The overestimated workweek? What time diary measures suggest

The comprehensive time-diary method allows analysts to distinguish work from nonwork activities; a comparison with workweek estimates reveals important and systematic differences

John P. Robinson
and
Ann Bostrom

The amount of time people spend at their jobs can be an important social and economic indicator of a society's quality of life. Measured hours devoted to work are important in many ways, as A. Mata-Greenwood describes, "... the regulation of working time is an aspect which has a direct and measurable impact on workers' health, level of strength and fatigue, on the establishment's productivity and costs, and on the society's general quality of life."[1] Thus, one of the central arguments of the rising quality of life in 20th century Western countries has been the reduction in the hours people spend at work.

Figures on hours spent at work allow analysts to see whether changes in productivity are attributable to changed production of outputs, or to changed time required to produce these outputs. They further allow analysts to gauge whether workers remain as adept in production as previously, whether workers in one industry are working more or fewer hours than those in other industries, or whether unionized workers work fewer hours than other workers.

This article describes problems that arise for respondents who are surveyed using the workweek estimate approach and comparable figures from the total time-diary approach to calculate hours of work. It also describes reliability and validity studies supporting the diary method. Finally, results showing deviations between the two approaches are presented, as well as the effects of other selected variables.

John P. Robinson is professor of sociology at the University of Maryland, College Park. Ann Bostrom is professor of Public Policy at Georgia Tech University, Atlanta.

Other measurement approaches

As central as issues related to hours at work are, it is surprising how little statistical effort has been expended examining the basic validity of these data. The oldest method uses data based on the jobs available in firms. Firms and industries provide data largely based on payroll accounts of employee work hours on jobs—or more precisely, the hours employees are paid for. These hours are calculated at the aggregate level of the firm on the basis of the job, not on the basis of the individual worker. Thus, comparisons cannot be made between employees who work fewer hours versus those who work more hours, or among workers who are of different ages, genders, and so forth. These figures also do not identify the total hours of individual workers who moonlight in additional jobs, have paid days off, and the like. Moreover, these figures, often, do not cover certain sectors of the economy, particularly the self-employed, unpaid family workers, and persons employed by very small or new firms.[2]

Analysts often rely on data gathered by government surveys, such as the Current Population Survey (CPS).[3] The CPS, conducted by the U.S. Bureau of the Census for the Bureau of Labor Statistics, asks respondents to report their hours at work per week at any and all jobs. Such "workweek estimate" questions are asked in other surveys, but they vary with each survey. Most of the surveys ask about the hours at work in the week preceding the study; others ask about normal, usual, average, or expected work hours.

In general, there is the implicit assumption in these estimate questions that respondents are able to answer them accurately. Thus, the question in the case of "last week" estimates assumes that respondents will recall for each day of the preceding week whether they worked or not, and if so, the number of hours they worked. Respondents have just a few seconds to answer the question, which is embedded in a longer survey. Then, it is presumed that the respondent will correctly add up the amounts for each day across all 7 days of the week.

There are several implicit assumptions in this estimate question. Even if respondents were able to reconstruct such information accurately from memory, would they reveal it to the interviewer? Would they worry that their estimate might not be what the interviewer or the research organization expected? Similarly, respondents might reasonably want to portray themselves as impressively as possible, either as very hardworking, or as not being too tied to the workplace (if there are young children in the family, or if they want to present themselves to interviewers as an "easy-going" person, not obsessed with making money). Moreover, if that last week was in fact atypical, might not these respondents attempt to "smooth out" their answers to conform to their "normal" workweek or to some societal norm? Anecdotal evidence from interviewers suggests that respondents spend less time answering this question each successive time they complete the survey, which increases the likelihood that a response will be a norm or average. (While the new CPS has a sequence of questions to help minimize the likelihood of this happening, this does not address the main differences described in this article.[4])

Another factor to consider is whether respondents clearly understand whether or not they should exclude or include the commute to work, work breaks, machine down time, changing clothes at work, the lunch break, work brought home, and the like. Even if they are given specific instructions, are they able to compartmentalize their work time neatly into the categories requested?

The potential problems with workweek estimates become more acute when one confronts the myriad definitions related to the interests of labor force analysts. For example, Mata-Greenwood has distinguished at least eight working time concepts in the literature, such as "time worked" (time actually worked); versus "time for work" (time scheduled for work); "contractual time" (time formally contracted to work); and "time paid" (the hourly basis for pay).[5] Each of these are dependent on such fine-grained distinctions that they are probably indiscernible to most respondents.

There is, moreover, the matter of how respondents define "workweek," which may be in terms of their contractual arrangement with their employer, rather than actual clock time. Hence, workweek answers tend to cluster at exactly 40 hours, the most prevalent workweek norm in this country. However, the "normal," "9-to-5" job only adds up to 40 hours, if the respondent works straight through the day without using time for lunch or other extended breaks. (Also, contracted hours for employment have not changed much, but the ability to undertake personal activities within this contractual period may have expanded greatly in the last 20 years.) Thus, the estimate question approach assumes that respondents will understand the timeframe of the question and the definitions of work and nonwork activities, retrieve the hours accurately from memory, sum the hours properly, and willingly disclose this accurate information to the interviewer. These assumptions are totally independent of any temptation to portray oneself in a socially undesirable or self-deprecatory light. Thus, the estimate approach appears to place great demands and expectations on the part of a typical respondent.

The time-diary approach

An alternative, more elaborate approach to measuring hours at work is the comprehensive time diary, one in which respondents recall all of their activities, work or nonwork, for a week or more — typically, the 24 hours of a single day. Rather than hours at work being the focus of the diary, and thus subject to unneeded emphasis in the reporting process, respondents have no cues about which activities the interviewer might be pleased about or interested in. Moreover, the task for respondents is focused on the sequence of activities and when they occurred, rather than their having much opportunity to project any larger image of themselves in their activity reports. While respondents do have a chance to give the answer that they want to provide, they are not told which activities are of survey interest, because all of the activities are.

Like other survey questions about behavior, work estimate questions, similar to those used in the CPS, usually examine people's activities in isolation from the natural temporal context in

which they are embedded. Thus, they ask respondents to compress their actual behavioral experiences by implicitly saying whether they "often" or "usually" do something. In contrast, time-diary accounts report activities as they naturally and sequentially occur in daily life. Thus, studies of time use allow examination of human activities in "real time"—as individuals are actually involved in the stream of daily behavior.

