Journal of Time Use Research
International Association for Time Use Research (IATUR)
jtur.iatur.org

Does Diary Mode Matter in Time-Use Research?

Stella Chatzitheochari[1] and Elena Mylona[2]

1. Corresponding author: Department of Sociology, University of Warwick, Coventry CV4 7AL, UK s.chatzitheochari@warwick.ac.uk

2.Warwick Medical School, University of Warwick

Abstract

Recent years have witnessed an increasing interest in the use of new technologies for time-use data collection, driven by their potential to reduce survey administration costs and improve data quality. However, despite the steady growth of studies that employ web and app time diaries, there is little research on their comparability with traditional paper-administered diaries that have long been regarded as the "gold standard" for measurement in time-use research. This paper investigates diary mode effects on data quality and measurement, drawing on data from a mixed-mode large-scale time diary study of adolescents in the United Kingdom. After controlling for observable characteristics associated with diary mode selection and adolescent time-use, we find that web and app diaries yield higher quality data than paper diaries, which attests to the potential of new technologies in facilitating diary completion. At the same time, our analysis of broad time-use domains does not find substantial mode effects on measurement for most daily activity categories. We conclude by discussing avenues for future methodological research and implications for time-use data collection.

Keywords: Diary Quality; Mode Effects; Mixed-Mode Surveys; New Technologies; Young People

https://doi.org/10.32797/jtur-2022-1

1. Introduction

Widely known as time-use research, the study of how people spend their time is of central importance for several social science disciplines including sociology, anthropology, psychology, economics, and public health research. Most time-use research depends on self-reported data from time diaries (Cornwell et al., 2019), recognised as the "gold standard" for the collection of information surrounding behavioural patterns of large populations (Bauman et al., 2019; Gershuny, 2003). Time diaries capture the full 24 hours of a day, typically gathering data on respondents' main and simultaneous activities, as well as contextual information such as location and co-presence of others. The diary method thus produces a sequential and comprehensive record of all the activities a respondent engages in, giving access to information on duration, participation, timing, and context. This is not the case for "stylized" time-use estimates produced by conventional social survey questionnaires, which collect data on participation and/or frequency/duration for a limited set of pre-specified activities.

Methodological research has shown that time diaries produce more reliable and valid time-use estimates than social survey questions, keeping social desirability and approximate answers at a minimum (Michelson, 2015; Robinson and Godbey, 2010). Most of this research focuses on "leave behind" paper-administered time diaries. This mode usually includes cross-checking of diary entries during a subsequent interviewer visit. Aside from the high cost of survey administration, post-fieldwork coding of paper diaries is also costly and time-consuming (Chatzitheochari et al., 2018). Additionally, the task of diary completion is burdensome, typically leading to a relatively high proportion of incomplete diary records and low response rates in stand-alone time-use surveys (Abraham et al., 2006; Gershuny, 2003). To achieve an optimal balance between sufficient time coverage and reasonable respondent burden, time diaries are usually collected for two randomly selected days, a weekday, and a weekend day (European Commission, 2004). However, time-use researchers have emphasized the need to collect diary data for longer periods to minimize problems of intra-individual variation of daily behaviour (Frazis and Stewart, 2012; Gershuny, 2003; Glorieux and Minnen, 2009).

It has been argued that the use of new technologies for time-use data collection could help overcome the abovementioned limitations of the paper diary, which remains the most frequently employed mode used by national statistical agencies (Cornwell et al., 2019). For example, web and app diaries could potentially reduce task complexity by using prompts and checks, improving response rates and data quality (Chatzitheochari et al., 2018). Less burdensome diary instruments would subsequently allow longer time periods to be surveyed. Additionally, the use of new technologies could automate diary coding procedures, eliminating coder bias, and enhancing research transparency (Minnen et al., 2014). The reduction of diary placement and interviewer costs would also contribute to an overall cost decrease, enabling more frequent waves of data collection. However, although there is a steady growth of self-administered web and app diary instruments (see, e.g., Bonke and Christensen, 2019; Chatzitheochari et al., 2018; Elevelt et al., 2021; Minnen et al., 2014), we know little about comparability across different modes of time diary administration.

