteristics as shown in steps 6 and 7. The SQL procedure matched each selected substitute with each potential donor. The data from one of the randomly selected donors was taken to represent the selected substitute's data. Overall this resulted in a match for 37,353 of the 38,197 (98%) cases on the selected substitute data file.

9. In the final step we appended the donor's time use estimates for the selected substitute onto the designated person data file with time use estimates.

Propensity Models

To examine the mechanism that might be used to select substitutes, we estimated response propensity models. Response propensity models estimate the probability of the designated person responding to the survey request; this is done using logistic regression models. The indicator for response was created from the final call disposition on ATUS case history file.

The estimation of response propensity models requires data on both respondents and nonrespondents. Two types of ATUS data files supply this type of information. First, the ATUS-CPS file includes several variables from the CPS on all members of the designated person's household. We included variables in the model based on previous research on response to the ATUS (e.g., Abraham, Maitland, and Bianchi, 2006; Fricker and Tourangeau, 2009). From the ATUS–CPS file we included variables such as the DPs sex, age, education, marital status, and work status. We also included information about the household from which the DP came from such as family income and the presence of children in the household. Last, we included socio-environmental information such as the region and urban environment where the DP lives.

Next, the ATUS case and call history files include survey methodological information about the process of recruiting the designated person to participate in the ATUS. As mentioned before, we used the final call disposition on the case history file to create the response indicator. We also used the call history file to calculate an indicator for whether or not the respondent was interviewed on the first assigned diary day.

We merged the response propensity onto each case in both the designated person file and the donor file.

At-Home Patterns

Those who are at home more often are easier to contact. Therefore, the at-home patterns of the DP and the selected substitute are an important piece of information to consider in a data collection strategy involving substitution. We calculated the at-home time of the cases in the designated person and donor data files using the ATUS activity file.

Scenarios

The analysis looks at the effect of substitution by examining different scenarios using the above variables. We examined the following scenarios in the analysis.

Scenario 1: ***Replace all designated persons with the selected substitute.***

Scenario 2: ***Replace the designated person with a substitute when the substitute spends more time at home than the designated person.*** In reality, we would not have any information about the at home patterns of the sampled households. However, we consider a plausible scenario that substitution leads to the selection of respondents who are at home more often and therefore easier to contact.

Scenario 3: ***Replace the designated person with a substitute if an interview is not completed on the first day.*** In order for substitution to improve data the probability of obtaining response, there will need to be a rule that allows for substitution fairly early in the field period. We use the outcome of the first day of calling in this scenario and those that follow as a mechanism to allow for substitution.

Scenario 4: ***Replace the designated person with a substitute if an interview is not completed on the first day and the designated person's day 2 response propensity is less than .5 and the substitute's response propensity is greater than .5.*** The propensity model for the designated person included an indicator created from the ATUS call history file for whether or not the designated person responded on day 1. This reflects the reality that on call 2 we have some additional information from day 1 about the designated person's response propensity.

Scenario 5: ***Replace the designated person with a substitute if an interview is not completed on the first day and the substitute's response propensity is greater than the designated.*** We first look at the differences and similarities between the time use patterns of the designated person and the selected donors. Then we show the results for the different scenarios described above.

Correlation Between Designated Person and Selected Substitute

Table 2-1 illustrates the overall similarities and differences between the cases in the designated person file and the donor file. These represent the original DPs and the possible substitutes for them. In the aggregate, the two groups look fairly similar. The overall means for the actual respondents and the selected substitutes are nearly equal over the main categories of time use.

Table 2-1. Difference between matched individuals in major time use categories

Activity	Designated person mean	Substitute mean	Percent equal	Mean absolute difference	Median absolute difference	Minimum absolute difference	Maximum absolute difference	Correlation
Personal care activities	570.67	572.68	1.6	137.41	105	0	1160	.10
Household Activities	123.83	125.06	7.3	131.54	90	0	1240	.11
Care for and helping household members	47.13	45.74	51.2	53.53	0	0	1035	.28
Caring for and helping non-household members	13.05	12.34	76.8	23.79	0	0	1095	.00
Work and work related activities	180.14	178.99	45.9	152.32	21	0	1315	.48
Education	20.88	21.21	89.2	22.71	0	0	1100	.48
Consumer Purchases	43.79	41.31	33.9	62.87	33	0	1320	.04
Professional and personal care services	6.12	5.78	86.2	11.38	0	0	1070	.02
Household services	1.26	1.05	96.4	2.29	0	0	735	.00
Government services and civic obligations	0.48	0.37	99	0.85	0	0	510	.01
Eating and drinking	77.27	76.35	3.9	60.75	45	0	1280	.04
Socializing, relaxing and leisure	286.09	287.36	1.1	196.44	155	0	1390	.25
Sports, exercise, and recreation	24.01	25.64	65.9	43.22	0	0	1073	.05
Religious and spiritual activities	14.00	14.18	80.2	23.85	0	0	1100	.09
Volunteer activities	11.37	10.97	85.6	21.38	0	0	1145	.01

