Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Daily rhythm of data quality: Evidence from the Survey of Unemployed Workers in New Jersey

by Jorge González Chapela

Release date: December 20, 2024

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

• Statistical Information Service	1-800-263-1136
• National telecommunications device for the hearing impaired	1-800-363-7629
• Fax line	1-514-283-9350

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, December 2024
Vol. 50, No. 2, pp. 513-541
Statistics Canada, Catalogue No. 12-001-X

Daily rhythm of data quality: Evidence from the Survey of Unemployed Workers in New Jersey

Jorge González Chapela[1]

Abstract

This paper investigates whether survey data quality fluctuates over the day. After laying out the argument theoretically, panel data from the Survey of Unemployed Workers in New Jersey are analyzed. Several indirect indicators of response error are investigated, including item nonresponse, interview completion time, rounding, and measures of the quality of time diary data. The evidence that we assemble for a time of day of interview effect is weak or nonexistent. Item nonresponse and the probability that interview completion time is among the 5% shortest appear to increase in the evening, but a more thorough assessment requires instrumental variables.

Key Words: Panel data; Survey data quality; Survey of Unemployed Workers in New Jersey; Time of day.

1. Introduction

That surveys are an essential tool for empirical research seems as indisputable as seems that measurement error can compromise the quality of survey data. Among the tenets which appear to underlie the measurement error literature is the principle that the survey respondent must perform a series of cognitive operations before answering a question (e.g., Tourangeau, Rips and Rasinski, 2000, Chapter 1). Each of those operations can be quite complex, involving a great deal of cognitive work (Krosnick, 1999). Extensive research (summarized among others by Schmidt, Collette, Cajochen and Peigneux, 2007) has shown that human performance on a wide range of cognitive tasks fluctuates over the day. Yet, the impact that these fluctuations may have on the quality of survey data remains largely ignored.

This paper attempts to identify problematic times of day for survey data quality by exploiting high-frequency longitudinal microdata from the Survey of Unemployed Workers in New Jersey (SUWNJ). The SUWNJ interviewed online every week for up to 24 weeks some 6,000 workers who were unemployed at the beginning of the survey in October 2009. Although SUWNJ respondents selected themselves to answer the survey at their most convenient times, the availability of repeated observations on each respondent makes it possible to remove the many unobserved factors that remained constant over the relatively short survey period of the SUWNJ (as compared with other large-scale longitudinal surveys).

The paper is organized as follows. Section 2 provides background and context to this research. Section 3 describes the data, the construction of the main variables, and the selection of the sample. Section 4 discusses the methodology. The results are presented in Section 5. Section 6 summarizes the findings and suggests directions for future research.

1. Jorge González Chapela, Centro Universitario de la Defensa de Zaragoza, Academia General Militar, Ctra. de Huesca s/n, 50090 Zaragoza, Spain. E-mail: jorgegc@unizar.es.

2. Background and context

2.1 Background

Psychologists and survey methodologists have characterized a series of cognitive steps in answering survey questions. Tourangeau et al. (2000, page 8) distinguish four steps (comprehension of the question, retrieval of relevant information, use of that information to make required judgments, and selection and reporting of an answer), and provide an illustrative list of mental processes that may be involved in the answering process. Attention and memory are part of that list, both of which have been shown to fluctuate over the day.

The search for time of day fluctuations in human cognitive performance has increasingly been based on the so-called two-process model of sleep-wake regulation (Blatter and Cajochen, 2007; Schmidt et al., 2007). This model postulates that the influence of time of day on cognitive performance is mediated by sleepiness, which in turn is determined by the interacting influences of two propensities. The homeostatic propensity for sleep continuously accumulates during time spent awake and continuously decreases during sleep. The nearly 24-hr (or circadian) oscillatory wake propensity balances the accumulated homeostatic sleep drive during wakefulness.

The circadian wake propensity, which is the result of an internal clock that is synchronized by signals created by the Earth's rotation (light, temperature, etc.) (Roenneberg, Kuehnle, Juda, Kantermann, Allebrandt, Gordijn and Merrow, 2007), reaches its maximum in the evening and its minimum in the early morning. So, for a person who usually sleeps from 23:00 to 07:00, cognitive performance would be at a lower level during nighttime and early morning, a better level occurs around noon, there is a decrease after lunch (e.g., Bes, Jobert and Schulz, 2009), and higher levels occur during afternoon and evening hours (Valdez, 2019). Yet, this time course can be modulated by the kind of task and inter-individual differences in task performance (Blatter and Cajochen, 2007).

The phase of the circadian wake propensity and that of the signals differ across individuals, creating a relationship between internal and external time called phase of entrainment. People who differ in the phase of entrainment are referred to as different chronotypes. The alignment between chronotype and time of day enhances a number of cognitive functions, giving rise to the so-called synchrony effect (e.g., Hasher, Goldstein and May, 2005; Hornik and Tal, 2010; Salehinejad, Wischnewski, Ghanavati, Mosayebi-Samani, Kuo and Nitsche, 2021; Guarana, Stevenson, Gish, Ryu and Crawley, 2022). Thus, if people responded to surveys during hours aligned with their chronotype (as the evidence in Fordsham, Moss, Krumholtz, Roggina, Robinson and Litman, 2019 suggests), the effect of the time of day would be positively moderated by the sorting of respondents into optimal times.

2.2 Context

A careful and comprehensive performance of each of the four steps in the survey answering process can require a substantial amount of mental effort. Hence, according to Krosnick's (1991) satisficing theory, survey respondents may simply provide a satisfactory answer, the likelihood of which decreases with respondent ability. This insight promoted studies investigating the link between cognitive ability and data quality, the former understood as a stable or slowly changing trait. See, e.g., Kaminska, McCutcheon and Billiet (2010), Kroh, Lüdtke, Düzel and Winter (2016), Gideon, Helppie-McFall and Hsu (2017), Olson, Smyth and Ganshert (2019), Truebner (2021), Angrisani and Couper (2022), Bais, Schouten and Toepoel (2022), and Phillips and Stenger (2022). As predicted by Krosnick (1991), cognitive ability and satisficing appear generally as inversely related.

Time of day fluctuations in cognitive performance may be another aspect of respondent ability related to satisficing. However, this potential link has been little studied. Ziniel (2008, Chapter 4) investigates whether the proportion of "don't knows" provided by respondents to the Health and Retirement Study is sensitive to the time of day, reaching a negative conclusion. Binder (2022) recruited participants from Amazon Mechanical Turk (MTurk) to examine whether inflation expectations and responses to questions with objectively correct answers differ depending on the time of day, finding little differences. On the other hand, a survey carried out on suppliers competing for public contracts in Ireland (Flynn, 2018) reveals that the time of day respondents started the survey predicts survey completeness.

A limitation of these previous studies is that respondents selected themselves to answer the survey at their most convenient times. Hence, and as recognized by Ziniel (2008, Chapter 4), inter-individual differences in cognitive ability or chronotype may interfere with potential time of day fluctuations in cognitive performance. To be sure that factors like these do not interfere with the time of day, Dickinson and McElroy (2010) randomize the survey response window, finding that the time of day (as represented by a binary variable equal to unity for response times from 1:00 to 5:00 a.m. and zero for response times from noon to 7:00 p.m.) has no effect on iterative reasoning.

Identifying problematic times of day for survey data quality is relevant first of all for survey practice, as further measures to reduce the extent of measurement error could be implemented. For example, the e-mailing of invitations/reminders for completing surveys or even the collection of data could be programmed at times of day that were best suited for the increases of data quality. However, forcing respondents to complete surveys at particular times of day could raise nonresponse error (e.g., Weeks, Kulka and Pierson, 1987; Durrant, D'Arrigo and Steele, 2011), so under the total survey error framework (e.g., Lyberg and Stukel, 2017) it would be necessary to study the tradeoff between measurement error and nonresponse error.

Besides the papers that we have already mentioned, our work is related to other strands of literature. Some studies have investigated the characteristics and behaviors of online survey participants as a function of the time of day of participation (e.g., Arechar, Kraft-Todd and Rand, 2017; Casey, Chandler, Levine, Proctor and Strolovitch, 2017; Binder, 2022). Although certain respondent characteristics may be associated with data quality, we focus on data quality and develop effects net of unobserved individual factors and optimal times of participation. The time-of-day fluctuations in cognitive performance have been blamed for the across-the-day variation in a wide spectrum of economic decisions and abilities; see, e.g., Carrell, Maghakian and West (2011), Dickinson and McElroy (2017), Williams and Shapiro (2018), Collinson, Mathmann and Chylinski (2020), Dickinson, Chaudhuri and Greenaway-McGrevy (2020), and Guarana et al. (2022). But whether survey data quality is modulated by the time of day remains largely ignored. Last, but not least, by exploiting start and end times of each interview, we relate to the literature using paradata to investigate measurement error (reviewed by Yan and Olson, 2013).

3. Data, measures, and sample selection

3.1 SUWNJ

The data for this study are taken from the SUWNJ, a longitudinal Internet-based survey of unemployed workers conducted by the Princeton University Survey Research Center between October 2009 and April 2010. Here, we describe the main features of this survey, referring to Krueger and Mueller (2010, 2011) for the survey questionnaire, the data set, and a more complete description of the SUWNJ. The Stata code needed to proceed from the raw data to the results is available from the author upon request.

3.1.1 Sampling and invitation

The individuals sampled were selected from the universe of unemployment insurance (UI) benefit recipients in the state of New Jersey as of September 28, 2009. During 2009 and 2010, New Jersey's unemployment rate closely mirrored the U.S. average, although its population of UI recipients was more female, older, and more educated than in the wider U.S. The sample was selected through stratified random sampling with strata defined by initial duration of unemployment and availability of an e-mail address. Those unemployed 60 weeks or longer and those with an e-mail address were oversampled.

The selected individuals were invited to participate in the survey for 12 consecutive weeks, although the long-term unemployed were invited to participate in an extended study for an additional 12 weeks. The initial invitation was sent by e-mail or (to those without e-mail address) physical letter. The e-mail (letter) contained a link to the online questionnaire. Individuals contacted by letter were required to enter a valid e-mail address in order for them to receive e-mail invitations for the follow-up weekly interviews. If a respondent did not have an e-mail address, he/she could nevertheless participate in the weekly interviews by logging into the same access web page. According to the October 2009 Current Population Survey, 15% of New Jersey's unemployed workers lived in households where no one used the Internet, but no further arrangements were made to secure the participation of Internet non-users. The invitation e-mails (sent in the morning) asked individuals to complete the survey within two days and even if they had already found a job.

3.1.2 Participation and weighting

The AAPOR (2023) RR6 response rate for the first interview was 9.7% (6,025 persons). These respondents completed an average of 4.1 follow-up interviews out of a maximum of 11 (excluding the longer-term follow-up), responding to 24,638 (37.2%) of the potential follow-up interviews. Only 302 individuals completed 12 interviews, so 95.0% attrited from the initial study. The RR6 response rate for the first interview of the extended study was larger, 56.8% (1,148 persons). These respondents completed an average of 6.4 follow-up interviews out of a maximum of 11, responding to 7,390 (58.5%) of the potential follow-up interviews. 115 individuals completed 12 interviews, so 90.0% attrited from the extended study. All this yields 39,201 interviews.

