DRAFT 11: FOR COMMENT ONLY

Workweek Estimate-Diary Differences and Regression to the Mean

Jonathan Gershuny and Kimberly Fisher, University of Oxford
John P. Robinson and Steven Martin, University of Maryland

June 2007

ABSTRACT

Using data from the 2003-05 American Time Use Study (ATUS), we replicate earlier results suggesting that "stylized" questionnaire time estimates consistently overestimate the time employed men and women spend doing paid work. We employ data from the Multinational Time Use Study (MTUS) to produce analogous results from six other Western countries. Drawing on diary studies from the UK, which contain a diary-type "work grid" (similar to the day diary but covering a continuous seven-day period), we find an asymmetry in the joint distributions of the two sorts of weekly work time measurement, with stylised questionnaire estimates more likely to exceed work grid-based estimates than *vice versa*. We then show that the "gap" between the diary or gride and estimate questions can be partly explained by the irregularity of the workweek (a phenomenon that cannot be directly observed in the US data), and the consequent difficulty that survey respondents face in answering stylized estimate questions. We conclude that differences between stylized estimate questions and diary-type measures of work time cannot be explained simply in terms of "regression to a mean".

Acknowledgement: Special appreciation is expressed to the US National Science Foundation for supporting the survey research that formed the basis for the present study, and to the UK Economic and Social Research Council which supported both the collection and the analysis of the work grid and diary data discussed here.

1 INTRODUCTION

Most evidence about the allocation of time to varying activities comes from "stylized" survey estimate questions that ask respondents to estimate how much time they spend on an activity during a particular period, usually a week or day (typically "last week" or "yesterday"). Examples include "How many hours a week do you typically spend working ?", or "How many hours a day do you usually watch television?" There is a rich body of historical data from American national samples that relies solely on the time-estimate approach - on time spent in paid work (from the Current Population Survey (CPS)), doing voluntary work (from the Independent Sector and other not-for-profit organizations), traveling (from the Census Bureau and the U.S. Department of Transportation), and watching television (from the Roper Organization and the General Social Survey). Putnam (2001) used a number of such questions to document his arguments about declining social capital in America.

The most widely-cited time estimates of US market work hours come from the Current Population Survey (CPS) (http://stats.bls.gov/cps/home.htm), where respondents report stylized estimates of how many hours they worked last week, in addition to estimating their "usual hours" of paid work per week. The CPS questions (similar to those used by central statistical agencies in other countries) are usually considered the "gold standard" for assessing the extent and changes in the work patterns of men and women. One of the great advantages of CPS-type estimate questions is that they are asked of very large samples with high response rates. The CPS surveys cover all workers in around 50,000 households every month across the full 12 months of each year. These market work estimates also have been tracked over a very long time period, extending back more than four decades. The CPS data thus make it possible to examine detailed

breakouts of work hours by gender, by marital status, by presence and ages of children, among other personal and household characteristics.

These estimate questions have drawbacks. Recalling details about time spent in an activity involves complicated calculations. Asking someone "How many hours per week do you usually work?" (or alternatively "…..did you work last week?") assumes that each respondent: interprets "work" the same way, searches memory for all episodes of work, chooses an appropriate sample of "usual weeks" to arrive at an appropriate "normal" week, and is able to properly add up and average all the episode lengths across the chosen set of weeks or across days in the last week. Obtaining completely accurate responses regarding time use becomes particularly difficult in the survey context, in which respondents are expected to provide on-the-spot answers to such questions in a few seconds. What seems at first to be a simple estimate task turns out to involve several steps that are quite difficult to perform, particularly for a respondent with irregular and unclear work hours and no established daily routine.

Time-diary instruments offer an alternative measurement method. Time diaries do not require respondents to make complex, vague or changing calculations about activities taken out of the context of the lived experience in which they were performed. Rather, diaries require the recall of specific activities in their daily sequence and in fuller context (who else was present, where activities took place) over a specific period in the recent past (usually the previous day), though sometimes more contemporaneously. In that way, it becomes possible to reduce the respondents' recall period and reporting task, in order to cover all daily activity and to ensure that the resulting account respects the "zero-sum" property of time -- in that respondents' daily activities must sum to the full 24 hours in a day. A full discussion of the data that underpin this paper is available in the documentation of the American Heritage Time Use Study (Fisher et. al.

2006, see www.timeuse.org/AHTUS), as well as in Fisher et. all. (2007). In addition to the AHTUS documentation, Gershuny (2000), Robinson and Gershuny (1994) and Robinson and Godbey (1997) further illustrate the advantages as well as the basic reliability and validity of the diary method. Variously collected diary accounts tend to produce results generally consistent with each other and with other ways of collecting time data by observation (e.g., "shadow" studies, on-site observation, or "beeper" studies, in which respondents report their activity at random moments during the day when a beeper goes off) (Gershuny 2000, Robinson and Godbey 1997; Kan and Pudney 2007).

That is not to say that the diary method is without flaws. Respondents can revise or distort accounts of their activities. When unable to recall their exact actions at a particular moment, some may well substitute activities they habitually engage in ot that would impress the interviewer. The method is also rather demanding of interviewer and respondent time, although survey respondents may well enjoy the task of recalling their own daily activities and accounting for where their time actually goes.

As much as an analyst might wish for fuller or more verifiable accounts of activity not based solely on self-reports, and more satisfactory ways of accounting for behavior (as perhaps when activities can be recorded by unobtrusive sensors or global positioning systems), the diary still presents us with a far richer and more persuasive estimates of activity than any presently available alternative (Michelson 2005).

Estimating Time in Paid Work

Some findings relating to paid work time in the USA drawn from time diaries challenge some existing beliefs, such as research that has shown that time-diary estimates of paid work

hours are typically lower than estimates derived from the CPS (Robinson and Bostrom 1994; Robinson and Godbey 1999).The GAP has usually defined as follows:

GAP = Grouped stylized weekly worktime estimate – (Daily diary work hours*7),

where each day of the week is equally represented. In part, this gap may arise from the assumptions the different stylized questions require respondents to make. Diaries constrain the reported work day within the actual 24 hours of the (usually previous) day, while time estimates may be inadvertently inflated to the length the hours may *seem* to take, or that employees are contracted or expected to work.