This paper rectifies this omission by investigating time diary mode effects on data quality and measurement. We analyse diary data from the sixth wave of the Millennium Cohort Study (hereafter MCS), a multidisciplinary longitudinal study following individuals born between 2000 and 2002 in the United Kingdom (University of London, Institute of Education, Centre for Longitudinal Studies, 2020). MCS employed an innovative mixed-mode time diary survey design (web, app, paper) to measure time allocation of cohort members at age 14. We analyse diaries from 3,982 cohort members and draw on a range of variables to adjust for observable characteristics associated with selection of diary mode and time-use allocation. Our analyses examine mode differences in key diary quality dimensions and time-use measures frequently used in social research. We contribute to the literature on mode effects in social survey research (see, e.g., Allum et al., 2018; Hox et al., 2017; Jäckle et al., 2010; Lugtig et al., 2011; Nandi and Platt, 2017;

Vannieuwenhuyze and Loosveldt, 2013), providing methodological insights for researchers and agencies interested in time-use data collection.

2. Background

The strength of the time diary technique lies in the fact that it "capitalizes on the most attractive measurement properties of the time variable" (Robinson and Godbey, 2010: 97). Capturing the full 24 hours of a day, time diaries respect the zero-sum property of time-use, namely that, if time in one (main) activity increases, it must be matched by decreases in other (main) activities. The time grid that respondents use to record their activities facilitates the provision of an accurate time allocation account by minimising recall difficulties, as it corresponds to the way daily events are stored in memory (Kelly et al., 2015; Robinson and Godbey, 2010). Time diaries usually divide the 24-hour period into ten or fifteen-minute increments. This feature additionally contributes to the high accuracy of obtained duration estimates, as respondents cannot manipulate their activity durations by using middle-range responses to counter their approximations (Krosnick, 1999).

Previous research using spousal activity reports, video records, shadow observations, and random-hour accounts has evidenced the validity and reliability of time-use estimates produced by paper diaries (Robinson and Godbey, 2010). Likewise, several studies comparing "stylized" questionnaire and diary estimates for different activities have established the higher validity and reliability of the latter, demonstrating social desirability and recall bias affecting the former (Juster et al., 2003; Kan and Pudney, 2008; Walthery and Gershuny, 2019). More recently, diary estimates have been successfully validated against objective measures from accelerometers and wearable cameras (Connor et al., 2016; Gershuny et al., 2020; Kelly et al., 2015).

There have been few attempts to collect diary data using web-based instruments. To our knowledge, all existing web diaries have followed a "light" format, providing respondents with an exhaustive list of activity codes to describe their daily time-use. One of the first web-administered time diaries was designed in the context of the nationally representative 2008/9 Danish time-use and consumption survey (Bonke and Fallesen, 2010). An analysis of two indicators of diary data quality, that is, mean number of different main activities and activity sequences reported during the surveyed day, suggested that web diaries produced higher quality data than paper diaries after controlling for respondent characteristics. The Modular Online Time Use Survey (MOTUS) from Belgium is a self-administered web instrument that employs soft and hard warnings during diary completion to ensure data quality (Minnen et al., 2014). Analyses of MOTUS data have shown that respondents reported a comparable number of main activities to those completing pen-and-paper "light" diary surveys. Additionally, the use of checks minimized missing diary time: a mean of 19 minutes in the first surveyed day dropped to solely 2 minutes in the second day, remaining similarly low for the entire week of data collection (Minnen et al., 2014).