The low response rates created noticeable differences between the universe of New Jersey UI recipients and the respondents. Krueger and Mueller (2011) created inverse propensity weights based on administrative data from the UI system. The weights labeled "current week weights" adjust for differential sampling probabilities and response rates over the 12 weeks of the survey (or 24 weeks, for those who participated in the extended study). The regressors utilized to create "current week weights" were strata indicator variables and time-invariant demographics.

3.1.3 Survey instrument

The SUWNJ questionnaire consists of two parts: an entry survey, administered in the first week, with demographic, income, and wealth questions, and a weekly survey, administered in the first and each subsequent week, with questions related to life satisfaction, food expenditure, job search activities, and time use. The time use information is for the day previous to the interview, and is collected by means of a self-completed time diary from 07:00 to 23:00 and two questions asking wake-up and going-to-bed times. To complete the diary, respondents could select up to 2 activities for each hour from a pre-designated list of 21 activities.

After initiating an interview, respondents could move back and forth through the questionnaire, as well as stop the interview and return to it later. Although completion via phone browser was possible, the questionnaire was not optimized to be taken on a mobile device. The data set includes the date and time (recorded to the second) that each interview was initiated and completed, plus the end time of the time-use section (the third of the five sections of the weekly survey).

3.2 Measures

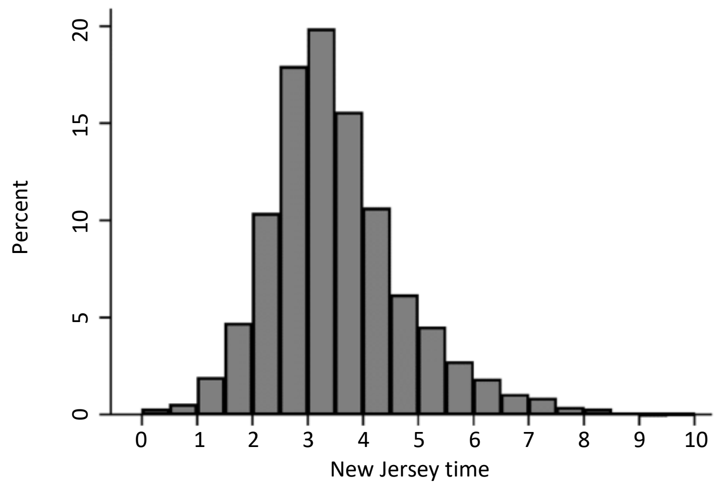
3.2.1 Time of day

Times are local times of New Jersey, measured continuously from midnight and expressed in hours (e.g., 9.5 for 09:30). The time of day of interview (denoted D) is approximated by the mid-time between the start and end times of the interview. In a robustness check, it will be approximated by randomly selected points within the start and end times of the interview (Ahn, Peng, Park and Jeon, 2012).

3.2.2 Chronotype

Roenneberg et al. (2007) use the Munich ChronoType Questionnaire to assess chronotype, measured as the half-way point between sleep-onset and sleep-end (or mid-sleep) on free days corrected for oversleep (MSF_{sc}). A proxy measure for chronotype can be constructed along those lines using the SUWNJ time-use information. Sleep duration is estimated as the time between going to bed and wake up, and its half-way point is averaged over free days. For individuals who sleep longer on free days than on workdays, the difference between sleep duration on free days and its weekly average (assuming a 5-day workweek) is subtracted from the mid-sleep on free days. The resulting measure is denoted MSF^{e}_{sc}. Sleep timing and sleep duration are essentially independent traits (Roenneberg et al., 2007). The correlation between MSF^{e}_{sc} and average sleep duration is 0.06 (although statistically different from zero at 5% level). Figure 3.1 shows the distribution of MSF^{e}_{sc} in the sample.

Figure 3.1 Chronotype (MSF^{e}_{sc}).



Source : SUWNJ.

3.2.3 Data quality

We analyze four sets of measures of data quality (Juster, 1986; Malhotra, 2008; Fricker and Tourangeau, 2010): i) the percent item nonresponse, ii) measures of the quality of time-diary data (the number and variety of activities and the number of hours not coded in the diary), iii) time to complete the interview, and iv) rounded values of mood at home, food expenditure at home, and expenditure on eating out. The SUWNJ questionnaire seems to contain insufficient items to investigate response errors caused by social desirability or extreme, midpoint, or nondifferentiated responding (see, e.g., Baumgartner and Steenkamp, 2001; Chang and Krosnick, 2009). The data file contains completed interviews, which precludes analyzing survey breakoff (e.g., Peytchev, 2009).

We define the percent item nonresponse (P_{INR}) as the percentage of missing values for questions administered to all respondents at a certain interview. This excludes follow-up questions plus questions that can be postponed to the next time the person is interviewed.

To count the number and variety of activities recorded in the time diary, we follow the convention that if an activity intervenes in the middle of some other activity (e.g., shopping on the way home from a job interview), the number of activities increases by two units and the variety of activities by one unit (Juster, 1986). We present results for the number of activities (denoted NumAct), as those for the variety follow the same patterns. If no activity is recorded in an hour, the hour is considered not coded. The variable counting the number of hours not coded is denoted HMissing. Juster (1986) notices that only weekday (Monday–Thursday) diaries suffer significant quality deterioration to the extent that they involve more than 24-hour recall, a finding that will be helpful for interpreting some of our results.

The relationship between interview completion time and data quality in Internet-based surveys is complex, as both short and long completion times may be a symptom of respondent inattention (Malhotra, 2008; Read, Wolters and Berinsky, 2021). Hence, besides a continuous measure of completion time (denoted IvDur), we analyze dummy variables for the 5% lowest and 5% highest completion times, denoted P_{IVDUR5L} and P_{IVDUR5H} respectively. These dummies are created by calculating the corresponding percentiles separately for first and subsequent interviews after removing outliers (see Section 3.3).

Information on mood at home is collected with the question: "Now we would like to know how you feel and what mood you are in when you are at home. When you are at home, what percentage of the time are you: in a bad mood, a little low or irritable, in a mildly pleasant mood, in a very good mood?" Respondents are asked to indicate the percentage of time that they experienced each mood category. We created dummy variables indicating respondents for whom *all* four reported percentages are multiples of 50 (leading to answers of 0, 50, or 100), 25, or 10. The three binary variables are denoted P_{MOOD50}, P_{MOOD25}, and P_{MOOD10}, respectively.

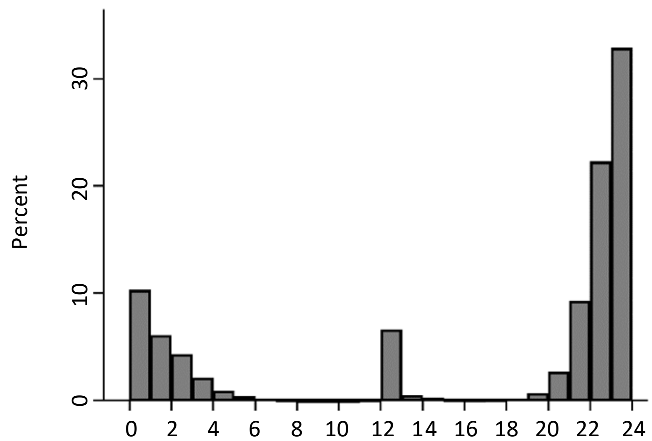
Two questions gather information on expenditure on food: "In the last 7 days, how much did you and anyone else in your family spend on food that you use at home? Please include food bought with food stamps", and "In the last 7 days, how much did you and anyone else in your family spend on eating out?" We created dummy variables indicating respondents for whom a certain expenditure is multiple of 100 or 50, denoted $P_{\text{FOODAH100}}$, P_{FOODAH50}, $P_{\text{EATING-OUT100}}$, and $P_{\text{EATING-OUT50}}$. Zero expenditure could be reflecting rounding, a corner solution, or infrequency of purchase. We present results assuming that zero expenditure reflects rounding, and analyze their robustness to assuming that zero expenditure does not reflect rounding.

3.3 Sample selection

The distribution of interview completion time is heavily right-skewed, with median (mean) completion time of 13.2 (144.3) minutes for first interviews and 11.8 (75.6) minutes for subsequent interviews. To avoid introducing much error into our measure of D, interviews with completion time greater than 60 minutes are discarded, representing 5.7 and 4.4% of first and subsequent interviews. Moreover, we discarded first or subsequent interviews with completion times below the corresponding 1st percentile (4.8 and 3.6 minutes, respectively). Based on our own reading time, this lower bound discards interviews in which the respondent cannot have read the questionnaire. (Including these interviews leaves the conclusions unchanged.)

We also discarded interviews presenting missing or inconsistent data in some variable used in this study. Here, an issue requires some discussion. Going-to-bed time was reported using three drop-down menus of hour, minute, and AM/PM period. The AM/PM menu was set by default to PM, and Figure 3.2 suggests that this may have induced error. While going to bed between 11:00 and 11:59 a.m. is reported in 0.04% of interviews, 6.6% report going to bed between 12:00 and 12:59 p.m. We probed the time diary for an inconsistent going-to-bed time when this was between 12:00 and 02:59 p.m. When an inconsistency was found, the interview was discarded. A total of 2,578 interviews were discarded for this reason. When assessing robustness, we shall include them in the sample assuming that the AM period applies, and a dummy indicating those cases (denoted $P_{\text{PM-AM}}$) will be analyzed for time of day of interview effects.

Figure 3.2 Going-to-bed time.



Source : SUWNJ.

Finally, the last interview of a person who completed 25 interviews is discarded because it is not clear whether he ended up attriting. All this leaves us with 5,531 persons and a total of 33,000 interviews. Figure 3.3 provides a histogram of the number of interviews contributed by each person. The mean (median) number of interviews is 6.0 (4). Table 3.1 provides descriptive statistics for the main variables used in this study.

Figure 3.3 Number of interviews contributed to the sample.

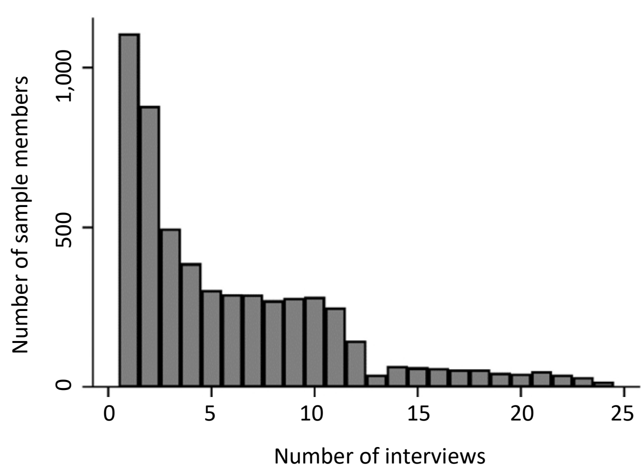


Table 3.1
Descriptive statistics.