These hours may for example overlap with other non-work activities, reflect excessive stress during work hours that may inflate them subjectively, or anticipate the start-up of a paid work episode may be added to the duration of the episode itself. There is no reason why parallel subjective experiences might not *shorten* the estimated work week—but our *a priori* expectations (which turn out to be consistent with the empirical evidence in previous studies and presented below) suggest that the net effects of such processes are on balance to overestimate paid work time. These estimate-diary gaps do seem to gaps vary by the length of the working schedules of respondents. Workers estimating the "more normal" range of 35-45 hour work weeks report relatively similar estimated and diary total hours of work Greater gaps emerge for people reporting longer work days and weeks (Robinson and Bostrom 1994; Robinson and Gershuny 1994), with workers estimating 60 to 80 hour work weeks showing the greatest gap,

suggesting a tendency of stylized time estimates to follow the pattern of "The higher the estimate, the greater the overestimate."[1]

More recently, Jacobs (1998, 2004) challenged the notion of inaccurate estimates, arguing that the gap was a result *simply* of the familiar "regression to the mean" phenomenon, and he produced statistical models that accounted for these gaps. Using more recent data from the ATUS survey which has now collected national diary data from more than 45,000 respondents, Frazis and Stewart (2004) found no significant difference between diary data and work estimate questions, also arguing that any gaps might result from regression to the mean. Bonke (2005) observed that in Danish data, diaries appear to offer lower and potentially more accurate hours, although the difference was slight enough for Bonke to conclude that the cost of diary surveys might not justify its increased accuracy.

Nevertheless, other research comparing estimates and diaries supports the conclusion that the estimate-diary gaps are endemic. Plainly, as Jacobs, Frazis and Stewart, Bonke and others contend, regression to the mean must play some role in generating the gaps. But it is less clear that the gaps can be entirely explained in this manner. Moreover, the key question remains—wherever the mean values of the diary and of estimates differ—that one is still left with the issue of *which is the more appropriate mean*. If the means of the two regularly show lower diary

[1] Similar (but more serious) overestimates are found with estimated time on the other "productive" activity of housework. Both Marini and Shelton (1993) and Press and Townsley (1998) found notably lower times on various housework tasks, like cooking and cleaning, in national time diaries than when time estimate questions from the 1984–86 National Survey of Families and Households (NSFH) were asked. These questions are of particular interest because they deal with unpaid work in society, which is also a "productive" area of daily activity with considerable economic importt, and because these NSFH questions had been extensively analyzed in the family studies literature to take this activity into account. There is an extensive European literature which finds gaps, but not as dramatic. While both the Marini and Shelton and Press and Townsley studies had to depend on data from separate time-diary and time-estimate surveys, a more recent 1998-2001 national diary study described in Bianchi et al. (2006), both the time-estimate questions and daily diary data were collected from the *same* respondents, making it possible to show that the discrepancy was not simply a result of other confounding factors.

means of paid work time than stylized question means, regression to *the* mean cannot be an adequate ground for dismissing the diary-estimate gap.

2 PRIOR RESEARCH EVIDENCE ABOUT THE GAP.

Among the arguments in the previous literature for the origin and prevalence of the gap are:

1) Respondent estimates across all, or most, activities overall sum to more than 168 hours per week. Some studies have asked respondents to estimate times on rather complete lists of daily activity. When asked to provide such daily and weekly estimates of several activities, survey respondents tend to give estimates that add up to considerably more than the 168 available hours of weekly time (e.g., Hawes, Talarzyk, and Blackwell 1975; Verbrugge and Gruber-Baldine 1993). In a similar way, Chase and Godbey (1983) asked members of swimming and tennis clubs in State College, Pennsylvania, how many times they had used the club during the last 12 months and checked their responses against the sign-in system each club had. In both cases, almost half of all respondents overestimated the actual number of times they participated by more than 100 percent.

2) The gap in work hours is found in many other countries. Robinson and Gershuny (1994) found consistent over-reporting of paid work hours by employed people, not only in the USA but also in ten other Western countries. More recently, it was also observed in newer diary studies conducted in the non-Western countries of Russia, China and Japan.

Figures 1a (for men) and 1b (for women) from the Multinational Time-Use Study (MTUS) update information from repeated comparable surveys across successive decades for seven European and North American countries between 1998 and 2003). While these figures