Time diaries can be seen as a prime example of the "micro-behavioral" approach to survey research.[6] This micro-behavioral approach recognizes the limited ability of respondents to report very complex behavior in a survey context. In the time diary, the survey question is limited to the most elementary experience about which respondents can accurately report. The micro-behavioral approach also provides researchers with a more basic and flexible data base from which to draw conclusions about human activity.

The time diary is a micro-behavioral technique for collecting self-reports in an open-ended fashion on an activity-by-activity basis. This technique capitalizes on the most attractive measurement properties of the time variable; namely, completeness, equal distribution, and understandability. Thus:

- All daily activity is potentially recorded over a 24-hour period.
- All 1,440 minutes of the day are equally distributed across respondents (thus allowing certain "trade-offs" between activities to be examined);
- Time accounts are mutually exclusive and exhaustive in that each of the 1,440 minutes is assigned to one and only one main activity; and
- Respondents are allowed to use a timeframe and accounting variable that is maximally understandable to them and accessible to memory. The open-ended nature of activity reporting means these activity reports are automatically geared to detecting new and unanticipated activities (for example, telecommuting, use of new communications technologies), as well as capturing the sequential context of how daily life is experienced.

In contrast, survey questions based on recall or estimates of time use implicitly assume that respondents will sort through and recall from memory only selected and partial behaviors. In attempting to recall only work time, respondents are likely to have trouble not only identifying, but recalling all their "work" episodes.

The time-diary method allows respondents to report the totality of their daily activity in a single account, one that for most people is consistent with the way events are sequentially organized in their experience and probably stored in episodic memory.

Thus, time diaries provide an ideal method for cross-person comparisons of daily behavior. Not only are respondents' daily reports standardized across groups, but the full daily context of their work experiences is recorded. Moreover, data on activities preceding work or following it at the end of the work day are recorded, resulting in more complete and systematic reports of daily behavior.

Time-diary data bases

The present analysis is based on national data available through the Survey Research Center at the University of Michigan and the University of Maryland. In this data series, national time-diary studies are available for more than 30 years. We focus on the 18–64 age group, included in time-use studies from 1965, 1975, and 1985. (See table 1.) In the 1985 study conducted by the Survey Research Center of the University of Maryland, single-day diaries were collected across the entire year. Three modes of diary collection were used for comparison: personal, mail-back, and telephone, with little difference in obtained time estimates. Respondents report each activity, as well as where they were, who they were with, and various other aspects. Methodological details on the 1965, 1975, and 1985 studies are provided in the appendix.

Prior to the 1985 national study with 5,358 total respondents aged 18 and older, two national time-diary studies had been conducted in 1965[7] and 1975[8], using this general approach. These open-ended diary entries were coded and arranged using the basic activity coding scheme developed for the 1965 Multinational Time Budget Research Project.[9] The main value of the open-ended diary approach is that activities can

Table 1. **Methodological features of national time diary studies, 1965, 1975, and 1985**

Characteristic	1965	1975	1985
Data collected by	University of Michigan	University of Michigan	University of Maryland
Sample size	1,244	2,409	5,358
Age range	18–64	18 and older	18 and older
Survey mode	Personal	Personal	Telephone, mail-back, and personal
Diary period (days)	1–2	1	1–2
Daily hours covered	24	24	24
Diary method	Tomorrow and additional 10 percent, yesterday	Yesterday	Tomorrow (72 percent) and yesterday (28 percent)

NOTE: Further description of study details can be found in the appendix.

be recorded or recombined, depending on the analyst's unique assumptions or purposes.

The activity code for the 1985 national study has some attractive features. First, it has been tested and found reliable in 25 countries, including virtually all Eastern and Western European countries, over the last two decades. Second, extensive, national data are comparable with earlier data. And third, the activity code can be easily adapted to include new or additional code categories of interest to various researchers.

Diary data, when aggregated, provide generalizable national estimates of the full range of alternative daily activities: from "contracted" time (work or the commute to work), to "committed" time (family care), to personal care (sleeping, eating, hygiene), and to all the types of activities that occur in free time. The multiple uses and perspective afforded by time-diary data have led to a recent proliferation of research and literature in this field.

Reliability. In the 1965 and 1975 studies, estimates from time diaries produced rather reliable and replicable results at the aggregate level. For example, there was a 0.95 correlation between time-use patterns found in the 1965–66 national time diaries (n=1,244) and the aggregate figures for the single site of Jackson, Michigan (n=788).[10] Similar high correspondence was found for the American data and for time-diary data from Canada, both in 1971 and in 1982.[11] A correlation of 0.85 was found between time expenditure patterns for the U.S.-Jackson, MI, time study in which respondents filled out diaries on the "day after" and a random tenth of the respondents also filled out a "day before" diary. A smaller replication study in Jackson in 1973 found an aggregate correlation of 0.88. (Unpublished reliability studies for the 1985 data show similar results.)

Validity. Almost all diary studies depend on the self-report method, rather than on some form of observation. Unfortunately, the data are open to questions based on validity. Can they be verified by some independent method of observation or report?

One such study concluded that standard television-rating service figures on the time participants spent watching TV provided higher estimates of viewing behavior than those recorded in the time diaries. The TV viewing behavior of a small sample of 20 households was monitored over a week's time by means of a video camera.[12] The results indicated that rating-service methods of TV exposure (audimeters and viewing diaries) produced estimates of viewing that were 20 to 50 percent higher than primary or secondary activities reported in time diaries.[13]

Three subsequent validity studies examined the full range of activities, not just television viewing, and employed larger, more representative samples. In the first study, a 1973 random sample of 60 residents of Ann Arbor and Jackson, Michigan kept beepers for a 1-day period and reported their activity whenever the beeper was activated (some 30 to 40 times during the day when most people are awake and active).[14] Averaged across all 60 respondents, the correlation of activity durations from the beeper and from the diaries was 0.81 for the Ann Arbor sample and 0.68 for the Jackson sample.

In a second study, a telephone sample of 249 respondents was interviewed as part of a 1973 national panel survey. Respondents were asked to report their activities for a particular "random hour" during which they were awake—with no hint from the interviewer about what they had previously reported for that hour in their diary.[15] An overall correlation of 0.81 was found between the two aggregate sets of data, that is between the activities reported in the random hours and in the diary entries for those same random hours.