	Observations	Mean	Standard deviation	Min	Max
P_{INR}[a]	33,000	2.64	6.58	0	60.87
NumAct	33,000	16.77	7.14	1	32
HMissing	33,000	0.53	2.14	0	15
IvDur (minutes)	33,000	13.95	8.19	3.57	59.95
P_{IVDUR5L}[a]	33,000	4.99			
P_{IVDUR5H}[a]	33,000	4.99			
P_{MOOD10}[a]	32,877	50.40			
P_{MOOD25}[a]	32,877	15.84			
P_{MOOD50}[a]	32,877	10.79			
P_{FOODAH50}[a]	31,949	51.64			
$P_{\text{FOODAH100}}$[a]	31,949	30.63			
$P_{\text{EATING-OUT50}}$[a]	29,084	45.74			
$P_{\text{EATING-OUT100}}$[a]	29,084	34.67			
Time of day of interview	33,000	12.94	4.80 (3.89) [3.45]	0	23.99
$\text{MSF}_{\text{sc}}^{\text{c}}$	5,531	3.56	1.65	0	23.99
Day of interview	33,000				
Monday[b]		8.56			
Tuesday[b]		23.39			
Wednesday[b]		16.18			
Thursday[b]		14.51			
Friday[b]		17.93			
Saturday[b]		12.76			
Sunday[b]		6.67			
Worked[b]	33,000	14.00			
Sleep duration (hours)	33,000	8.35	2.10	0.50	23.58
No. of previous interviews	33,000	5.34	5.00	0	23
Weeks between $t-2$ and $t-1$	33,000	1.43	1.40	0	16

Notes: The data pertain to 5,531 individuals. The sample variation of time of day of interview is made up of "within" (or time series) variation (shown in parentheses) and "between" (cross-section) variation (shown in brackets). Worked and Sleep duration are for the diary day. [a]: Binary indicator for the outcome given in the name's subscript scaled as a percentage. [b]: Binary indicator scaled as a percentage.

4. Methods

4.1 Baseline specification

As respondents select themselves to answer surveys at their most convenient times, it is unlikely that a simple comparison of data quality outcomes by time of day of interview can identify a causal effect. The availability of repeated observations on each SUWNJ respondent allows us to control for unobserved time-constant factors such as cognitive ability or chronotype. Measurement error (as defined in Biemer, Groves, Lyberg, Mathiowetz and Sudman, 2004, page xvii) also arises from the method of data collection and the questionnaire, but since these features are fixed across interviews, they cannot interfere with our estimates.

The following unobserved effects panel data model (Wooldridge, 2010, Chapter 10) is estimated:

$$y_{it} = \alpha\left(D_{it}\right) + \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it} \tag{4.1}$$

where y_{it} denotes some data quality measure for individual i $\left(i = 1, 2, \ldots, N\right)$ at interview number t $\left(t = 1, 2, \ldots, T_i\right)$, $\alpha\left(D_{it}\right)$ is a scalar function of time of day of interview, $\mathbf{x}_{it}$ is a vector of interview-variant observable controls, $\boldsymbol{\beta}$ is a vector of unknown parameters, c_i is an unobserved individual effect arbitrarily correlated with D_{it} and $\mathbf{x}_{it}$, and u_{it} is an idiosyncratic error term.

Besides an intercept, and following Binder (2022) and Juster (1986), included in $\mathbf{x}_{it}$ are dummy variables for day of week of interview and a dummy for whether the respondent worked on the diary day (this information is not available for the day of interview). Cumulative insufficient sleep (e.g., Lowe, Safati, and Hall, 2017) and synchrony effects could also affect y_{it}. Hence, sleep duration on the diary day and the interaction between $\mathrm{MSF}_{\mathrm{sc}}^{\mathrm{e}}$ and single-hour dummies for D_{it} are included in $\mathbf{x}_{it}$. The single-hour dummies are constructed by rounding D_{it} to the nearest integer hour, producing 24 dummies to be interacted with $\mathrm{MSF}_{\mathrm{sc}}^{\mathrm{e}}$. Yet, one dummy is excluded because of collinearity with c_i. The median $\mathrm{MSF}_{\mathrm{sc}}^{\mathrm{e}}$ is subtracted from $\mathrm{MSF}_{\mathrm{sc}}^{\mathrm{e}}$ so $\alpha\left(D_{it}\right)$ represents the median chronotype.

Panel conditioning effects can operate in a longitudinal survey, which may entail positive or negative consequences for data quality (e.g., Bach, 2021). Respondents may gain a better understanding of the meaning of the questions with repeated administration of the questionnaire, increasing the reliability of their responses (Kroh, Winter and Schupp, 2016). On the other hand, respondents may learn to falsely respond some questions to skip follow-up questions, lowering the quality of the data (e.g., Davis, 2011). To account for panel conditioning effects, a complete set of dummy variables for the number of previous interviews is included in $\mathbf{x}_{it}$. This number can be 0, 1, 2, …, 23, producing 24 dummies. Yet, one dummy is excluded because of collinearity with the intercept.

Let $\mathbf{z}_{it}\boldsymbol{\theta} \equiv \alpha(D_{it}) + \mathbf{x}_{it}\boldsymbol{\beta}$ with $K = \dim(\boldsymbol{\theta})$. Under the strict exogeneity assumption $E(u_{it} \mid \mathbf{z}_{i1}, \mathbf{z}_{i2}, \ldots, \mathbf{z}_{iT_i}, c_i) = 0$, $\boldsymbol{\theta}$ can be estimated by ordinary least squares (OLS) of

$$\Delta y_{it} = \Delta \mathbf{z}_{it}\boldsymbol{\theta} + e_{it}, \qquad t = 2, 3, \ldots, T_i, \tag{4.2}$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$, $\Delta \mathbf{z}_{it} = \mathbf{z}_{it} - \mathbf{z}_{i,t-1}$, and $e_{it} = u_{it} - u_{i,t-1}$. The e_{it} are assumed to be independently distributed across individuals but no restrictions are placed on the form of the autocovariances for a given individual. Heteroskedasticity and serial correlation consistent standard errors are obtained from the following variance matrix estimator (Wooldridge, 2010, pages 172 and 318):

$$\hat{\mathrm{V}}\left(\hat{\boldsymbol{\theta}}\right) = \left(\sum_{i=1}^{N} \Delta \mathbf{Z}_i' \Delta \mathbf{Z}_i\right)^{-1} \left(\sum_{i=1}^{N} \Delta \mathbf{Z}_i' \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i' \Delta \mathbf{Z}_i\right) \left(\sum_{i=1}^{N} \Delta \mathbf{Z}_i' \Delta \mathbf{Z}_i\right)^{-1} \tag{4.3}$$

where $\Delta \mathbf{Z}_i$ is the $(T_i - 1) \times K$ matrix obtained by stacking $\Delta \mathbf{z}_{it}$ from $t = 2, 3, \ldots, T_i$ and $\hat{\mathbf{e}}_i$ is the $(T_i - 1) \times 1$ vector of OLS residuals $\hat{e}_{it}$, $t = 2, 3, \ldots, T_i$. Alternatively, a working correlation matrix for modeling within-individual correlations can be specified, and the resulting model can be estimated by population-averaged methods, called feasible generalized least squares (FGLS) estimators in econometrics. We provide the results of two FGLS estimators in a separate supplement (González Chapela, 2024). They reveal essentially the same patterns reported here.

To assess the strict exogeneity of $\{D_{it}\colon t = 1, \ldots, T_i\}$, $\alpha\left(D_{it}\right)$ will be added to equation (4.2) and then its statistical significance tested using the Wald test (Wooldridge, 2010, page 325).

4.2 Model types and model selection

Our objective is to arrive at a reasonable, parsimonious representation of $\alpha(D_{it})$. Hence, an information criterion is employed to select a model for $\alpha(D_{it})$ out of three linear-in-parameters model types: piecewise constant functions (specifically, those of Arechar et al., 2017; Binder, 2022; Durrant et al., 2011; Flynn, 2018; Valdez, 2019 and Weeks et al., 1987), a polynomial of degree three, and the cosinor model

$$\alpha(D_{it}) = \alpha_1 \sin(D_{it} \times 2\pi/24) + \alpha_2 \cos(D_{it} \times 2\pi/24), \tag{4.4}$$

where α_1 and α_2 are unknown parameters.

The cosinor model is a type of Fourier series representation in which sines and cosines are used to approximate complex mathematical waveforms (Brown and Czeisler, 1992; Cornelissen, 2014). Given the waveform character of the homeostatic and circadian propensities for sleep, cosinor may provide an appropriate representation of $\alpha(D_{it})$. The cosinor model has 1 peak and 1 trough separated by 12 hours and equal in amplitude and width, the locations of which are determined by α_1 and α_2. Twice the amplitude of the cosinor wave, or $2 \times \sqrt{\alpha_1^2 + \alpha_2^2}$, provides a measure of the extent of predictable change within the day.

The degree three polynomial is less restrictive than cosinor because the peak and the trough may not be separated by 12 hours and the amplitude and width of the peak may differ from those of the trough. On the other hand, a polynomial may not be periodic, i.e. its values may not repeat themselves every 24 hours. To ensure periodicity, the restriction $\alpha(0) = \alpha(24)$ is imposed, yielding

$$\alpha(D_{it}) = \alpha_1 D_{it}\left(1 - (D_{it}/24)^2\right) + \alpha_2 D_{it}^2 (1 - D_{it}/24). \tag{4.5}$$

To select among models, Schwarz's (1978) Bayesian information criterion

$$\text{BIC} = \ln \text{SSR} + \frac{K \ln\left(\sum_{i=1}^{N}(T_i - 1)\right)}{\left(\sum_{i=1}^{N}(T_i - 1)\right)} \tag{4.6}$$

is used, where SSR denotes a model's sum of squared residuals. BIC is preferred to other popular criteria when some modelling alternatives are nested (Nishii, 1988). The specification of $\mathbf{x}_{it}$ is kept the same throughout the selection process. Schwarz (1978) establishes the validity of BIC for independent and identically distributed observations. To guard against possible biases created by correlated e_{it}, the BIC values were recalculated using N in place of $\sum_{i=1}^{N}(T_i - 1)$ (StataCorp, 2019, page 104), producing the same selection of models.

4.3 Attrition

If attrition is driven by unobserved factors that do not change over the survey period, then removing c_i would correct for attrition bias. Nevertheless, one might still be concerned about attrition as a consequence of unobserved interview-variant factors. We use a variant of the procedure proposed by Wooldridge (2010, page 837) to test and correct for attrition bias, though we note that this procedure does not correct for individuals selected to participate in the SUWNJ who never responded. As the data for each individual are organized by interview number, attrition is an absorbing state.

Let s_{it} denote the interview completion indicator, with $s_{it}=1$ if individual i completed the t interview and $s_{it}=0$ if i abandoned the survey right after the $t-1$ interview. The completion equation for interview t conditional on $s_{i,t-1}=1$ is

$$s_{it}=1\left[\mathbf{w}_{it}\boldsymbol{\delta}+v_{it}>0\right], \qquad t=2,3,\ldots,T_i, \tag{4.7}$$

where $1[\cdot]$ is the indicator function, $\mathbf{w}_{it}$ is a set of variables that are observed whether or not the individual attrited, $\boldsymbol{\delta}$ is a vector of unknown parameters, and v_{it} is a standard normal error term assumed independent of $\left(\Delta\mathbf{z}_{it},\mathbf{w}_{it},s_{i,t-1}=1\right)$. Nonrandom attrition occurs when v_{it} and e_{it} are correlated.

