

SOEPpapers

on Multidisciplinary Panel Data Research

SOEP – The German Socio-Economic Panel Study at DIW Berlin

390-2011

Measuring Time Use in Surveys – How valid are time use questions in surveys? Concordance of survey and experience sampling measures

Bettina Sonnenberg, Michaela Riediger, Cornelia Wrzus, Gert G. Wagner

SOEPPapers on Multidisciplinary Panel Data Research at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel Study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPPapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPPapers are available at
<http://www.diw.de/soeppapers>

Editors:

Joachim R. **Frick** (Empirical Economics)

Jürgen **Schupp** (Sociology, Vice Dean DIW Graduate Center)

Gert G. **Wagner** (Social Sciences)

Conchita **D'Ambrosio** (Public Economics)

Denis **Gerstorff** (Psychology, DIW Research Professor)

Elke **Holst** (Gender Studies)

Frauke **Kreuter** (Survey Methodology, DIW Research Professor)

Martin **Kroh** (Political Science and Survey Methodology)

Frieder R. **Lang** (Psychology, DIW Research Professor)

Henning **Lohmann** (Sociology, DIW Research Professor)

Jörg-Peter **Schräpler** (Survey Methodology, DIW Research Professor)

Thomas **Siedler** (Empirical Economics, DIW Graduate Center)

C. Katharina **Spieß** (Empirical Economics and Educational Science)

ISSN: 1864-6689 (online)

German Socio-Economic Panel Study (SOEP)

DIW Berlin

Mohrenstrasse 58

10117 Berlin, Germany

Contact: Uta Rahmann | soeppapers@diw.de

Measuring Time Use in Surveys –

How valid are time use questions in surveys? Concordance of survey and experience sampling measures

Bettina Sonnenberg

LIFE Research School (Max Planck Institute for Human Development, Berlin) & SOEP (German Institute for Economic Research), Berlin

Michaela Riediger

Max Planck Institute for Human Development, Berlin & German Institute for Economic Research, Berlin

Cornelia Wrzus

Max Planck Institute for Human Development, Berlin

Gert G. Wagner

German Institute for Economic Research, Berlin & Max Planck Institute for Human Development, Berlin

Abstract

Since it is still unclear to what extent time allocation retrospectively reported in questionnaires, reflects people's actual behavior, examining the accuracy of responses to time use survey questions is of crucial importance. We analyze the congruence of time use information assessed through retrospective questionnaires and through experience sampling methodology. The sample comprised 433 individuals ranging in age from 14 to 86 years. Participants completed standard survey questions on time allocation. In addition, a mobile-phone based experience sampling technology was used over a period of three weeks to obtain snapshots of, on average, 54 momentary activities in which participants participated while pursuing their normal daily routines. Experience sampling assessments were scheduled six times a day over at least nine days, including workdays, Saturdays, and Sundays.

Results indicate that the congruence between time allocation assessed with survey questions (i.e. in SOEP) and time allocation assessed with experience sampling methodology depends on the characteristics of the respective activities. Associations between standard survey questions and experience sampling methods are quite substantial for long-lasting and externally structured activities, such as paid work on workdays. In contrast, associations between survey and experience sampling methods are somewhat weaker, though highly statistically significant, for less externally structured, short-term and infrequent activities, such as errands, housework, and leisure. These moderate and relatively small correlations may indicate either an error-prone estimation of the prevalence of short-term and infrequent activities by experience sampling or respondents' overrating of sporadic and short activities in survey questions.

We conclude that activities with a long duration, such as paid work, can be measured in a satisfactory manner using short survey questions. Future research is necessary to elucidate which method (experience sampling method or survey questions) delivers more reliable and valid measures for short-term and sporadic activities. Day Reconstruction Methods (DRM) should be included in this future methodological research.

JEL Classifications: C21, C83, (C93), D10

Keywords: survey methods; experience sampling method; validity; time use; paid work; housework; leisure; German Socio-Economic Panel Study; MMAA; SOEP

1 Introduction

Time use has become an important field of research in a variety of different disciplines. Sociologists and economists consider time allocation in terms of working and leisure time, for instance, to be an important indicator of quality of life and well-being, household production, social participation, or gender inequality (see, for instance, Ahn, Jimeno, & Ugidos 2005; Fisher et al. 2007; Gershuny 2000; Isengard 2005). Moreover, on the macro level, time allocation has been analyzed with regard to social change and social modernization.¹ Researchers analyzing time use can access a range of national and international time use data surveyed by means of different measurement techniques. Popular survey methods capturing the time people spend on different activities are diary methods and survey questions. Short survey questions, in particular, are commonly used in large-scale surveys because they are considered to be a relatively low-cost, but nevertheless valid, alternative method to time diaries.

However, although time allocation data have been extensively collected using different measurement strategies for the past 50 years, detailed knowledge about the quality of the data produced by different methods of surveying time allocation remains rather rare and one-sided. Existing literature predominantly compares the performance of survey questions and time diaries, both of which have specific advantages and shortcomings. For instance, time diaries allow the collection of comprehensive information for whole days, but are mostly restricted to only one or two days as filling in these diaries is rather burdensome for participants. Survey questions, on the other hand, are easy to answer and can be applied in longitudinal surveys, but are restricted to only a limited number of activities. Furthermore, survey questions have often been suspected of providing incorrect data since people's responses rely on their cognitive ability to estimate the approximate time they spend, on average, on certain activities. An additional method of obtaining information on people's activities is the experience sampling method (ESM). ESM combines the advantages of time diaries and survey questions by collecting people's time use in real time, covering a wide range of activities as well as assessing a longer time period than only a few days.²

Time use data collected by experience sampling is, thus, of particular interest in comparisons with time use information obtained by asking survey questions. This paper, therefore, aims to

¹ For a good overview of different research topics related to time use, see Andorka (1987). A synopsis of different time use research fields is also provided by Merz (2010).

² A fourth method which combines details of diaries with a more respondent-friendly answering method, the Day Reconstruction Method (DRM), will be discussed in the outlook.

investigate the association between time use data derived from both measurement techniques. Specifically, we compare time use information assessed by both traditional time use survey questions and experience sampling assessments over a 21-day period. Experience sampling assessments were pseudo-randomly scheduled several times a day and implemented using cell phones as assessment instruments. Our analyses focus on time allocation reports for paid work, errands, housework, and leisure. Given the structural differences in the time allocation between workdays and weekends (Friedberg & Webb 2005; Kramer 2004), for each activity we conducted separate analyses for weekdays and weekends.

Overall, our analyses indicate strong associations between both measurement methods for time spent on market work, whereas the results show somewhat weaker associations between survey questions and ESM for errands, housework, and leisure time activities. Given ESM's advantage that, unlike traditional survey questionnaires, it does not rely on bias-prone retrospective generalizations, our results may be interpreted as indicating that common survey questions perform well with regard to time spend on paid work, whereas they may perform less well with regard to errands, housework, leisure. However, as far as design issues in the ESM-Study are concerned, more research is needed to draw clear cut conclusions on the performance of survey questions.

The paper is organized as follows: in Section 2, the main time use measurement techniques are discussed and the empirical literature concerning the quality of data produced by standard survey questions is outlined. After deducing our leading hypotheses from the existing literature, we will describe the data and define our estimation strategies in Section 3. Section 4 provides summary statistics and descriptive results, before we present our multivariate empirical results in Sections 5 and 6. Section 7 consists of a brief summary of our findings and our conclusion.

2 Measurement instruments – a literature review

People's time allocation has been studied since the beginning of the twentieth century (Bevans 1913; Jahoda, Lazarsfeld, & Zeisel 1975 [1933]) cumulating in national and international time use surveys (e.g., the American and German Time Use Survey and Harmonized European Time Use Survey). Whilst some of the main starting points for looking at time use have been people's day structures and living conditions in the event of unemployment and increasing material hardship, time use has increasingly served as a

multiple indicator for factors ranging from aging, gender inequality and social change to well-being and welfare. Parallel to an increase in surveys on people's time allocation, several measurement instruments were also developed, each with different advantages and disadvantages. Two widely used measurement techniques are survey questions, also known as stylized measures, and time diaries.

Using survey questions respondents are normally provided with a short list of activities and asked to report the amount of hours they spend on each of these activities on a normal day. Survey questions, thus, constitute a rather brief measurement instrument used in large-scale, particularly household panel, surveys, where time use only represents one topic among a set of others (e.g., labor market participation, satisfaction with life, or household composition) and time diaries are typically not feasible. Short survey questions are, therefore, used as a time-efficient method for assessing time use. Activities included in survey questions normally cover labor market activities, errands, housework, childcare, schooling and training, and leisure. Due to structural differences some distinguish between workdays and weekends.