A third study used the 1975–76 diaries to compare answers to the question, "With whom?" between respondents and their spouses.[16] In more than 80 percent of the diary entries, these independently obtained husband and wife diaries agreed that their spouses were present or absent. A separate analysis of these 1975–76 data found a 0.93 correlation between time spent on various home energy-related activities (such as lighted homes or appliance use) and aggregate time-of-day patterns of energy use derived from utility meters.[17]

In conjunction with the reliability studies, then, the data from these validity studies provide some assurance about the basic generalizability of time-diary data. This has been the case as well in methodological studies conducted in other countries.[18] Nonetheless, a definitive well-controlled study needs to be conducted, particularly for specific types of locations and activities of interest to labor analysts.

Limitations of the data sets

One can imagine several reasons the more detailed diary approach would provide lower estimates of work time. First, like most activities, work can be combined with other activities, in the sense that one can take care of personal business (such as paying bills), socializing (as in taking off early with work colleagues to go to a restaurant or bar), or attending to the mass media during scheduled work hours. While most workers might report this simply as work, other workers might report it for what it is—household work, social life, or TV viewing.

Second, some persons who work more than the usual hours might be subject to distorted perception, particularly because extensive hours could involve less regular work schedules in relation to those of other workers. To these respondents, the retrospective reporting of hours-at-work could be more difficult because they have fewer solid "anchor points" or time markers around which to base their estimates. They work during hours of the day and week when others do not, and, therefore, they might be least likely to include accurate estimates. Moreover, they may feel deprived by having to work when others do not, so that work time seems longer because of their social isolation from the mainstream of society that is not engaged in work at these points in time. If such work time is also subject to unscheduled interruptions and distractions, that would further add to the sense of longer time. These conditions can easily be seen to lead not only to distortion of where time goes, but to the lack of convenient and solid anchor points from which to make accurate estimates.

A further factor leading to more reported work hours involves the well-known statistical phenomenon of "regression to the mean" which arises when the estimating procedure asks respondents to estimate their hours worked "last week." To the extent that regression toward the mean is in operation, people who worked unusually longer hours in the week before the study period are likely to work fewer hours during the week or day reported in the diary. This would also lead to higher estimates than diary workweeks among those estimating longer workweeks (as well as longer diary hours for those estimating fewer workweeks "last week").[19]

It should be noted, however, that comparisons between the diary and estimated work hours in our studies are far from ideal. First, the workweek estimate questions often do not have the same time referent as the diary survey. Second, the estimate questions used in these diary studies differ from those used in government surveys and vary between each survey. Third, and related to these two points, the diary studies were simply not designed to match with the estimate questions, nor even to elicit precise data on time spent at work—but rather as general purpose procedures to measure time spent on all different kinds of activities, including nonwork activities that might occur during scheduled work time.

Consequently, diary data cannot be expected to capture any fine-grained distinctions in work time. Individual respondents sensitive to these distinctions might report various rest periods or travel during work in their diary accounts, but the diary instructions do not ask respondents to report such episodes at work on a systematic basis. The diary accounts in this analysis are no more sensitive to reporting work activities than to any other daily activity, be it travel, household work, sleep, or TV viewing, which are all subject to the same reporting uncertainties.

A further problem is that the diary data are not available for the week, but for only a single day. That means that we can only construct "synthetic weeks" for groups of respondents by adding together equal proportions of Monday diaries, Tuesday diaries, (and so forth), and weekend diaries to estimate work hours across the week.

In brief, there are many ways in which both the diary and the estimate data fall short of ideal comparison. The studies simply were not designed for that purpose. Nonetheless, we shall examine both measurements to see if the pattern of results is consistent enough to warrant consideration of the diary approach for future, broad-based analysis of workweek measurements.

Methodology

The first step in making the data comparable is to merge the diary data for the 18–64 age group for 1965, 1975, and 1985. That gives a total of more than 7,000 diary respondents across these three studies, a sufficient number of respondents for the following categories of workweeks used in our analyses: 0 (all persons who did no work at all in the survey period, *including* those "with a job, not at work" due to: sickness, labor dispute, vacation); 1–19 hours of work per week (midpoint 10 hours); 20–29 (midpoint 25) hours; 30–34 (32) hours; 35–39 (37) hours; 40–44 (42) hours; 45–49 (47) hours; 50–54 (52) hours; 55–59 (57) hours; 60–64 (62) hours; 65–74 (70) hours; and 75 hours or more. (The midpoints are used as the points in chart 1.)

These 12 categories are the prime independent estimate variable, with which we compare the hours of paid work as reported in the diary. Our dependent variable thus becomes the difference between the two where: *Diff= Estimate–Diary.* Thus, the estimate-diary difference takes on a value of 0 when the two measures of the workweek are identical. Positive values of the estimate-diary difference occur when the average estimated workweek exceeds the amount of work hours reported in the diary over a week's time. Negative values indicate more work hours extracted from the diary surveys than those implied in the estimate response.

In addition to differences in hours between the CPS estimated data and diary data, we examine the differences between the two measurements separately for men and for women. There are differences in reported work hours between men and women in virtually all surveys.

Results

First, a comparison of the CPS distribution of work hours for those working 20 or more hours a week for 1985 with those for the 1985 University of Maryland Americans' Use of Time Project sample shows a similar percent distribution to the estimate question on weekly hours at work:

	CPS estimate question	*University of Maryland estimate question*
Work hours, total	100.0	100.0
20–29	8.8	7.3
30–34	6.3	5.5
35–39	6.6	8.0
40–44	42.4	43.7
45–49	10.3	10.1
50–54	11.8	9.8
55–59	3.4	4.2
60–64	5.9	5.7
65–74	2.9	3.2
75 or more	1.5	2.5

This indicates that the 1985 time diary sample was rather similar to the larger CPS sample in its response to estimated work hours' question to what the larger CPS sample reported.

Table 2 shows the calculations of diary work hours from the aggregated diaries of each level of estimated work hours for 1965, 1975, and 1985. The first column shows the average values of diary work hours across all 3 years combined and weighted equally. It can be seen that values of diary work hours do rise steadily with the estimated hours, indicating that those who estimate more work hours do report more work hours in their diaries, as expected. Moreover, the relation is close to monotonic, rising from 2.8 hours for those estimating no work to 55.2 hours for those estimating 65–74 hours per week, then 54.9 hours for those reporting 75 or more hours per week. That last figure, being slightly lower, provides the only departure from overall monotonicity in this column.