Assuming that e_{it} is independent of $\left(\Delta\mathbf{z}_{it},\mathbf{w}_{it}\right)$ and that $\mathrm{E}\left(e_{it}\mid v_{it},s_{i,t-1}=1\right)=\rho_t v_{it}$, ρ_t being an unknown parameter, the unknown parameters of equation (4.1) can be estimated by OLS of

$$\Delta y_{it}=\Delta\mathbf{z}_{it}\boldsymbol{\theta}+\rho_2 d2_t\hat{\lambda}_{it}+\ldots+\rho_{24}d24_t\hat{\lambda}_{it}+\varepsilon_{it}, \qquad t=2,3,\ldots,T_i. \tag{4.8}$$

In this expression, $d2_t,\ldots,d24_t$ are interview dummies so that $dj_t=1$ if $t=j$ and $dj_t=0$ if $t\neq j$, $\hat{\lambda}_{it}\equiv\lambda\left(\mathbf{w}_{it}\hat{\boldsymbol{\delta}}\right)=\phi\left(\mathbf{w}_{it}\hat{\boldsymbol{\delta}}\right)\big/\Phi\left(\mathbf{w}_{it}\hat{\boldsymbol{\delta}}\right)$, where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the pdf and cdf of the standard normal distribution, is the estimated inverse Mills ratio, and ε_{it} is an error term.

An estimator of $\boldsymbol{\delta}$ is available from pooled probit estimation of the interview completion equation:

$$\mathrm{P}\left(s_{it}=1\middle|\mathbf{w}_{it},s_{i,t-1}=1\right)=\Phi\left(\mathbf{w}_{it}\boldsymbol{\delta}\right), \qquad t=2,3,\ldots,T_i. \tag{4.9}$$

We use pooled probit because $\boldsymbol{\delta}$ is assumed to be constant across interviews. If $\boldsymbol{\delta}$ was allowed to change (as in Wooldridge's original formulation), a probit would be estimated for each t. However, this approach is problematic because in many occasions the variables included in $\mathbf{w}_{it}$ perfectly predict one of the outcomes. The vector $\mathbf{w}_{it}$ comprises single-hour dummies for $D_{i,t-1}$, $\mathbf{x}_{i,t-1}$, and the number of weeks passed between $t-2$ and $t-1$. (For $t=2$, we count the number of weeks between the week when the initial invitations to participate in the survey were sent and the week of the first interview.)

Attrition bias can be tested by a joint test of H_0: $\rho_t=0$, $t\geq 2$, in equation (4.8). If H_0 is rejected, standard errors are corrected for the presence of estimated parameters in $\hat{\lambda}_{it}$ drawing upon Arellano and Meghir (1992).

4.4 Weighting

Since the regressors utilized to create "current week weights" are absorbed in c_i, model (4.1) includes all the design variables and thus the sampling design can be considered ignorable (Pfeffermann, 1993). Hence, the main analysis is conducted without sampling weights. However, reporting weighted estimates is useful as a misspecification check, as the failure to model heterogenous effects can generate significant contrasts between weighted and unweighted estimates (e.g., Solon, Haider and Wooldridge, 2015). Hence, equation (4.2) will be re-estimated by weighted least squares (WLS).

4.5 Multiple inference

Nearly all of our groupings of data quality measures contain more than one measure. Consequently, significant effects may emerge by chance for some measure even if no effect on the grouping exists. To control for this, Bonferroni corrections are performed and significance is declared at level $0.05/M$, M being the number of measures in the grouping.

5. Results

5.1 Model selection

Table 5.1 lists the best-fitting models of $\alpha\left(D_{it}\right)$. The cosinor model is the preferred option for analyzing most of the data quality measures. However, Binder's (2022) piecewise constant function (indicators for 06:00 to 11:59, 12:00 to 18:59, and 19:00 to 05:59) is the best fitting alternative for the number of hours not coded in the diary (HMissing), the probability of reporting all mood at home categories in multiples of 50 (P_{MOOD50}), and the probability of reporting expenditure on eating out in multiples of 50 ($P_{EATING\text{-}OUT50}$). The degree three polynomial is favored for the probability of being among the 5% highest completion times ($P_{IVDUR5H}$) and the probability of reporting expenditure on eating out in multiples of 100 ($P_{EATING\text{-}OUT100}$). For the probability of reporting all mood at home categories in multiples of 25 (P_{MOOD25}), Durrant et al.'s (2011) piecewise constant function (indicators for 00:00 to 11:59, 12:00 to 16:59, and 17:00 to 23:59) is preferred.

Table 5.1
Model selected for $\alpha(D_{it})$.

Dependent variable	Model	BIC value
P_{INR}	Cosinor	13.259
NumAct	Cosinor	13.585
HMissing	Piecewise constant (Binder, 2022)	11.081
IvDur	Cosinor	14.327
$P_{IVDUR5L}$	Cosinor	16.476
$P_{IVDUR5H}$	Degree 3 polynomial	16.716
P_{MOOD10}	Cosinor	18.031
P_{MOOD25}	Piecewise constant (Durrant et al., 2011)	17.063
P_{MOOD50}	Piecewise constant (Binder, 2022)	16.670
$P_{FOODAH50}$	Cosinor	18.096
$P_{FOODAH100}$	Cosinor	18.018
$P_{EATING\text{-}OUT50}$	Piecewise constant (Binder, 2022)	17.956
$P_{EATING\text{-}OUT100}$	Degree 3 polynomial	17.759

5.2 Baseline results

The results of estimating equation (4.2) with the functional forms listed in Table 5.1 are presented in Tables 5.2 and 5.3. Table 5.2 shows the results for the percent item nonresponse (P_{INR}), the time-diary measures, and interview completion time. Table 5.3 gathers the results for the indicators of rounding. The bottom rows of both tables list the p-values for the tests of significance of $\alpha(D_{it})$ and strict exogeneity of $\{D_{it}: t = 1, \ldots, T_i\}$.

A statistically significant $\alpha(D_{it})$ is detected in some of the regressions, which suggests the existence of some effects on data quality of D_{it}. In a p-value sense, the strongest evidence is found in the regressions for the number of activities (NumAct) and the probability of being among the 5% lowest completion times ($P_{IVDUR5L}$). The null of no effect is also rejected at 5% in the regressions for P_{INR} and $P_{EATING\text{-}OUT100}$. No statistically significant effect is detected in the remaining cases.

In the case of $P_{EATING\text{-}OUT100}$, the rejection of the null does not hold if zero expenditure (reported in 28% of the interviews) is assumed not to reflect rounding (p-value 0.55). In addition, the effect on $P_{EATING\text{-}OUT100}$ does not survive a Bonferroni correction for two simultaneous tests in the group of measures assessing expenditure on eating out, which would require p-value $<$ 0.025.

Table 5.2
Time of day of interview effects on data quality.

	(1)		(2)		(3)		(4)		(5)		(6)	
	P_{INR}		NumAct		HMissing		IvDur (min)		$P_{IVDUR5L}$		$P_{IVDUR5H}$	
Explanatory variables	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**
12:00–18:59					0.049	0.029						
19:00–05:59					0.048	0.039						
$D_{it}(1-(D_{it}/24)^2)$											-0.169	0.238
$D_{it}^2(1-D_{it}/24)$											0.039	0.024
$\sin(D_{it} \times 2\pi/24)$	-0.147*	0.070	0.259*	0.087			-0.020	0.135	-0.405	0.361		
$\cos(D_{it} \times 2\pi/24)$	0.040	0.064	-0.191*	0.086			-0.282*	0.138	0.814*	0.373		
Tuesday	-0.190	0.108	1.480*	0.137	-0.103*	0.039	0.733*	0.192	-0.251	0.491	1.013	0.624
Wednesday	-0.157	0.118	1.207*	0.150	-0.079	0.041	0.430*	0.204	-0.960	0.547	-0.046	0.656
Thursday	0.015	0.131	1.233*	0.159	-0.048	0.043	0.653*	0.226	-0.380	0.573	1.146	0.762
Friday	-0.008	0.118	1.132*	0.152	-0.075	0.042	0.272	0.206	0.188	0.505	-0.138	0.656
Saturday	0.012	0.124	0.755*	0.153	-0.027	0.041	0.387	0.246	0.514	0.605	0.140	0.781
Sunday	0.289*	0.140	0.054	0.171	0.001	0.045	-0.380	0.242	0.151	0.660	-0.882	0.759
Worked	0.408*	0.137	-1.711*	0.158	-0.120*	0.037	-0.174	0.173	1.378*	0.609	0.623	0.555
Sleep duration	-0.024	0.018	-0.080*	0.025	-0.016*	0.007	-0.131*	0.028	0.400*	0.089	-0.178	0.095
Significance of $\alpha(D_{it})$	[0.04]		[0.00]		[0.22]		[0.10]		[0.01]		[0.14]	
Strict exogeneity of $\{D_{it}\}$	[0.01]		[0.73]		[0.10]		[0.03]		[0.03]		[0.79]	
Observations	25,184		25,184		25,184		25,184		25,184		25,184	

Notes: Estimations are conducted using first differencing, and include complete sets of first-differenced dummies for number of previous interviews and first-differenced single-hour dummies for D_{it} interacted with MSF_{sc}^{c}. The dependent variables whose name start with P are binary indicators for the outcome given in the name's subscript scaled as a percentage. Standard errors take account of heteroskedasticity and clustering at individual level. Probability values are in brackets. *: Significant at 5%.

The estimated effects on P_{INR}, NumAct, and $P_{IVDUR5L}$, calculated by zeroing out all the controls and varying D_{it}, are depicted in Figure 5.1. The three graphs tell a rather consistent story: The quality of the data peaks in the early morning and is worst in the evening. The estimated change within the day is 0.30

percentage points (pps) for P_{INR}, 0.64 activities for NumAct, and 1.82 pps for P_{IVDUR5L}, representing 11, 4, and 36% of the corresponding mean.

Table 5.3
Time of day of interview effects on data quality.