Even though the overall means look similar, there are differences between the members of individual pairs. The far right column shows the correlation between the designated person's time use and that of their selected substitute. The correlations are the largest for activities like work and

education. Activities such as care for and helping household members and leisure show weak to moderate correlations. The remainder of the correlations are near zero.

The fourth column of Table 2-1 shows the percentage of the matched pairs that had time use estimates that were exactly equal to each other. In general, activities in which relatively few respondents engage in such as civic obligations and household services had the highest percentage of cases where the designated person and the selected substitute spent the same amount of time in the activity, with both spending zero time. In contrast, relatively few cases matched on more prevalent activities such as personal care activities and household activities.

Comparisons Between Scenarios for Substitution

Table 2-2 shows the differences in main activity levels by the scenarios that were described earlier in the methods section. The means for the designated respondents are shown in column 1. Recall that scenario 1 replaces all designated persons with a selected substitute. Hence column 3 is repeated from Table 2-1 and shows very few differences from the mean of the designated persons.

Scenario 2 replaces the designated person with a substitute when the substitute is at home more often than the designated person. As mentioned in the methods section, we typically do not have any knowledge of a sample household's at home patterns. However, for the purposes of this exercise, we can see that scenario 2 has the largest differences with the mean of the designated person. Replacing designated persons with substitutes who are at home more often leads predictably to higher reports of activities such as household activities, care for household members, and leisure. Conversely, it leads to lower reports of activities performed outside the home such as work related activity. Scenario 2 represents a likely outcome from implementing Scenario 1 or some of the other scenarios, since a likely outcome of any substitution strategy is an increase in the representation of people who are at home more often.

The remaining scenarios are based on the response propensities of the designated person and selected substitute. In general, the estimates for these scenarios move in the same direction as scenario 2 does from the mean of the designated persons. However, the time use means for the actual DPs is much more different from those of scenario 2 than from any of the other scenarios. This finding is not entirely surprising given the relatively weak correlation that has been found between time use estimates and the results of response propensity models of the ATUS data. The

largest correlations between time use and response propensities found in this analysis were between .1 and .2.

Table 2-2. Mean activity levels across different scenarios of substitution

Activity	Designated person mean	Scenario 1 mean	Scenario 2 mean	Scenario 3 mean	Scenario 4 mean	Scenario 5 mean
Personal care activities	570.67	572.68	559.86	571.40	567.92	568.87
Household Activities	123.83	125.06	155.24	124.75	125.08	125.19
Care for and helping household members	47.13	45.74	52.90	47.38	48.83	48.37
Caring for and helping non-household members	13.05	12.34	10.19	12.53	12.83	12.62
Work and work related activities	180.14	178.99	145.27	177.86	179.31	178.12
Education	20.88	21.21	20.84	21.41	21.24	21.35
Consumer Purchases	43.79	41.31	37.17	42.78	43.92	43.56
Professional and personal care services	6.12	5.78	4.92	6.10	6.11	6.17
Household services	1.26	1.05	1.17	1.21	1.24	1.24
Government services and civic obligations	0.48	0.37	0.40	0.44	0.43	0.43
Eating and drinking	77.27	76.35	73.12	76.75	78.15	77.53
Socializing, relaxing and leisure	286.09	287.36	319.94	287.56	283.89	285.61
Sports, exercise, and recreation	24.01	25.64	19.05	24.40	25.12	25.01
Religious and spiritual activities	14.00	14.18	10.74	13.83	13.55	13.69
Volunteer activities	11.37	10.97	8.56	11.46	11.98	11.80
Average relative difference from designated person mean		5.0%	15.0%	1.9%	2.4%	2.3%

Simulation 3

Chapter 2 demonstrated that the effect of substitution on key time use estimates can vary by the mechanism that leads to substitution. In this chapter, we present the results of a simulation that further explored how DP substitution would affect response rates, sample composition, and time use estimates.