The middle columns of table 2 show that the relation essentially holds for each of the 3 survey years of study, although many more examples of non-monotonicity are found within each year — such as the 46.2-hour figure for those estimating 75 hours or more in 1965 and the 57.9-hour figure for those estimating 55–59 hours in 1975.

Column 5 shows the resulting values of the difference between the estimate and the diary workweek from these comparisons. Because this difference is based on actual hours and not the ranges of estimated hours, the values may be slightly different from what would be calculated from the table itself. Thus, for the 60–64 hour category, the 14-hour value of the difference between the estimated and diary hours is larger than the 11-hour figure that would result from the 50.7-hour value of total diary work hours (1965, 1975, and 1985) being subtracted from the midpoint of the estimate responses (62 hours).

Nonetheless, it is clear that values of the estimate-diary difference do rise as values of estimate responses increase, being –3 hours for the zero-hour category, 2 hours for the 40–44-hour category and 25 hours for the 75 hours or more category. There is a major departure for the 35–39-hour category however, where the 7-hour figure is larger than either 2-hour figure for the 30–34- or 40–44-hour groups. Otherwise, it is clear that values for the difference between the two methods increase as the estimate response increases, indicating greater overestimation among those working more hours, as hypothesized.

Another pattern in table 2 is that values of the difference are lowest in 1965 (1 hour), higher in 1975 (4 hours), and highest in 1985 (7 hours). This suggests that, over the 20-year-period, respondents were becoming progressively more inaccurate in more recent surveys.

Multivariate analysis

To control for third variables that could account for these differences, the data were entered into a multivariate analysis of variance. (See table 3.)[20] The attractive feature of this program is a multiple classification analysis which provides adjustments in different values of categorical variables, based on the statistical contributions of the other variables of interest.[21] In other words, the multiple classification analysis shows the effects of each independent variable if other variables were equal.

In the present analysis, we want to equalize the effects of survey-year differences in years, gender, and days of the week on these different values of the estimate responses. This would ensure that the differences in table 2 are not attributable to disproportionate numbers of weekends, women, or 1985 diary days, for example, in the calculations. Multiple classification analysis also provides results that have been adjusted for such differences.

The multiple classification analysis results shown in table 3 do perform some form of a corrective role. We restrict the sample to those working 20 hours or more, and the first group that shows positive values of the difference between the estimated work hours and the diary hours in table 2. For this group of workers, the total sample value of the difference is about 5 hours a week, both before and after the multiple classification analysis. That is 5 hours lower than the estimated average 43-hour workweek (shown at the bottom of table 2) for this sample and puts

Table 2. **Diary workhours per week by estimated workweek hours, 1965, 1975, and 1985 time diary surveys**

Estimated workweek hours	1965–85 average	Diary hours at work			Estimate diary difference (*DIFF*)	1965–85 diary hour difference
		1965	1975	1985		
0	2.8	0.5	1.7	6.2	–3	5.7
1–19	17.2	20.5	14.6	16.6	–6	–3.9
20–29	24.3	27.1	24.5	21.3	2	–5.8
30–34	30.1	30.9	30.0	29.4	2	–1.5
35–39	30.8	31.6	32.6	28.1	7	–3.5
40–44	38.6	41.3	38.2	36.2	2	–5.3
45–49	44.3	49.8	41.5	41.7	3	–8.1
50–54	44.6	49.9	42.4	41.6	9	–8.3
55–59	47.9	42.5	57.9	43.2	10	0.7
60–64	50.7	55.7	52.1	44.2	14	–11.5
65–74	55.2	57.6	55.1	52.8	15	–4.8
75 or more	54.9	46.2	63.5	54.9	25	8.7
Average estimated workweek (20 hours or more)						
Men	...	47.1	46.5	46.4	...	...
Women	...	39.9	39.2	40.6	...	...

the overestimate closer to 12 percent (43/38–1) for those working 20 hours or more.

The first set of data in table 3 indicate that the basic table 2 results not only hold up, but are slightly enhanced by the multiple classification adjustments for survey year, gender, and day of week. Before adjustment, the estimate-diary difference between 20–29-hour work groups and those working 75 hours or more is 29 hours (29–0), and after adjustment that figure is 31 hours (30-[-1]). That is reflected in the rise of the correlation coefficient (Eta) from 0.17 to 0.19 after adjustment. Also, the "after adjustment" results have fewer irregularities in the progression from 20- to 75-hours-or-more work groups than the "before adjustment" results do. Thus, the multiple classification analysis does provide slightly more consistent evidence to support the hypothesis.

Another example of the effects of the multiple classification analysis shows that the aforementioned increase in the estimate-diary difference between 1965 and 1985 not only is reflected in the 1-hour versus 7-hour gap between 1965 and 1985, but in the slightly reduced 1-hour versus 6-hour difference after the multiple classification analysis adjustment. That statistically significant difference indicates that workers in 1985 were more likely to overestimate their workweeks relative to the diary than were workers in 1965. Among the potential reasons for this increased difference are the increase in service jobs with no fixed hourly schedule, the rise in the amount of flexible work schedules in general, increased pressures for family and personal business during work hours, and the increased blending of work and nonwork time.

Gender differences indicate the presence of a significant gap in workweek estimates. What makes these differences surprising is that women are more likely to overestimate their workweeks, and work shorter workweeks than men (and as noted at the top of table 2, shorter workweeks are associated with lower values of the estimate-diary difference). Thus, the 2-hour gap between men and women doubles to 4 hours after the multiple classification analysis adjustment, again a difference that is statistically significant.

By virtually any measure of work time, women work fewer hours than men on their paid jobs. Thus, although the basic relation remains robust across years and across days of the week of the diary interview, gender is another important correlate of work time.