However, the accuracy of information collected by survey questions requires that people are able to correctly estimate the average time they spend on different activities. This is a strong assumption. Particularly with regard to synchronous activities, respondents may have difficulty accurately estimating time spent on certain activities, resulting in an overestimation or underestimation. Furthermore, individuals' time allocation to different activities typically varies over time. In response to survey time use items, respondents, therefore, have to provide an estimate of their 'average' time use, which may be error-prone (Blair & Burton 1987; Schwarz 2007; Schwarz & Oyserman 2001; Tourangeau, Rips, & Rasinski 2000). However, by asking respondents to answer questions about a typical week instead of asking about their time allocation on a specific day, it may be possible to avoid issues regarding the representativeness of information for only one or two days.

Unlike survey questions, time diaries typically collect data for just one or two days (Harvey 1993: 207ff.). However, more detailed data are captured for this short period of time, as participants are asked to fill in time diaries dividing the day in approximately 15-minute blocks. Time diaries, thus, acquire information on, not only the overall amount of time spent on different activities on that day, but also the duration of each activity, the frequency, and the sequence of different activities. Apart from this detailed information, primary and secondary activities can be covered using diary methods. For these reasons and since respondents are requested to fill in the diary immediately when one activity ends and a new one begins, time

diaries are often considered to be the most exhaustive, as well as the most elaborate, method of surveying people's time use (Kitterød 2001: 146f.; Niemi 1993: 233, 242; Robinson 1999: 47). Time diaries are typically seen as an accurate representation of people's real behavior. Accordingly, previous research has focused on the comparison of time use information collected using common survey questions and time diaries (Kan & Pudney 2007; Plewis, Creeser, & Mooney 1990).

However, like survey questions, the time diary method also has shortcomings. Completing the diaries requires time-consuming effort, particularly as people are expected to provide this information immediately in order to avoid retrospective response biases. Also, as time diaries often only cover one day, they do not account for time use differences between work days and weekends on an intra-individual level.

Despite the advantages, it has been shown that the amount of time reported to be spent on different activities varies according to data collection procedures used in time diaries (Kitterød 2001; Harvey 1993). For instance, response patterns with respect to main activities differ according to whether or not parallel second activities can be reported (Kitterød 2001). Moreover, previous empirical literature reveals failure to ask about concurrent activities results in an overestimation of certain activities, such as housework (Kan 2007; Kitterød 2001: 169; Kitterød & Lyngstad 2005: 14f; Robinson 1985: 46ff.).

Furthermore, time diaries are assumed by some authors to produce systematic incorrect data with regard to unpaid work in comparison to paid market work, as short breaks from work are considered as leisure in the case of unpaid work, while short breaks are incorporated in paid work (Rydenstam 2001 according to Kitterød & Lyngstad 2005). Therefore, time diaries seem to systematically underestimate hours spent on housework, in comparison with paid work, which is especially relevant for women due to the fact that they carry out a greater share of housework.

However, other studies indicate that survey questions provide less accurate data compared to time diaries. Taking time diaries as the superior measurement technique to obtain valid data, various studies indicate that activities which can easily be distinguished from others may be accurately captured by short survey questions, whereas survey questions produce invalid data for rather less distinct or externally structured activities (Marini & Shelton 1993; Niemi 1993; Juster, Ono, & Stafford 2003; Schulz & Grunow 2007). For example, Niemi (1993) finds high correspondences of hours reported on paid work, unpaid work, and leisure (library visits,

movie visits, and physical exercise) collected by survey questions and by time diaries, while this is not true for occasional or less externally set activities. Specifically, she finds that women seem to overrate the amount of physical exercise in surveys as compared to time diaries. Focusing on paid and unpaid work, Bonke (2005) also reveals complex differences between the data obtained by survey questions and time diaries: people who spend many hours on paid work seem to report more working hours when responding to survey questions compared to time diaries whereas those less engaged in labor market activities tend to underestimate the actual time worked (see also Otterbach & Sousa-Poza 2010). Bonke's findings also indicate the same trend for unpaid work, i.e., housework (2005: 366). However, Kitterød & Lyngstad (2005) find only slight differences between reported housework captured by surveys and time diaries for the population in general. Nevertheless, their analysis reveals some deviations for different age groups. In particular, their findings indicate that housework is adequately represented by survey questions for people in the middle-aged group, but produces somewhat incongruent data for younger and older people (2005:13).

Conversely, other investigations confirm high correspondences of survey questions with regard to paid work. For instance, comparing hours engaged in labor market activities surveyed in the Panel Study of Income Dynamics (PSID) as well as in the Swedish Panel Study Market and Nonmarket Activities (HUS) by both, time diaries and stylized measures, Juster, Ono & Stafford (2003) show that average hours for paid work are very similar for both measurement methods (2003: 35).

Overall, focusing on the literature comparing time use data obtained by survey questions and time diaries provides mixed and rather inconclusive results. Having regarded detailed time diaries as the superior method, incongruences between both measures have predominantly been interpreted as revealing shortcomings of survey questions. Considering this rather inconclusive empirical literature as well as how frequently survey questions are used in common small and national surveys, more in-depth analyses on the performance of survey questions are needed. One first approach may be to compare with an alternative measurement technique that avoids the shortcomings associated with survey questions and time diaries. Thus, we use experience sampling method which we explain in more detail in the following section.

3 Experience Sampling Method

Mostly used in psychological research, the experience sampling method (ESM) is a measurement technique that aims at obtaining repeated snapshots of people's *momentary* experiences and behaviors without having to rely on people's recollection of past states (Riediger 2010). In order to survey people in their natural environment, respondents are provided with pagers – or more recently cell phones – that are used to contact them over a certain period of time and report their current situation and psychological state each time they are contacted (Csikszentmihalyi 1987: 527; Mannel & Zuzanek 1991: 339). Normally, the assessment time covers several days during which participants are contacted multiple times per day covering most of people's awake time. Assessment points are randomly assigned within certain time periods, for example, about every two hours resulting in repeating observations for each individual per day.

With regard to avoiding retrospectively collected data, ESM has one main characteristic in common with time diaries which is assumed to lead to the collection of more valid data. Since ESM does not rely on recalling past activities, ESM data is very likely to be less biased by respondents' socio-economic characteristics, such as educational attainment (Juster, Ono, & Stafford 2003: 33). Moreover, in contrast with time diaries, ESM incorporates a longer time span investigated and, therefore, avoids one of the main limitations of time diaries. Of particular importance for time use, covering a longer time span makes it possible to distinguish between workdays and weekends, and, accordingly, factors in structural differences of time use (Friedberg & Webb 2005; Kramer 2004).

Despite the benefits of ESM, there are only a few studies that have applied this method to analyze time use to this date (Csikszentmihalyi & Hunter 2003; Klumb & Baltes 1999). The main reasons for the small number of ESM studies are the resource intensity and the effort needed by participants to comply with the requirement of always carrying the pager or cell phone and being available to be contacted several times a day for a number of subsequent days (Riediger 2010: 585f.).

3.1 Data and method

To examine the performance of survey questions we draw on the ESM Study developed by researchers at the Max Planck Institute for Human Development in cooperation with the

German Socio-Economic Panel Study (SOEP). The study belongs to the „Multi-Method Ambulatory Assessment“-Project (MMAA)³ (Riediger et. al 2009, 2011) and was conducted in three German cities, Berlin, Düsseldorf, and Munich, from 2006 to 2009. It includes two waves and contains two subsamples.

The first wave, conducted in 2006 and 2007, comprises an age- and gender-stratified sample of 378 participants. In the second wave, which was carried out in 2009, a subsample of first-wave adolescent and young adult participants was investigated again and 55 additional participants were included. Since the recruitment procedures do not select the participants randomly, inferences to the general population are limited.

However, the ESM Study elicits participants' time use using survey questions as well as experience sampling and, therefore, allows a direct comparison of both measures. In the standard questionnaire part of the study, preceding the ESM section, socio-demographic and socio-economic information about the participants was collected in addition to responses to survey questions on time use.