Response Propensities

The first step in the simulation was to estimate the probability DPs would complete the survey during each of the eight weeks in the field period. To do so, we first link the paradata files with the ATUS-CPS file for 2004-2011 in order to construct a person-period dataset with one record for each week that a DP was actively called. Each record contains the DP's demographic predictor variables taken from the CPS (listed below), and an indicator for whether or not the DP completed the survey in a given week (0=No, 1=Yes). We then perform a discrete time hazard model to estimate the probability that each case will respond during a given week conditional upon their demographic characteristics and the week in the field period. This allows us to observe how response propensities change as the field period progresses. We use the following variables to estimate the model:

1. Week in the field period (1-8) plus a quadratic term (week2);
2. Region (1=Northeast, 2=Midwest, 3=South, 4=West);
3. If the household was indicated as having a telephone in the CPS (0=Phone, 1=No Phone);
4. Whether or not the DP is Hispanic (Hispanic, Non-Hispanic);
5. Sex (0=Male, 1= Female);
6. Marital Status (1=Married Spouse Present, 2=Married Spouse Absent, 3=Widowed, 4=Divorced, 5=Separated, 6=Never Married);
7. Number of hours worked each week (1=Under 35 hours, 2=35-44 hours, 3=45 or more hours, 4=Hours vary, 5=None);

8. Whether or not there were children in the household (0=No, 1=Yes);

9. Age (1=15-30 years old, 2=31-45 years old, 3=46-55 years old, 4=56-64 years old, 5=65 years old or older);

10. Race (1=White, 2=Black, 3=American Indian, 4=Asian/Pacific Islander, 5=Multiple Races);

11. Education (1=Less than high school, 2=High school graduate, 3=Some college, 4=College graduate, 5=Post-graduate);

12. Metropolitan area (1=Metropolitan Area, 2=Non-Metropolitan area, 3=Unknown);

13. Housing status (1=Owned, 2=Rented, 3=Occupied);

14. Family Income (1=Under $20,000, 2=$20,000 to under $40,000, 3=$40,000 to under $75,000, 4=$75,000 or more, 5=Missing); and

15. Year of CPS participation (2004-2011).

We then use this model to calculate response propensities for each person in the ATUS-CPS file, both DPs and other household members, for each of the eight weeks in the field period.

Identifying Substitutes

For the purposes of this simulation, we identify one substitute in each household containing more than one eligible person. In households with more than one eligible substitute, we select the person with the highest response propensity in week 1. Single person households remain in the data set, but are not candidates for substitution.

Imputing Time Use Data

We next identify cases for which time use data must be imputed. This set includes DPs who were non-respondents as well as substitutes. We perform an unconstrained k-nearest-neighbor hot-deck imputation to identify donor cases with similar characteristics (D'Orazio et al 2006). Each case was assigned to an imputation cell based on the following variables:

1. Sex (0=Male, 1=Female);

2. Sampled day of the week for the DP in the household (1=Monday-Thursday, 2=Friday, 3=Saturday, 4=Sunday);

3. Whether or not they are a student (0=Not Student, 1=Student);

4. Whether or not they are married (0=No Spouse, 1=Spouse);

5. Age (1=15-30 years old, 2=41-64 years old, 3=65 years old or older);

6. Whether or not they are employed (0=Not Employed, 1=Employed); and

7. Whether or not children under the age of 18 are present in the household (0=None Present, 1=Children Present).

This cross-classification yields 384 possible categories, although only 256 of these categories are actually present in the data. Of the 211,078 cases requiring imputation, there were 2 cases for which no donor was available. Because there were so few, they were simply removed from the analysis.

For each case requiring imputation, we calculate the rectilinear distance (or city block distance) between the recipient and each donor in the cell using the following variables.

1. Year (2004-2011);

2. Spouse's employment status (1=Spouse Employed, 2=Spouse Not Employed);

3. Number of eligible persons in the household (1, 2, 3+); and

4. Region (1=Northeast, 2=Midwest, 3=South, 4=West).

We then impute time-use variables from the nearest donor within each cell to the recipient case. If multiple donors are of equal distance from the recipient, one is chosen at random.