The data from table 2 have been subdivided into separate figures for men and for women and are presented in chart 1 for the 1985 data. It can be seen that the two lines diverge notably, except in the low to normal workweek categories (namely the 20- to 34-hour workweek categories and the 45- to 49-hour category). Among men and women reporting workweeks less than 20 hours per week (including the zero-hour category of the nonemployed), it is the men who underreport work hours, while women report diary work hours that are rather consistent with their estimated hours. For categories of more than 35 hours per week, however, women's values of the estimate-diary difference become clearly and consistently higher than men's values (with the exception of the 45–49 hour category as noted earlier), particularly past the 50-hour workweek, for which women's values are almost double those of men.

These results could be explained in terms of general traditional role expectations and experiences of men and women. Women are more likely

Table 3. **Difference in estimated versus diary work hours (Diff), by estimated workweek and other selected variables, 1965, 1975, and 1985 data combined**

Variable	*DIFF*	Before multiple classification analysis	After multiple classification adjustment
		DIFF = + 5 hours	*DIFF* = + 5 hours
Estimated work hours per week:			
20–29	(296)	0	–1
30–34	(248)	1	0
35–39	(394)	6	4
40–44	(2,083)	3	3
45–49	(533)	3	4
50–54	(466)	7	8
55–59	(183)	9	9
60–64	(262)	12	14
65–74	(149)	15	14
75 and over	(110)	29	30
Correlation (Eta)		[1].17	[1].19
Year:			
1965	(816)	1	1
1975	(1,305)	4	4
1985	(2,602)	7	6
Correlation (Eta)		[1].07	[1].06
Gender:			
Men	(2,702)	4	3
Women	(2,022)	6	7
Correlation (Eta)		[1].04	[1].08
Day:			
Weekday	(3,336)	–4	–4
Saturday	(720)	23	23
Sunday	(668)	30	30
Correlation (Eta)		[1].52	[1].52

[1] Difference statistically significant at 0.001 level as measured by the correlation coefficient, Eta.

to have had part-time jobs and have more familiarity with the hourly requirements of fitting life around such schedules. When women take on jobs requiring more hours than the usual 40 hours per week, they may still be expected to fulfill other family support roles as well, making it more likely that their actual work hours would be interrupted, irregular, and perhaps shorter than expected, compared with those of men in the same jobs with fewer outside expectations. The more irregular the schedule, the more difficult the estimation task. There are perhaps stronger social expectations that men be employed and not have short-workweek jobs. This might partly explain their larger values for the estimate-diary difference under nonemployment and short workweek conditions. It may also be that their hours are likely to be more variable under such conditions, making the estimation task more difficult. Another explanation is that for some women, having a part-time job (or working less than 35–40 hours per week) might appear to them as being not employed at all, or that they define and see themselves as unemployed for reasons related to "unemployment" benefits. Whatever the reason, it seems inappropriate to assign them to a zero-work category in calculations of work time for entire populations.

The final variable in table 3 is day of the week, with the expected result that weekend diaries significantly underestimate the workweek, while weekday diaries overestimate it. Nonetheless, the day of the week is a crucial variable to control and adjust when making comparisons across groups and survey years.

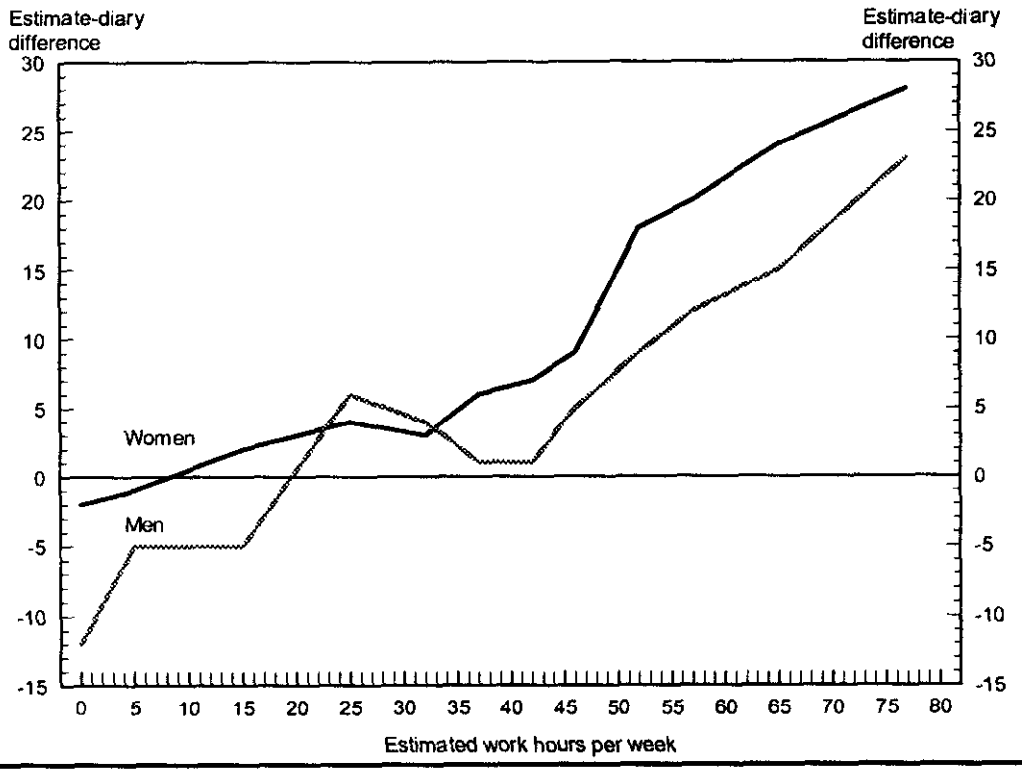
Conclusions and recommendations

We have found systematic and significant deviations from the workweeks that people estimate and the time devoted to work that these same people report in total time diaries that they keep. We interpret the differences between the estimated hours and the diary recorded time devoted to work as reflecting the greater accuracy of the diary or as reflecting the diary's greater detail from which we can extract more precise work activities; and we have presented several explanations for the overestimate of actual time at work, generated from the estimated workweek approach. The gap remains rather robust across the 3 years studied, and thus does not seem to be a function of the way the estimate question was worded, the reference period of the question, the type of time-diary format (personal versus telephone) or the day of the week that the diary was kept. The gap is larger in more recent years, possibly as a result of the increase in service occupations.