Figure 1a: Gap in estimated hours/week: men

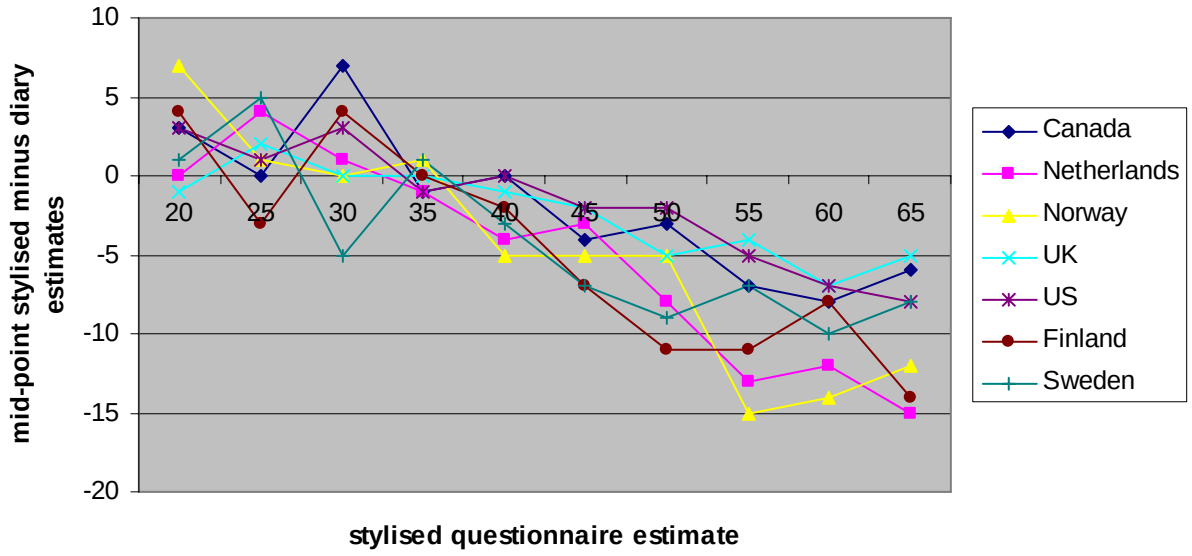
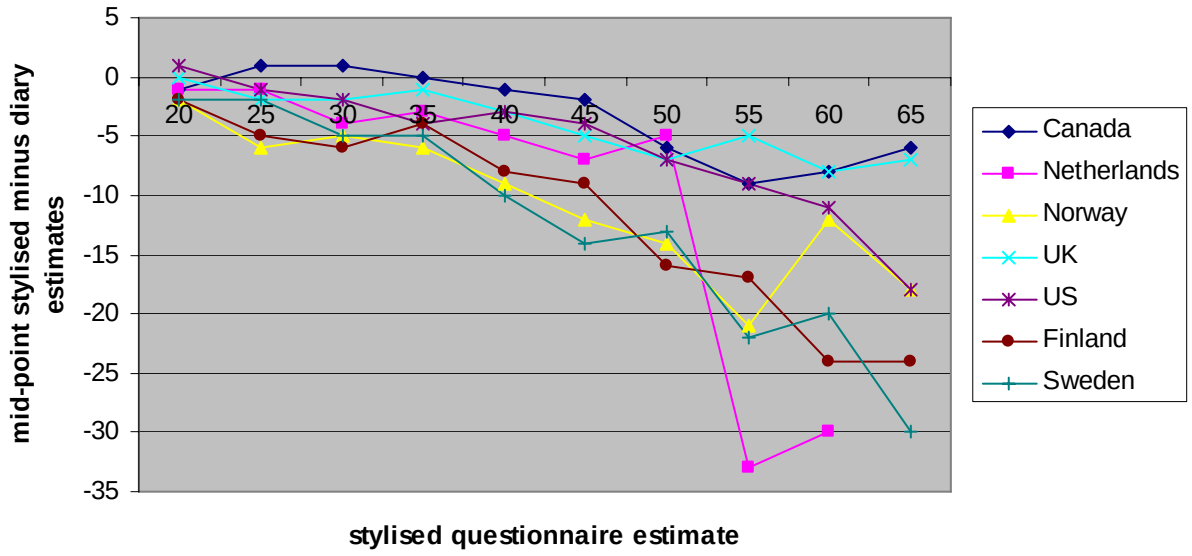


Figure 1b: Gap in estimated hours/week: women



show considerable variation in the extent and monotonicity of the gap across countries, it is clear that the overall pattern is maintained, namely one of underestimates for lower estimated workweeks, lowest discrepancies for those with more "normal" workweeks (30-50 hour weeks) and increasing overestimates for those estimating increasingly longer workweeks. The variations for women are greater than for men, with greater overestimation, particularly for those working more than 50 hours (here possibly due to smaller sample sizes for women working longer hours). Nonetheless, with the exception of Canada (for women) and Finland (for men), their overall pattern tends the follow the "greater estimate-greater overestimate" rule,

3) *The gaps show some increase during recent historical time.* Robinson and Bostrom (1994) observed the discrepancy in the first 1965 national United States diary study, but its small magntude resulted in few initial analysts commenting on the gap. However, over time, this gap has tended to increase, although hardly at a constant rate. In the 1965 US study, the gap was only 1.3 hours, but it rose in 1975 to 3.6 hours and in 1985 to 6.2 hours. In 1993-95 diaries, the gap then decreased to 2.7 hours, but it rose again in the 1998-2001 national diary studies to 3.7 hours (Robinson and Bostrom 1994; Bianchi et. al. 2006).

Figures 2a and 2b, again derived from the MTUS archive, provides supporting evidence of parallel increases in the gap in more recent diary studies conducted in Europe and other Western countries. The increasing gap is particularly evident for those reporting longest work estimates, who can of course also skew the overall estimates of workhours for the entire workforce. One possible reason for the increasing discrepancy in Figure 2 to be explored here (see Figure 3 below) may be the greater variety and irregularity of work hours as the service sector increasingly dominates paid employment.

Figure 2a: stylised vs diary estimates of weekly paid workhours, MTUS men

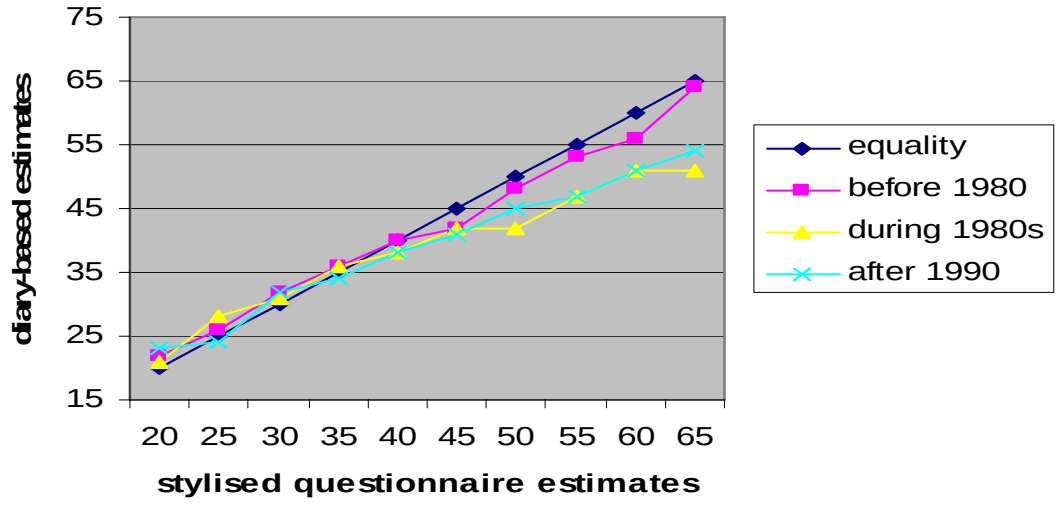
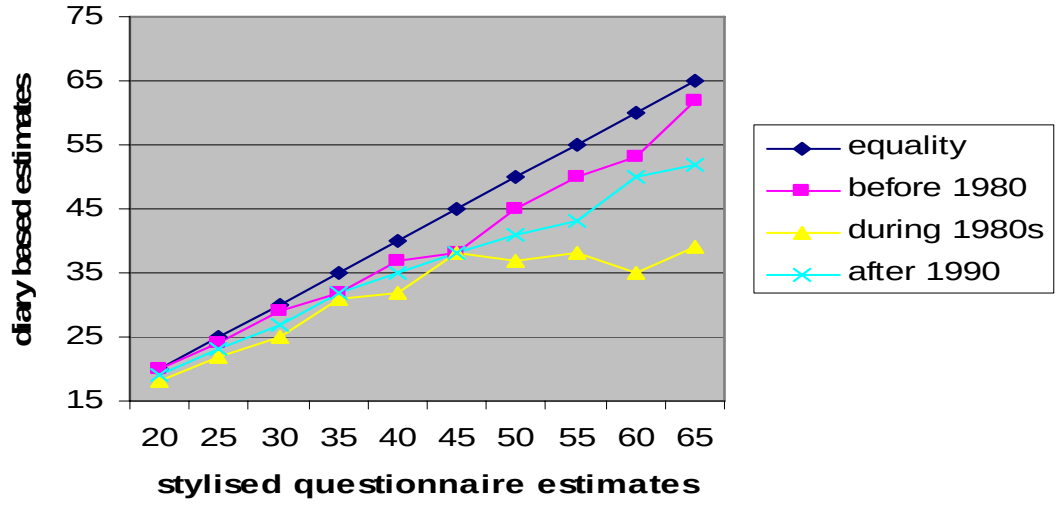


