

Working Paper 2007-05

Comparing Time Diary Data and
Stylized Time Use Estimates

Florian Schulz

Daniela Grunow

Yale University
New Haven, December 2007

Biographical Information

Florian Schulz is a junior researcher at the Institute for Family Research (ifb) and the University of Bamberg, Germany. He is currently working for his PhD in sociology in the fields of family and life course research. He is studying processes of mate selection and the division of housework in German couples in a (quantitative and qualitative) longitudinal perspective.

Daniela Grunow is a Postdoctoral Associate at the Center for Research on Inequalities and the Life Course. She defended her Ph.D. thesis ‘Convergence, Persistence and Diversity in Male and Female Careers’ at Otto-Friedrich University Bamberg (Germany) in April 2006. Her research addresses the interaction of market work, domestic work and gender relations in divergent welfare regimes from a life-course perspective, applying quantitative as well as qualitative methods.

Comparing Time Diary Data and Stylized Time Use Estimates*

Florian Schulz and Daniela Grunow

Dipl.-Soz. Florian Schulz

University of Bamberg, State Institute for Family Research (ifb)

Lichtenhaidestraße 11

D – 96045 Bamberg, Germany

e-mail: florian.schulz@ifb.uni-bamberg.de

Dr. Daniela Grunow

Center for Research on Inequalities and the Life Course (CIQLE)

Yale University

New Haven, CT 06520-8265, USA

e-mail: daniela.grunow@yale.edu

Abstract

We compare data obtained through time diaries and survey questions on time use. Based on data from a unique German study designed for assessing methodological concerns of time use measurement, we assess the exchangeability of time diaries and time estimates for several statistical procedures. We find that both measures yield rather consistent findings in the aggregate. However, they produce significantly different results in the independent variables when applied to OLS regression analyses. When studying time use as a dependent variable, we conclude that it cannot be taken for granted that both types of data lead to similar conclusions.

JEL Key words: C81, D13, J16, R20

Own keywords: Time use, time diary, time estimation, housework, comparison of measurement techniques.

* This paper was presented at the 28th Annual Conference of the International Association of Time Use Research (IATUR), August 16 – 18, 2006, in Copenhagen (DK), and at the Spring Meeting of the Methodology Section of the American Sociological Association, March 3, 2007, in New Haven (USA). We would like to thank the participants of these conferences for their helpful comments. – An earlier version of this paper was published in German in the "Zeitschrift für Familienforschung" (Schulz and Grunow 2007); yet, this paper goes beyond the scope of the German one in several major aspects.

1 Introduction

This paper assesses and compares the potential of the two dominating strategies of obtaining time use data: time diaries and survey questions on time use. Even though there are many different ways to collect time use data today, these two techniques have proved to be the most feasible in the last decades. While time diaries are still viewed as the most reliable sources of time budgets (e. g. Marini and Shelton 1993; Juster et al. 2003), these data are considerably more complex, more expensive and more time-consuming to conduct and to analyze than survey questions (e. g. Bonke 2005; Kitterød and Lyngstad 2005). Despite the long research tradition and popularity of the questionnaire and diary studies (Juster and Stafford 1991), a major drawback of current time use research appears to be that the survey instruments have not frequently been evaluated and adjusted to the actual needs and analytical potential of today's improved data analysis techniques (Kan and Pudney 2007). While there are many studies which address theoretical or methodological concerns of measuring time use (just two examples: Robinson 1999; Juster et al. 2003), very few involve empirical comparisons. To our knowledge, there are hardly any data sets available which would allow for a *direct* comparison of stylized and diary estimates for the very same respondents. This situation makes it difficult to fundamentally criticize the different measurement techniques, even though it is possible to compare the results of different studies with regard to the content. Thus, it is often assumed that time diaries and survey questions measure the same dimensions of time use in a commensurable way. Yet, recent studies from Great Britain (Kan 2006; Kan and Pudney 2007), Denmark (Bonke 2005) and Norway (Kitterød and Lyngstad 2005) make us doubt that this assumption is true.

The aim of this paper is, thus, to compare the data obtained by time diaries and survey questions on time use directly and empirically, using housework time as an example. Our analysis is based on data which were specifically obtained for this purpose. We gathered information on the every day time use of 286 German couples, who provided us with detailed survey-based estimates of their time use on normal work days and weekends, and, in addition, subsequently completed a seven-day time diary. In the our paper, we summarize the main arguments of the theoretical and methodological discourse of current time use research and illustrate some important features of the data set we use, before turning to the empirical results of our time use data comparison.

2 Time diaries and survey questions on time use – a review

Scholars interested in the study of household labor consider time use data to be the best way to obtain information on the structure of housework activities. For many research applications in this field, time diary and time estimation techniques are in principle equally suited. Ideally, both methods should provide us with equivalent information about how much time the members of a population spend on certain activities. Because

conducting a time use survey is usually cheaper and less time-consuming than conducting a time diary study, the former is often preferred when it comes to designing an empirical study on time use. However, one may easily overlook the fact that both methods are not perfect substitutes and that there are important theoretical and methodological problems which can be addressed appropriately only by the one *or* the other (recently Kan 2006; Kan & Pudney 2007; Bonke 2005; Kitterød and Lyngstad 2005).

In general, the *time diary method* enjoys the reputation of being an exceptionally valid, highly reliable instrument (Marini and Shelton 1993). Robinson argues that time diaries may cover individual behavior in a very sophisticated manner (1999: 56). In practice, the respondents record all their activities for a whole day (24 hours), or for the relevant intervals of a day, chronologically, alongside a given time axis and/or certain given activities. Depending on the degree of standardization, the information gathered in this procedure can be more or less detailed. In an extreme example, the respondents record the course of their day freely by writing down all facts perceived as important (activities *and* time) in their own words. Based on this detailed information, one can draw the picture of a day as it is subjectively seen by the individual. This very "open" method of data collection is, however, associated with quite a complex procedure of data interpretation, thus entailing problems of analytical consistency. That is why standardized diaries are often preferred in large scale data collection; especially in cases where the interest in a predefined set of activities, as well as their exact sequence and timing, is preconditioned by the research problem (for a recent example of standardized diaries in Germany cf. Schulz et al. 2005).