Simulation

At this point, each DP has a set of values for time use variables that correspond either to their original values or to values imputed from a similar respondent. In households where a substitute is available, the substitute has values that were imputed from matched respondent. We next create a dataset that contains one record for each household, and contains demographics, weekly response propensities and time use variables for the original DP as well as the possible substitute if one exists.

We wish to see how the effects of allowing DP substitution vary according to the rule permitting a substitution to be made. Therefore, we perform the simulation 1000 times for each of 9 conditions based on the number of weeks during which substitution could be allowed, ranging from 0 weeks (meaning no substitution allowed) to 8 weeks (substitution always allowed). In between, this is the number of weeks from the end of the field period during which substitution will be allowed.

For each household, we step through each week in sequence, starting with week 1. For each week, we generate a random number between 0 and 1. If this number is less than or equal to the DP's response propensity for that week, we code that household as complete with the DP and move on to the next case. If the number is greater than the DP's response propensity, we check to see whether substitution is allowed and a substitute is available. If both of these conditions are met, we generate a second random number to see if it is below the substitute's response propensity for that week. If it is, we code the case as completed by the substitute and move on to the next case. If substitution is not allowed, there is no available substitute, or the random number exceeds the substitute's response propensity for that week, we move on to the next week and repeat the process. This process continues for each week until there is a complete with the respondent, a complete with the substitute, or we have gone through 8 weeks without a complete. Once all households have been resolved, we calculate demographic and time use statistics for the resulting set of complete cases.

Results

Figure 3-1 summarizes the mean proportion of households in which an interview would have occurred under the simulation conditions. The proportion of households yielding an interview diminishes as we allow fewer weeks of substitution. Furthermore, the relationship is nonlinear showing diminishing returns as we restrict the number of weeks in which substitution is allowed. The share of completes obtained from substitutes also declines more rapidly than the proportion of completes overall, indicating that early on, some of the substitutes interviewed are replacing DPs who would likely have been interviewed eventually had substitution not been allowed.

Figure 3-1. Proportion complete by number of weeks with substitution allowed

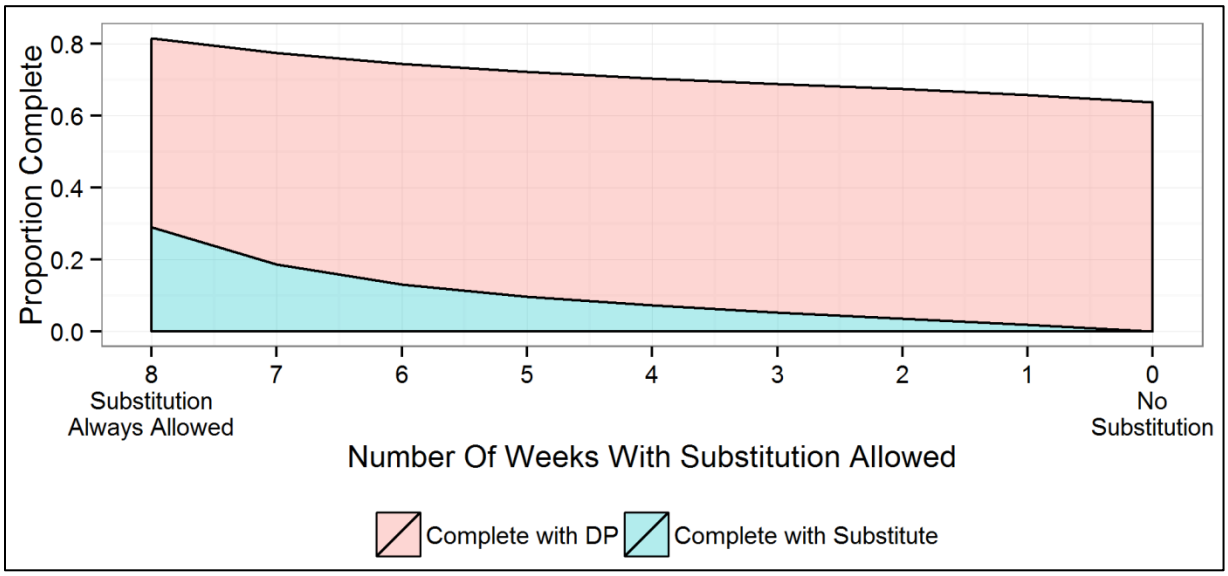


Figure 3-2 shows how the demographic composition of ATUS respondents changes according to the week in which substitution becomes allowed. Each panel in Figure 3-2 shows the differences between the proportion of respondents in that demographic category under each substitution condition (1-8 weeks) and the no-substitution condition (0 weeks). The number of completed interviews in each demographic category rises as more substitution is allowed, however the number of completes in each group increases at different rates. For example, seniors aged 65 or older represent a lower proportion of completed interviews with greater substitution; this is because their number increases more slowly in absolute terms than 15 to 30 year olds or 31 to 45 year olds.