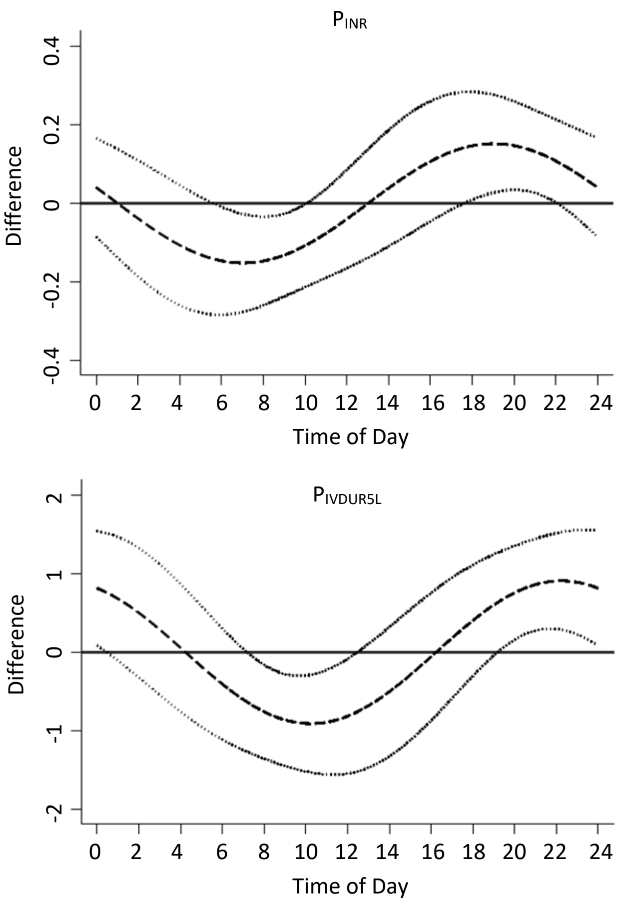
	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
	P_{MOOD10}		P_{MOOD25}		P_{MOOD50}		P_{FOODAH50}		$P_{\text{FOODAH100}}$		$P_{\text{EATING-OUT50}}$		$P_{\text{EATING-OUT100}}$	
Explanatory variables	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**
12:00–18:59					-0.109	0.500					-1.594	1.108		
19:00–05:59					-0.023	0.675					-1.740	1.446		
12:00–16:59			0.589	0.589										
17:00–23:59			-0.502	0.733										
$D_{it}(1-(D_{it}/24)^2)$													0.839	0.449
$D_{it}^2(1-D_{it}/24)$													-0.032	0.045
$\sin(D_{it}\times 2\pi/24)$	-0.714	0.828					-0.277	0.898	-0.280	0.822				
$\cos(D_{it}\times 2\pi/24)$	0.820	0.820					-1.598	0.837	-0.883	0.799				
Tuesday	-0.531	1.210	0.176	0.770	-0.706	0.662	-1.212	1.276	0.833	1.239	-0.501	1.378	0.273	1.267
Wednesday	-1.326	1.361	-0.768	0.842	-1.661*	0.722	-1.028	1.433	0.720	1.389	1.332	1.552	3.122*	1.417
Thursday	-0.220	1.358	0.374	0.867	-0.873	0.737	-1.294	1.413	-0.873	1.388	0.669	1.575	1.909	1.416
Friday	-0.444	1.277	0.243	0.826	-0.455	0.663	-1.570	1.374	0.163	1.338	1.603	1.485	2.055	1.353
Saturday	-0.162	1.394	1.094	0.885	-0.068	0.721	-0.116	1.476	0.282	1.447	0.380	1.554	0.971	1.424
Sunday	-0.907	1.544	-0.408	0.920	-1.711*	0.776	-3.351*	1.569	-2.001	1.539	0.264	1.692	2.627	1.529
Worked	-2.307*	1.103	0.721	0.678	-0.160	0.573	-1.491	1.272	-2.048	1.238	-2.277	1.360	-0.664	1.201
Sleep duration	0.215	0.182	0.223	0.117	0.166	0.097	-0.378	0.194	-0.326	0.192	0.193	0.216	0.244	0.196
Significance of $\alpha(D_{it})$	[0.28]		[0.28]		[0.97]		[0.16]		[0.54]		[0.30]		[0.04]	
Strict exogeneity of $\{D_{it}\}$	[0.87]		[0.91]		[0.42]		[0.25]		[0.55]		[0.27]		[0.55]	
Observations	25,083		25,083		25,083		23,957		23,957		20,874		20,874	

Notes: See notes to Table 5.2.

The number of activities might be lower when the diary is completed in the evening due to the longer period of recall. To disentangle the effect of D_{it} from that of the recall period, the sample is split into weekday (Monday–Thursday) and weekend (Friday–Sunday) diaries. The results of re-estimating the equation for NumAct in each of the two subsamples of diaries are presented in Table 5.4. (Remember that the day indicated in the tables is the interview day.) $\alpha(D_{it})$ becomes insignificant in the subsample of weekend diaries, although this conclusion is partly driven by the imprecision of the estimates. Moreover, the extent of change within a weekend day comes out much smaller than within a weekday: 0.43 vs. 1.01 activities, representing 2.7 and 5.9% of the corresponding mean. Thus, a large extent of the daily rhythm of NumAct is driven by the period of recall.

As for the effects of the controls, the number of activities is higher in Monday–Thursday diaries, and interviews appear to be longer on Tuesdays, Wednesdays, and Thursdays. Working and sleeping longer on the diary day have contradictory effects on the quality of time-diary data, as they tend to reduce both the number of activities and the number of hours not coded. These effects are probably reflecting that working and sleeping longer reduce the time available for other activities, and the reduction of activities facilitates their recalling. Working on the diary day increases the likelihood that the interview is among the 5% shortest by 1.4 pps (or 28%).

Figure 5.1 Time of day of interview effects on data quality.



Notes: The effects (dashed lines) are calculated from the corresponding estimations in Table 5.2. Dotted lines delimit the 95% confidence interval.

Table 5.4
Time of day of interview effects on NumAct, by diary day.

	(1) Monday–Thursday diaries		(2) Friday–Sunday diaries	
Explanatory variables	**Coef**	**S.E.**	**Coef**	**S.E.**
$\sin(D_{it} \times 2\pi/24)$	0.476*	0.110	0.213	0.218
$\cos(D_{it} \times 2\pi/24)$	-0.171	0.112	-0.028	0.209
Monday			Ref.	
Tuesday	0.236	0.165		
Wednesday	0.099	0.172		
Thursday	0.088	0.165		
Friday	Ref.			
Saturday			0.238	0.263
Sunday			-0.415	0.304
Worked	-1.626*	0.213	-1.374*	0.345
Sleep duration	-0.093*	0.030	-0.021	0.057
Significance of $\alpha(D_{it})$	[0.00]		[0.59]	
Observations	14,904		3,073	

Notes: Estimations are conducted using first differencing, and include complete sets of first-differenced dummies for number of previous interviews and first-differenced single-hour dummies for D_{it} interacted with MSF_{sc}^{e}. Standard errors take account of heteroskedasticity and clustering at individual level. Probability values are in brackets. *: Significant at 5%.

5.3 Supplementary analyses

5.3.1 Strict exogeneity

We have been assuming that the variation in D_{it} within respondents is strictly exogenous. This assumption would be questioned if, for example, respondents rush through the survey or become distracted at times of day when the opportunity cost of completing the interview is highest. The p-value for the test of strict exogeneity of $\{D_{it}: t = 1, \ldots, T_i\}$ is shown in the next-to-last row of Tables 5.2 and 5.3. At 5% level, exogeneity is questioned in the regressions for P_{INR}, completion time (IvDur), and P_{IVDUR5L}. Since the within (or fixed effects) estimator tends to be more robust to the violation of strict exogeneity, we re-estimated equation (4.1) with the OLS estimator from the regression

$$y_{it} - \bar{y}_i = \left(\alpha\left(D_{it}\right) - \overline{\alpha\left(D_i\right)}\right) + \left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right)\boldsymbol{\beta} + u_{it} - \bar{u}_i \tag{5.1}$$

where $\bar{y}_i = T_i^{-1}\sum_{t=1}^{T_i} y_{it}$, $\overline{\alpha\left(D_i\right)} = T_i^{-1}\sum_{t=1}^{T_i}\alpha\left(D_{it}\right)$, and so on. The null hypothesis $\text{H}_0\text{: } \alpha\left(D_{it}\right) = 0$ is rejected in the regression for P_{INR} (p-value 0.01), but not rejected in the regressions for IvDur and P_{IVDUR5L} (p-value 0.39 in both cases). Note, however, that both the first-difference and the within estimators may be biased when strict exogeneity fails.

5.3.2 Robustness

The estimates change little when sleep duration is excluded from $\mathbf{x}_{it}$, or when D_{it} is approximated by the end time of the time-use section of the questionnaire or by randomly selected points within the start and end times of the interview (results not shown). When the 2,578 interviews presenting inconsistent going-to-bed time are included in the sample, the preferred model for $\alpha\left(D_{it}\right)$ changes in some cases (Table A.1 in the Appendix). A statistically significant $\alpha\left(D_{it}\right)$ is detected in the regressions for P_{INR}, NumAct, HMissing, and IvDur, whereas $\alpha\left(D_{it}\right)$ becomes insignificant in the regression for P_{IVDUR5L} (Tables A.2 and A.3 in the appendix). When an effect is detected, it suggests that data quality peaks in the early morning.

5.3.3 Attrition

Table 5.5 presents probit estimation output for the decision to complete an interview. It shows selected $\boldsymbol{\delta}$ coefficients plus average marginal effects (AMEs) calculated by averaging marginal effects across observations. Completing the $t-1$ interview on Tuesday–Saturday increases the probability of completing the t interview. Working on the diary day increases that probability by 1.8 pps, whereas one more hour of sleep reduces it by 0.6 pps. The number of weeks passed between $t-2$ and $t-1$ is a strong predictor for completing the t interview, whose likelihood reduces by 3.0 pps with every week passed. None of the single-hour dummies for $D_{i,t-1}$ attains significance at 5% (not shown).

After correcting for nonrandom attrition, the cosinor model becomes the preferred option for analyzing P_{MOOD50}, while Binder's (2022) piecewise constant function comes out as the best fitting alternative for the probability of reporting expenditure on food at home in multiples of 100 ($P_{\text{FOODAH100}}$). The null hypothesis of no attrition bias is questioned in the regressions for IvDur, P_{IVDUR5L}, and $P_{\text{FOODAH100}}$. However, the attrition-corrected estimates (reported in Tables A.4 and A.5 in the appendix) reveal essentially the same patterns as the non-attrition-corrected ones. The correction for nonrandom attrition makes less dubious the assumption of strict exogeneity of $\alpha\left(D_{it}\right)$ in the regressions for IvDur and $\mathrm{P}_{\mathrm{IVDUR5L}}$ (p -value 0.12 in both cases).

Table 5.5
Probit for interview completion.

	Dependent variable: s_{it}, $t \geq 2$			
Explanatory variables ($t-1$)	**Coef.**	**S.E.**	**AME**	**S.E.**
Tuesday	0.154*	0.034	0.034*	0.008
Wednesday	0.175*	0.038	0.039*	0.008
Thursday	0.184*	0.038	0.040*	0.008
Friday	0.215*	0.037	0.047*	0.008
Saturday	0.174*	0.039	0.038*	0.009
Sunday	0.058	0.042	0.014	0.010
Worked	0.089*	0.028	0.018*	0.005
Sleep duration	-0.027*	0.004	-0.006*	0.001
Weeks between $t-2$ and $t-1$	-0.145*	0.006	-0.030*	0.001
Intercept	1.121*	0.105		
R-squared	0.070			
Observations	32,779			
Mean of s_{it}	0.859			

Notes: Observations for the last interview are excluded because individuals did certainly not continue in the survey. Includes single-hour dummies for $D_{i,t-1}$, dummies for number of previous interviews, and $\mathrm{MSF}_{\mathrm{sc}}^{\mathrm{e}}$ interacted with single-hour dummies for $D_{i,t-1}$. Standard errors take account of heteroskedasticity and clustering at individual level. R-squared equals one minus the ratio of the log likelihood of the fitted function to the log likelihood of a function with only an intercept. *: Significant at 5%.

5.3.4 Weights

Tables 5.6 and 5.7 present the WLS estimates. A statistically significant $\alpha\left(D_{it}\right)$ is not detected in most of the regressions shown. While in some cases (e.g., the regression for NumAct), the WLS estimated coefficients are smaller than the OLS ones, in most cases the inference is driven by the larger standard errors. A statistically significant $\alpha\left(D_{it}\right)$ is detected in the regression for $\mathrm{P}_{\mathrm{IVDUR5H}}$ (p -value 0.03), but this effect does not survive a Bonferroni correction for simultaneous tests in the group of measures assessing completion time. The null of no effect is also rejected at 5% in the regressions for $\mathrm{P}_{\mathrm{EATING\text{-}OUT50}}$ and $\mathrm{P}_{\mathrm{EATING\text{-}OUT100}}$ (p -value 0.01 in both cases), but in both cases the rejection of the null does not hold if zero expenditure is assumed not to reflect rounding (p -values 0.43 and 0.35 respectively).