Survey questions on time use follow a different logic: respondents are asked how much time they "normally" spend for given activities in a certain space of time, usually a day, week or month (Juster et al. 2003: 22). Survey questions ask for aggregate time budgets, which the respondents are required to calculate spontaneously during an interview. Based on this technique, it is not possible, however, to draw conclusions concerning the course of the respondents' days or the sequencing of certain tasks. Moreover, the questioning procedure is problematic insofar as it usually remains unclear what is meant exactly by the pre-defined activities such as "housework", or by terms such as "normal day" or "normal week" (Robinson 1999: 57). The latter problem at least can be solved in time diary studies by asking the respondents if the recorded day was a "normal" day in their minds; for the analyses then, one may only use the days labelled as "normal" (cf. Bonke 2005: 353).

One very important reason for the occurring differences between time diary and survey data with respect to housework time arises from the fact that "housework" is used to summarize a set of tasks which are performed erratically and for varying durations of time. Thus, Marini and Shelton (1993) mention random and recall errors as major reasons for the lack of accuracy of time estimates. Furthermore, they indicate that a direct measurement in the form of standardized questions usually does not allow one to differentiate between primary and secondary activities; thus, they assume that some time intervals may be recorded twice or not at all. In those cases where respondents do consider parallel activities in their estimations, Marini and Shelton (1993) expect an overassessment of the actual time budget.

Other differences between time diaries and time estimates arise from the construction of the dependent variable (Baxter and Bittman 1995; Kitterød and Lyngstad 2005). Two problems are of specific relevance in this respect. Firstly, in the case of "free recording," which is an option used exclusively in time diaries, not time estimations, it might be difficult to code the raw information in an intersubjectively consistent manner. Secondly, in the case of standardized recording, which is an option for both diary and survey methods, all respondents may not understand the given categories in the same way. Additionally, effects of social desirability may occur, which may systematically bias the responses. Although this problem is well known for survey questions, it has not been discussed in the case of time diaries. Anyway, one can assume that a longer observation period may also have an impact on the actual behavior of the respondents, at least for the period of survey collection (Künzler 1994: 58).

In this context, Press and Townsley (1998) conclude from their study that in interview situations women and men both tend to reproduce social roles through their answers on direct time use questions; they may, for example, do so in order to avoid further inquiry or to provide a consistent picture of themselves. Depending on age and education, which are commonly used as proxy indicators for liberal attitudes, men and women over-report their contribution to the every day household work. An example: Today young and highly educated men with supposedly "liberal" gender role attitudes are expected to help their partners with the daily household chores. Thus, against the background of the normative gender attitudes, these men might tend to overestimate their time spent on housework, compared to, for example, older and lower educated women (cf. Kitterød and Lyngstad 2005: 16). Furthermore, the pretests of our own study (described below) indicated that couples, when interviewed together, often ended up "bargaining" over their individual time use estimates for household chores. This is obviously due to different individual perceptions concerning their own and their partners' time use on household tasks.

An overview of some of the most important studies of the last 20 years (e. g. Baxter and Bittman 1995; Robinson 1985; Marini and Shelton 1993; Niemi 1993; Press and Townsley 1998) reveals that women usually overestimate their time for housework activities to a greater extent than men do. Bonke (2005) confirms this result with Danish data, but Kitterød and Lyngstad (2005) do not find a significant difference between women's and men's time use estimates. Kitterød and Lyngstad (2005) mention two possible reasons for these partially contradictory findings: On the one hand, women might overestimate their general time use for housework because they still do the largest share of it (Marini and Shelton 1993); on the other hand, women are said to have a more precise knowledge concerning the structure and time binds of housework, and thus should be able to estimate these time budgets more precisely (Press and Townsley 1998). Precise estimation of the total time spent on housework is difficult because one usually has to handle household duties in several, typically short, episodes spread over the course of the day. This problem does not exist in time diary surveys (Bonke 2005). However, Kitterød and Lyngstad (2005) conclude that despite the pronounced differences between time diary and time estimations, both measurement techniques provide relatively similar patterns of variation between different subgroups. Hence, they consider it appropriate to compare both measures on an ordinal scale. Based on this argument, one can assume that

multivariate analyses should produce similar patterns for the independent variables, regardless of how the dependent variable was collected (Bonke 2005).

3 Data and Methods

Institutionally supported by the project *The Division of Domestic Labor as a Process* (funded by the German Research Foundation), we developed our own time diary instrument, accompanied by a set of validation instruments, which are being applied in a new German panel study on intimate relationships and family development (Schulz forthcoming 2008). In addition, and in combination with our research and teaching on the household division of labor at the University of Bamberg, we designed a class program to involve our students in the data collection of a unique set of comparative time use measurement data on couples. With this cumulative survey, we obtained data that enable us to compare different measurement techniques directly, particularly time diaries and survey questions on time use (Grunow et al. 2006). The analyses presented in this paper draw on this comparative data set of time use measures.

Our data set consists of married and cohabiting couples residing in Germany at the time of the interview. Interviews focus on the individual time use for housework and childcare (if applicable). The student interviewers recruited the respondents from within their circle of acquaintances. Hence, the sample drawn is a convenience rather than a random sample, which severely limits the possibilities of content-based analyses. As far as methodological issues are concerned, however, there should be no sampling-related limitations; even less so as almost all planned interviews were realized, so that we do not have to concern ourselves with the problem of dropouts. Additionally, our data contains extremely few cases of item non-response, which clearly is a result of the convenience sampling strategy.

The core of our study is the *ifb time diary* (Schulz et al. 2005), which every respondent kept for seven consecutive days. The time diary divides the 24 hours of the day into 96 intervals of 15 minutes each. The respondents are asked to record all their activities chronologically in each time interval along 27 standardized categories. Activities that are performed simultaneously, e. g. cooking and taking care of the children, are to be recorded as parallel activities; initial analyses have shown, however, that this option is rarely used.¹ In addition to the specific activities performed, respondents are asked to note during which time intervals their partner was present. Thus, we can draw a circumstantial picture of the everyday life of individuals in partnerships for a whole week, at least for primary activities.