However, it should be noted that these studies draw on a set of indicators that is not sufficient to evidence the quality of time-use accounts produced by web-based instruments. Aside from number of reported activities and levels of unspecified diary time, diary quality also depends on the report of key daily activity domains such as personal care and eating/drinking (Fisher and Gershuny, 2013). It could be argued that soft and hard warnings employed in web diaries can ensure appropriate recording of these activities. However, there are also challenges linked to screen size limitations: Although web diaries can support the use of a time grid, they cannot visually display all activity categories like paper diaries do. Instead, web instruments often display broad time-use categories that "unfold" to reveal detailed activity codes (Chatzitheochari et al., 2015). This may be associated with a lower degree of familiarity with the range of available activity categories during diary completion, and a subsequent narrower and less detailed description of the surveyed day, alongside a higher use of non-substantive "other" activity codes. However, this hypothesis can only be tested with a systematic comparison of time-use measures produced by web and paper diaries.

While web-based time diaries remain relatively rare, a growing body of studies use smartphone apps to collect diary data (see, e.g., Bonke and Christensen, 2019; Chatzitheochari et al., 2018; Daum et al., 2019; Elevelt et al., 2021; Fernee and Sonck, 2013). This mode of data collection seems particularly promising, given the unprecedented diffusion of internet-enabled mobile devices in advanced economies. The widespread use of commercial apps to record daily habits (e.g., diet, fitness) may render respondents more familiar with smartphone-based diary completion. At the same time, the proximity to one's body and the sheer frequency of daily smartphone usage may facilitate more frequent and accurate recording of time-use patterns (Klasnja and Pratt, 2012). Finally, the combination of self-reported diary data with passive sensor data (e.g., from GPS) can increase the validity of diary records whilst providing novel insights into daily behaviour (Cornwell et al., 2019; Elevelt et al., 2021; Zeni et al., 2020).

The main challenge of app diaries relates to the small size of smartphone screens (Couper and Peterson, 2017; Link et al., 2014), which necessitates the use of a substantially different format compared to that used in paper and web diaries. The use of a time grid is particularly difficult to employ in this mode, which means that app respondents do not have access to the same visual representation of their surveyed day like paper diarists do. Indeed, most existing diary apps use sequential question-based approaches, inviting respondents to specify start and end times of activities (see, e.g., see Chatzitheochari et al., 2015, Fernee and Sonck, 2013). At the same time, regardless of variations in existing formats, "light" app diaries always entail the use of scrolling down and/or swiping to locate the activity of choice. It could thus be argued that the diary task requires more cognitive effort in app modes compared to paper and web grid-based modes. This may lead to a slower learning curve at the beginning of diary completion, which may involve more frequent use of "other" activity codes. Similarly, it could be associated with higher survey fatigue and satisficing effects (Chatzitheochari and Mylona, 2021; Krosnick, 1991). The impact of app diary formats on the level of detail and quality of obtained time-use accounts has not yet been investigated, as most studies have relied on small and non-representative samples, while there have only been a few attempts to combine app diaries with paper/web diaries within a single study. A notable exception is the 2017/8 Danish time-use survey that offered respondents the choice to fill in a diary app in a smartphone or a PC (web version) or to complete a time diary interview. Mode differences in data quality indicators were attenuated after controlling for selection effects (Bonke and Christensen, 2019). However, this study did not scrutinize time-use measures produced by different diary modes.

Our paper responds to the lack of systematic research on time diary mode effects by exploring the comparability of paper, web, and app modes of diary administration. Using MCS data, we address the following questions:

1) Are there differences in data quality of time diaries collected by paper, web, and app modes?

2) Are there differences in time-use accounts produced by paper, web, and app diaries?

Our analysis provides vital evidence for future diary data collection, ascertaining the feasibility of mixed-mode diary designs to reduce costs and improve coverage in large-scale time-use surveys (Couper, 2017).

3. Method

3.1 The Millennium Cohort Study Time-Use Record

The *Millennium Cohort Study* is a multidisciplinary longitudinal study that follows the lives of approximately 19,000 children born between 2000 and 2002 in the United Kingdom (Connelly and Platt, 2014). Millennium Cohort Study data can be accessed from the UK Data Service (www.ukdataservice.ac.uk). Seven survey waves have been carried out so far – at age 9 months, 3, 5, 7, 11, 14, and 17 years. At age 14 (Wave 6), cohort members were invited to complete 24-hour time diaries for two randomly allocated days within the space of 10 days, a weekday, and a weekend day. In line with the Harmonised European Time Use Study guidelines (European Commission, 2004), the retrospective MCS diary started at 4.00am and finished at 3.59am the next day. A "light" diary was used, providing young people with 44 age-appropriate activity categories to describe their daily time-use. This activity code scheme corresponds to 12 broad time-use categories (see Table 1). Contextual diary columns collected information on location, co-presence, and enjoyment throughout the 24-hour period.