Values of the estimate-diary deviation do vary in systematically and approximately linear fashion with the estimated length of the workweek, most clearly in the aggregate across surveys, but also rather systematically within each survey year. Values were negative among those claiming to be unemployed or estimating less than 20 hours of work per week. They were slightly above average for those estimating 20–44-hour workweeks and became progressively higher among those claiming workweeks of 45 hours and higher. Among workers claiming to work more than 55 hours per week, the gap was often more than 10 hours per week, indicating reports considerably above the actual hours worked. We further generally found values of the estimate-diary difference to be higher among women than men.

The generally linear relation of the estimate-diary difference in work hours generally rules out an explanation of the results in terms of simple regression toward the mean, because for that to occur, we should find more curvilinearity in the data. That is, we should also find above-average values of the difference for those working fewer hours (last week). In contrast, the data show below-average values of the difference.

Chart 1. **Values of the difference between estimated and diary work hours for men and women, 1985 data (in hours per week)**



The higher values of the estimate-diary difference among workers who estimate more hours per workweek have important implications, particularly with higher proportions of the U.S. work force in this category.[22] It would appear that simply taking these estimates at face value and averaging them would lead to serious overestimates.

The diary data suggest that only rare individuals put in more than a 55–60-hour workweek, with those estimating 60 or more hours on the job averaging closer to 53-hour weeks. In these high-hour workweek categories, the ratio of estimate-diary difference to actual hours worked is as high as 50 percent. Among those in normal 35–44-hour categories, overestimation is not nearly as high—closer to 10 percent. Indeed, this is a level that could be explained by workers' including their lunch hours or their work commutes as part of their hours of work. Without specific prompting or monitoring by an interviewer, it would not be unreasonable for workers to consider this part of the workweek, especially in the stereotypical 9-to-5 job.

What these results have yet to show is why these differences occur. It is well to remember that the diary data employed in this analysis were not designed nor intended to uncover the discrepancies we have described. Nor were the estimate questions in the CPS originally designed to estimate hours at work. (The CPS questions perform many functions, too numerous to list here.) What the data do clearly demonstrate and support, however, is the need to conduct such a well-designed methodological experiment—one in which questions and interview instructions on the workweek from government surveys would be strictly followed, along with time diaries that would carefully delineate some of the types of work time distinctions of interest to labor analysts.

It is also possible that some improvements to the current collection methods could be feasibly incorporated into government surveys like the CPS, by focusing respondent attention on a shorter reporting period. For example, CPS respondents could be asked to report their work hours "yesterday." For those who did work yesterday, follow-up questions could ascertain when the respondents actually started work, when they took lunch, or other breaks—and if they tended to other nonwork matters between the beginning and ending times of work. They could also be asked about work brought home yesterday, or work done in other nonwork locations.

Thus, this research suggests that a more detailed measurement strategy, like the total time diary method, is useful for capturing the com-

plexity of people's daily worklives. Obviously, more attention is needed on ways to recapture what takes place on a workday. What is recommended now are studies employing observational and diary methods that would provide appropriately complex records against which to measure and understand problems that arise when respondents report their time at work. □

Footnotes

[1] J. Mata-Greenwood, *An Integrated Framework for the Measurement of Working Time.* Working Papers No. 92-2 (Geneva, Switzerland, STAT, International Labour Organization, 1992).

[2] The establishment surveys that collect data on hours at work are the Current Employment Statistics survey (CES, also known as the 790), and the Hours at Work Survey. The list of establishments from which the samples are selected is a composite of State Unemployment Insurance reports for the 50 States and the District of Columbia and covers about 98 percent of employees on nonfarm payrolls in the U.S.

[3] The Current Population Survey collects monthly data from 60,000 households in the United States. The households are selected to represent the U.S. population 16 years of age and older. Personal interviews are conducted on a rotating basis so that three-fourths of the sample is the same for any two consecutive months.

[4] Beginning with data for January 1994, changes have been introduced into the CPS questionnaire and collection methodology. For details on the new CPS questions, see Anne Polivka and Jennifer Rothgeb, "Redesigning the questionnaire," *Monthly Labor Review,* September, 1993, pp. 10–28.

[5] Mata-Greenwood, *Integrated Framework.*

[6] John Robinson, "The Validity and Reliability of Diaries Versus Alternative Time Use Measures," in F. Thomas Juster and Frank Stafford, eds., *Time, Goods, and Well-Being* (Ann Arbor, MI, Institute for Social Research, The University of Michigan, 1985), pp. 33–62.

[7] The study, conducted in 1965 by the Survey Research Center, University of Michigan, had 1,244 adult respondents aged 18–24 to keep a single-day diary of activities, mainly in the fall of that year; respondents living in rural and nonemployed households were excluded. See John Robinson, *How Americans Use Time* (New York, Praeger, 1977) and "Changes in Americans' Use of Time, 1965–1975" (Cleveland, OH, Communication Research Center, 1976). See appendix for more details.

[8] The 1975 study conducted by the Survey Research Center, University of Michigan had 1,519 respondents aged 18 and older to keep diaries for a single day in the fall of that year; see Robinson, "Changes in Americans' Use of Time." See appendix for more details; in addition, diaries were obtained from 788 spouses of these designated respondents. See Juster and Stafford, eds., *Time, Goods, and Well-Being.* Further details are listed in the appendix.

[9] Alexander Szalai and others, *The Use of Time* (The Hague, Netherlands, Mouton, 1972).

[10] Robinson, *How Americans Use Time.*

[11] Andrew Harvey and David Elliot, *Time and Time Again* (Ottawa-Hull, Canada, Employment and Immigration Commission, 1983).

[12] R. Bechtel, C. Achepohl and R. Akers, "Correlates Between Observed Behavior and Questionnaire Responses, in Television Viewing," in E.A. Rubenstein, G.A. Comstock, and J.P. Murray, eds., *Television in Day-to-Day Life: Patterns and Use* (Washington, DC, U.S. Government Printing Office, 1972).

[13] Charles Allen, "Photographing the TV Audience," *Journal of Advertising Research,* 1968, 8(1), pp. 2–8; and Bechtel and others, "Correlates Between Observed Behavior."

[14] Robinson , "The Validity and Reliability."

[15] *Ibid.*

[16] Thomas F. Juster, "The Validity and Reliability of Diaries versus Alternative Time Use Measures," in F. Thomas Juster and Frank Stafford, eds., *Time, Goods, and Well-Being* (Ann Arbor, MI, Institute for Social Research, The University of Michigan, 1985).