¹ In current time use research, it is a widely recognized fact that a valid measurement of individual time budgets depends a lot on the possibility to obtain information on all sorts of secondary activities. Thus, it might be seen as a drawback of our study, that secondary activities – especially childcare – are not covered in a circumstantial manner. Yet, reported time budgets for housework activities, the central variable with respect to our cognitive interest here, are very similar to the results of representative time use studies (cf. Künzler et al. 2001, Statistisches Bundesamt 2004).

Prior to the time diary, which stays with the respondents for a week, a standardized interview is conducted at the beginning of the survey (for the questionnaires see Grunow et al. 2006). Besides obtaining socio-demographic variables, the respondents are asked to estimate their every day time use on "normal" weekdays as well as on "normal" Saturdays and Sundays. The 27 standardized categories of the time estimation correspond exactly to those used in the time diary; thus, we are able to compare the single dimensions of time use without any intermediary steps of data aggregation or interpretation.

There is an important difference in the range of reference of both time use instruments, though: The time diary records a *particular* week, the survey questions refer to a *hypothetical* week. Yet, we consider the direct comparison of both measures as fruitful, because both measurement techniques are commonly used in time use studies, and the substantial findings are in fact being referred and related to one another. It is therefore in practice assumed that both techniques provide substantially commensurable data. Whereas normally only one of these measures is available in a given data set, our data enable us to check whether this assumption is sustainable indeed.

Our current sample consists of 572 individuals, 290 of them female and 282 of them male (we obtained data from 280 hetero- and six homo-sexual couples). 21 % of the respondents have completed lower secondary schooling ("Hauptschule"), 31 % have obtained an intermediate level of secondary schooling ("Realschule") and 48 % have attained higher secondary education ("Abitur"). At the time of the interview, 57 % of the respondents were working full-time and 22 % part-time; 21 % reported being non-employed, most of them being still enrolled in further education. The youngest person in the sample is 19, the oldest 75 years old. The mean age is 40 years; the mean age difference between the partners of a couple is slightly more than two years. According to the sample requirements, the partners were all cohabiting at the time of the interview. 193 of 286 couples have children; the average number of children is about 1.9 per couple. 39 % of the couples included in our study live together with their underaged children. The age range of these children is between one year and 17 years. 67 % of the couples are married and 27 % live together in a nonmarital union (for 20 couples, the degree of institutionalization is not known).

As we have noted, our sample is not a random sample. Compared to the national distribution of socio-economic characteristics, higher educated, not employed and non-married respondents are overrepresented. The sample does not include any single person households. The average age of the respondents and the mean number of children, however, are similar to the national averages (cf. Statistisches Bundesamt 2006a, 2006b).

Out of the 27 activities contained in our data set, we investigate the respondent's estimated (survey) and reported (diary) time spent on the following eight activities, which are generally seen as the most important household tasks (cf. Lee and Waite 2005): (1) washing the dishes and setting the table; (2) cooking, food preparation; (3) cleaning and tidying up; (4) doing laundry and ironing; (5) shopping for groceries and basic commodities; (6) coordinating with other family members; (7) doing repair work, gardening; maintaining the car; (8) doing family paperwork.

The following analyses then compare the stylized time estimates for these tasks obtained through survey questions with the corresponding reported time according to the time diaries. Parallel activities as recorded in the diaries are added up. Due to the small sample size and possible sample selectivity, we do not seek to produce substantial conclusions concerning time use. The specific aim of this paper is rather to answer the methodologically motivated question of congruence or discrepancy of time diary and survey question data on time use. We use simple statistical methods (means, OLS regression models) in our analyses and control for different subgroups according to theoretically important socio-demographic variables to provide deeper insight into the structure of our data. The calculations are restricted to the time use information on "normal weekdays," that is Monday to Friday, because weekends are often characterized by a special time use structure. In order to make our analyses comparable to current studies on this methodologically interesting topic, we harmonized our proceedings according to the work of Kan (2006), Bonke (2005) and Kitterød and Lyngstad (2005) wherever possible.

Comparing reported time interval data with estimated continuous time data

It is widely acknowledged that the specific data collection technique may influence the results. We therefore have to discuss whether observed discrepancies between time diary and time estimation data may result from the fact that the instruments applied in our study allow for different degrees of accuracy in reported versus estimated time use. Whereas the time axis of the diaries is split into intervals of 15 minutes, meaning that 15 minutes is the shortest duration that can be reported for a given activity, there are in principle no time interval restrictions in the time estimation instrument. The time estimation technique applied in our study allows respondents to report any estimated time budget between one minute and 24 hours. Thus, we have to acknowledge that observed discrepancies may result from the fact that the outline of the diaries does not allow for recording the "exact" time, thereby forcing the respondents to give imprecise answers. This problem should be especially relevant when considering tasks that are performed infrequently and when considering actions that take very little time (i.e. less than 15 minutes).²

Theoretically, one may tackle this problem by assuming that the inaccuracies, namely "over- or underestimation", are randomly distributed in the sample and thus compensate for each other (resulting in an expected error of zero). However, it is not possible to directly test this assumption empirically. Nevertheless, we found rough empirical evidence which legitimates our proceeding: An analysis of the frequency distributions of the estimated time budgets for the single activities shows that the vast majority of respondents reported estimated time budgets on the basis of 15-minute-intervals. In other words, the option to give a minutely precise estimate of the time use for a certain task

² Eventually, the problem of short activities may only be solved by allowing the respondents to keep the diaries along their individually perceived time axis. For our study, though, we have to assume that the daily household chores are covered adequately by the time intervals given in our diary. Still, one has to bear this aspect in mind when interpreting our results substantially.

was rarely used in practice and the time categories chosen by the respondents correspond well with the given time intervals in the diary.

There are two more noteworthy aspects. For one, it becomes clear that survey questions on time use have a certain information potential – namely as far as *precision* of the time estimates is concerned – which cannot be provided in diaries yet. At the same time, diaries have a much greater information potential as far as timing, sequencing, and, typically, the number of activities performed is concerned, therefore allowing for a much greater range of possible research questions to be answered. Secondly, it follows that current research faces the problem that the ideal data which would allow for a flawless theoretical or methodological analysis cannot realistically be gathered. While one can imagine a time use diary without a standardized time axis or standardized activity lists, such a diary would be too error-prone to be feasible, because of, for example, recall problems or effects of social desirability. Therefore, the decision whether to use a time diary or survey questions on time use is still a choice between two suboptimal measures.