Cohort members were asked to choose between an app and a web diary. The web diary could only be completed on desktops, laptops, or netbooks, and the app diary on tablets or Apple/Android smartphones. Those who did not have internet/device access or refused to complete either of these two instruments were offered a paper diary.

The paper diary divided the 24-hour period in 144 ten-minute slots. Activity and contextual codes were found across both sides of the 24-hour grid. Diarists recorded activities by drawing a line across the grid. The web diary followed the same format and is largely comparable to the paper instrument, with the exception that activity codes were nested within broader time-use domains and could only be displayed by "unfolding". The smartphone app differed in three dimensions: First, it employed a question-based rather than a grid measurement approach. Second, instead of following a ten-minute slot approach, it required diarists to assign ending times to each recorded activity. Third, app diary dimensions were coterminous, which means that it was not possible for cohort members to record changes in contextual dimensions (e.g., enjoyment, location) during a single activity.

Table 1: Millennium Cohort Study Age 14 Time-Use Record Activity Coding Scheme

Time-Use Domain	Activity Codes
Sleep and personal care	Sleeping and resting (including sick in bed); Personal care (including taking a shower/bath, grooming, getting dressed etc.)
School, homework, and education	Homework; In class; School breaks; Schools clubs; Detention
Paid work	Paid work (including paid babysitting and paid work for the family)
Unpaid work	Unpaid work for family or other non-household members (e.g., help in family business)
Chores, housework, and looking after people or animals	Cooking, cleaning, and shopping for the household; Fixing things around the house, fixing bike, gardening; Looking after brothers, sisters, other children in the household; Looking after parent or other adult in the household (medical or personal care); Looking after animals
Eating and Drinking	Eating or drinking in a restaurant or café; Eating a meal; Eating a snack or having a drink
Physical exercise and Sports	Cycling; Individual ball games and training (e.g., tennis, badminton); Jogging, running, walking, hiking; Team ball games and training (e.g., football, hockey); Swimming and other water sports; Other physical exercise (e.g., dancing, keeping fit) and other sports (e.g., skateboarding, gymnastics)
Travelling (including walking to school)	Travel by bus, taxi, tube, plane; Travel by car, van (including vehicles owned by friends and family); Travel by physically active means (walk, bike etc.)
Social time and family time	Attending live sporting events; Cinema, theatre, performance, gig etc.; Exhibition, museum, library, other cultural events; Shopping (including window shopping, hanging out at shopping centre); Speaking on the phone (including Skype, video calls); Speaking, socializing face-to-face
Internet, TV, and Digital Media	Answering emails, instant messaging, texting; Browsing and updating social networking sites (e.g., Twitter, Facebook, BBM, Snapchat); General Internet browsing, programming (not time on social networking sites); Listening to music, radio, iPod, other audio content; Playing electronic games and Apps Watching TV, DVDs, downloaded videos
Volunteering and religious activities	Volunteering; religious activities (including going to places of worship, praying etc.)
Hobbies and other free time	Hobbies and other free time activities; Did nothing, just relaxing, bored, waiting; Hobbies, arts and crafts, musical activities, writing stories, poetry etc.; Reading (not for school)
Any other activity	Other activities not listed

Note: Table presents activity codes as presented in the MCS time-use record.

Web and app diarists were also provided with an “aide-memoire” that resembled a school notebook. This allowed them to take notes of their activities during times when they did not have access to their PC/smartphone (e.g., school time)[1]. Cohort members who agreed to the task received mode-specific instructions. Further help was available via phone, SMS, or text throughout data collection.