5.3.5 Subpopulations

Finally, we split the sample by educational attainment (at most some college vs. college diploma) to investigate time of day of interview effects with certain types of individuals. Although cognitive abilities are important predictors of educational attainment, we do not expect to find big differences between demographic groups as our estimates are net of synchrony and cognitive ability effects. Indeed, although the best-fitting model of $\alpha\left(D_{it}\right)$ changes for most of the dependent variables in both subpopulations, the main conclusions are preserved (results not shown).

Table 5.6
Time of day of interview effects on data quality. Weighted estimates.

	(1)		(2)		(3)		(4)		(5)		(6)	
	P_{INR}		NumAct		HMissing		IvDur (min)		P_{IVDUR5L}		P_{IVDUR5H}	
Explanatory variables	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**
12:00–18:59					0.224	0.124						
19:00–05:59					0.063	0.113						
$D_{it}(1-(D_{it}/24)^2)$											-0.993*	0.474
$D_{it}^2(1-D_{it}/24)$											0.153*	0.058
$\sin(D_{it}\times 2\pi/24)$	-0.298	0.249	0.061	0.207			-0.581*	0.295	0.363	0.662		
$\cos(D_{it}\times 2\pi/24)$	0.133	0.139	-0.111	0.157			-0.314	0.293	0.082	1.023		
Tuesday	-0.241	0.242	1.222*	0.260	-0.133	0.093	1.341*	0.374	-0.299	1.306	1.439	1.417
Wednesday	-0.124	0.334	1.248*	0.359	-0.185	0.138	1.028*	0.438	0.077	2.241	0.812	1.507
Thursday	0.483	0.331	0.881*	0.312	0.020	0.098	1.025*	0.409	0.543	1.653	1.879	1.619
Friday	0.512	0.336	0.703*	0.291	0.104	0.127	0.621	0.431	1.067	1.422	0.262	1.533
Saturday	0.354	0.291	-0.009	0.411	0.038	0.128	-0.243	0.483	1.911	2.031	-1.101	1.531
Sunday	-0.034	0.358	0.109	0.339	-0.075	0.130	-0.959*	0.485	0.381	1.556	-3.991*	1.663
Worked	0.200	0.250	-2.582*	0.260	-0.158	0.084	-0.706*	0.320	3.005*	1.494	-1.619	1.025
Sleep duration	-0.057	0.043	-0.077	0.046	-0.030	0.017	-0.097	0.062	0.656*	0.209	0.040	0.276
Significance of $\alpha(D_{it})$	[0.38]		[0.69]		[0.14]		[0.14]		[0.84]		[0.03]	
Strict exogeneity of $\{D_{it}\}$	[0.26]		[0.92]		[0.36]		[0.02]		[0.25]		[0.23]	
Observations	25,184		25,184		25,184		25,184		25,184		25,184	

Notes: Estimations are conducted using first differencing, and include complete sets of first-differenced dummies for number of previous interviews and first-differenced single-hour dummies for D_{it} interacted with $\text{MSF}_{\text{sc}}^{\text{c}}$. The dependent variables whose name start with P are binary indicators for the outcome given in the name's subscript scaled as a percentage. Standard errors take account of heteroskedasticity and clustering at individual level. Probability values are in brackets. *: Significant at 5%.

Table 5.7
Time of day of interview effects on data quality. Weighted estimates.

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
	P_{MOOD10}		P_{MOOD25}		P_{MOOD50}		P_{FOODAH50}		$P_{\text{FOODAH100}}$		$P_{\text{EATING-OUT50}}$		$P_{\text{EATING-OUT100}}$	
Explanatory variables	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**
12:00–18:59					-1.514	1.575					-1.255	2.295		
19:00–05:59					0.127	1.500					-8.426*	3.040		
12:00–16:59			-1.515	1.582										
17:00–23:59			-3.035*	1.436										
$D_{it}(1-(D_{it}/24)^2)$													1.074	0.889
$D_{it}^2(1-D_{it}/24)$													0.037	0.096
$\sin(D_{it}\times 2\pi/24)$	-1.143	1.758					-2.348	2.000	-2.029	2.399				
$\cos(D_{it}\times 2\pi/24)$	0.236	2.192					-4.367	2.250	-4.183	2.698				
Tuesday	4.451	2.827	0.293	1.508	-0.836	1.199	2.645	2.530	3.310	2.620	-5.478	2.900	-2.843	2.732
Wednesday	1.520	2.894	-0.776	1.728	-0.807	1.502	1.089	3.120	0.796	3.130	-0.336	3.387	3.425	3.026
Thursday	5.568	3.929	0.631	1.738	-0.993	1.292	3.577	2.871	2.227	3.014	0.251	4.177	1.842	3.670
Friday	7.726*	3.824	-0.968	1.819	-1.234	1.152	2.891	3.447	3.313	3.682	4.003	3.041	2.739	2.822
Saturday	3.580	3.349	0.399	1.659	-0.234	1.321	3.190	3.127	3.793	3.692	-0.065	3.192	1.163	3.009
Sunday	5.240	3.578	-0.992	1.842	-2.572	1.450	2.211	3.327	0.165	3.645	-1.928	3.554	0.044	3.236
Worked	-3.964	2.072	0.514	1.111	-1.931*	0.918	-0.547	2.809	-0.775	2.856	-3.397	2.839	-2.032	2.383
Sleep duration	0.282	0.361	0.174	0.219	0.149	0.178	-0.855*	0.368	-0.576	0.440	0.330	0.441	0.280	0.432
Significance of $\alpha(D_{it})$	[0.77]		[0.11]		[0.35]		[0.09]		[0.28]		[0.01]		[0.01]	
Strict exogeneity of $\{D_{it}\}$	[0.60]		[0.68]		[0.97]		[0.13]		[0.34]		[0.41]		[0.92]	
Observations	25,083		25,083		25,083		23,957		23,957		20,874		20,874	

Notes: See notes to Table 5.6.

6. Summary and discussion

The analysis of high-frequency longitudinal microdata from the SUWNJ reveals no evidence of a time of day of interview effect on the quality of time-diary data (beyond the effect exerted by the length of the recall period), or on the tendency to report rounded values of subjective probabilities or food expenditure. As regards the period of recall, we found that self-completing a yesterday diary in the evening reduces the number of activities reported, whereas the amount of time not coded suffers no meaningful daily fluctuation. Thus, it appears that some activities are underreported and the duration of others is overestimated, introducing error in the measurement of the use of time. All these findings have been developed accounting for inter-individual differences in cognitive ability and synchrony effects, which may explain why they persist across education groups. They also appear to be robust to a range of alternative specifications assessing the impact of nonrandom attrition, unmodeled heterogenous effects, and different measures of time of day of interview. Although there is some evidence to indicate that item nonresponse and the probability that interview completion time is among the 5% shortest increase when the survey is completed in the evening, a more thorough assessment requires instrumental variables.

Our most reliable results support the conclusion of previous research that survey data quality is insensitive to the time of day of interview (Ziniel, 2008; Dickinson and McElroy, 2010; Binder, 2022), but disagree with those of Flynn (2018), who found that respondents who start a survey in the evening answer significantly more questions than those who start it in the morning/afternoon. Yet, Flynn's (2018) sample is made up of firm representatives, and completing a survey outside of regular office hours might benefit from reduced time pressures. As the unemployed (as compared to the employed) do not have to adhere to the limitations of work hours, their time of day of interview can be more evenly spread over the 24 hours, facilitating the identification of effects around the clock. It is also worth noting that, in contrast to MTurk samples (e.g., Binder, 2022), interviews appear to be longer on Thursdays (plus Tuesdays and Wednesdays), and that the number of activities reported is higher in Monday–Thursday diaries as in Juster (1986).

Overall, therefore, it appears that beyond the effect exerted by the length of the recall period, inducing respondents to complete surveys at specific times of the day might have limited impacts on measurement error. Thus, survey practitioners should not worry much about the consequences for measurement error of seeking to interview subjects at times of the day they are most likely to be contactable.

All that said, we recognize some limitations of this study. As regards the question of whether we uncover causal effects for the population being studied, it must be noted that we lack data on the situational context in which the interviews were completed (e.g., where the respondent was and what he/she was doing), and as argued by Bison and Zhao (2023) the temporal and situational contexts might be correlated. However, it is difficult to suggest instrumental variables sufficiently correlated with time of day of interview but uncorrelated with idiosyncratic errors, as most variables in the SUWNJ refer to days other than the interview day. Also, although the percentage of SUWNJ interviews completed from a mobile device must have been low (Callegaro, 2010, for example, reports that among all respondents who attempted to complete an online customer satisfaction survey conducted in North America in June 2010, 2.6% did so from a mobile device), if completing an interview from a mobile device affects the quality of the data (as the evidence reviewed in Toninelli and Revilla, 2020 suggests) and depends on the time of day, our results might contain bias. As regards the predictive value of our findings in a different context, it must be noted that the results obtained for the unemployed might not be representative for broader populations if, for example, the activities conducted before completing the survey interact with sleepiness/fatigue.

In addition, insufficient data prevented us from investigating the existence of time of day of interview effects on alternative measures of data quality, such as survey breakoff and response errors caused by social desirability or extreme, midpoint, or nondifferentiated responding. As regards the effects of the length of the recall period, it seems worth investigating whether the administration of a yesterday diary by an interviewer (who could foster respondents' attention and motivation), or the "own words" reporting of activities by respondents (which avoids the process of mapping the answer onto the appropriate response option), could improve the quality of time-diary data.

Acknowledgements

This paper has greatly benefited from the comments and suggestions of Jean-Francois Beaumont, Kristen Olson, and, especially, an anonymous Associate Editor and several anonymous Referees. Thanks also to Andreas Mueller for assistance with SUWNJ data. This study was supported by the Government of Aragón, grant S32-23R.

Appendix

Table A.1
Model selected for $\alpha(D_{it})$. Including observations with inconsistent going-to-bed time.

Dependent variable	Model	BIC value
P_{INR}	Piecewise constant (Binder, 2022)	13.405
NumAct	Cosinor	13.731
HMissing	Piecewise constant (Binder, 2022)	11.271
IvDur	Piecewise constant (Binder, 2022)	14.460
$P_{IVDUR5L}$	Cosinor	16.590
$P_{IVDUR5H}$	Degree 3 polynomial	16.849
P_{MOOD10}	Piecewise constant (Binder, 2022)	18.162
P_{MOOD25}	Piecewise constant (Durrant et al., 2011)	17.211
P_{MOOD50}	Cosinor	16.820
$P_{FOODAH50}$	Degree 3 polynomial	18.226
$P_{FOODAH100}$	Piecewise constant (Binder, 2022)	18.145
$P_{EATING\text{-}OUT50}$	Piecewise constant (Binder, 2022)	18.080
$P_{EATING\text{-}OUT100}$	Degree 3 polynomial	17.887

Table A.2
Time of day of interview effects on data quality. Including observations with inconsistent going-to-bed time.