4 Empirical Results

Total time for housework

Table 1 shows the time budgets for respondent's time spent on housework on a "normal" weekday (Monday to Friday, weekends and weekdays indicated as "special" are excluded). The figure shows means and standard deviations for time diary and time estimation data for the whole sample and separately for women and men.

On average, the respondents included in our sample spend approximately 139 minutes per day on household tasks, according to the time diary information. Based on the stylized estimates, the mean time for housework is 191 minutes, which is almost one hour more when compared to the time diary information. In sum, the average time for household labor in our sample is very similar to the results from larger representative studies. For the German case, Künzler et al. (2001) as well as the Federal Statistical Office of Germany (2004) report slightly higher weekly time budgets for housework. These estimates may correspond better to the actual population average because of the random sampling strategies used in these studies. However, differences may also stem from the fact that we did not include weekend data in our calculations, whereas most of the larger studies do.

Meanwhile, t-tests for different subgroups show that the means of the time diary data differ significantly from the means of the survey questions. Still, both variables are highly correlated, meaning that a person with a greater reported time budget in the time diary also tends to report a greater estimated time budget in the survey questions.

A similar pattern of overestimated time budgets can be found for the sex specific means of housework time: women on average tend to overestimate their time use by 61 minutes

per day, men by 43 minutes. However, women's reported housework time is almost twice as high as men's.³

Table 1: Average time spent on housework on a "normal" weekday in minutes

	All		Women		Men	
	Diary	Survey	Diary	Survey	Diary	Survey
Average time for housework	139 (112)	191 (137)	186 (113)	247 (139)	91 (90)	134 (108)
N	572	572	290	290	282	282

Means, standard deviations in parentheses, rounded to the next even

Source: ifb-Time-Budget-Study 2005/2006/2007 (Mondays to Fridays), original calculations.

On this highly aggregated level, our data confirm the findings reported by Bonke (2005: 361), who states that the stylized estimates provided by men differ less from the respective diary information than those provided by women. However, relative to the absolute time budget spent, women's estimates are actually more consistent. According to this logic, women tend to overestimate their time by 33 %, and men by 47 %.⁴ The findings by Kitterød and Lyngstad (2005: 18) are not supported, as their study suggests no differences between the measurement techniques on this level of analysis.

Further analyses for distinct subgroups show to some extent noticeable differences between age groups, educational groups, and family status groups with respect to reported time budgets (Table 2). However, similar systematic tendencies of overreporting occur between the groups when comparing reported and estimated time budgets. As we demonstrate later on, the reported time budget differences between the subgroups can be reproduced in multivariate analysis, in the form of higher regression coefficients for survey data compared to diary data.

Table 2 shows that time diary and time estimation data differ significantly on this high level of aggregation, both in general and also when controlling for different subgroups. The average "estimation error" ranges from 45 to 60 minutes per day. We therefore conclude that one cannot simply substitute both measurement techniques, at least not if one aims to compare absolute time budgets on a metric level. For a comparison based on an ordinal scale, though, both measurement techniques produce similar results, whereas it is then not possible to quantify the differences.

³ As a matter of convenience, we follow Bonke's (2005) terminology by using the terms "over-/underreporting" or "over-/underestimation". Yet, to be exact, both types of data, diaries *and* stylized estimates, have errors and it is only an assumption that diary estimates are more accurate than stylized ones (cf. Kan 2006: 4).

⁴ These proportions are calculated as follows: The average "estimation error" is divided by the average time spent on housework as it is reported in the time diary; for women, this is for example 61 minutes divided by 186 minutes equals 0.327, which is approximately 33 % (for men analogously).

Table 2: Average time spent on housework on a "normal" weekday in minutes for several subgroups

	Women		Men	
	Diary	Survey	Diary	Survey
Age: 30 to 39 years	206 (117)	266 (115)	72 (63)	114 (89)
Age: 40 to 49 years	197 (104)	262 (128)	95 (95)	136 (102)
Age: 50 + years	242 (129)	308 (166)	125 (116)	165 (138)
N	274	274	265	265
Education: low	213 (103)	284 (154)	112 (119)	162 (133)
Education: middle	219 (115)	282 (137)	96 (91)	148 (120)
Education: high	152 (107)	218 (122)	84 (75)	126 (83)
N	268	268	254	254
Non-marital union	114 (67)	176 (74)	71 (58)	120 (63)
Marriage	221 (114)	289 (144)	100 (99)	144 (113)
N	270	270	261	261
Childless	110 (67)	166 (110)	78 (69)	124 (101)
With children	222 (112)	286 (135)	98 (98)	139 (110)
N	290	290	282	282

Means, standard deviations in parentheses, rounded to the next even.

Source: ifb-Time-Budget-Study 2005/2006/2007 (Mondays to Fridays), original calculations.

Time use for selected household activities

Different from Kitterød and Lyngstad (2005) and Bonke (2005), who analyzed total housework time, we collected more detailed information on several specific housework chores which appear to be especially important according to the body of time use literature. This approach is more geared to the cognition of the respondents compared to questions like "How much time do you usually spend on housework on a normal weekday?," which address this issue on a much more abstract level. We assume that a

respondent tends to think of the actual activities, such as "cooking", "cleaning" or "watching TV," rather than thinking in abstract categories like "housework" or "leisure". Methodologically speaking, in our survey the respondents do not need to solve the challenging task of reconstructing all the episodes corresponding to an abstract category (Robinson 1999: 56 – 58 discusses this problem in more detail). For the previous analyses the total time for "housework"⁵ is calculated as a sum of the eight single household tasks mentioned earlier:

$$t_{Total} = t_{Dishes} + t_{Cooking} + t_{Cleaning} + t_{Laundry} + t_{Shopping} + t_{Coordination} + t_{Maintenance} + t_{Paperwork}$$

Table 3 shows the time budgets for the eight single domestic tasks, both, for time diaries and survey estimates. Considering the single activities, it becomes obvious that the mean differences between the reported and estimated time are quite marginal. Thus, the marked differences observed for the total time spent on housework are to some extent a product of the statistical modelling, as they result from summing up all the little inconsistencies for the single activities that constitute our "housework" category. We therefore conclude that what can be seen as a negligible "blur" on the level of single tasks results in significantly different means as soon as these activities are added up.