Surprisingly, as the number of substitution weeks increases, we see an increase in the proportion of completes belonging to lower propensity categories such as men, 15-30 year olds, blacks and renters. This appears to result from the fact that higher propensity DPs are more likely to complete an interview than their substitutes, while at the same time, response propensities for DPs and their substitutes show a modest positive correlation of .36. The result is that allowing within household substitution has a disproportionate effect on households with low propensity DPs, and substitutes in these households are likely to share at least some demographic characteristics with the DP.

Figure 3-2. Effect of substitution on demographic composition. Mean proportion for each demographic category is the mean over 1000 simulations for each weekly substitution condition

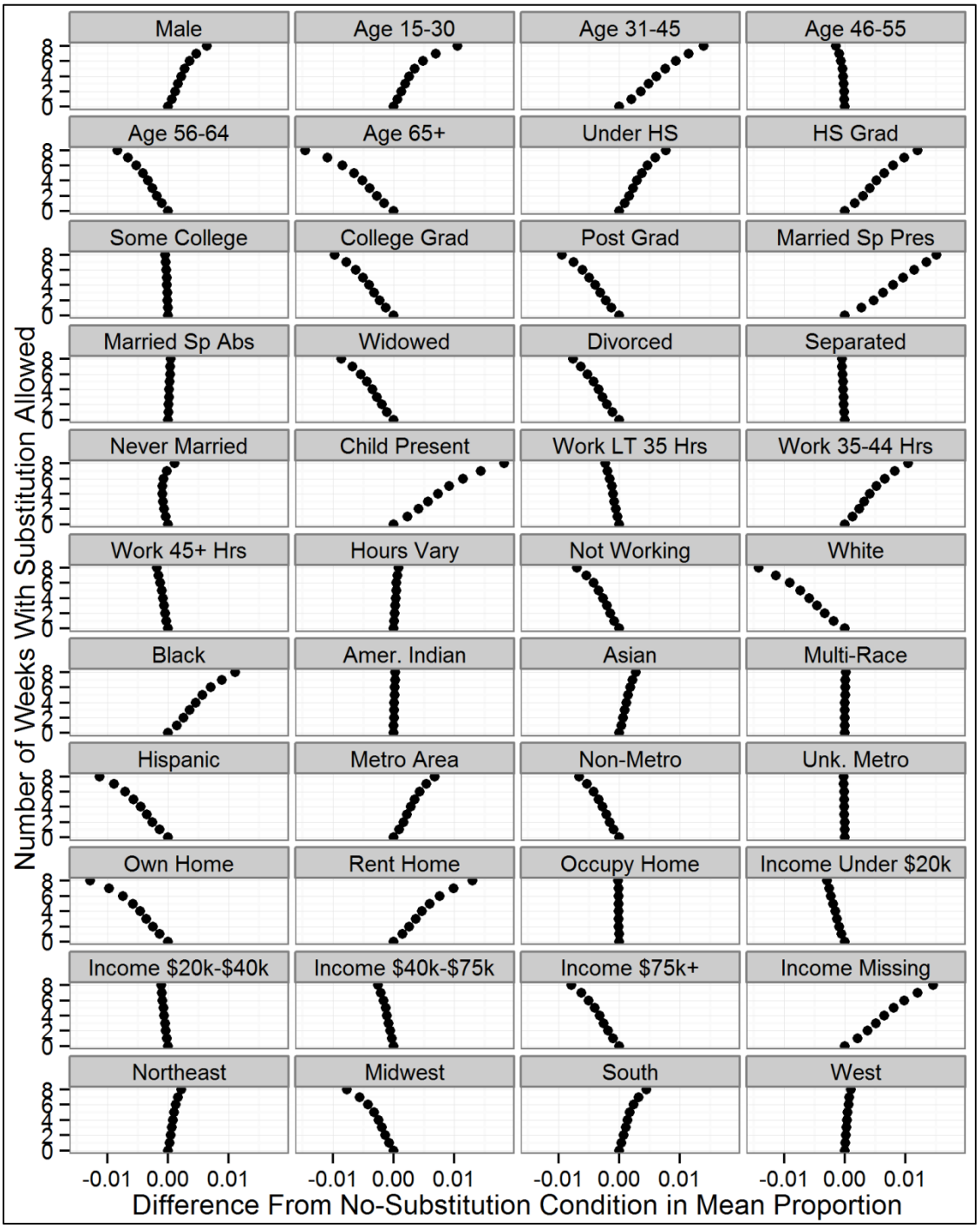
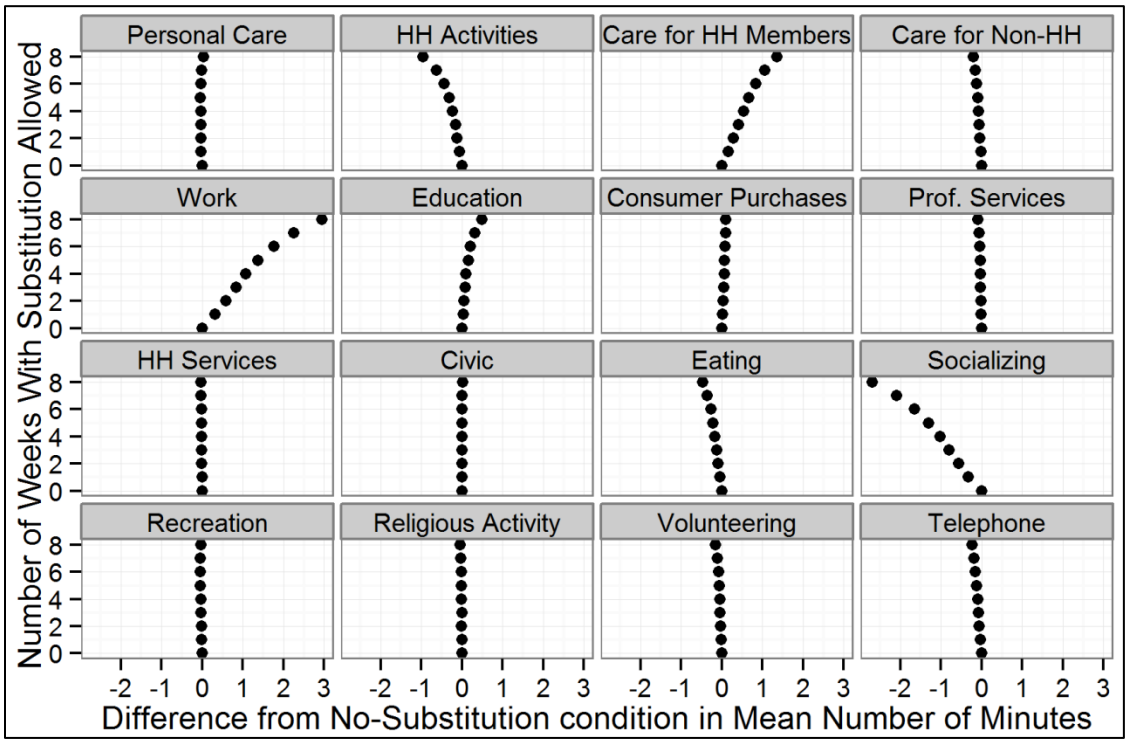


Figure 3-3 shows the degree to which time use estimates are sensitive to DP substitution in this simulation. It depicts the difference in the mean number of minutes spent on each of 16 time use categories between each substitution condition (1-8 weeks) and the no-substitution condition (0-weeks).

Figure 3-3. Effect of substitution on time use estimates



Care for household members, and work both show a clear positive association with increased substitution, whereas the mean number of minutes spent socializing declines with each additional week of substitution. It makes sense that individuals who spend more time at home caring for household members or less time socializing would be easier to reach and therefore more likely to be substituted. The positive relationship between substitution and minutes spent working seems counterintuitive, however it is likely the case that in households with multiple adults who work, at least one of them is more likely to be interviewed when substitution is allowed. When substitution is not allowed, those households that disproportionately benefit from substitution are more likely to result in nonresponse. Although a definitive explanation of the causal factors at work requires further study, it is clear that some demographic groups and time use categories are disproportionately affected by allowing DP substitution.