Figure 2b: stylised vs diary estimates of weekly paid workhours, MTUS women



The historical change in the magnitude of the discrepancy reduces the plausibility of explanations simply based on "regression to the mean". We hypothesize that this growth may instead reflect the progressive movement of the labor force either into service occupations (with more irregular work hours having no "time clock" to punch or other concrete memoty aid) into self-employment, or into a "portfolio" of intermittent or part-time jobs. Each of these workplace changes and developments provide workers with fewer convenient temporal benchmarks to estimate their workweek accurately.

4) There are <u>at least</u> two separate means – one for (different forms of) stylized estimates, vs. another for diaries. While there is legitimate concern over different wordings of stylized work estimate questions (as between "last week" vs. "a usual week"), there nevertheless appears to be a systematic relationship between the various sorts of estimate-based means and the time diary mean, as shown in Table 1 from the 2003-2005 ATUS It can be seen that figures from three separate estimate questions: 1) "Usual hours" in the ATUS (variable TEHRUSLT), 2) "Usual hours" in the earlier Wave 8 CPS study (PEHRUSLT) and 3). "Actual hours last week " also in the Wave 8 CPS study (PEHRACTT) vary by less than an hour and a half a week.

In contrast, the diary work figures from different years also tend to vary little, usually by less than an hour a week. Indeed, these estimate vs diary comparisons show virtually no overlap across years in Table 1, with the diary figures consistently being one to five hours lower than the estimates. In the Table 1 figures for all workers, for example, three different estimate questions produced means of 40.5, 40.1 and 39.4 hours, while the diary work time for the year 2003 study (not shown in Table 1) was 36.9,hours, for 2004 35.9 hours and for 2005 36.1 hours -- the three again being within an hour of each other. Kan and Pudney (2007) similarly found evidence that estimates of time in housework and housework time reported in diaries in the UK may also

Table 1: Estimated vs. Diary Hours at Work: 2003-05 ATUS Data

		0+ WORK HOURS			20+ WORK HOUR			35+ WORK HOURS		
WOMEN	Estimate	Estimated	Diary	*Est-Diary*	Estimated	Diary	*Est-Diary*	Estimated	Diary	*Est-Diary*
	Usual hours from ATUS	36.8	32.6	+4.2	39.7	35.1	+4.6	43.2	37.9	+5.3
	Usual hours from CPS	37.2	32.6	+4.6	39.2	35.1	+4.1	42.1	37.3	+4.8
	Actual hours last week, CPS	36.1	32.6	+3.5	38.8	35.1	+3.7	42.8	37.7	+5.1
MEN										
	Usual hours from ATUS	44.3	40.4	+3.9	45.8	41.5	+4.3	47.4	42.7	+4.7
	Usual hours from CPS	43.1	40.4	+3.7	43.8	41.5	+2.3	44.9	43.4	+1.5
	Actual hours last week, CPS	42.4	40.4	+2.0	44.1	41.5	+2.6	46.3	43.4	+2.9
ALL										
	Usual hours from ATUS	40.5	36.7	+3.8	42.7	38.3	+4.4	45.5	40.5	+5.0
	Usual hours from CPS	40.1	36.7	+3.4	41.5	38.3	+3.2	43.6	40.2	+3.3
	Actual hours last week, CPS	39.3	36.7	+2.6	41.5	38.3	+3.2	44.7	40.8	+3.9

generate different means.

3 COMPARING DIARY AND WEEKLY "WORK GRID" ESTIMATES.

The central problem that has thus far prevented serious progress in this area, is a shortage of weekly diary data to compare directly with the weekly estimates. Virtually all of the analysis has focused on stylized questions about weekly work hours from samples of individuals who also complete time diaries for a single day. Because the sampling is randomized across the days (and indeed the appropriate representations of days of the week can be ensured through weighting), the two sorts of estimates are compared in Table 1 and Figures 1-2, by grouping respondents according to their questionnaire responses, and then setting the mean diary work-times for each group against the appropriate mid-point or mean of the range for that group (e.g., 30 hours for those estimating between 27.5 and 32.4 hours). Ideally we would wish to compare the work-week estimates with week-long diaries, which would enable a more symmetrical treatment of the competing work-hours measurements.

Week-long time diaries (and the similar weekly "work grid" measures, first described in Marsh 1987 and in Chenu and Robinson (2003) more recently) are indeed rare, and may be subject to increased refusal and other and other non-response biases due to respondent burden, but they do exist. We find that the same sort of Figure 1 gap emerges from the week-long UK grid measures, (as well as in weeklong diaries collected in the UK, the Netherlands and elsewhere). Respondents in the UK Harmonised European Time Use Study (UK-HETUS) were also asked to complete the "work grid", which detailed for each quarter hour of each day of a week, whether or not they were working at a paid job. The UK Office of National Statistics 2001 HETUS microdata are used here. (Chenu and Robinson 2003 have analyzed the data from the similar French HETUS weekly work grid, which produced similar results to what follows.)