On average, the survey question data (column "Survey") differ from the time diary data (column "Diary") by a maximum of 19 minutes per day. The differences between time diary and survey questions are slightly higher for women than for men, particularly with respect to the so-called "female" tasks. For both sexes, the greatest differences can be found for those everyday tasks which cannot be easily postponed, for instance cooking or cleaning. With respect to more selective chores, e. g. paperwork or coordination, the differences are negligible, as they do not exceed three minutes per day on average. For all activities, the mean time estimated based on survey questions is higher than the mean time reported in diaries. That is, the respondents tend to overestimate their time budgets. Analyses of several subgroups show similar relations of the differences within these groups (as demonstrated in Table 2), therefore allowing for ordinal-scale comparisons between both measurement techniques.

How can these findings be interpreted? It seems that the respondents are able to estimate their time for precisely specified activities quite well, though there is a tendency to overestimate time budgets by a maximum average of 19 minutes per day. These findings for single activities relativize the previous findings of great differences between diaries and surveys on the aggregated level. Furthermore, it has to be noted that we barely find any precise estimation of time budgets in our sample, except for those respondents who do not perform a certain activity at all.

⁵ Due to our cognitive interest, we collect very detailed information on specific housework activities in our time use study. Hence, that enables us to make an in-depth analysis of housework in German couples. Two major drawbacks of this detailed data collection technique, however, are that the interviews are time consuming and that respondents face the challenge to recall their housework episodes in this detailed manner (cf. Belli et al. 2000).

Table 3: Average time for single household tasks on a "normal" weekday in minutes

	All		Women		Men	
	Diary	Survey	Diary	Survey	Diary	Survey
Washing the dishes, setting the table	18 (18)	25 (23)	25 (20)	33 (25)	11 (14)	17 (18)
Cooking, food preparation	26 (28)	39 (38)	38 (29)	57 (38)	13 (20)	21 (30)
Cleaning, tidying up	26 (33)	36 (41)	42 (35)	55 (43)	10 (19)	17 (27)
Doing the laundry, ironing	14 (22)	20 (25)	25 (25)	34 (27)	3 (9)	5 (10)
Shopping groceries and basic commodities	23 (22)	26 (25)	30 (22)	32 (25)	15 (20)	20 (24)
Coordination with other family members	7 (12)	11 (29)	7 (12)	12 (38)	6 (11)	9 (13)
Repair work, gardening, DIY, maintaining the car	21 (47)	28 (48)	13 (35)	18 (37)	29 (55)	38 (55)
Family paperwork	4 (11)	6 (12)	4 (10)	5 (10)	4 (12)	7 (14)
N	572	572	290	290	282	282

Means, standard deviations in parentheses, rounded to the next even.

Source: ifb-Time-Budget-Study 2005/2006/2007 (Mondays to Fridays), original calculations.

Average deviation and average absolute deviation

In order to examine the differences between estimated and reported time more closely, we now shift the focus from the average deviation to the absolute deviation. In the previous analyses we calculated the deviance as the arithmetic mean:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n (Estimate_i - Diary_i)$$

The nature of this measure is to neutralize the distances of each observed value to the mean's value, because the sign of this difference is taken into account. As a consequence, "estimation errors" with positive and negative signs tend to cancel each other out. Thus, the ordinary mean tells us something about the differences between two different distributions. It does not tell us anything about the average *absolute* deviation between diaries and survey questions for the single respondents. Due to this, one cannot assess the *precision* of the respondents' estimates. Against this background we assume that the actual individual "estimation errors" which underlie the mean value should be much larger at a second glance.

The absolute deviation is given by:

$$|\bar{x}| = \frac{1}{n} \sum_{i=1}^n |Estimate_i - Diary_i|$$

Table 4 compares the average deviation and average absolute deviation for the eight single housework activities, for all respondents and separately for men and women.

Table 4: Deviation of survey data from time diary data in minutes

	All		Women		Men	
	S – D	S – D	S – D	S – D	S – D	S – D
Washing the dishes, setting the table	7 (22)	15 (17)	8 (25)	19 (19)	5 (18)	12 (15)
Cooking, food preparation	13 (32)	22 (27)	19 (34)	28 (27)	8 (29)	16 (25)
Cleaning, tidying up	10 (35)	22 (29)	12 (42)	29 (32)	7 (28)	15 (25)
Doing the laundry, ironing	6 (22)	13 (18)	9 (28)	21 (21)	2 (10)	5 (9)
Shopping groceries and basic commodities	4 (28)	17 (22)	2 (30)	19 (24)	5 (26)	15 (21)
Coordination with other family members	4 (30)	10 (29)	5 (40)	11 (38)	3 (15)	8 (13)
Repair work, gardening, DIY, maintaining the car	7 (43)	22 (37)	5 (38)	16 (34)	10 (47)	28 (40)
Family paperwork	2 (14)	7 (12)	1 (12)	6 (10)	3 (15)	8 (13)
N	572	572	290	290	282	282

Means, standard deviations in parentheses, rounded to the next even.

Note: S = Survey question data; D = Time diary data.

Source: ifb-Time-Budget-Study 2005/2006/2007 (Mondays to Fridays), original calculations.

The findings confirm our assumption that the application of the mean suggests "better" estimates than the application of the absolute deviation. The actual absolute deviation is much higher in most cases when compared to the mean values. We can also find slightly different results by gender, according to specific tasks. Again, subgroups (by age, education, employment status, parenthood, family status) were tested analogously, but no systematic differences could be found. Thus, these findings corroborate the deep impact of the statistical modelling on the time use results.

We conclude that, even on the level of single activities, a precise estimation of time is not very likely to be obtained, thereby assuming that time diaries measure the actual time budgets more validly. The respondents' time use estimates clearly differ from the diary

data. This result does not usually occur when using the arithmetic mean because this measure neutralizes the underlying variation. It might be of crucial relevance, though, when considering other statistical procedures. In the following section we assess the relevance of these findings for the actual content-based research on the division of household labor by experimenting with several multivariate analyses.