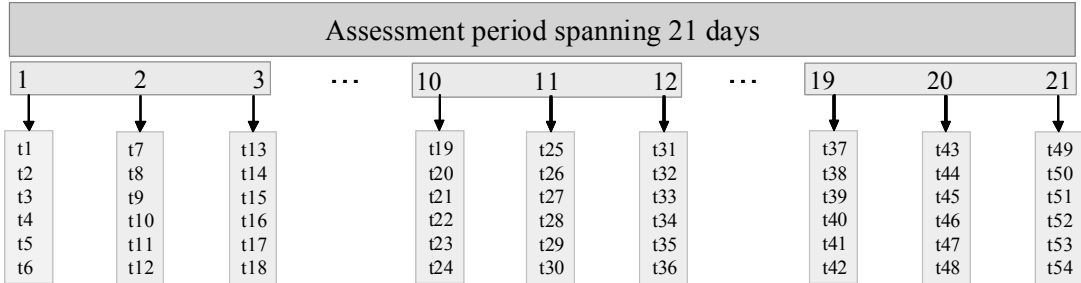
For the experience sampling section, participants were given special cell phones ('measurement devices') which they carried with them during a period of three weeks. Within this time period, there were three experience sampling phases (see Figure 1). In advance of the survey period, participants stated at what time of the day the cell phone questionnaire was allowed to start. From this arranged time onward, the participants were contacted automatically and randomly within a 12-hour time window each day. This time period was subdivided into 6 two-hour time windows in which people were contacted randomly via their cell phones. Two consecutive assessment points had to be at least 15 minutes apart. When the participants were contacted, they had 15 minutes to answer the questions on their cell phones, otherwise the assessment occasion was lost. If participants missed more than one assessment occasion on one day, the experience sampling phase was prolonged by an additional day.

Between the experience sampling phases, participants were given a break of six days. In order to initiate the random signal and capture the data during the experience sampling period, "[mobile] phones [...] signaled the participants to complete an assessment instrument referring to his or her momentary experiences" (Riediger 2010: 588f). By stating their

³ For more information on the project, visit <http://www.mpib-berlin.mpg.de/en/research/max-planck-research-groups/mprg-affect-across-the-lifespan/projects/multi-method-ambulatory>.

momentary experiences, respondents also specified their activities by reporting where they currently were, who was with them and what they were doing.

Figure 1: Assessment period for Experience Sampling Method



Source: Own illustration.

For our analysis we only consider those individuals who participated in the ESM Study for the first time in each wave because in the subsequent wave time use was not captured by survey questions again. Our sample, thus, contains 433 individuals. For those 433 individuals 4,350 days were assessed with 23,842 measurement occasions. For 2,971 of these 4,350 days 6 assessment points could be realized, while 895 days cover 5 time points. For about 484 (11.1 percent) of the assessment days participants provided up to four of the six assessment occasions. Since missing more than one of the requested contacts resulted in a prolonged experience sampling phase, a considerable share of the sample exceeds the minimum of 9 assessment days. In fact, 215 participants, accounting for 49.7 percent of the sample, had 9 measurement days. For about 36.7 percent the experience sampling period was prolonged by one or two days, while 6.9 percent of the sample had more than two additional assessment days.

In addition to the ESM Study, we also use the German Socio-Economic Panel Study for 2007 as reference data (Siedler et al. 2009). Since the questions on time use included in the ESM Study are taken from the SOEP, it is feasible to investigate the reliability of the survey questions⁴. Since the SOEP focuses on adults of at least 18 years of age, for comparison, we restrict our sample to those 336 participants in the ESM Study of this age. Potential incongruences between reported hours devoted to different activities in the ESM Study and

⁴ The SOEP has collected information on people’s time use since its first wave in 1984 (see Wagner, Frick, & Schupp 2007). Christof Helberger, Hilmar Schneider, and Joachim Merz developed the basic and broader concept of the time use questions (for an in-depth analysis, see Merz & Wolff 1993). Joachim Merz developed the concept that was implemented in the SOEP questionnaire (see Merz & Kirsten 1999 for an analysis of these data).

the SOEP 2007 could result from sample composition effects, so we control for the sample composition of the ESM data with respect to gender, age, and employment status by using a redressment weight (see Appendix A, Table A1).

3.2 Variables and measures

The survey questions in the ESM Study were taken from the German Socio-Economic Panel Study, an annual household panel survey using the same time use questions since 1984, and read as follows: “What is a typical day like for you? How many hours do you spend on the following activities on a typical weekday, Saturday, and Sunday?” The activities listed were 1) job, apprenticeship, second job, 2) errands, 3) housework, 4) childcare, 5) care and support for persons in need of care, 6) education or further training, 7) repairs on and around the house, car repairs, gardening and 8) hobbies and other free-time activities. Respondents were asked to state the number of hours per day, where zero hours indicated the activity as not being appropriate, for instance working time for those who were unemployed.

Using the experience sampling information, people’s time allocation was derived from responses to three questions: the first one assessed participants’ momentary location by asking “Where are you?” with the following available response categories a) at home, b) at work, in school, at university, c) in public institutions, d) visiting someone, e) en route, or f) somewhere else. A second question addressed participants’ momentary social partners. Participants could choose multiple categories from the set of a) nobody, b) partner, c) family, d) colleagues, e) strangers and/or f) other people. The last question asked participants to classify their activities themselves. In response to the question “What are you currently doing?” they could choose several activities from a) working, studying, training, b) doing housework, errands, c) doing leisure activities, d) nothing, sleeping, or watching TV, e) in consultation, using governmental services, f) chatting, visiting, being visited and/or g) other activities.

Although both survey questions and experience sampling items address people’s time use, they differ in crucial ways: the SOEP survey questions on people’s time allocation used in the ESM Study collect hours ranging from 0 to a maximum of 24 hours. Conversely, the experience sampling section captures current engagement in different activities, not accounting for the duration of the different activities. Additionally, questions eliciting

people's time use deviate considerably between the two techniques with regard to the activities captured and the possibility of reporting concurrent activities.

The main challenges for the analyses were, therefore, first to generate corresponding activities for both measures ex post and second to deal with the two distinct kinds of information provided by both techniques: hours for survey questions and specific activities at different points in time captured by the cell phone method.

The first challenge was surmounted by classifying comparable activities from the experience sampling section relying on the three questions presented above. Although we would have preferred to exploit the data on all activities assessed by the survey questions, information included in the experience sampling section was not detailed and/or distinct enough to include other activities, such as child care or education. Thus, we focus on those activities captured in sufficient detail within the experience sampling framework, namely paid work, errands, housework, and leisure. These activities were derived from the three questions in the experience sampling sections as follows:

Paid work: Participants had to state that they were working, studying or training. They were not allowed to report being in a consultation, using governmental services, or doing housework or errands at the same time. Because the ESM questions do not distinguish between labor market work and training, we recoded to zero those individuals who were still in school or further education. Otherwise, we would have possibly included non-work-related activities.

Errands: When people reported doing housework or errands and also reported being somewhere else other than at their workplace, school, university, or at home, activities were coded as errands. Attending a consultation or receiving governmental services was also categorized under errands. However, if participants were accompanied by colleagues, we did not classify this activity under 'errands'.

Housework: We classified activities under housework when respondents were at home and regarded their current activity as housework or errands. If they also stated doing other things like studying or leisure activities, we did not consider this housework due to possible fuzziness of the underlying activities.

Leisure: If people stated they were doing leisure activities and did not simultaneously classify their activity as working, studying, or doing housework, or errands, we defined those activities as leisure.

While inconsistencies with regard to the survey questions may only show up with different activities exceeding 24 hours – assuming exclusive and not overlapping concurrent activities – data from the experience sampling section enables us to investigate simultaneously occurring activities as well as inconsistent responses. In our analyses, we only included consistent responses. Thus, concurrent activities which are mutually exclusive or contradictory were excluded from the analysis. For instance, with respect to housework we do not consider occasions where participants reported concurrently doing housework and seeing the doctor or receiving government services at the same time.

The second challenge for our analysis resulted from the fact that survey questions provided information on hours spent on different activities, while experience sampling captured the momentary activity of the participant at multiple points in time. The experience sampling, therefore, does not consider the duration or number of certain activities or their definite sequence but does provide information on the frequency. Even if participants reported engaging in the same type of activity on two consecutive measurement points, one cannot conclude that they did so all the time in between. For instance, though a person may state leisure activities at two or three measurement points in a row, the possibility cannot be ruled out that this person had been engaged in different activities in between the measurement points. Additionally, estimating a number of hours from the frequency corresponding to the proportion of the overall amount of time covered by ESM contacts may also be error prone. Thus, in our analyses, we rely on the pure frequency and do not estimate durations.

As an estimate of the frequency of activities, we calculated the average daily frequency of reported activity types by summing up the frequencies for each activity (labor market, errands, housework, and leisure) for each day and then divided this frequency by the number of measurement points per day. In a second step, we calculated the sum of these daily means and divided that sum by the number of days assessed. By using this procedure, we account for varying measurement occasions per day. Since previous empirical findings found structural differences between time allocation on workdays and on weekends, and measurement points in the ESM study accounted for both, we distinguished between weekends (Saturdays and Sundays) and workdays (Monday to Friday). To give a short example of this procedure: imagine a respondent carrying out housework on weekends as follows: on Saturday the respondent reported having done housework on twice out of four measurement occasions and on Sunday, the same person stated having done housework on just one out of six assessment points. For the first day, we calculated a mean of 0.5, while we get 0.167 for the second day.

Assuming that both days constitute the only weekend days for which data has been collected, we get an overall mean of about 0.33.