[17] Daniel Hill, Implications of Home Production and Inventory Adjustment Processes for Time-of-Day Demand for Electricity," in F. Thomas Juster and Frank Stafford, eds. *Time, Goods, and Well-Being.*

[18] Gershuny and others, "Time Budgets: Preliminary Analyses of a National Survey," *Quarterly Journal of Social Affairs,* (2), 1985; and William Michelson, *Public Policy in Temporal Perspective* (The Hague, Mouton, 1978).

[19] Overestimation of the 40-hour week also arises when holidays, sicktime and other leave benefits might not have been recalled or precluded from the report. Because of true differences in the occurrence of lost and extra hours, this yields a net overestimate in the hours worked. See W.C. Edwards, R. Levine, and S. Cohany, "Procedures for validating reports of hours worked and for classifying discrepancies between questionnaire reports and validation totals, *Proceedings of the American Statistical Association: Section on Survey Research Methods* (Washington, DC, American Statistical Association, 1989), pp. 496–501.

[20] The Analysis of Variance program of Statistical Packages for the Social Sciences for Personal Computers, Norius, 1985.

[21] For more information, see F. Andrews, J. Morgan, J. Sonquist, and L. Klem, *Multiple Classification Analysis* (Ann Arbor, MI, Institute for Social Research, 1973).

[22] Juliet Schor, *The Overworked American* (New York, Basic Books, 1991).

APPENDIX: Background of the studies

Methodology of the 1985 study

The 1985 Americans' Use of Time study employed the same basic open-ended diary approach as the 1965 and 1975 national studies. In the 1985 study, however, an explicit attempt was made to spread the collection of diary days across the entire calendar year—from January through December of 1985. Data were retrieved from three survey modes using the diary and estimated workweek questions.

Mail-back sample. The data for the main (mail-back) study were collected from a sample of Americans who were first contacted by telephone, using the random-digit-dial method of selecting telephone numbers. All calls were made from the central telephone facility at the Survey Research Center of the University of Maryland, College Park.

Once a working-telephone household was contacted, one respondent aged 18 and older in each

household was selected at random. That person was given a brief (2–5 minutes) orientation interview, followed by an invitation to participate in the diary/mail-out part of the study. If that respondent agreed, diaries were then mailed out for each member of the participating household aged 12 and older to complete for a particular day in the subsequent week.

Followup calls interviews were made 4 to 6 days later to ensure that respondents had received these materials and understood how to complete them. After respondents had completed these diaries, they then mailed all their completed forms back to College Park for coding and analysis. Some 3,349 diaries from 99 households were returned using this mail-out procedure during the 12 months of 1985. It is the diaries obtained from adults aged 18 and older, however, that form the data base for the analyses described in this article. Other 1985 data included parallel diary data from 809 additional respondents interviewed in a separate personal interview sample in the fall and winter of 1985–86, and from an additional 1,210 "yesterday" diaries obtained by telephone as part of the initial contact for the mail-back diaries.

Collection of the mail-back data, then, was obtained using basically the same "tomorrow" approach as employed in the 1965–66 study. The main procedural difference was that a personal interviewer was not present to check on the adequacy of diary entries. This check was instead done by telephone as soon as the diaries had been mailed back for coding and analysis. If any discrepancies were detected (for example, significant gaps of missing times or indecipherable diary entries), the respondent involved was recontacted by telephone to clarify any ambiguities.

Households were given special monetary incentives and gifts (a pen with a digital watch) to ensure that other family members in the selected households participated in keeping a diary. This also ensured that the sample would be approximately self-weighing, as well as covering approximately an entire year's activities.

In addition to the estimates of daily time use from the diary, the study also obtained information on the employment status, age, education, race, and gender from each member of the household. Additional questions ascertained the stock of certain technology available in the household, as well as certain physical characteristics of the dwelling unit.

The sample was designed to represent all telephone households in the coterminous United States. The sample first covered 173 area codes/three-digit prefixes selected at random from a master random-digit-dial sample frame of five base numbers prepared by the Sampling Department of the Institute for Social Research at the University of Michigan. If that base number located a working household telephone number, it was then used to generate additional clusters of random numbers within that area code and prefix. The initial list of 500 numbers had been stratified by geographical region of the country. That ensured that the sample telephone numbers had an adequate representation from all regions of the country. The sample was designed to yield about 1,800 households (and 4,000 individuals) during the calendar year.

Telephone sample. The telephone sample consisted of the random sample of the population who were contacted in the first phase of the random-digit-dial sample. This consisted of the randomly selected adult (aged 18 or older) who responded to the first interview. Some 67 percent of respondents contacted by telephone, however, did complete a day-before diary over the telephone. This was the highest response rate for any of the three data collection modes.

Personal sample. In addition to the mail-back and telephone diaries, a separate national sample of 809 diaries were collected by personal in-home interviews. This sample was drawn from a subset of 20 primary sampling units developed by random probability methods for the continuing national samples of the Institute for Survey Research at Temple University in Philadelphia. That stratified sample was further stratified and subjected to a "controlled selection" to ensure that the subset of 20 primary sampling units retained sufficient representation by rural-urban character within each of the four regions of the country.

Respondents in this sample were asked to follow much the same procedures as for the initial telephone sample. One adult selected at random was asked to complete a retrospective diary from memory for the previous day. The interviewer then left diaries for all adult respondents in the household to complete for the following day. The interviewer then returned the day following the initial survey day to collect the diaries and to ensure they were filled out adequately and accurately. For example, if the interviewer contacted the household on a Tuesday, the random adult respondent first filled out a retrospective diary for Monday; the interviewer then left diary forms for that respondent and other household adults to fill out for Wednesday, and the interviewer returned to collect those completed forms and ask additional questions about the household on Thursday. As in mail-back diary procedure, respondents were given monetary and other incentives for participating.

Diary coding. In the 1985 time-diary form each respondent is expected to write out each *primary*

activity in which they engaged, the time that the activity began and ended, where it took place, who was present during the activity, and what other activities were performed during this same time. In this way, the diary form remained basically the same as that used in the 1965 and 1975 studies.