Determinants of time use – multivariate analyses

Bonke (2005), as well as Kitterød and Lyngstad (2005) argue that an ordinal comparability of time budgets within different subgroups would lead to similar results when applying multivariate analyses to the data. In order to tackle this assumption, we apply to our sample simple multivariate models with two different dependent variables. The first dependent time use variable is based on the stylized time estimates, the second is based on the reported time use in the diaries. Since both calculations are based on the same set of respondents, both models contain identical independent variables. This is a crucial advantage of our data set over previous studies. We refer to two major hypotheses generated by economic theory (e. g. Becker 1998; Ott 1992) in order to assess the theoretical relevance of our methodological comparison. We address the following hypothesis: Individuals with a higher level of human capital resources, relative to their partners', spend less time on housework than individuals with equal or lower relative human capital resources; and, vice versa, individuals with a lower relative level of human capital resources spend more time on housework than individuals with equal or higher relative human capital resources.

We estimate the total time for housework using an OLS regression. Currently, there is an ongoing debate in the field of time use research on whether using OLS or tobit regression is more appropriate for time use data. At least some comparable studies used tobit regression models to estimate the time spent on housework to avoid possible problems, especially with men reporting zero hours. Following Gershuny and Egerton (2007), though, we think that using tobit regression is neither necessary nor appropriate in our case, because with respect to time use, not doing housework at all, and therefore reporting zero hours, is an empirically relevant pattern for men. This means we interpret reported zero hours as real quantities, not as a result of sample selection by truncation or censoring.

According to economic theory, we include the relative employment status and the relative individual income of the respondents into our model. "Woman < Man" stands for a couple where the woman's resources (employment hours or income) are considerably lower than the man's, "Woman = Man" then stands for a couple with approximately equal resources and "Woman > Man" means that the woman's resources clearly exceed the man's.⁶ To control for the composition of our sample, we include the respondent's age, a

⁶ The variable representing the *relative employment hours* was derived from a variable measuring the current employment status in terms of "full time employed", "part time employed", or "not employed". The couple's actual relation was then derived as a combination of the two individual answers using the relation "not employed" < "part time employed" < "full time employed". The *relative income variable* was constructed as follows: "woman < man" means that the woman's share of the total household income is lower than 40 %; "woman > man" means that the woman's share is higher than 60 %; and "woman = man"

dummy variable that distinguishes between married and cohabiting persons, and a dummy variable indicating parenthood.⁷ The models are calculated separately for the sexes; the homosexual couples were excluded from the analysis because otherwise we would have needed arbitrary assumptions on the assignment of sex categories for calculating relative resources.

To be sure, these models do not provide a serious test of economic theory, but this is not our intention here: *In fact, we aim to discover whether the regression analyses produce similar or different effects of the independent variables, according to which dependent variable is used, diary-based time budgets or survey estimates.* Although the number of cases available for our analysis is still quite low, the results presented proved to be stable for various model specifications and provide an interesting insight in the structure of our data.

Table 5: OLS regression of total time for housework (relative employment)

	Women						Men					
	Diary			Survey			Diary			Survey		
	B	s.e.	P	B	s.e.	p	B	s.e.	p	B	s.e.	p
Intercept	94	22	0.00	168	29	0.00	18	20	0.35	78	22	0.01
Age	2	1	0.00	2	1	0.08	1	1	0.10	-0	1	0.80
Not married (Ref.)		-----			-----			-----			-----	
Married	24	20	0.29	53	26	0.04	4	17	0.80	27	19	0.16
No children (Ref.)		-----			-----			-----			-----	
Children	30	22	0.16	11	28	0.69	12	18	0.52	16	20	0.45
<i>Relative employment</i>												
Woman < Man (Ref.)		-----			-----			-----			-----	
Woman = Man	-60	15	0.00	-55	19	0.00	33	12	0.01	61	14	0.00
Woman > Man	-69	23	0.00	-16	30	0.61	41	20	0.05	69	23	0.00
N		242			242			242			242	
Adjusted R-Square		0.286			0.162			0.046			0.079	

B-Coefficients rounded to the next even; listwise exclusion of missing values.

Note: If $0 < |B| < 0.5$ the B-Coefficient is labelled as '0'.

Source: ifb-Time-Budget-Study 2005/2006/2007 (Mondays to Fridays), original calculations.

Table 5 shows the B-coefficients, the standard errors (s. e.) and the corresponding level of statistical significance (p) of the OLS regression. Interestingly, the specified model for the total estimated time for housework fits the data for women better than for men (see

symbolizes a constellation, in which the woman's share is between 40 and 60 %; cases with missing values were excluded.

⁷ Actually, these three control variables are highly correlated. This incidence of multi-collinearity is the reason that these coefficients do not show the well-known effects we would expect from the current research literature (e. g. a positive and significant effect of parenthood for women). When including marriage and parenthood separately, the coefficients become significant (models not shown). The effects of the theoretically relevant variables representing employment or income are not affected by the inclusion of the control variables.

the adjusted R-square values). In general, the patterns of the independent variables are very similar for women across both models. Merely one of the two coefficients for relative employment, indicating that the woman is spending more time in paid work than her partner (compared to "Woman < Man") is no longer statistically significant in the case of time estimates. Yet, this variable is a pivotal indicator with respect to the cited economic theories: what would be interpreted as support for the economic theory in one case makes us think about rejecting the same theory in the other case.

The regression of men's total time for housework shows a quite different picture. First, the model fit is not as good as it is for women. Secondly, we do not observe any major variation of the theoretically important relative employment variables.

Table 6: OLS regression of total time for housework (relative income)

	Women						Men					
	Diary			Survey			Diary			Survey		
	B	s.e.	p	B	s.e.	P	B	s.e.	p	B	s.e.	p
Intercept	85	24	0.00	163	36	0.00	7	24	0.75	59	26	0.03
Age	3	1	0.00	1	1	0.32	2	1	0.00	2	1	0.03
Not married (Ref.)	-----			-----			-----			-----		
Married	5	23	0.81	67	33	0.05	12	21	0.57	16	24	0.49
No children (Ref.)	-----			-----			-----			-----		
Children	26	23	0.27	15	34	0.67	-35	21	0.10	-26	24	0.27
<i>Relative income</i>												
Woman < Man (Ref.)	-----			-----			-----			-----		
Woman = Man	-57	16	0.00	-36	24	0.13	13	15	0.39	24	17	0.15
Woman > Man	-65	27	0.02	23	40	0.56	22	26	0.38	89	29	0.00
N	138			138			137			137		
Adjusted R-Square	0.330			0.165			0.081			0.083		

B-Coefficients rounded to the next even; listwise exclusion of missing values.