Conclusion 4

This report examines the effect of substitution of the designated person on the ATUS estimates. We first looked at the similarity between the designated persons and likely substitutes. We found large individual differences between the designated persons and their likely substitutes. In contrast, the overall means were not substantively very different from each other.

We then examined some likely scenarios under which substitution might occur. We once again found small differences in overall means across several of the scenarios. However, the effect of substitution on key time use estimates can vary by the mechanism that leads to substitution. As both the matching and simulation exercise demonstrate, substitution appears likely to produce small, but systematic changes in sample composition and time use estimates.

These results indicate that caution is warranted in pursuing a strategy of substitution for DPs. Substitution would move the ATUS from a purely design-based strategy with known probabilities of selection to a model-based strategy. The assumptions of this model-based strategy may be difficult to evaluate. For example, it will be nearly impossible to know beforehand if a substitution is occurring primarily because the respondent is persistently more at-home and available. As our analysis shows, if this is the case there is a potential for the ATUS estimates to be impacted by this change in the selection of the substitute.

References

Abraham, K.G., Helms, S., and Presser, S. (2009). How Social Processes Distort Measurement: The Impact of Survey Nonresponse on Estimates of Volunteer Work in the United States. *American Journal of Sociology* 114(4): 1129-1165.

Abraham, K.G., Maitland, A., and Bianchi, S. (2006). "Nonresponse in the American Time Use Survey: Who is Missing from the Data and How Much Does It Matter?" *Public Opinion Quarterly* 70(5):676-703.

Brick, J.M., and Williams, D. (2013). "Explaining Rising Nonresponse Rates in Cross-Sectional Surveys" *The Annals of the American Academy* 645:36-59.

D'Orazio, M., Di Zio, M., and Scanu, M. (2006). Statistical Matching, Theory and Practice. Wiley, Chichester.

Fricker, S., and Tourangeau, R. (2010). "Examining the Relationship Between Nonresponse Propensity and Data Quality in Two National Household Surveys." *Public Opinion Quarterly* 74(5): 934-955.

General Conclusion

This report presented the results of Westat's investigation into allowing substitution of the diary day or designated person. Westat's research included a review of the existing literature, analyses of existing ATUS microdata, and the development of an experimental design to evaluate the effect of allowing substitution of the diary day.

Findings for Day-of-Week Substitution

A few recurring themes emerged throughout the literature review. First, there is general consensus that a designated day-of-week approach is needed to minimize bias in the time use estimates. Second, the literature makes a major distinction between weekdays and weekends in terms of the type of activity in which respondents engage. However, even days like Friday at the end of the work week sometimes appear to be slightly different from Tuesday – Thursday. Saturday and Sunday have their own unique patterns of time use. Finally, some of the conclusions from the literature are based on relatively little published research. Although much of the literature is suggestive, there are very few studies directly addressing important issues such as whether a convenient day approach leads to significantly different estimates from the designated day approach. There is also very little research on the acceptable length of the recall period in time use surveys. These gaps in the literature leave open the possibility for some much needed methodological research that could help inform the design of the ATUS and future time use surveys.

Our analysis of the existing ATUS microdata and other relevant data sources found some differences in time use across days of the week. As expected, weekdays differ from weekends. Respondents appear to have slightly different time use patterns on Friday compared to Monday, Tuesday, Wednesday, and Thursday. Our analysis of time spent in main activities showed that Monday may also have some differences with other weekdays; however, the dissimilarity indexes indicated that differences between weekdays are clearest between Friday and the remaining weekdays.

We also examined differences in response propensities by day of the week. Response propensity is slightly higher at the beginning of the week and then declines the later in the week that someone is scheduled to be interviewed. We did find that cases that require substitution (i.e. those interviewed after the first scheduled interview day) engage in somewhat different levels of activity than those that

do not require substitution. For example, one might expect that those who are most likely to substitute days are those who engage in more activities such as work and other related activities outside the home.

Based on these findings from the literature review and our analysis of the data, we designed an experiment that tests the effect of substituting either the call day or the diary day. Both approaches have the advantage of increasing the flexibility of call scheduling, which should lead to more efficiency, less effort, and a higher success rate in contacting households.