This variable, therefore, represents a continuous ratio-scaled measure, more specifically a proportional frequency ranging from 0 to 1, where zero indicates that participants reported having never done the corresponding activity, whereas 1 would imply 100 percent of the assessment occasions, i.e., that the person would have done the activity whenever they were required to state their current activity. Since participants were allowed to report further activities (such as being at home with the children, doing nothing, sleeping, or watching TV) that were not included in the present analyses, the cumulated frequencies of all of the four activities may undercut the total of 1, thus 100 percent, per day.

While our key independent variables in the following analyses constitute the prevalence of activities as derived from the experience sampling section, we included some additional variables in our models. In particular, two methodological as well as key socio-demographic and socio-economic variables have been included. Since the probability of capturing certain activities may depend on the time pagers have been allowed to start contacting the participants, we include this information as one methodological control variable. For instance, participants defining 11.00 a.m. as the start time may be less likely to be contacted while they are working compared to respondents who set the start to 6.30 a.m.

Furthermore, systematic differences might result from missing observations. For example, one could assume that those people with missing assessment points differ in some unobserved aspects from those having always been successfully contacted. Therefore, we also include a variable capturing whether or not the participant was given additional assessment days as a result of repeated missing responses to cell phone requests. This variable is subdivided into three categories, 0 for those having no additional day, those with one day and those with more than one additional day. With 0 being the reference category, thus, two dummy variables accounted for missing responses.

Furthermore, socio-demographic and socio-economic variables such as gender, age, education, employment status and the presence of children under the age of 6 and children from 7 to 14 years are incorporated in the multivariate analyses, as time use patterns differ according to these key characteristics (Beaujot & Liu 2005; Kitterød & Lyngstad 2005; Schulz & Grunow 2007).

3.3 Estimation method

First, we present descriptive results for labor market, errands, housework, and leisure. Specifically, we compare the distributions of time allocation elicited by standard survey questions used in the ESM Study (which has a small sample size and is a convenience sample instead of a random sample) to that of the SOEP 2007. Whereas a high congruence can be taken as initial evidence that the SOEP survey questions perform well, we further examine the correspondences of time spent on the four activities with the frequency variable obtained by the experience sampling section of the ESM Study, i.e., referring to the same respondents. On this descriptive level, we take the rank of activities in their relative importance as an indicator of the correspondences of both measurement techniques.

In a further step, we employ OLS estimations to examine the congruence of both measurement instruments. While the dependent variable in our multivariate analyses constitutes hours as obtained by survey questions, our key explanatory variable is the overage proportional frequency of activities per day generated from the cell phone data. For every activity, we separately run a baseline specification, including only the ESM frequency measure and the two core methodological variables described above. A second model incorporates further control variables such as gender, age, education, etc. The second specification aims at detecting whether certain population groups systematically differ in their reported time spent on activities not entirely captured by the ESM frequency.

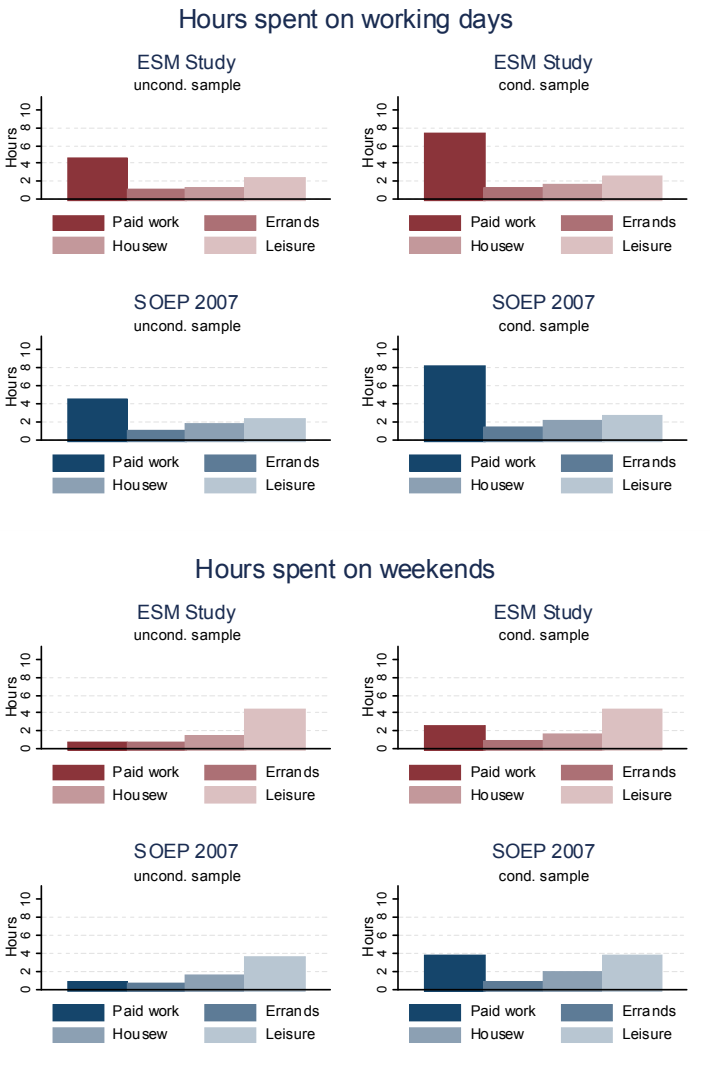
OLS estimation strategies rely on the assumption that the variables are normally distributed and the correlation between explanatory and outcome variables is linear. As an alternative model we run Tobit regressions as robustness checks. Since hours are censored to the left – as well as to the right – by excluding negative values, the Tobit technique may serve as the better estimation strategy for our outcome variable.

4 Summary statistics and descriptive results

In this section we present the descriptive results on hours spent on paid work, errands, housework, and leisure activities. First, we focus on a brief comparison between hours spent on these four activities in the ESM Study compared to the SOEP 2007. As depicted in the upper left part of Figure 2 we see a high congruence in reported hours spent for paid work, errands, housework and leisure for working days elicited by survey questions in the SOEP

2007 and the ESM Study. In both studies people stated that they dedicated about 4.5 hours to market work. This is in line with previous research (Kramer 2004; McGinnity & Russell 2007) and, thus, our data confirms the important role of paid work in people’s time allocation. Leisure demonstrates the second most important activity. Respondents dedicated about 2.5 hours to leisure time activities on workdays. In contrast, errands and housework play a rather subordinated role, since only about one hour and 1.5 hours were devoted to errands and free time activities respectively. It is clear that, according to the ESM Study, people engaged

Figure 2: Comparison of average hours spent on different activities compared according to survey questions in the ESM Study and the SOEP



Notes: The unconditional sample includes the full samples, whereas in the conditional sample those participants are excluded that report to never engage in the activity of interest, i.e., zero hours.
 Source: SOEP 2007 V25, ESM Study 2006/07 & 2009, N= 20,886 SOEP, N= 336 ESM Study. Data weighted. Own calculations.

somewhat less in housework and slightly more in leisure time activities. These differences appear to be statistically significant. Looking at the conditional sample, i.e., those people who stated that they were at least sometimes occupied with the activity of interest, considerable similarities with hours spent on paid work and housework are also manifest. However, the rather small incongruences for errands are significant.

At first glance, looking at weekends in the lower part of Figure 2 suggests very high correspondences with regard to the unconditional samples. Minor differences can only be seen for working hours and leisure activities. In particular, according to the ESM Study, paid work seems to play a stronger role, while leisure time is somewhat less important compared to the SOEP 2007. These differences, and also the difference that people spent more time on errands compared to the SOEP 2007, are significant. The same is true for the conditional population doing the activities to at least some extent on weekends, although the differences are overall rather small. Despite the t-statistics indicating significant differences between time allocation patterns in the ESM Study and the SOEP 2007, we would nevertheless argue that the hours spent on those activities reveal considerable similarities across both samples.

While we referred to the ESM data for the adult population in the previous comparison of hours between the ESM Study and the SOEP 2007, we now use the full ESM Study sample to investigate the correspondences between the survey questions and the experience sampling section.