When the UK-HETUS diary and work grid work-hours totals are compared with the UK-HETUS "hours worked last week including overtime" stylized estimate question, they again exhibit the same general Figure 1 "diary-estimate gap" for the single day-based UK-HETUS diary totals plotted against the UK-HETUS questionnaire estimates, and the week-based UK-HETUS work-grid totals plotted against the UK-HETUS stylized estimates.

Thus, if the regression to the mean argument were to provide a definitive or even effective explanation of these gaps, then the cross-tabulation of the stylized estimate against the weekly diary or grid estimates in Table 2 (as estimated from the UK-HETUS study, with both estimates grouped into the same five-hour intervals) should produce a symmetrical joint distribution around the major diagonal. However, the joint distribution in Table 2 is clearly quite *asymmetrical*, with 46% of the stylized/schedule based pairings above the major diagonal (i.e., cases where the stylized estimate substantially exceed the schedule estimate) whereas only 26% lie below the diagonal[2].

[2] The equivalent cross-distribution of the stylized work time estimates from the 1999-2001 UK Home on Line (HoL) against the HoL diary work time totals shows a similar (though much less extreme) non-symmetrical pattern around the major diagonal.

Table 2 Joint distribution of HETUS questionnaire and weekly work schedule hours (UK 2001) (% of entire sample)

	Grouped stylized work hours →													
	<=20	25	30	35	40	45	50	55	60	>=65 Hrs.	Sum	*Stylized under*	*Stylized over*	Ratio
Grouped schedule work hours														
<=20	***9.4***	1.7	1.3	2.1	2.5	1.2	0.7	0.2	0.6	0.1			*10.5*	
25	1.7	***1.4***	0.8	1.0	1.4	0.8	0.4	0.3	0.3	0.0		*1.7*	*5.0*	**0.3**
30	0.8	0.8	***1.2***	2.4	2.4	1.1	0.8	0.5	0.2	0.1		*1.7*	*7.5*	**0.2**
35	0.6	0.4	0.9	***4.4***	5.3	1.8	1.1	0.3	0.4	0.1		*1.9*	*9.0*	**0.2**
40	0.7	0.4	0.5	3.3	***6.1***	3.3	1.7	0.8	0.6	0.3		*4.9*	*6.6*	**0.7**
45	0.5	0.1	0.2	2.0	3.5	***2.4***	2.1	0.9	1.1	0.2		*6.3*	*4.4*	**1.5**
50	0.2	0.0	0.2	0.3	1.4	1.2	***1.7***	1.1	0.8	0.2		*3.4*	*2.0*	**1.7**
55	0.2	0.1	0.0	0.2	0.8	0.6	0.7	***0.4***	0.6	0.1		*2.5*	*0.7*	**3.7**
60	0.1	0.0	0.0	0.2	0.4	0.2	0.7	0.2	***0.8***	0.1		*1.7*	*0.1*	**14.3**
>=65	0.0	0.0	0.1	0.0	0.4	0.3	0.4	0.3	0.5	***0.2***		*2.0*		
Sum												=26.1	=45.8	
Grid. Under		1.7	2.2	5.5	11.7	8.2	6.8	3.9	4.5	1.3	=45.8			
Grid Over	4.8	1.9	1.9	6.0	6.4	2.4	1.7	0.5	0.5		=26.1			
Ratio		**0.9**	**1.1**	**0.9**	**1.8**	**3.5**	**4.0**	**7.8**	**9.0**				(N	6755)

4. Modelling the stylized estimates

One can advance various hypotheses to explain the systematic differences between stylized and diary work time estimates. First there are hypotheses that relate to the nature of the estimate question ("last week or "usual week"), and the effect of occupation or status in employment (eg whether hourly paid, subject to time-clock, etc), and if the question is targeted to a specific period --whether or not the target period is or is not representative of normal work

patterns. Each of these issues might in principle be directly investigated using ancillary data from the survey, but they are not pursued further in this article.

Second are those hypotheses which concern the relationship of the "true" daily work hours and practices of the respondents to their knowledge of their own total work-time. Irregular work patterns make the sorts of instant respondent calculations necessary to answer stylized work hours estimate questions more complex and hence unreliable; and the combination of long hours of work with particularly irregular work patterns may introduce notable systematic biases in the resulting errors. This second category of hypotheses imply that diary or similar approaches, requiring respondents to list their work timings in some detail, leaving calculations of total working time as a separate step (incidentally enabling the analyst rather than the respondent to decide which activities are actually to be included *as* work), may be expected on *a priori* grounds to be superior to the less explicit stylized approach., as explored below.

Consider for example the effect of occasional and unplanned interruptions to a regular work-hours job, such as home plumbing repairs or accompanying sick chilren to medical facilities. The longer the regular hours of work—and given that the times of availability of such service are often restricted to something like a 9am to 5pm "normal" working day—the more likely it is that satisfying these sorts of domestic/family requirements will cause interruptions to the worker's regular work schedules. How do survey respondents factor these sorts of irregularities into their stylized estimates of work paid time? One might suspect that, (given the high social esteem usually attached to long hours of work, or perhaps just because of their occasional nature) such work-time-reducing interruptions will often fail to be included in accounts of "usual" or "last week" work times, while by contrast overtime episodes which increase work time (which are after all themselves paid work) may be more often remembered.

We would expect that such interruptions will be registered in diaries or work schedule instruments in the form of *irregularities in the starting and finishing times of paid work* through the work week. We might hypothesize that these irregularities would be associated with larger-than normal gaps between the diary or schedule estimates and the stylized estimates.

We can use the UK-HETUS to test this proposition, since the "last week" work grid registers seven consecutive days—which allow us to calculate, for each respondent, various relevant characteristics of the workweek. We should note that while these characteristics are straightforwardly derivable from the weekly grid, and while each is of substantial interest for labor market research, only the first of six work parameters derivable from the grid , reflects standard work hours as measured by the stylised estimate question, namely.

1) W_w the length of the work week as estimated from the schedule instrument.