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
	P_{INR}		NumAct		HMissing		IvDur (min)		P_{IVDUR5L}		P_{IVDUR5H}		$P_{\text{PM-AM}}$	
Explanatory variables	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**
12:00–18:59	0.203*	0.073			0.070*	0.027	-0.368*	0.140						
19:00–05:59	0.221*	0.101			0.054	0.036	-0.461*	0.189						
$D_{it}(1-(D_{it}/24)^2)$											-0.106	0.219		
$D_{it}^2(1-D_{it}/24)$											0.027	0.022		
$\sin(D_{it}\times 2\pi/24)$			0.264*	0.084					-0.099	0.355			0.611	0.499
$\cos(D_{it}\times 2\pi/24)$			-0.184*	0.080					0.427	0.333			-1.069*	0.495
Tuesday	-0.177	0.101	1.420*	0.131	-0.086*	0.037	0.772*	0.184	-0.160	0.447	1.114	0.602	-2.738*	0.696
Wednesday	-0.145	0.109	1.180*	0.141	-0.056	0.039	0.434*	0.196	-0.672	0.502	0.075	0.635	-2.417*	0.763
Thursday	0.024	0.121	1.247*	0.150	-0.011	0.041	0.780*	0.213	-0.469	0.519	1.255	0.724	-1.503	0.788
Friday	-0.009	0.113	1.091*	0.145	-0.057	0.042	0.355	0.196	0.095	0.476	0.241	0.626	-2.161*	0.748
Saturday	0.072	0.121	0.755*	0.147	-0.002	0.041	0.507*	0.233	0.838	0.561	0.482	0.755	-1.156	0.815
Sunday	0.270*	0.136	-0.025	0.163	0.029	0.046	-0.283	0.228	-0.199	0.613	-0.538	0.728	-0.759	0.879
Worked	0.369*	0.129	-1.687*	0.152	-0.123*	0.035	-0.206	0.164	1.401*	0.565	0.454	0.531	-2.210*	0.659
Sleep duration	-0.028	0.018	-0.056*	0.023	-0.015*	0.007	-0.105*	0.026	0.361*	0.082	-0.131	0.090	-1.803*	0.124
Significance of $\alpha(D_{it})$	[0.01]		[0.00]		[0.03]		[0.01]		[0.35]		[0.30]		[0.02]	
Strict exogeneity of $\{D_{it}\}$	[0.02]		[0.61]		[0.17]		[0.57]		[0.01]		[0.64]		[0.65]	
Observations	28,576		28,576		28,576		28,576		28,576		28,576		28,576	

Notes: Estimations are conducted using first differencing, and include complete sets of first-differenced dummies for number of previous interviews and first-differenced single-hour dummies for D_{it} interacted with $\text{MSF}_{\text{sc}}^{\text{e}}$. The dependent variables whose name start with P are binary indicators for the outcome given in the name's subscript scaled as a percentage. Standard errors take account of heteroskedasticity and clustering at individual level. Probability values are in brackets. *: Significant at 5%.

A

Table A.3
Time of day of interview effects on data quality. Including observations with inconsistent going-to-bed time.

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
	P_{MOOD10}		P_{MOOD25}		P_{MOOD50}		P_{FOODAH50}		$P_{\text{FOODAH100}}$		$P_{\text{EATING-OUT50}}$		$P_{\text{EATING-OUT100}}$	
Explanatory variables	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**
12:00–18:59	1.065	0.895							-0.600	0.911	-1.238	1.043		
19:00–05:59	2.380*	1.197							-0.187	1.204	-0.950	1.373		
12:00–16:59			0.525	0.573										
17:00–23:59			0.206	0.686										
$D_{it}(1-(D_{it}/24)^2)$							0.283	0.451					0.578	0.411
$D_{it}^2(1-D_{it}/24)$							0.022	0.047					-0.022	0.041
$\sin(D_{it}\times 2\pi/24)$					0.111	0.440								
$\cos(D_{it}\times 2\pi/24)$					0.511	0.374								
Tuesday	0.208	1.142	-0.282	0.723	-0.864	0.619	-0.697	1.205	0.729	1.168	-0.468	1.312	0.398	1.194
Wednesday	-0.735	1.277	-0.951	0.791	-1.703*	0.672	-0.453	1.363	0.947	1.320	1.463	1.471	3.295*	1.339
Thursday	-0.176	1.270	0.191	0.814	-0.966	0.692	-1.183	1.351	-1.337	1.305	-0.120	1.473	1.251	1.322
Friday	0.034	1.207	-0.163	0.782	-0.636	0.623	-1.346	1.313	0.104	1.266	1.540	1.405	2.252	1.279
Saturday	0.078	1.300	0.353	0.840	-0.369	0.677	-0.448	1.392	-0.298	1.352	0.382	1.471	1.028	1.342
Sunday	-0.489	1.440	-0.529	0.876	-1.590*	0.727	-2.606	1.462	-1.390	1.451	0.611	1.607	2.820	1.441
Worked	-2.167*	1.067	0.377	0.659	-0.363	0.542	-1.344	1.200	-2.024	1.162	-2.238	1.277	-0.372	1.147
Sleep duration	0.152	0.171	0.200	0.110	0.115	0.095	-0.401*	0.179	-0.336	0.178	0.139	0.204	0.190	0.187
Significance of $\alpha(D_{it})$	[0.13]		[0.65]		[0.39]		[0.10]		[0.79]		[0.49]		[0.17]	
Strict exogeneity of $\{D_{it}\}$	[0.64]		[0.99]		[0.79]		[0.25]		[0.25]		[0.42]		[0.93]	
Observations	28,461		28,461		28,461		27,187		27,187		23,612		23,612	

Notes: See notes to Table A.2.

Table A.4
Time of day of interview effects on data quality. Attrition-corrected estimates.

	(1)		(2)		(3)		(4)		(5)		(6)	
	PINR		NumAct		HMissing		IvDur (min)		PIVDUR5L		PIVDUR5H	
Explanatory variables	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**
12:00–18:59					0.048	0.029						
19:00–05:59					0.046	0.039						
$D_{it}(1-(D_{it}/24)^2)$											-0.157	0.238
$D_{it}^2(1-D_{it}/24)$											0.038	0.024
$\sin(D_{it}\times 2\pi/24)$	-0.143*	0.070	0.256*	0.087			0.026	0.135	-0.457	0.364		
$\cos(D_{it}\times 2\pi/24)$	0.041	0.064	-0.191*	0.086			-0.293*	0.138	0.827*	0.374		
Tuesday	-0.175	0.113	1.475*	0.142	-0.096*	0.040	0.944*	0.202	-0.505	0.501	1.100	0.658
Wednesday	-0.147	0.121	1.209*	0.154	-0.074	0.043	0.645*	0.211	-1.195*	0.559	0.048	0.680
Thursday	0.033	0.136	1.231*	0.163	-0.040	0.044	0.858*	0.231	-0.681	0.576	1.243	0.774
Friday	0.003	0.123	1.130*	0.155	-0.070	0.044	0.470*	0.212	-0.052	0.512	-0.055	0.673
Saturday	0.023	0.127	0.755*	0.155	-0.022	0.042	0.542*	0.249	0.319	0.610	0.187	0.795
Sunday	0.289*	0.140	0.058	0.172	0.003	0.045	-0.336	0.242	0.121	0.661	-0.896	0.762
Worked	0.414*	0.137	-1.720*	0.158	-0.118*	0.037	-0.117	0.174	1.282*	0.608	0.717	0.556
Sleep duration	-0.025	0.018	-0.079*	0.025	-0.016*	0.007	-0.142*	0.028	0.416*	0.088	-0.188*	0.095
Attrition bias	[0.91]		[0.20]		[0.29]		[0.00]		[0.01]		[0.38]	
Significance of $\alpha(D_{it})$	[0.04]		[0.00]		[0.24]		[0.07]		[0.01]		[0.16]	
Strict exogeneity of $\{D_{it}\}$	[0.00]		[0.54]		[0.08]		[0.12]		[0.12]		[0.95]	
Observations	25,184		25,184		25,184		25,184		25,184		25,184	

Notes: Estimations are conducted using first differencing, and include a complete set of first-differenced dummies for number of previous interviews, first-differenced single-hour dummies for D_{it} interacted with $\mathrm{MSF}_{\mathrm{sc}}^{\mathrm{e}}$, and the inverse Mills ratio interacted with dummies for interview number. The dependent variables whose name start with P are binary indicators for the outcome given in the name's subscript scaled as a percentage. Standard errors take account of heteroskedasticity and clustering at individual level and correct for generated regressors. Probability values are in brackets. *: Significant at 5%.

Table A.5
Time of day of interview effects on data quality. Attrition-corrected estimates.

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
	PMOOD10		PMOOD25		PMOOD50		PFOODAH50		PFOODAH100		PEATING-OUT50		PEATING-OUT100	
Explanatory variables	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**	**Coef**	**S.E.**
12:00–18:59									-1.027	0.994	-1.642	1.113		
19:00–05:59									-1.107	1.281	-1.824	1.446		
12:00–16:59			0.621	0.589										
17:00–23:59			-0.467	0.734										
$D_{it}(1-(D_{it}/24)^2)$													0.836	0.448
$D_{it}^2(1-D_{it}/24)$													-0.031	0.045
$\sin(D_{it}\times 2\pi/24)$	-0.741	0.831			-0.039	0.458	-0.316	0.899						
$\cos(D_{it}\times 2\pi/24)$	0.827	0.820			0.084	0.400	-1.588	0.836						
Tuesday	-0.758	1.250	0.140	0.803	-0.844	0.685	-1.400	1.328	0.792	1.290	-0.348	1.429	0.201	1.304
Wednesday	-1.539	1.394	-0.823	0.874	-1.789*	0.744	-1.175	1.486	0.745	1.443	1.505	1.609	3.053*	1.459
Thursday	-0.437	1.387	0.324	0.888	-1.011	0.749	-1.447	1.445	-0.919	1.429	0.822	1.620	1.867	1.451
Friday	-0.621	1.305	0.183	0.850	-0.570	0.679	-1.740	1.416	0.180	1.379	1.759	1.534	2.022	1.389
Saturday	-0.342	1.409	1.065	0.903	-0.162	0.733	-0.280	1.501	0.240	1.475	0.494	1.589	0.907	1.446
Sunday	-0.976	1.544	-0.396	0.926	-1.741*	0.779	-3.374*	1.575	-2.103	1.543	0.286	1.702	2.611	1.533
Worked	-2.281*	1.106	0.653	0.684	-0.214	0.576	-1.511	1.277	-1.992	1.241	-2.132	1.369	-0.491	1.203
Sleep duration	0.221	0.184	0.232*	0.117	0.176	0.097	-0.366	0.195	-0.333	0.193	0.176	0.216	0.247	0.195
Attrition bias	[0.79]		[0.30]		[0.26]		[0.67]		[0.04]		[0.96]		[0.08]	
Significance of $\alpha(D_{it})$	[0.27]		[0.28]		[0.97]		[0.16]		[0.54]		[0.28]		[0.04]	
Strict exogeneity of $\{D_{it}\}$	[0.85]		[0.98]		[0.68]		[0.25]		[0.61]		[0.33]		[0.50]	
Observations	25,083		25,083		25,083		23,957		23,957		20,874		20,874	

Notes: See notes to Table A.4.

References

Ahn, J., Peng, M., Park, C. and Jeon, Y. (2012). A resampling approach for interval-valued data regression. *Statistical Analysis and Data Mining*, 5, 336-348.

American Association for Public Opinion Research (AAPOR) (2023). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys, 10th Edition.* AAPOR.

Angrisani, M., and Couper, M. (2022). A simple question goes a long way: A wording experiment on bank account ownership. *Journal of Survey Statistics and Methodology*, 10, 1172-1182.