Table 1: Average number of reported activities in the ESM sample for men and women

	Men		Women	
	Workdays	Weekends	Workdays	Weekends
Paid work	1.96	0.36	1.54	0.24
Errands	0.22	0.20	0.25	0.15
Housework	0.55	0.56	0.73	0.82
Leisure	0.94	1.45	0.80	1.27

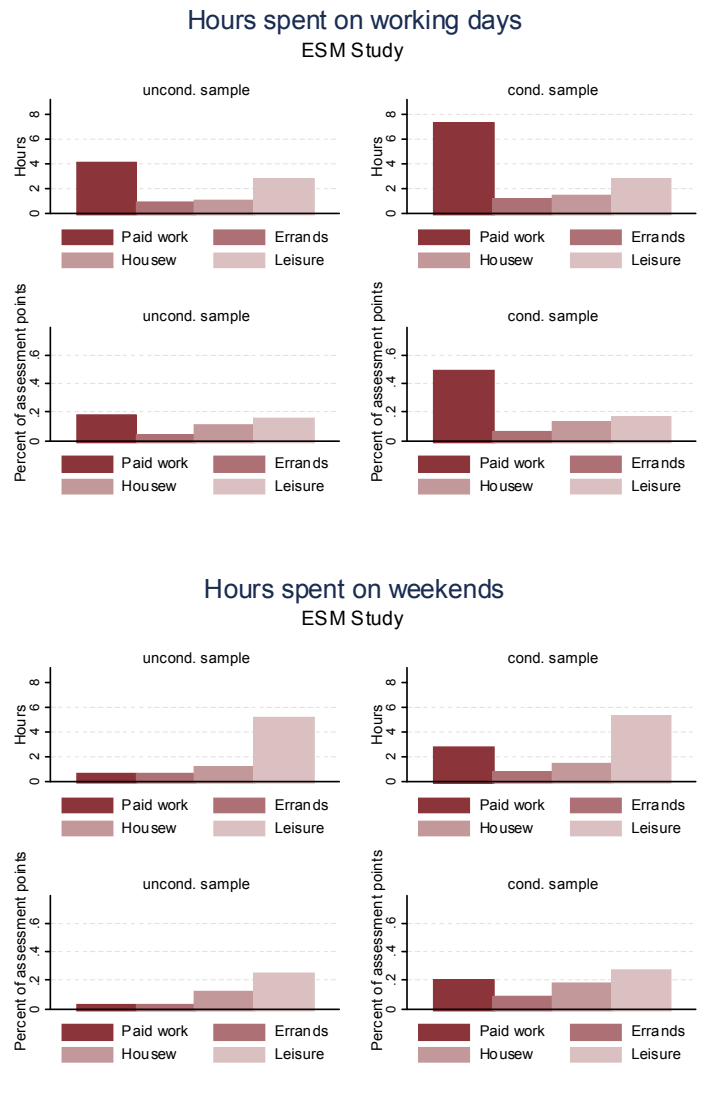
Notes: Participants were contacted 6 times a day. For paid work, those in education are not considered, which reduces the sample to 387 participants.

Source: ESM Study 2006/07 & 2009, N=433. Own calculations.

Before turning to this comparison, we would like to draw attention to Table 1 where the average number of reports on the different activities in the ESM sample for workdays and weekends separately for men and women is illustrated. The highest number of reports is given for gainful employment (working time) which is reported 1.96 times out of a maximum of 6 times for men and 1.54 times for women on workdays. The other activities are, however, considerably less often reported. Housework, for instance, is reported only 0.55 times by men and 0.73 times by women. On the weekend, leisure represents the activity most often recorded using the cell phone method. Out of the maximum 6 time points, males state as being engaged in leisure activities about 1.45 times, whereas the equivalent statistic for females is 1.27 time points. It is apparently rather the short-term activities in which only small numbers report to engage.

Given the different scales for both measures, the daily number of hours versus the daily proportional frequency, we cannot compare the absolute values of the scales, but are restricted to the rank. The upper part of Figure 3 shows the distribution of time allocation captured with both instruments for working days. Concentrating on the unconditional sample on the left, Figure 3 indicates that the rank of activities is similar. While paid work ranks in the highest position, leisure time activities take the second position, followed by housework and errands. However, comparing the time allocation derived from the experience sampling section reveals that the latter varies rather strongly, while with respect to the proportional frequencies considerably less variation occurs. Specifically, with regard to working time as well as to leisure activities, comparing the graphs indicates rather significant inconsistencies. When assessed using survey questions, paid work and leisure time activities both seem to play a more important role than other activities when compared with equivalent results from experience sampling. The graph for the conditional sample, on the other hand, implies a higher congruence with regard to working time as that activity takes a more prominent role. However, in addition to the repeatedly consistent ranking, we also see similar proportional frequency with regard to housework and leisure time activities. Also, looking at correlations shows that both measures are significantly correlated on at least a 5 percent level. However, the strength of correlation varies from about 0.18 to 0.74, being lowest for errands and leisure, moderate for housework and highest for working time (not shown).

Figure 3: Comparison of time use captured by survey questions and experience sampling in the ESM Study



Notes: The unconditional sample includes the full samples, whereas the conditional sample excludes those participants who were never accessed while being engaged in the activity of interest, i.e., having a frequency of zero.
Source: ESM Study 2006/07 & 2009, N= 433. Own calculations.

Turning to weekends, the comparison in the bottom part of Figure 3 again illustrates that, if we look at overall importance, the ranks of activities do not vary between the two measurement techniques. However, the proportion of leisure time activities compared to other activities stands out more clearly when we use survey questions, while its importance seems somewhat less distinct using experience sampling. This is particularly true with regard to the distribution of activities in the conditional sample shown in top left hand corner. Overall, the correlations are lower for weekends and less consistently significant regarding the

unconditional and particularly the conditional samples. However, since both measurement techniques provide different scales, opportunities to analyze the differences from a bivariate perspective are limited. We, therefore, address the question of statistical associations in more depth in the following section.

5 Multivariate results

In line with previous research, we expect individual differences in labor market activities to be captured correctly by survey questions, particularly since these activities are presumably carried out on a regular basis and are normally externally structured by institutions. On the other hand, we assume less efficient performance of survey questions for errands, housework, and leisure as these are less distinctive or prominent. The question of whether different subgroups such as labor market groups, for instance, respond to experience sampling and stylized measures in a similar manner is also of interest to us.

However, while we interpret strong significant associations as being in favor of good survey question performance, it is hard to interpret inconsistencies with regard to the data quality of survey questions and experience sampling, respectively. Although ESM accounts for some weaknesses associated with survey questions, by capturing activities in real time, it remains unclear to what extent ESM data is a fully reliable method due to its small number of measurement points for infrequent activities. Given this limitation, we restrict our analyses to mere associations and do not draw causal conclusions.

Table 2 shows OLS models for hours spent on paid work, errands, housework and leisure. For each of the activities we run two models. While our baseline model includes only the methodological variables and our main explaining variable, which is the corresponding daily proportional frequency deduced from the experience sampling section, our second models additionally incorporate basic socio-demographic and socio-economic characteristics. The latter enables us to reveal systematic differences in stated survey questions not fully captured by the ESM frequency measure. For better interpretation of the association between the ESM frequency measure and reported hours we rescaled the frequency variable from 0 to 10, so that a one unit change increase in this variable represents an increase of being contacted while undertaking a certain activity of 10 percent.

In the first two columns of Table 2, our results for time devoted to the labor market show high correspondence for paid work. A ten percent increase in having been contacted while doing paid work is associated with having reported an additional 1.2 hours of paid work in our baseline model. This slightly decreases in our extended model to about one hour, where we also see significant associations for females, age and educational attainment. Women tend to report about 40 fewer minutes spent on paid work, while increasing age is negatively correlated with time devoted to paid work by about 2.4 minutes per year. With respect to education, model 2 reveals that, compared to people having completed general secondary school, those with intermediate education report about 1.6 more hours spent on labor market work, while those with an upper secondary school certificate report about 1.9 more hours. These effects can be translated into absolute working hours per work day when we add the constant of about 3.1 hours to the effects for the different predictors. The overall relatively low absolute working time results from the fact that the estimations account for the total sample, only excluding those that are currently in education (see also section 6 for robustness checks).

The significant correlations between time use and control variables imply that different time allocation patterns regarding paid work might not be adequately captured either by survey questions or by experience sampling. Specifically, it could either imply that working time has been underrepresented using the cell phone measure or that people tend to miscalculate the time they dedicate to market work. However, also according to the high explained variance, amounting to 54.5 percent in model 1 (compared to the only slightly higher R-squared of 58.9 in model 2 which includes the controls), both measures are highly correlated suggesting an overall good performance of survey question on working hours.