The other two characteristics related to the *amounts* of paid work done last week are not derivable from the stykised estimates, namely:

2) N_w the number of days during the week in which paid work is undertaken; and

3) L_w the mean length of work time across the N_w work days

The second group of three characteristics available from the grid relate the *variability* of the length of the working day *through* the work-week[3], consisting of

4) S_w the variability (standard deviation) of L_w across the N_w work days

5) C_w the coefficient of variation of the length of the working day:

$$C_w = S_w/L_w$$

…which, unlike the measure of the variability of the length of the working day D_w , could in principle vary quite independently of the absolute length of the day.

[3] We have also experimented with using the variability of the timing of the start and end of paid work through the week, producing similar models to those which follow.

6) P_w the product of the length of the workweek and the coefficient of variation of the length of the working day

$$P_w = C_w * W_w$$

Using these work grid variables, it is then possible to estimate a straightforward OLS regression to test our hypothesis that the long hours of work *combined with* work-time interruptions can explain (at least part of) the gap between the diary and stylised estimates. Where T_w is the questionnaire estimate-based measure of work time, b_1 is a vector of regression coefficients relating this to the three work schedule-derived measures of *amounts of work time* through the week, and b_2 is a vector of regression coefficients relating this to the three measures of the *variability* of work time through the week, we can estimate (for the entire sample registering any paid work time in their work grid), the following straightford OLS regression:equation:

$$T_w = a + b_1 (W_w, N_w, L_w) + b_2 (S_w, C_w, P_w)$$

Table 3 provides estimates of the coefficients of this model, first of all for the entire working population (with the addition of an extra variable representing female respondents), and then separately for men and for women. Table 3 shows that all but one of the regression coefficients is highly significant, implying, other things equal, strong support for the proposition that irregularity in the length of the working day (or variations in the start and finish times through the week) is positively associated with the stylized estimate, quite independently of its overall association with the grid-based measure of work time.

Table 3: Predicting Questionnaire-based Paid Work Estimates

(Regression coefficients)	TOTAL		Men		Women	
Length of work week from schedule W_w	0.619	**	0.566	**	0.646	**
Number of workdays N_w	-2.531	**	-3.354	**	-1.508	**
Mean length of work days L_w	0.944	**	0.715	**	1.416	**
Variability of length of work days S_w	-3.006	**	-3.64	**	-1.28	
Coefficient of variation of length of work days $C_w = S_w/L_w$	17.358	**	9.416	**	22.868	**
Length of workweek* coefficient of variation $W_w * C_w$	0.277	**	0.71	**	-0.372	**
Woman	-7.659	**				
(Constant)	32.717	**	32.487	**	9.325	**
Multiple R	0.611		0.451		0.594	

** Significant at .005

We can see the nature of this relationship (between work-grid irregularity and the stylized estimate) more clearly by visualizing a simple statistical experiment corresponding to the question; "What do the coefficients in Table 3 tell us about the effect on the stylized estimates of a reduction in the *variability of the work-week*?". We can simulate the answer to this question quite straightforwardly by using the Table 3 coefficients to derive a predicted value for the stylized estimate for each respondent in the UK-HETUS dataset, having reset the value of the variability of length of work days (S_w) for each case to zero—so that the vector of b_2 coefficients has no effect on the prediction.

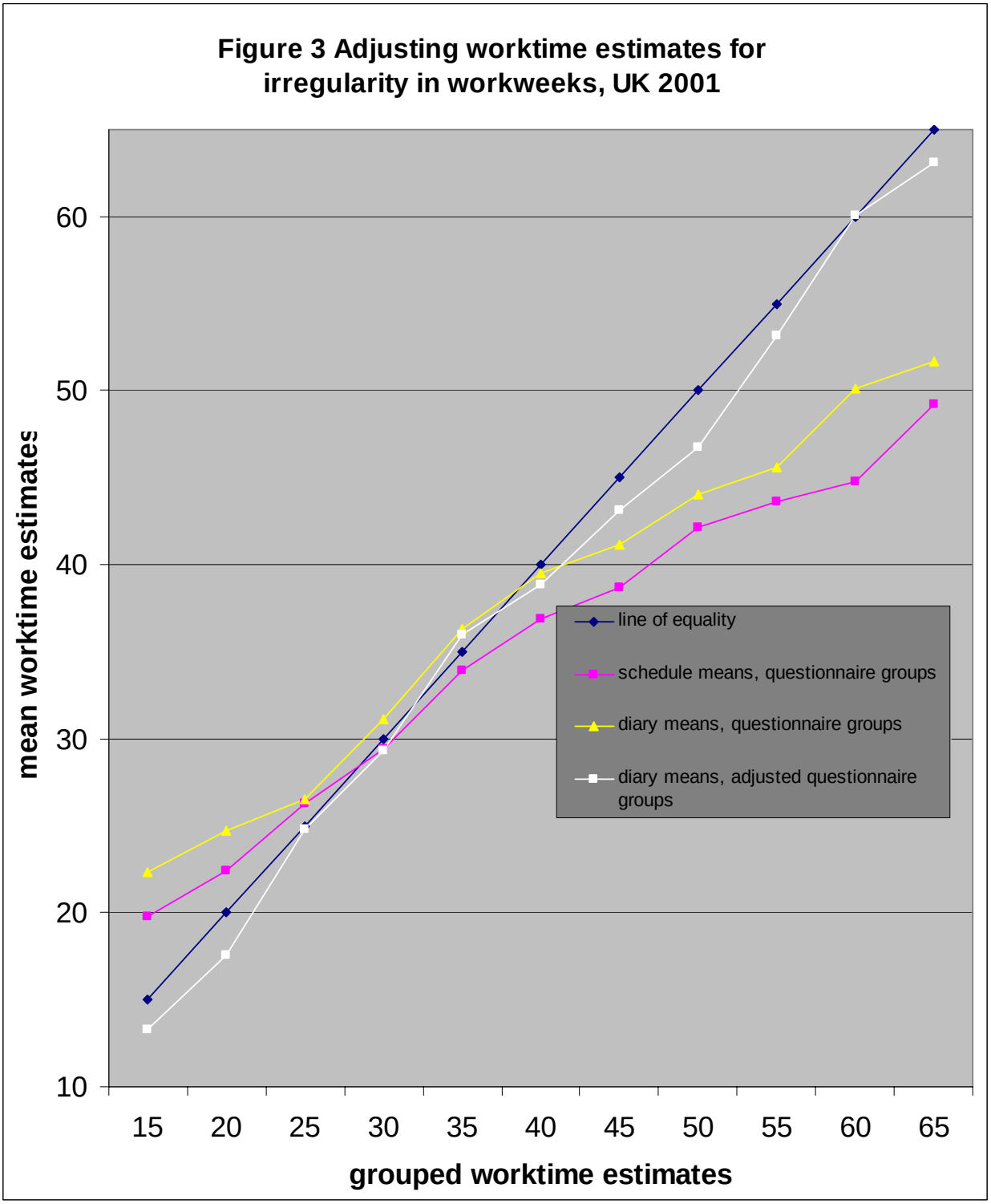
Figure 3 thus shows first, in the line marked "grid means, questionnaire groups", the equivalent to the previous "gap" plots: the group of respondents who estimate around 30 hours of paid work per week (in fact 27.5 to 32.4 hours) in their stylized responses do also show about 35 hours of paid work in their work grids (or "schedules"). For each subsequent five-hour increment in their stylized response, the corresponding schedule mean rises by around 2 hours. The means of the diary responses for the same groups (based on randomly sampled single days