The two approaches also have the potential to increase error in time use estimates, which needs to be carefully evaluated. For example, substituting the diary day may increase the probability of contacting individuals on days when they are more likely to be at home. This in turn could result in the respondent being more likely to report on activities they do when they are not at home if they are reporting about a previous day when initial contact attempts were unsuccessful. Substitution of the call day does not entail much of risk for contact and activity bias. However, extending the call day further away from the designated diary day does increases the risk of errors from memory decay on time use estimates.

We propose a limited day-of-week substitution procedure with the diary day. The conditions in this experiment are shown in Tables 1 and 2. The literature review and our analysis indicated that Monday – Thursday were the most substitutable days. Hence, we allow for substitution of the diary day between these days. One of the conditions allows for those who do not complete the interview by week two to be randomly assigned to another diary day between Monday and Thursday. This condition should minimize the impact of activity bias. Another condition allows the respondent to complete the diary about any day Monday to Thursday. Finally, we include two conditions for substituting the call day and extending the recall period if the respondent has not completed the interview by week two. In one of these conditions the respondent is asked to recall their activity from the day before yesterday and in the other condition the respondent is asked to recall their activity from two days before yesterday. In addition to providing more thorough evidence about the length of the recall period in time use studies, substitution of the call day is a viable option for weekend days since there really is not another compatible day to which someone can be reassigned if they are originally assigned a weekend day. We also included two designs for a preliminary study that would provide ATUS staff with insight about the feasibility of extending the recall period prior to full-scale implementation of the experiment.

Table 1. Summary of experimental design for weekday (Monday – Thursday) cases

		Weekday cases = 9,200			
Condition	**Control**	**Condition 1**	**Condition 2**	**Condition 3**	**Condition 4**
Description	Current ATUS Design	Substitute diary day randomly between Monday and Thursday	Substitute any diary day between Monday and Thursday	Substitute call day – ask about day before yesterday	Substitute call day – ask about 2 days before yesterday
Example		Tuesday switched to Wednesday	Tuesday switched to any Tuesday-Friday	Wednesday asking about Monday	Thursday asking about Monday
Details		After week 2 – sample person is assigned a new call day between Tuesday and Friday	After week 2 – sample person is switched to a protocol with contact attempts made between Tuesday and Friday	After week 2 – sample person keeps the same diary day, but is switched to a protocol with reporting beginning 4 AM on day before yesterday	After week 2 – sample person keeps the same diary day, but is switched to a protocol with reporting beginning 4 AM 2 days before yesterday
Sample	10,400	2,300	2,300	2,300	2,300
Eligible	9,360	2,070	2,070	2,070	2,070
n	5,200	1,155	1,155	1,155	1,155

Table 2. Summary of experimental design for weekend (Saturday and Sunday) cases

		Weekend cases = 4,600	
Condition	**Control**	**Condition 5**	**Condition 6**
Description	Current ATUS Design	Substitute call day – ask about day before yesterday	Substitute call day – ask about 2 days before yesterday
Example		Monday asking about Saturday	Tuesday asking about Saturday
Details		After week 2 – sample person keeps the same diary day, but is switched to a protocol with reporting beginning 4 AM on day before yesterday	After week 2 – sample person keeps the same diary day, but is switched to a protocol with reporting beginning 4 AM 2 days before yesterday
Sample	13,000	2,300	2,300
Eligible	11,170	2,070	2,070
n	6,400	1,155	1,155

Findings for Substitution of the Designated Person

We first looked at the similarity between the designated persons and likely substitutes within the household. We found large individual differences between the designated persons and their likely substitutes. In contrast, the overall means were not substantively very different from each other.

We then examined some likely scenarios under which substitution might occur. We once again found small differences in overall means across several of the scenarios. However, the effect of substitution on key time use estimates can vary by the mechanism that leads to substitution. As both the matching and simulation exercise demonstrate, substitution appears likely to produce small, but systematic changes in sample composition and time use estimates.

These results indicate that caution is warranted in pursuing a strategy of substitution for DPs. Substitution would move the ATUS from a purely design-based strategy with known probabilities of selection to a model-based strategy. The assumptions of this model-based strategy may be difficult to evaluate. For example, it will be nearly impossible to know beforehand if a substitution is occurring primarily because the respondent is persistently more at-home and available. As our analysis shows, if this is the case there is a potential for the ATUS estimates to be impacted by this change in the selection of the substitute.