Turning to errands, seen in column 3 and column 4, our results indicate that time spent on errands captured by experience sampling is positively correlated with respondents' reported number of hours spent on this activity. Specifically, a 10 percent increase in having been contacted while doing errands is associated with an additional 20 minutes reported in survey questions. However, the association is much lower for errands than for labor market work, which is indicated by the relatively low R-squared, amounting to 8.4 percent in our baseline model. Including further socio-demographic and socio-economic variables in model 2 results in a drop in the effect size of the daily average frequency of about 10 minutes. However, we find significant positive associations for females, being unemployed or retired and negative correlations for those having not finished school education. The extended model is also much

Table 2: Time spent on paid work, errands, housework, and leisure on workdays (OLS estimation with robust standard errors)

Dependent variable	Paid work		Errands in hours		Housework in hours		Leisure in hours	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Daily frequency of reporting the relevant activity	1.183** (20.539)	1.056** (17.258)	0.313** (4.461)	0.152* (2.194)	0.302** (7.826)	0.188** (4.584)	0.378** (4.883)	0.174* (2.150)
Female		-0.671* (-2.222)		0.224** (3.408)		0.592** (8.188)		-0.269 (-1.525)
Age		-0.040** (-4.715)		-0.004 (-1.154)		-0.005 (-1.243)		-0.008 (-0.750)
Still in school (Ref: general secondary school)		(excluded)		-0.483** (-2.999)		-0.415* (-2.346)		0.284 (0.664)
Intermediate school		1.591** (3.702)		-0.050 (-0.449)		-0.096 (-0.795)		-0.222 (-0.757)
Upper secondary school		1.916** (4.786)		0.021 (0.202)		-0.137 (-1.192)		-0.121 (-0.433)
Employed full-time (Ref: part-time)				-0.107 (-0.921)		-0.334** (-2.613)		-0.828** (-2.649)
Unemployed				0.723** (4.134)		0.697** (3.610)		0.101 (0.213)
Retired				0.423** (2.777)		0.082 (0.487)		0.667 (1.625)
Still in education				0.049 (0.284)		-0.229 (-1.218)		-0.136 (-0.295)
Intercept	0.751 (0.610)	3.062* (2.224)	1.194** (4.430)	1.011** (2.937)	1.204** (3.784)	1.524** (4.077)	1.500* (2.187)	3.087** (3.376)
N	375	375	403	403	404	404	406	406
R-square	0.545	0.589	0.084	0.235	0.177	0.379	0.074	0.183

Notes: t statistics in parentheses, * significant at 5%, ** significant at 1%, *** significant at 0.1%.

Dependent variable is time use in hours stated for workdays (Mondays to Fridays) as measured by survey questions on paid work, errands, housework, and leisure. The main explanatory variable is the ESM daily proportional frequency. For a better interpretation this variable was rescaled ranging from 0 to 10. A one unit change, therefore, represents an increase of being contacted while engaged in the corresponding activity of 10 percent. The first specification always includes the start time of ESM contacts and dummy variables indicating whether people have additional assessment days. The second specification also controls for having children under the ages of 6 or 14.

Source: ESM-Study 2006/2007 & 2009. Own calculations.

better for explaining the variance in people's time spent on errands, increasing to 23.5 percent. The significant effects of the controls as well as the increased explanatory power of model 2 indicate that time use of women and people not engaged in the labor market is less properly captured by either of the instruments.

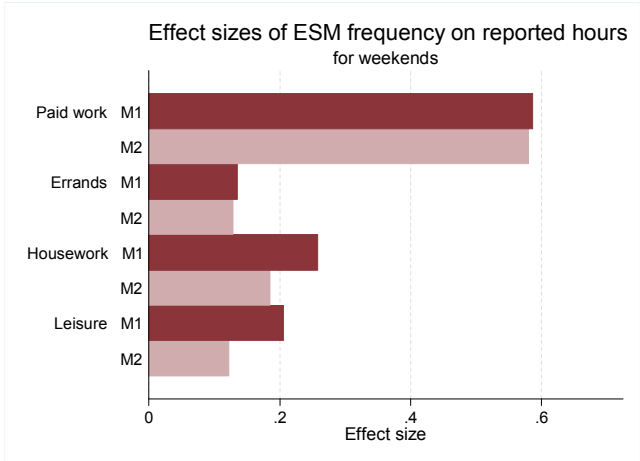
Almost the same pattern appears when looking at housework in columns 5 and 6 in Table 2. While in the baseline model a ten percent change in the average daily frequency results in about 20 more minutes reported for housework, the effect almost bisects in value in the extended model. Women tend to report spending significantly more time on housework than has been captured via cell phones, whereas those still in school education report significantly less time than experience sampling indicates. Specifically, women spend an additional 35 minutes on housework and pupils 25 fewer minutes. Showing no significant correlation for retired people, model 2 also reveals that unemployed respondents report spending 42 minutes more on housework compared to part-time employed. Those in full-time employment, on the other hand, report significantly less time, about 20 minutes, spent on housework. The substantial rise in explained variance from model 1 to model 2 also suggests that, though the associations between both measurement instruments are significant, time spent on housework, as measured in surveys, cannot be properly 'explained' by experience sampling measures alone.

Finally, if we shift our attention to leisure time in Table 2, model 1 in column 7 shows that, though significantly correlated, the daily proportional frequency of being assessed while doing leisure activities only partially accounts for the hours people report as spending on leisure activities: The R-squared only amounts to 7.4 percent. Turning to model 2, the effect size of the proportional frequency is reduced by half when we include additional variables. Interestingly, apart from being full-time employed which is negatively correlated with hours spent on leisure time, none of the socio-demographic and socio-economic characteristics included are significantly associated with this activity.

Similar to our estimations for time spent on working days, we also find significant correlations between both measures for weekends. Figure 4 shows the effect sizes of the associations for the four activities of interest, separately for our baseline model and the extended model. Similar to working days, the effects for paid work are considerably stronger than for the other activities. While for labor market activities they amount to about 0.6, they range from 0.12 to 0.26 for the other activities. Overall, effect sizes reveal to be somewhat lower for weekends than for working days. However, apart from leisure time activities in the

second model, which only meet the 10 percent criteria, all effects are significant on the 5 percent level. Significant differences due to socio-demographic and socio-economic characteristics are overall lower compared to the models for workdays (results not shown in detail).

Figure 4: Effects of time use elicited by experience sampling on hours spent on weekends (OLS estimation with robust standard errors)



Notes: Effects coefficients of daily frequency on reported hours estimated using OLS estimations with robust standard errors. Dependent variable is time use in hours stated for weekends as measured by survey questions asking on paid work, errands, housework, and leisure. Models 1 control for start time of ESM contacts and dummy variables indicating whether people have additional assessment days. Models 2 additionally control for age, gender, education, labor market status, and having young children in the household.
 Source: ESM Study 2006/07 & 2009, N= 433. Own calculations.

As we have seen, for all activities reported hours using survey questions are significantly correlated with daily frequencies collected by the experience sampling method. This is equally true for working days as for weekends. However, our analyses also imply considerable differences in the strength of association. While we find strong correlations for paid work, those for errands, housework and leisure are moderate or rather small. Given the relatively high amount of time dedicated to leisure, about 2.8 hours on workdays and 4.5 hours on weekend days, we would have expected higher effect sizes for this activity in particular.

Yet, the rather small correlations for errands, housework and leisure time activities are in line with prior research literature and, thus, our hypotheses. Since these activities are less prominent and, as opposed to labor market activities, less externally structured, our findings

may nudge us towards the previously suggested notion that they are harder to accurately capture using survey questions. On the other hand, the weaker correlations could also reveal shortcomings in assessing these activities using experience sampling.

6 Robustness checks

So far, we have used ordinary least squares (OLS) to analyze the data quality of time use captured by survey questions. OLS estimations require normal distributed errors of the outcome variables to measure the statistical significance of the explanatory variables correctly. Since hours spent on different activities are left-censored, the essential assumption may not hold for our outcome variable. Tobit estimations may be a more appropriate technique since they account for censorship (Greene 2003: 764ff.). Therefore, we also run our different specifications for the activities analyzed on workdays and weekends using Tobit regression models.

Overall, the estimations (see appendix Table B1) confirm the findings presented in the previous section. The Tobit models also show a statistically significant correlation between time use data elicited through survey questions and experience sampling techniques found using ordinary least square estimations. For paid work the Tobit models even increase the associations between both measures. While in Table 2 a ten percent change in the frequency of labor market activity on working days resulted in about an hour increase, accounting for censorship raises the correlation to 1.4 hours. For errands, housework, and leisure the associations decrease slightly, but are still significant. Furthermore, comparing the Tobit models for working days with the OLS models in Table 2 demonstrates high correspondences in the effects, although the effects for education on reported working hours, in particular, are considerably larger in the Tobit model. Overall the Tobit estimations strongly support the results based on OLS estimations presented above.

7 Conclusion

Previous literature on the data quality of time use information has compared stylized survey measures to time diaries. Resting on the assumption that time diaries provide correct data on time use, differences in the amount of time reported by participants has mostly been

interpreted as poor survey question performance. These comparisons have implied that survey questions perform rather poorly with regard to diffuse, sporadic and less prominent activities. Discussing this literature and their findings, we address the question of survey question performance by firstly, looking at the reliability of different samples drawing on identical survey questions. Secondly, we use an alternative measurement method for comparisons. Unlike previous studies, we apply experience sampling data in order to examine the survey question performance with regard to paid work, errands, housework, and leisure.