of data, but multiplied by 7 to produce weekly estimates) are quite similar. The plot of these means against the questionnaire estimate groups again closely resembles the results in Figure 1.

Figure 3 also plots the same diary means, but this time against groups formed, not on the basis of the stylized estimates but rather from the groups formed from our experimental simulation of the stylized responses that might have been forthcoming *if the respondents had no irregularity* in their work-weeks By inspection, one can conclude that this last plot corresponds much more closely to the diagonal "line of equality" that represents complete agreement with the (on an *a priori* basis more accurate) diary data than do the equivalent plots against the grouped actual stylized responses. In short, by removing the results of the variability in the workweek, the stylized estimates much more closely resemble the diary-based estimates.

Table 4 further sets out the means for the various different estimates of the work week. For men, the working time variability adjustment only slightly moves the stylised estimate towards the diary and grid estimates, whereas for women the adjustment brings the estimate below the diary estimate, but still somewhat above the higher of the two grid estimates. Overall the adjustment for workweek irregularity alone seems to move the stylized estimate just under half way between the original stylized estimate and the diary estimate.

Figure 3 Adjusting worktime estimates for irregularity in workweeks, UK 2001



.

Table 4: Comparison of work week estimates (In hours/week)

	Men	Women	TOTAL
Stylised estimates	45.6	32.0	39.3
Adjusted stylised estimate	44.4	31.4	38.5
Diary estimate	42.0	31.9	37.3
Grid estimate with travel	41.4	30.4	36.3
Grid estimate, no travel	39.3	29.9	34.9

4 SUMMARY AND CONCLUSIONS.

The conclusion of the above analyses is not of course to suggest substituting some form of seven-day diary instrument for the traditional,, large-scale collection of paid work hours data currently collected through stylized estimate questions in CPS surveys or the standardized "Labor Force Surveys" collected in every member state of the EU. The evidence discussed above is however sufficient to warrant a serious reconsideration of the accuracy of stylized estimates measures, and, at the least, some attempt to calibrate the errors.

The first part of our discussion, comparing the weekly estimate data with daily diary data, concerns issues that are ultimately irresolvable. One may wish to argue on *a priori* grounds that methods that require respondents to set out explicitly and in detail how many hours they spent working across a particular period, must necessarily by preferred to methods which allow

respondents simply to estimate hours worked without any sort of explication. The only plausible way one can test this proposition, is by comparing, and making inferences from, the individual distributions of errors. At the same time, one can only compare estimates at a grouped level in the manner of Figures 1-2 above, which gives rise to the suggestion of the "regression to the mean", which is essentially untestable with stylized questions.

Whole weekly diaries, where they exist, provide the best opportunity for some limited form of test. The distribution of marginal values in Table 2 reveal an asymmetry, insofar as the marginal counts of respondents claiming long hours of work in their stylized responses, provide larger numbers of much lower week-schedule-based estimates than the corresponding count of lower stylized estimates for those with higher grid-based estimates. Plainly there are *some* random errors of the sort that would correspond to a "regression to the mean" explanation of the "estimation gap" phenomenon—but, Table 2 suggests some of the estimation gap does reflect a systematic upwards bias for the higher end of the stylized estimates.

Moreover, whole-week diaries or diary-like measures do themselves contain at least some of the information necessary to understand the nature of the bias, enough in particular to demonstrate that the bias largely disappears if one of its main causes is removed. Variability in the length of the working day plainly makes it difficult for respondents to make accurate stylized estimates. This does not of course help us directly to improve the stylized estimates, since work-day length can be in reality quite variable. However, it does suggest that if one is to measure the exact extent of the work week, one must find a way to calibrate the stylized estimate question—presumably by employing some sort of diary instrument.

This line of argument points principally in the short term to a new agenda of needed research activities: (1) collecting and analyzing some form of whole-week diary instrument for some subsample of the CPS respondents, and (2) bringing together results of this sort of data collection with an investigation of other related questions, about the "normality" or regularity of the estimated week, and about the relationship of work-hours estimations of a target week to estimations of work hours across the whole year including holidays. Moreover, the whole-week measures are themselves of substantive interest to labor analysts, providing evidence not only on the length of the working day, and the number of working days per week, but also on the variation in the length of working days, on the timing of work during the working day and on systematic differences in these characteristics across different sorts of workers.

The whole week measures also provide potential sources of evidence on wider aspects of public policy: They shed light on the temporal accessibility of services—for example, on childcare provision across the working day, on shopping hours when not at work, on optimal sleep hours—and, as the most plausible source of information on the temporal availability of population groups for the purposes of sociability and of informal/unpaid caring activities through the week. They further also provide much-needed evidence on other major but neglected issues of psychological wellbeing and social exclusion, such as sleep deprivation and insufficient free time.

Appendix 1 Time Use Diary Materials

The current (2007) release of the **The Multinational Time Use Study (MTUS)** is in the form of a series of national data files referred to collectively as WORLD5.5. This release currently consists of 46 random sampled national surveys from 15 countries, providing 460,000 days of time-diary data. At least 7 further surveys and three new countries will be added by the end of 2007 (and approximately 12 other surveys are available in the previous WORLD5.0 format). WORLD5.5 national data files represent the full age range of the national populations (excluding children below the age of 10); they provide 40 aggregated primary time-use activity categories (summing to the 1440 minutes of the sampled day), together with 30 socio-demographic classifiers of individual and household characteristics, and with equal-selection-probability weight variables that also produce properly balanced distributions of days of the week.