Source: ifb-Time-Budget-Study 2005/2006/2007 (Mondays to Fridays), original calculations.

In table 6 we present more results that address our economic resources hypothesis, now using the relative post-government income as the key explanatory variable. While this measure is theoretically well suited to test our hypothesis, our case numbers here are lower, due to item non-response. Interestingly, the two models estimated for women now differ substantially with respect to the economic resource variables. The dummies indicating resource equality ("Woman = Man") and women's higher relative income ("Woman > Man") both differ significantly from the reference category ("Woman < Man") for diary data but not for time estimates. Again, we would come to different conclusions concerning the economic theories. Also for men we find a difference between the diary based and the survey question based estimates with respect to relative income. Thus, the findings vary according to the dependent variable chosen. In addition to the presented results, we applied T-tests to examine whether the regression coefficients for the single variables differ significantly between the models for diary and survey

question data (separately for women and men). Since this is indeed the case for each of the theoretically relevant variables, these findings further support our assumption that the choice of the dependent variable may indeed matter when analyzing time use processes. We would have come to a similar conclusion here by replicating Bonke's (2005: 360, 364) regression models using the time use differentials (survey minus diary data) as a dependent variable. We therefore conclude that the dependent variable could be substituted for neither women nor men, as both methods of measuring time use produce different patterns in the independent variables for the very same people. Nonetheless, we have to draw larger samples to see whether these findings are substantially stable. Still, our analyses show that in contrast to the common assumption, time diaries and survey questions on time use may not be arbitrarily substituted.

5 Discussion

This paper analyzed whether and for which statistical procedures diary-based time budget data and survey questions on time use produce similar results and thus may be substituted by each other. In general, time diary studies are considered to be richer, more valid and more reliable, but compared to survey questions they are also more expensive and time-consuming to conduct and to analyze.

We began with the assumption that a major goal of modern empirical research is to obtain data of optimal quality while exerting the minimum amount of effort necessary, and that time use scholars thus need well-founded empirical knowledge about the quality of time estimates as compared to time diaries. Therefore we conducted and analyzed our own time use data set of married and cohabiting couples in order to compare the two methods for the same group of respondents.

The most important finding of our analyses is that time budgets may noticeably differ according to the measurement technique employed. On the highly aggregated level of total time spent on housework, both strategies, time diaries and survey estimates, produce significantly different means for several subgroups. Whereas the means for single household tasks suggest a relatively high concord, a deeper analysis of absolute deviations illustrates that the estimates in fact tend to be much higher or lower than the respective time diary data; completely congruent estimates virtually do not occur. The high level of correspondence between inter-individual diary and estimated time use data, which is reported in several papers on time use (for the German case see e. g. Künzler 1999) therefore appears to result from the statistical characteristic of the mean value, which neutralizes the variation, rather than from the actual correspondence of both measurement techniques. Thus, not only the measurement technique, but also the statistical modelling of the data strongly influences the results. In addition, we applied the two data sources, time diaries and survey estimates, as dependent variables to regression analysis, as is frequently done in studies of gendered labor division. We thereby investigated the sensitivity of identically specified regression models to the measurement of the dependent variable. For women in particular the effect sizes and significance levels vary, based on the time use measure used to create the dependent variable. For men, the differences are less pronounced, but still notable. With respect to our theoretical

hypotheses, we would have reached quite different conclusions both for women and men, depending on the time use measurement technique.

Conclusion: The findings presented in this paper provide original evidence from Germany that time use data obtained from diaries and survey questions may not be substituted arbitrarily. This problem becomes particularly important when using statistical methods which seek to investigate theory driven research questions (e. g. regression models). Rather than simply applying survey questions on time use in large, newly set up social scientific studies, it would be worthwhile to invest the time and money necessary to obtain empirical data to clarify why survey estimates and time diaries differ markedly, and what can be done to improve the quality and power of both techniques (cf. Kan and Pudney 2007). Otherwise time use research must remain in constant doubt as to whether available data can provide commensurable time budgets for answering research questions. After all, our results indicate that using survey questions on single household activities instead of a general "housework" category may lead to more congruent estimates, at least when comparing arithmetical means. Future research should examine closely the cognitive processes underlying respondents' estimates of their time use. A goal would therefore be to better harmonize survey questions to the interviewee's perception of everyday (inter)action. In general, time use scholars would be well advised to carefully select their time use measurement techniques depending on theoretical needs rather than on custom or convenience, because with today's knowledge, strictly speaking, only the combination of several methods allows for the creation of a circumstantial picture of peoples' time use, namely by mutually emphasizing and relativizing each method's advantages and measurement errors.

Even though our results and conclusions are similar to those of recent studies, they should be interpreted carefully, given the limitations of our data set: the case numbers are low and our sample is not a random sample. Nevertheless, our data provide both interesting and relatively stable tendencies and relations, which may serve as a heuristic "seed" for future research on this timeless problem.