In our study, as well as asking people about their time allocation using survey questions, experience sampling was used to collect data on people's activities over a period of 12 hours per day for several days. Using these rich data, our analyses show that there are highly significant associations between survey questions and the actual activities people perform during their awake time. This is equally true for working days as for weekends, although the associations are somewhat stronger for working days. Regarding the different activities, we find particularly strong correlations for paid work, while those for errands, housework and leisure reveal as being somewhat smaller. Despite the significant associations between both measures with respect to errands and leisure, the number of hours reported for these two activities seems less strongly connected with people's daily frequency, indicated by the low explanatory power of the corresponding models. With respect to errands and leisure the frequency of being engaged in these activities elicited by ESM does not 'explain' the number of hours people report in survey questions. This is also captured by the fact that the models get more effective when we include further socio-demographic and socio-economic characteristics. Incorporating further characteristics also reveals that subgroups systematically differ in the hours they report on the four activities – even after 'controlling' for the ESM frequency measure. The latter indicates that the accuracy of time use data varies somewhat for different population groups.

Overall, in line with previous findings, we find that survey questions tend to satisfactorily capture longer activities that are externally settled, while the rather low correspondence with regard to more short-term and diffusive activities may indicate that survey questions are less suitable for these activities. Yet, this interpretation relies on the assumption of the ESM's superior performance. Due to the scarcity of empirical evidence on the performance of ESM, the differences between both measures could also be the result of error-prone estimations of short-term activities using the cell phone method.

In order to further examine the data quality of survey questions compared to ESM, more empirical evidence is required on its data quality. We, therefore, plan to compare a fully fletched Day Reconstruction Method survey (DRM, see also Kahnemann et al. 2004) and experience sampling in a future study. If people's time use constitutes the main focus of a study, it may be likely that the full DRM approach delivers the best results.⁵ Firstly, it is relatively easy for respondents to mentally reconstruct the activities undertaken on the day the data is collected, and secondly, DRM may be less error prone with regard to short-term activities.

For the time being, until there is more methodological evidence for the experience sampling method, our comparison of ESM and common survey questions as used in the German Socio-Economic Panel Study (SOEP) indicates that the stylized measures for important activities such as gainful employment are effective, whereas they seem to be somewhat less suitable for collecting accurate information on errands and, to some extent, also on leisure.

⁵ However, if researchers are interested in the actual correlation of time use and the mood of respondents at different points over a period of time, there cannot be any doubt that there is no alternative survey technique to ESM (Riediger et al. 2009).

References

- Ahn, Namkee, Juan F. Jimeno, and Arantza Ugidos (2005): 'Mondays in the Sun': Unemployment, Time Use, and Consumption Patterns in Spain. In: Hamermesh, Daniel S. and Gerard A. Pfann (Eds.) (2005): *The Economics of Time Use*. Emerald Group Pub. 237-259.
- Andorka, Rudolf (1987): Time Budgets and Time Uses. *Annual Review of Sociology* 13: 149-164.
- Beaujot, Roderic and Jianye Liu (2005): Models of Time Use in Paid and Unpaid Work. *Journal of Family Issues* 26: 924-946.
- Bevans, George Esdras (1913): *How Workingmen Spend Their Spare Time*. New York: Columbia University Press. Available from: <http://www.archive.org/details/howworkingmenspe00bevarich>.
- Blair, Edward and Scot Burton (1987): Cognitive Processes Used by Survey Respondents to Answer Behavioral Frequency Questions. *Journal of Consumer Research* 14: 280-288.
- Bonke, Jens (2005): Paid Work and Unpaid Work: Diary Information versus Questionnaire Information. *Social Indicators Research* 70: 349-368.
- Csikszentmihalyi, Mihaly (1987): Validity and Reliability of Experience-Sampling Method. *The Journal of Nervous and Mental Disease* 175: 526-536.
- Csikszentmihalyi, Mihaly and Jeremy Hunter (2003): Happiness in Everyday Life: Uses of Experience Sampling. *Journal of Happiness Studies* 4: 185-199.
- 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: 1-33.
- Friedberg, Leora and Anthony Webb (2005): *The Chore Wars: Household Bargaining and Leisure Time*. University of Virginia. Available from: <http://atususers.umd.edu/papers/friedberg.pdf>.
- Gershuny, Jonathan I (Ed.) (2000): *Changing Times. Work and Leisure in Postindustrial Society*. Oxford: Oxford University Press.
- Greene, William H. (2003): *Econometric Analysis*. Upper Saddle River, New Jersey: Pearson Education.
- Harvey, Andrew S. (1993): Guidelines for Time Use Data Collection. *Social Indicators Research* 30: 197-228.
- Isengard, Bettina (2005): Freizeitverhalten als Ausdruck sozialer Ungleichheiten oder Ergebnis individualisierter Lebensführung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 57: 254-277.
- Jahoda, Marie, Paul F. Lazarsfeld, and Hans Zeisel (1971 [1933]). *Marienthal: The Sociography of an Unemployed Community*. Aldine Atherton, Chicago.
- Juster, F. Thomas, Hiromi Ono, and Frank P. Stafford (2003): An Assessment of Alternative Measures of Time Use. *Sociological Methodology* 33: 19-54.
- Kahnemann, Daniel, Alan B. Krueger, David A. Schkade, Norbert Schwarz, and Arthur A. Stone (2004): A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method. *Science* 306: 1776-1780.
- Kan, Man Yee (2007): Measuring Housework Participation: The Gap between „Stylized“ Questionnaire Estimates and Diary-based Estimates. *Social Indicators Research* 86: 381-400.
- Kan, Man Yee and Stephen Pudney (2007): Measurement Error in Stylised and Diary Data on Time Use. ISER Working Paper 2007-03. Institute for Social and Economic Research, University of Essex. Available from: <http://www.iser.essex.ac.uk/publications/working-papers/iser/2007-03.pdf>.

- Kitterød, Ragni H. (2001): Does the Recording of Parallel Activities in Time Use Diaries Affect the Way People Report Their Main Activities? *Social Indicators Research* 56: 145-178.
- Kitterød, Ragni H. and Torkild H. Lyngstad (2005): Diary versus Questionnaire Information on Time Spent on Housework – The Case of Norway. *electronic International Journal of Time Use Research* 2: 13-32.
- Klumb, Petra L. and Margret M. Baltes (1999): Validity of Retrospective Time-Use Reports in Old Age. *Applied Cognitive Psychology* 13: 527-539.
- Kramer, Caroline (2004): Zeitverwendung. In: Statistisches Bundesamt (Ed.): *Datenreport 2004. Zahlen und Fakten über die Bundesrepublik Deutschland*. Bonn: Bundeszentrale für politische Bildung. 545-554.
- Mannell, Roger C. and Jiri Zuzanek (1991): The Nature and Variability of Leisure Constraints in daily Life: The Case of the Physically Active Leisure of Older Adults. *Leisure Sciences* 13: 337 – 351.
- Marini, Margaret M. and Beth A. Shelton (1993): Measuring Household Work. Recent Experience in the United States. *Social Science Research* 22: 361-382.
- McGinnity, Frances and Helen Russell (2007): Work Rich, Time Poor. Time-Use of Women and Men in Ireland. *The Economic and Social Review* 38: 323-354.
- Merz, Joachim (2010): Time Use and Time Budgets. In: German Data Forum (RatSWD) (Ed.): *Building on progress: Expanding the research infrastructure for the social, economic, and behavioral sciences*. Opladen: Budrich UniPress. Vol. 1. 413-449.
- Merz, Joachim and Dagmar Kirsten (1999): Extended Income Dynamics: Valuing Household Production and its Distributional Impacts -- A Ten Years Panelanalysis for Professions, Entrepreneurs and Employees. In: Merz, Joachim and Manfred Ehling (Eds.): *Time Use -- Research, Data and Policy*. Baden-Baden: Nomos Verlag. 351-382.
- Merz, Joachim and Klaus G. Wolff (1993): The Shadow Economy: Illicit Work and Household Production - A Microanalysis of West Germany. *The Review of Income and Wealth* 39: 177-194.
- Niemi, Iliris (1993): Systematic Error in Behavioral Measurement: Comparing Results from Interview and Time Budget Studies. *Social Indicators Research* 30: 229-244.
- Otterbach, Steffen and Alfonso Sousa-Poza (2010): How Accurate are German Work-time Data? A Comparison of Time-diary Reports and Stylized Estimates. *Social Indicators Research* 97: 325-339.
- Plewis, Ian, Rosemary Creeser, and Ann Mooney (1990): Reliability and Validity of Time Budget Data: Children's Activities Outside School. *Journal of Official Statistics* 6: 411-419.
- Riediger, Michaela (2010). Experience sampling. In: German Data Forum (RatSWD) (Ed.): *Building on progress: Expanding the research infrastructure for the social, economic, and behavioral sciences*. Opladen: Budrich UniPress. Vol. 1. 581-594.
- Riediger, Michaela, Florian Schmiedek, Gert G. Wagner, and Ulman Lindenberger (2009): Seeking Pleasure and Seeking Pain: Differences in Pro- and Contra-Hedonic Motivation from Adolescence to Old Age. *Psychological Science* 20: 1529-1535.
- Riediger, Michaela, Cornelia Wrzus, Florian Schmiedek, Gert G. Wagner, and Ulman Lindenberger (2011): Is seeking bad mood cognitively demanding? Contra-hedonic orientation and working-memory capacity in everyday life. *Emotion* 11: 656-665.
- Robinson, John P (1999): The Time-Diary Method: Structure and Uses. In: Pentland, Wendy E. et al (Eds.): *Time Use research in the Social Sciences*. New York et al: Kluwer Academic/Plenum Publishers. 47-89.
- Robinson, John P. (1985): The Validity and Reliability of Diaries versus Alternative Time Use Measures. In: Juster, F. Thomas and Frank P. Stafford (Eds.): *Time, Goods, and Well-being*. Ann Arbor, MI: The University of Michigan. 33-62.