Sample sizes for components of MTUS WORLD5.5 (Release May 2007)
() indicates survey currently being processed for inclusion

N of days	**1961-69**	**1970-75**	**1976-84**	**1985-89**	**1990-94**	**1995-99**	**2000-04**
Canada		2138	2682	9618	8936	10726	()
Denmark	4069			2389	()		()
France	2898	4633		()		15318	
Neth'lands		1292	2727	3263	3158	3227	1649
Norway		6516	6068		6129		7675
UK	9292	17507		18060		1906	19400
USA	2021	7010		4935	9386	1151	20340
Finland			11908	15219		10076	
Italy			2116	37764			()
Australia				3181	13806	14071	()
Sweden					7065		7747
Germany	3687				25775		()
Austria					25162		
S. Africa							14217
Slovenia							12273
Total							**457,135**

Available in WORLD5.0 format: Belgium, Hungary, Czech, Yugoslavia, Israel.
Recent Spain, Portugal surveys also currently being processed for inclusion.

Comprehensive documentation and quality profiles, including algorithms used to transform microdata from the original to the harmonized form is provided online at

< www.timeuse.org/mtus/>

The previously issued form of the Multinational Time Use Study was WORLD5.0 which represented more restricted populations (aged 20-60), providing the same 40 aggregated time-use activity classification, and 15 less comprehensively harmonised socio-demographic classifiers. Currently in the early stages of design and consultation is WORLD6.0 which will have a revised (and considerably more detailed) activity classification. It will also for the first time provide a cross-nationally harmonised version of the activity sequence data from the original diaries, allowing effective use of the whole of the time use diaries (including simultaneous "secondary" activities, as well as presence of others, location and transport mode indicators). A partial preliminary study version of WORLD6.0, the American Heritage Time Use Study (AHTUS) has been prepared, using only surveys collected in the USA (with the support of the Glaser Progress Foundation). These materials may be downloaded from:

<www.timeuse.org/ahtus/>

A detailed description of the UK Harmonised European Time Use Study (UK HETUS) is provided at:

<www.timeuse.org/information/studies/data/uk-2000-01a.php>

References

Bianchi, Suzanne M., John P. Robinson, and Melissa A. Milkie (2006) *Changing Rhythms of American Family Life* New York: Russell Sage Foundation.

Bonke, Jens. (2005) "Paid Work and Unpaid Work: Diary Information Versus Questionnaire Information" *Social Indicators Research* 70: 349-368.

Chase, D.R., and Geoffrey C. Godbey. (1983) "Accuracy of Self-Reported Participation Rates: Research Notes" *Leisure Studies* 2(2): 231-35.

Chenu, Alain, and Robinson John P., (2002)"Synchronicity in the Work Schedules of Working Couples" *Monthly Labor Review* 125 April: 56-83

Fisher, Kimberly, Muriel Egerton, Jonathan I. Gershuny, and John P. Robinson. (2007) "Gender Convergence in the American Heritage Time Use Study (AHTUS)" *Social Indicators Research* 82, 2007, 1-33.

Fisher, Kimberly, Muriel Egerton, Nuno Torres, Andreas Pollmann and Jonathan Gershuny, with contributions from John Robinson and Anne Gauthier. (2006) "American Heritage Time Use Study (AHTUS) Codebook" Oxford: Centre for Time Use Research. http://www.timeuse.org/ahtus/documentation/

Frazis, Harley and Jay Stewart. (2004) "What Can Time-Use Data Tell Us About Hours of Work" *Monthly Labor Review* December 2004: 3-9.

Gershuny, Jonathan. (2000) *Changing Times: Work and Leisure in Postindustrial Society*. Oxford: Oxford University Press.

Hawes, D., W. Talarzyk, and R. Blackwell. (1975) "Consumer Satisfactions from Leisure Time Pursuits." In M. Schlinger, ed., *Advances in Consumer Research* Chicago: Association for Consumer Research.

Jacobs, Jerry A. (1998) "Measuring Time at Work: Are Self-Reports Accurate?" *Monthly Labor Review* December 1998: 42-53.

Kan, Man Yee, and Pudney Steve. 2007 "Measurement Error in Stylised and Diary data on Time." Sociology Working Papers 2007-05, Department of Sociology, University of Oxford, <www.sociology.ox.ac.uk/swp.html>

Krosnick, Jon (1999) Maximizing Questionnaire Quality" in J Robinson et al. (eds.) *Measures of Political Attitudes*. San Diego CA: Academic Press

Marini, Margaret M., and Beth A. Shelton (1993) "Measuring Household Work: Recent Experiences in the United States" *Social Science Research* 22(4): 361-82.

Michelson, William. (2005) *Time Use: Expanding Explanation in the Social Sciences* Boulder, Colorado: Paradigm Press.

Press, Julie and Eleanor Townsley (1998) "Wives' and Husbands' Housework reporting: Gender, Class and Social Desirability" *Gender and Society* 12(2):188-218

Putnam, Robert. (2001) *Bowling Alone: The Collapse and Revival of American Community* New York: Simon & Schuster Ltd.

Robinson, John P. and Ann Bostrom. 1994. "The Overestimated Workweek? What Time Diary Measures Suggest." *Monthly Labor Review* 117(8):11–23.

Robinson, John P. , Alain Chenu and Anthony Alvarez (2002) Measuring the Complexity of Hours at Work: The Weekly Work Grid *Monthly Labor Review* 125 April: 45-54

Robinson, John P., and Geoffrey Godbey. (1997) *Time for Life: The Surprising Ways Americans Use Their Time* University Park, PA: Pennsylvania State University Press.

Robinson, John P. and Jonathan Gershuny. (1994) “Measuring Hours of Paid Work: Time-Diary vs Estimate Questions” *Bulletin of Labour Statistics*: xi-xvii.
Verbrugge, L. and D. Gruber-Baldine. (1993) *Baltimore Study of Activity Patterns*. Ann Arbor: Institute of Gerontology, University of Michigan.