To illustrate to respondents the types of activities and level of detail expected of them to complete diaries, an example of a complete diary form was enclosed in each packet mailed to the household (or left behind during the home visit). The example form was filled out in considerable detail, with several hand-written comments by the presumed "diary keeper" to help the interpretation of unusual diary entries (for example, going home during work; caring for children while playing sports). In general, the example form was intended to ensure that respondents would include enough detail in their diaries; this seemed successful because mailed-back diaries contained about the same number of primary activities (about 25 per day) as found in the 1965 "tomorrow" diaries.

Once received and checked, these diaries were then entered on a computer by trained coding staff, using the direct data entry features of the University of California at Berkeley Computer Assisted Telephone Interviewing (CATI) system. Activities were coded into one of more than 250 activity codes elaborated from the 174 categories developed at the Universityof Michigan for the 1975 data; this in turn represented an elaboration of the 96 basic code categories that had been developed for the 1965 Multinational Time-Use Project.[1]

The University of Maryland used the same complete document of coding conventions that were developed by the Survey Research Center at the University of Michigan for its 1975 time diary project. Each activity in the diary was coded descriptively as a separate block of 21 digits in length. This comprised the primary activity (a three-digit code) during the period, the time the activity began and ended (each coded in 4-digit military time, for example, 8 AM = 0800; 8 PM = 2000) location (1 digit), social partners (2 digits), secondary activity (3 digits), enjoyment level (1 digit) and media use (3 digits). When this 21-digit entry for all activities in the diary was entered and computed, the totals were programmed into the machine to ensure that each day's diary entries added to exactly 1,440 minutes (24.0 hours). These "variable-field" data (that is, varying depending on the number of activities reported) were then processed by a special computer program that generated "fixed-field" compilations of diary time for each of the 94 activities across the day, that is, total daily minutes spent working, cooking, watching TV, and so forth, for that respondent for that day.

The averages of these fixed field totals are presented in the analytic tables in this article. The weekly hour data in the tables are also weighted slightly by day of the week and by five major demographic factors (education, sex, race, marital status, and work hours) to ensure that all days of the week are equally represented and that the overall sample figure corresponds with the U.S. Bureau of the Census figures for these important demographic variables.

Methodology of the 1975 national study

The data for the 1975 study were collected from a sample of Americans interviewed in person during October-November, 1975, as part of the 1975 fall omnibus study conducted by the Institute for Social Research at the University of Michigan. The respondents in the 1975 omnibus were chosen to form a representative sample of American adults 18 years of age and older, living in the coterminous United States. As part of the time-use measurement effort, spouses of the respondents were interviewed as well.

The sample was designed to represent units in the coterminous United States exclusive of those on military reservations. The 74 sample points, located in 37 States and the District of Columbia, included the New York-Northeastern New Jersey and the Chicago-Northwestern Indiana consolidated areas, the 10 largest standard metropolitan statistical areas (SMSA's) outside of the two standard consolidated areas, 32 other SMSA's, and 30 counties or county groups representing the nonmetropolitan and less urban portions of the country. In this multi-stage area probability sample, first-stage stratification of SMSA's and counties was carried out independently within each of the four major geographical regions—Northeast, North Central, South and West—each of which received representation in proportion to its population. Probability selection was enforced at all stages of sampling; the interviewers had no freedom of choice among housing units or among household members within a sample dwelling.

Data processing. The data were obtained in the field through personal interviews. In processing the data, several innovations were used which added to data analysis capabilities. Telephone reinterviews on the study were conducted from Ann Arbor rather than from the field, allowing better quality control over the conduct of the interview.

Considerable use was made of direct data entry capabilities, enabling the staff to bypass the preparation of code sheets and punch cards. Time

diaries for the respondent and spouse were entered on line to direct access magnetic disk. Information from edited diaries was entered via computer terminal with standard check-coding procedures being performed at this time. For the time diaries, certain consistency checks (that is, ending time of one activity must be compatible with starting time of the next activity; all activities must add up to 1,440 minutes) were built into the entry program.

Methodology of the 1965 study

The study involved a sample of over 2,000 American adults aged 18–65 who kept complete diaries of their activities for a single day—mainly between November 1 and December 15, 1965, but also in the winter and spring of 1966. The sample was deliberately chosen to be an urban and employed one, conforming to the guidelines of the multinational study of which it was a part.[2] Thus, residents of non-SMSA's (namely, areas with cities of fewer than 50,000 persons) were excluded, as well as residents of households in which no member aged 18–65 was part of the labor force; and farmers. Respondents were randomly assigned to fill out diaries on a weekday or on a weekend.

Of the total sample, 1,244 adults were part of the national urban sample; another 788 came from the city of Jackson, MI, and its environs. The Jackson data are not reported here.

The field procedures involved the "tomorrow" approach, that is, the interviewer contacted the respondent and conducted a brief "warm-up" interview on the first day and left the diary for the respondent to enter the next day's activities. The interviewer returned to the respondent's home on the subsequent day (that is, the day after "tomorrow") to ensure that the diary had been filled out correctly and to fill in any missing parts.

When the diaries were returned to the Survey Research Center, they were edited to ensure completeness and consistency. Missing time periods were noted, as well as trip estimates where these were not pointed out by respondents. Primary activities were coded into one of the 96 activity categories.[3] These durations were then summarized, and deviations of greater than 10 minutes from the 1,440 minutes total were noted and the diaries recorded to be within that 10-minute limit. Deviations of less than 10 minutes were added to or subtracted from the activity of maximum duration, that activity usually being sleep.

Comparison of the studies

All three studies were based on strict probability sampling methods across the Nation. Only the 1985 study was spread across the entire year. Moreover, the 1985 national data were mainly collected by prospective mail-back diaries, while the 1975 study employed the retrospective recall of activities done "yesterday." The 1965 and 1975 studies had somewhat higher overall response rates (72 percent each), although not much higher than the telephone portion of the national study (67 percent). The 1985 study had more than twice the number of adult respondents over age 18 than the 1975 study (n=5,358, versus 2,409).

The 1985 national study was more evenly spread across the year and across days of the week, while the 1975 study oversampled Sundays and undersampled Saturdays. All studies used open-end diary entries across the full 24 hours of a single day and the same basic code for diary activities—although the 1975 and 1985 studies employed more than twice as many activity codes.

Footnotes to the appendix

[1] Alexander Szalai and others, *The Use of Time* (The Hague, Netherlands, Mouton, 1972).

[2] *Ibid.*

[3] *Ibid.*