- Rydenstam, Klas (2001): Paid and Unpaid Work. The Apples and Pears of Time Use Statistics. Paper prepared for the 2001 IATUR Conference in Oslo, October 3-5, 2001. Available from: http://www.ssb.no/english/about_ssb/conference/iatur_2001/rydenstam.pdf .
- Schulz, Florian and Daniela Grunow (2007): Tagebuch versus Zeitschätzung. Ein Vergleich zweier unterschiedlicher Methoden zur Messung der Zeitverwendung im Haushalt. *Zeitschrift für Familienforschung* 19: 106-128.
- Schwarz, Norbert (2007): Retrospective and Concurrent Self-Reports: The Rationale for Real-Time Data Capture. In: Stone, Arthur; Saul Shiffman, Audie Atienza, and Linda Nebeling (Eds.) (2007). *The science of real-time data capture: Self-reports in health research*. New York: Oxford University Press. 11-26.
- Schwarz, Norbert and Daphna Oyserman (2001): Asking Questions About Behavior: Cognition, Communication, and Questionnaire Construction. *American Journal of Evaluation* 22: 127-160.
- Siedler, Thomas, Jürgen Schupp, C. Katharina Spiëß, and Gert G. Wagner (2009): The German Socio-Economic Panel as Reference Data Set. *Schmollers Jahrbuch* 129: 367 – 374.
- Tourangeau, Roger, Land J. Rips, and Kenneth Rasinski 2000: *The Psychology of Survey Response*. Cambridge, NY: Cambridge University Press.
- Wagner, Gert G., Joachim R. Frick, and Jürgen Schupp (2007): The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. *Schmollers Jahrbuch* 137: 139-169. Available from: [http://schmollersjahrbuch.diw.de/schmollersjahrbuch/webcontent/2007/Wagner et al.pdf](http://schmollersjahrbuch.diw.de/schmollersjahrbuch/webcontent/2007/Wagner%20et%20al.pdf) .

APPENDIX

A Sample comparison of the ESM Study and SOEP 2007

Table A1: Summary statistics and representativeness of the sample

	ESM Study		SOEP 2007	
	Frequency	Percent	Frequency	Percent
Gender				
Female	169	50.3	10,695	51.9
Male	167	49.7	9,928	48.1
Total	336	100.0	20,623	100.0
Age				
18-29	79	23.5	3,473	16.8
30-49	115	34.2	7,247	35.1
50-64	75	22.3	4,667	22.6
65+	67	20.0	5,236	25.4
Total	336	100.0	20,623	99.9
Employment Status				
Employed full-time	131	39.0	7,175	34.8
Employed part-time	56	16.7	4,308	20.9
Not employed	149	44.3	9,140	44.3
Total	336	100.0	20,623	100.0
Educational attainment				
Still in school	18	5.5	285	1.5
Left without graduating	0	0.0	534	2.9
Secondary general school level	50	15.4	7,548	40.3
Intermediate school	105	32.3	5,708	30.5
Upper secondary school	152	46.8	4,646	24.8
Total	315	100.0	18,721	100.0

Notes: ESM sample and SOEP participants aged 18 or over

Source: ESM-Study 2006/2007 & 2009, SOEP 2007 (data weighted). Own calculations.

Comparing our ESM sample to that of the German Socio-Economic Panel Study (SOEP) 2007, which is - after proper re-weighting - representative of the German population with regard to the main socio-demographic and socio-economic characteristics, indicates that the ESM sample represents the German population well with regard to gender, age, and employment status, but not with respect to education. Despite the fact that the majority of people in Germany has completed intermediate school, in our ESM sample, the proportion is only 15 percent whereas about 79 percent in the sample attained higher educational qualifications, with almost 50% having the general qualification for university entrance. Since the educational bias violates the representativeness of our sample considerably, we apply a self-computed redressment weight for educational attainment for descriptive analyses which puts less weight on participants with high educational attainments and more weight on those with secondary general school level education.

Table B1: Time spent on paid work, errands, housework and leisure on workdays (Tobit estimation)

Dependent variable	Paid work		Errands in hours		Housework in hours		Leisure in hours	
	Workdays	Weekends	Workdays	Weekends	Workdays	Weekends	Workdays	Weekends
Daily frequency of reporting the relevant activity	1.427** (12.837)	1.988** (6.184)	0.184* (2.126)	0.151* (2.494)	0.222** (4.351)	0.205** (5.025)	0.180* (2.172)	0.124+ (1.679)
Female	-1.385* (-2.430)	-0.866 (-1.264)	0.299** (3.581)	0.119+ (1.815)	0.762** (8.353)	0.660** (6.646)	-0.258 (-1.438)	-0.363 (-1.365)
Age	-0.076** (-4.363)	-0.025 (-1.157)	-0.005 (-1.106)	0.007+ (1.821)	-0.007 (-1.280)	0.000 (0.082)	-0.008 (-0.733)	-0.037* (-2.356)
Still in school (Ref: general secondary school)	--	--	-0.661** (-3.254)	-0.405* (-2.568)	-0.554* (-2.516)	-0.867** (-3.684)	0.350 (0.803)	0.836 (1.282)
Intermediate school	3.331** (4.032)	3.123** (2.989)	-0.059 (-0.428)	-0.131 (-1.211)	-0.103 (-0.689)	-0.099 (-0.619)	-0.186 (-0.621)	0.908* (2.028)
Upper secondary school	3.828** (4.938)	4.156** (4.061)	0.018 (0.136)	-0.116 (-1.142)	-0.157 (-1.100)	-0.129 (-0.850)	-0.091 (-0.319)	0.578 (1.363)
Employed full-time (Ref: part-time)			-0.109 (-0.748)	0.360** (3.099)	-0.358* (-2.255)	0.352* (2.098)	-0.922** (-2.894)	-0.489 (-1.039)
Unemployed			0.783** (3.611)	0.169 (0.955)	0.740** (3.105)	0.008 (0.030)	0.090 (0.188)	-0.570 (-0.794)
Retired			0.511** (2.666)	-0.273+ (-1.831)	0.106 (0.508)	-0.187 (-0.842)	0.648 (1.554)	0.337 (0.546)
Still in education			0.048 (0.224)	0.339* (2.020)	-0.224 (-0.967)	0.246 (0.995)	-0.179 (-0.380)	0.229 (0.330)
Intercept	0.350 (0.132)	-3.887 (-1.218)	0.904* (2.067)	0.248 (0.707)	1.476** (3.161)	0.985+ (1.941)	3.034** (3.262)	6.635** (4.765)
Sigma	4.748** (17.445)	3.878** (8.983)	0.791** (23.292)	0.610** (23.315)	0.861** (23.470)	0.917** (24.491)	1.744** (27.614)	2.610** (28.182)
N	375	378	403	385	404	393	406	407
Pseudo R-square	0.157	0.158	0.103	0.080	0.155	0.138	0.050	0.040

Notes: t statistics in parentheses, * significant at 5%, ** significant at 1%, *** significant at 0.1%.

Dependent variable is time use in hours stated for workdays (Mondays to Fridays) as measured by survey questions about paid work, errands, housework, and leisure. The main explanatory variable is the ESM daily proportional frequency. For a better interpretation this variable was rescaled ranging from 0 to 10. A one unit change, therefore, represents an increase of being contacted while engaged in the corresponding activity of 10 percent. The first specification always includes the start time of ESM contacts and dummy variables indicating whether people have additional assessment days. The second specification also controls for having children under the ages of 6 or 14.

Source: ESM-Study 2006/2007 & 2009. Own calculations.