Arechar, A., Kraft-Todd, G. and Rand, D. (2017). Turking overtime: how participant characteristics and behavior vary over time and day on Amazon Mechanical Turk. *Journal of the Economic Science Association*, 3, 1-11.

Arellano, M., and Meghir, C. (1992). Female labour supply and on-the-job search: An empirical model estimated using complementary data sets. *Review of Economic Studies*, 59, 537-557.

Bach, R. (2021). A methodological framework for the analysis of panel conditioning effects. In *Measurement Error in Longitudinal Data*, (Eds., Alexandru Cernat and Joseph Sakshaug), 19-42. Oxford: OUP.

Bais, F., Schouten, B. and Toepoel, V. (2022). Is undesirable answer behaviour consistent across surveys? An investigation into respondent characteristics. *Survey Methodology*, 48, 1, 191-224. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2022001/article/00001-eng.pdf.

Baumgartner, H., and Steenkamp, J.-B. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, 38(2), 143-156.

Bes, F., Jobert, M. and Schulz, H. (2009). Modeling napping, post-lunch dip, and other variations in human sleep propensity. *Sleep*, 32(3), 392-398.

Biemer, P., Groves, R., Lyberg, L. Mathiowetz, N. and Sudman, S. (eds.) (2004). *Measurement Errors in Surveys*. Hoboken, NJ: John Wiley & Sons, Inc.

Binder, C. (2022). Time-of-day and day-of-week variations in Amazon Mechanical Turk survey responses. *Journal of Macroeconomics*, 71, Article 103378.

Bison, I., and Zhao, H. (2023). Factors impacting the quality of user answers on smartphones. *CEUR Workshop Proceedings*, 3456, 208-213.

Blatter, K., and Cajochen, C. (2007). Circadian rhythms in cognitive performance: Methodological constraints, protocols, theoretical underpinnings. *Physiology and Behavior*, 90, 196-208.

Brown, E., and Czeisler, C. (1992). The statistical analysis of circadian phase and amplitude in constant-routine core-temperature data. *Journal of Biological Rhythms*, 7(3), 177-202.

Callegaro, M. (2010). Do you know which device your respondent has used to take your online survey? *Survey Practice*, 3(6). https://doi.org/10.29115/SP-2010-0028.

Carrell, S., Maghakian, T. and West, J. (2011). A's from zzzz's? The causal effect of school start time on the academic achievement of adolescents. *American Economic Journal: Economic Policy*, 3, 62-81.

Casey, L., Chandler, J., Levine, A., Proctor, A. and Strolovitch, D. (2017). Intertemporal differences among MTurk workers: Time-based sample variations and implications for online data collection. *SAGE Open*, 7(2), 1-15.

Chang, L., and Krosnick, J. (2009). National surveys via RDD telephone interviewing versus the Internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly*, 73(4), 641-678.

Collinson, J., Mathmann, F. and Chylinski, M. (2020). Time is money: Field evidence for the effect of time of day and product name on product purchase. *Journal of Retailing and Consumer Services*, 54, Article 102064.

Cornelissen, G. (2014). Cosinor-based rhythmometry. *Theoretical Biology and Medical Modelling*, 11, Article 16.

Davis, S. (2011). Comment on: Krueger, A., and A. Mueller. Job search, emotional well-being and job finding in a period of mass unemployment: Evidence from high-frequency longitudinal data. *Brookings Papers on Economic Activity*, 42(1), 58-70.

Dickinson, D., and McElroy, T. (2010). Rationality around the clock: Sleep and time-of-day effects on guessing game responses. *Economics Letters*, 108, 245-248.

Dickinson, D., and McElroy, T. (2017). Sleep restriction and circadian effects on social decisions. *European Economic Review*, 97, 57-71.

Dickinson, D., Chaudhuri, A.·and Greenaway-McGrevy, R. (2020). Trading while sleepy? Circadian mismatch and mispricing in a global experimental asset market. *Experimental Economics*, 23, 526-553.

Durrant, G., D'Arrigo, J. and Steele, F. (2011). Using paradata to predict best times of contact, conditioning on household and interviewer influences. *Journal of the Royal Statistical Society. Series A*, 174(4), 1029-1049.

Flynn, A. (2018). e-Surveying and respondent behaviour: Insights from the public procurement field. *Electronic Journal of Business Research Methods*, 16(1), 38-53.

Fordsham, N., Moss, A., Krumholtz, S., Roggina, T., Robinson, J. and Litman, L. (2019). Variation among Mechanical Turk workers across time of day presents an opportunity and a challenge for research. PsyArXiv. doi:10.31234/osf.io/p8bns.

Fricker, S., and Tourangeau, R. (2010). Examining the relationship between nonresponse propensity and data quality in two national household surveys. *Public Opinion Quarterly*, 74(5), 934-955.

Gideon, M., Helppie-McFall, B. and Hsu, J. (2017). Heaping at round numbers on financial questions: The role of satisficing. *Survey Research Methods*, 11(2), 189-214.

González Chapela, J. (2024). Supplement to "Daily rhythm of data quality: Evidence from the survey of unemployed workers in New Jersey". Available at https://drive.google.com/file/d/14YPt9BmXlxFfuURatCQY0ak-OW9z1c7B/view?usp=sharing.

Guarana, C., Stevenson, R. Gish, J. Ryu, J.W. and Crawley, R. (2022). Owls, larks, or investment sharks? The role of circadian process in early-stage investment decisions. *Journal of Business Venturing*, 37, Article 106165.

Hasher, L., Goldstein, D. and May, C. (2005). It's about time: Circadian rhythms, memory, and aging. In *Human Learning and Memory: Advances in Theory and Applications*, (Eds., Chizuko Izawa and Nobuo Ohta), 199-217. New York: Lawrence Erlbaum Associates Publishers.

Hornik, J., and Tal, A. (2010). The effect of synchronizing consumers' diurnal preferences with time of response on data reliability. *Marketing Letters*, 21, 1-15.

Juster, T. (1986). Response errors in the measurement of time use. *Journal of the American Statistical Association*, 81(394), 390-402.

Kaminska, O., McCutcheon, A. and Billiet, J. (2010). Satisficing among reluctant respondents in a cross-national context. *Public Opinion Quarterly*, 74(5), 956-984.

Kroh, M., Winter, F. and Schupp, J. (2016). Using person-fit measures to assess the impact of panel conditioning on reliability. *Public Opinion Quarterly*, 80(4), 914-942.

Kroh, M., Lüdtke, D., Düzel, S. and Winter, F. (2016). Response error in a web survey and a mailed questionnaire: The role of cognitive functioning. SOEPpapers on Multidisciplinary Panel Data Research, No. 888. DIW, Berlin.

Krosnick, J. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213-236.

Krosnick, J. (1999). Survey research. *Annual Review of Psychology*, 50, 537-567.

Krueger, A., and Mueller, A. (2010). *Survey of Unemployed Workers in New Jersey* (version Nov. 12, 2013) [Data set]. Data Archive at the Office of Population Research, Princeton University. https://oprdata.princeton.edu/archive/njui/.

Krueger, A., and Mueller, A. (2011). Job search, emotional well-being and job finding in a period of mass unemployment: Evidence from high-frequency longitudinal data. *Brookings Papers on Economic Activity*, 42(1), 1-57.

Lowe, C., Safati, A., and Hall, P. (2017). The neurocognitive consequences of sleep restriction: A meta-analytic review. *Neuroscience and Biobehavioral Reviews*, 80, 586-604.

Lyberg, L., and Stukel, D. (2017). The roots and evolution of the total survey error concept. In *Total Survey Error in Practice*, (Eds., Paul Biemer, Edith de Leeuw, Stephanie Eckman, Brad Edwards, Frauke Kreuter, Lars Lyberg, Clyde Tucker and Brady West), 1-22. Hoboken, NJ: John Wiley & Sons, Inc.

Malhotra, N. (2008). Completion time and response order effects in web surveys. *Public Opinion Quarterly*, 72(5), 914-934.

Nishii, R. (1988). Maximum likelihood principle and model selection when the true model is unspecified. *Journal of Multivariate Analysis*, 27, 392-403.

Olson, K., Smyth, J. and Ganshert, A. (2019). The effects of respondent and question characteristics on respondent answering behaviors in telephone interviews. *Journal of Survey Statistics and Methodology*, 7, 275-308.

Peytchev, A. (2009). Survey breakoff. *Public Opinion Quarterly*, 73(1), 74-97.

Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. *International Statistical Review*, 61(2), 317-337.

Phillips, A., and Stenger, R. (2022). The effect of burdensome survey questions on data quality in an omnibus survey. *Journal of Official Statistics*, 38(4), 1019-1050.

Read, B., Wolters, L. and Berinsky, A. (2021). Racing the clock: Using response time as a proxy for attentiveness on self-administered surveys. *Political Analysis*. Available at https://doi.org/10.1017/pan.2021.32.

Roenneberg, T., Kuehnle, T., Juda, M., Kantermann, T., Allebrandt, K., Gordijn, M. and Merrow, M. (2007). Epidemiology of the human circadian clock. *Sleep Medicine Reviews*, 11, 429-438.

Salehinejad, M., Wischnewski, M., Ghanavati, E., Mosayebi-Samani, M., Kuo, M.-F. and Nitsche, M. (2021). Cognitive functions and underlying parameters of human brain physiology are associated with chronotype. *Nature Communications*, 12, Article 4672.

Schmidt, C., Collette, F., Cajochen, C. and Peigneux, P. (2007). A time to think: Circadian rhythms in human cognition. *Cognitive Neuropsychology*, 24(7), 755-789.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461-464.

Solon, G., Haider, S. and Wooldridge, J. (2015). What are we weighting for? *Journal of Human Resources*, 50(2), 301-316.

StataCorp (2019). *Stata Base Reference Manual. Release 16*. College Station, TX: Stata Press.

Toninelli, D., and Revilla, M. (2020). How mobile device screen size affects data collected in web surveys. In *Advances in Questionnaire Design, Development, Evaluation and Testing*, (Eds., Paul Beatty, Debbie Collins, Lyn Kaye, Jose Luis Padilla, Gordon Willis and Amanda Wilmot), 349-373. Hoboken, NJ: John Wiley & Sons, Inc.

Tourangeau, R., Rips, L. and Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge, UK: Cambridge University Press.

Truebner, M. (2021). The dynamics of “neither agree nor disagree” answers in attitudinal questions. *Journal of Survey Statistics and Methodology*, 9, 51-72.

Valdez, P. (2019). Homeostatic and circadian regulation of cognitive performance. *Biological Rhythm Research*, 50, 85-93.

Weeks, M., Kulka, R. and Pierson, S. (1987). Optimal call scheduling for a telephone survey. *Public Opinion Quarterly*, 51, 540-549.

Williams, K., and Shapiro, T.M. (2018). Academic achievement across the day: Evidence from randomized class schedules. *Economics of Education Review*, 67, 158-170.

Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data*. Second edition. Cambridge, MA: MIT Press.

Yan, T., and Olson, K. (2013). Analyzing paradata to investigate measurement error. In *Improving Surveys with Paradata: Analytic Uses of Process Information*, (Ed., Frauke Kreuter), 73-95. Hoboken, NJ: John Wiley & Sons, Inc.

Ziniel, S. (2008). *Cognitive Aging and Survey Measurement*. PhD dissertation, University of Michigan.