References

- Baxter, J. and M. Bittman (1995), Measuring time spent on housework: A comparison of two approaches, in: *Australian Journal of Social Research*, Vol. 1, No. 1, 21 – 46.
- Becker, G. S. (1998), *A Treatise on the Family*, Enlarged Edition, Fourth Printing, Cambridge, London, Harvard University Press.
- Belli, R. F., Schwarz, N., Singer, E. and J. Talarico (2000), Decomposition can harm the accuracy of behavioral frequency reports, in: *Applied Cognitive Psychology*, Vol. 14, No. 4, 295 – 308.
- Bonke, J. (2005), Paid Work and Unpaid Work. Diary Information Versus Questionnaire Information, in: *Social Indicators Research*, Vol. 70. No. 3, 349 – 368.
- Gershuny, J. and M. Egerton (2007), *Time and participation: Modelling unpaid work using time diary data - A framework for investigating the suitability of OLS for modelling diary-derived time-use dependent variables*. Manuscript submitted for publication.
- Grunow, D., Schulz, F., Rost, H. and M. Rupp (2006), *Zeitverwendung im Alltag. Zusatzinstrumente zum ifb-Zeitverwendungstagebuch*. ifb-Materialien 4-2006, Bamberg, Staatsinstitut für Familienforschung.
- Juster, F. T. and F. P. Stafford (1991), The Allocation of Time. Empirical findings, Behavioral Models, and Problems of Measurement, in: *Journal of Economic Literature*, Vol. 29, No. 2, 471 – 522.
- Juster, F. T., Ono, H. and F. P. Stafford (2003), An Assessment of Alternative Measures of Time Use, in: *Sociological Methodology*, Vol. 33, No. 1, 19 – 54.
- Kan, M. Y. (2006), *Measuring Housework Participation: The Gap between "Stylised" Questionnaire Estimates and Diary-based Estimates*, ISER Working Paper 2006-11, University of Essex (UK), download at <http://www.iser.Essex.ac.uk/pubs/workpaps/04.07.2006>.
- Kan, M. Y. and S. Pudney (2007), *Measurement errors in stylised and diary data on time use*, Sociology Working Papers 2007-05, University of Oxford (UK), download at <http://www.sociology.ox.ac.uk/research/electronic%20resources.asp> (12.10.2007).
- Kitterød, R. H. and T. H. Lyngstad (2005), Diary versus questionnaire information on time spent on housework – The case of Norway, in: *electronic International Journal of Time Use Research*, Vol. 2, No 1, 13 – 32.
- Künzler, J. (1994), *Familiale Arbeitsteilung: Die Beteiligung von Männern an der Hausarbeit*, Bielefeld, Kleine.
- Künzler, J. (1999), Arbeitsteilung in Ehen und nichtehelichen Lebensgemeinschaften, in: Klein, Thomas/Lauterbach, Wolfgang (eds.), *Nichteheliche Lebensgemeinschaften. Analysen zum Wandel partnerschaftlicher Lebensformen*, Opladen, Leske + Budrich, 235 – 268.
- Künzler, J., Walter, W., Reichart, E. and G. Pfister (2001), *Gender division of labour in unified Germany*, Tilburg, Tilburg University Press.
- Lee, Y.-S. and L. J. Waite (2005), Husbands' and Wives' Time Spent On Housework: A Comparison of Measures, in: *Journal of Marriage and Family*, Vol. 67, No. 2, 328 – 336.
- Marini, M. M. and B. A. Shelton (1993), Measuring Household Work: Recent Experience in the United States, in: *Social Science Research*, Vol. 22, No. 4, 361 – 382.
- Niemi, I. (1993), Systematic Error in Behavioural Measurement: Comparing Results from Interview and Time Budget Studies, in: *Social Indicators Research*, Vol. 30. No. 2/3, 229 – 244.
- Ott, N. (1992), *Intrafamily Bargaining and Household Decisions*, Berlin, Springer.

Schulz, Grunow: Comparing time diary data and stylized time use estimates

- Press, J. E. and E. Townsley (1998), Wives' and husband's housework reporting. Gender, class, and social desirability, in: *Gender and Society*, Vol. 12, No. 2, 118 – 218.
- Robinson, J. P. (1985), The Validity and Reliability of Diaries versus Alternative Time Use Measures, In: Juster, F. T. and F. P. Stafford (eds.): *Time, Goods and Well-Being*, Ann Arbor, The University Press of Michigan.
- Robinson, J. P. (1999), The Time-Diary Method. Structure and Uses, in: Pentland, W. E., Harvey, A. S., Lawton, M. P and M. A. McColl (eds.), *Time Use in the Social Sciences*, New York, Kluwer Academic Publishers, 47 – 89.
- Schulz, Florian (forthcoming 2008): Tagein, tagaus. Zur Erhebung der Zeitverwendung im Rahmen der Vorstudie zum Beziehungs- und Familienentwicklungs-Panel, in: Huinink, J. and M. Feldhaus (eds.), *Neuere Entwicklungen in der Beziehungs- und Familienforschung. Vorstudien zum Beziehungs- und Familienentwicklungspanel (PAIRFAM)*, Würzburg, Ergon.
- Schulz, F. and D. Grunow (2007): Tagebuch versus Zeitschätzung. Ein Vergleich zweier unterschiedlicher Methoden zur Messung der Zeitverwendung für Hausarbeit, in: *Zeitschrift für Familienforschung*, Vol. 19, No. 1, 106 – 128.
- Schulz, F., Grunow, D., Rost, H. and M. Rupp (2005), *Zeitverwendung im Alltag. Das ifb-Zeitverwendungstagebuch*, ifb-Materialien 3-2005, Bamberg, Staatsinstitut für Familienforschung.
- Statistisches Bundesamt (ed.) (2004), *Alltag in Deutschland. Analysen zur Zeitverwendung*, Forum der Bundesstatistik, Band 43, Wiesbaden.
- Statistisches Bundesamt (ed.) (2006a), *Datenreport 2006. Zahlen und Fakten über die Bundesrepublik Deutschland*, Bonn, Bundeszentrale für politische Bildung.
- Statistisches Bundesamt (ed.) (2006b), *Statistical Yearbook 2006 for the Federal Republic of Germany*, Wiesbaden.

Appendix

Table A: Descriptive statistics of the regressors from tables 5 and 6

		All	Women	Men
Age	\bar{x}	40.25	39.10	41.39
	s_x	12.54	12.15	12.85
	N	524	262	262
Marital Union	\bar{x}	0.73	0.72	0.73
	s_x	0.45	0.45	0.44
	N	520	260	260
Parenthood	\bar{x}	0.69	0.69	0.68
	s_x	0.46	0.46	0.46
	N	557	278	279
Relative Employment Status Woman = Man	\bar{x}	0.39	0.39	0.40
	s_x	0.49	0.49	0.49
	N	489	244	245
Relative Employment Status Woman > Man	\bar{x}	0.08	0.08	0.07
	s_x	0.26	0.27	0.26
	N	489	244	245
Relative Income Woman = Man	\bar{x}	0.40	0.40	0.40
	s_x	0.49	0.49	0.49
	N	278	139	139
Relative Income Woman > Man	\bar{x}	0.08	0.08	0.08
	s_x	0.27	0.27	0.27
	N	278	139	139