Reminder texts were sent to those who had given consent to be contacted for such purposes. It is worth noting that paper diarists filled in their records independently without the help/visit of an interviewer. Proxy diaries were not acceptable. Detailed information on the format and administration of all three diary instruments and other accompanying materials is provided in Chatzitheochari et al. (2015).

Both web and app diary instruments made balanced use of mode-specific soft and hard checks to reduce unspecified time and improve diary quality. A soft check was triggered when respondents recorded an activity other than school and sleep that lasted more than three hours, asking them whether they are sure that the activity lasted this long. It was not possible to submit a blank diary as a hard check was triggered. Additionally, soft checks were triggered when respondents attempted to submit incomplete diaries. Web diarists could ignore sort checks and leave blank time slots in the main activity diary column. In contrast, app diarists had to assign an activity code to any given time period in order to continue recording subsequent activity episodes. This means that app diaries may have lower reports of unspecified time and higher reports of “any other activity” than web and paper diaries. App diarists could still click “submit” before completing a full 24-hour record.

All cohort members in Wales, Scotland, and Northern Ireland were invited to complete the diary. A random subsample of households (81%) was selected in England. Among eligible cohort members, 48% completed the time diary for designated Day 1 and 41% for designated Day 2 (Ipsos MORI, 2016). These response rates are considerably higher than the 33% achieved by the latest UK Time Survey that only employed paper diaries (Morris et al., 2016).

3.2 Outcome Variables

We examine diary data quality by focusing on two indicators: *average number of episode changes* and *proportion of “good quality” diaries*.

Corresponding to number of recorded main activities, diary episodes constitute an accepted indicator of overall diary data quality (Glorieux and Minnen, 2009; Väisänen and Finland, 2006). Time-use researchers frequently employ a broader diary episode definition than the one we use in this study, focusing on changes across all diary columns rather than main activities only (Rydenstam and Wadeskog, 1998). However, this measure does not allow comparisons across studies that often differ in contextual diary dimensions. Considering that diary dimensions in the MCS app diary were also coterminous, we argue that solely focusing on changes in main activities is more appropriate to understand diary quality and mode equivalence in this study.

There is no accepted definition of what constitutes a “good quality” time diary, which can be understood as a productive diary in social survey terms (Chatzitheochari et al., 2018). We draw

[1] Aide-memoires were not returned to the survey office. We therefore have no information on whether/how they were used by web and app diarists.

on Fisher and Gershuny (2013), who suggest that the quality of a time diary should be judged upon three criteria: missing data on the main activity column, number of diary episodes, and report of basic activity domains (sleep, personal care, travel, and eating/drinking). We thus define good quality diaries as those diaries that report at least six diary episodes, one episode of personal care (including sleep) and one episode of eating or drinking, and less than 90 minutes of missing data in the activity diary column. We note that these criteria are less strict than those employed for adult diaries, as we do not consider travelling, which is less frequently reported in children's diaries.

We then focus on *total time spent in each broad activity category at the sample level* (i.e., all diarists, including non-participants reporting zero time in an activity category), which remains the most employed measure in time-use research (Gershuny, 2003; Michelson, 2015).

3.3 Control Variables

As cohort members were not randomly assigned into diary modes, we need to control for sample composition of the three mode groups. This is achieved through use of a model that aims to capture correlations between diary mode and outcomes variables of interest arising from common socio-demographic variables (Vannieuwenhuyze and Loosveldt, 2013). We employ a wide range of *mode-insensitive* variables associated with diary mode choice and daily time use. More specifically, we examine the influence of sex (male/female), ethnicity (White/Mixed/Indian/Pakistani or Bangladeshi/Black), and subjective health status (excellent-very good/good/fair-poor). We additionally control for seasonal influences (Autumn/Winter/Spring/Summer), using completion date information from the diary file. Our analyses also include a set of socio-demographic variables from parental interviews: We employ a variable on parental educational attainment (Degree or higher/A Levels/O Levels/Level 1-CSE/no qualifications[2]), using the highest qualification reported by either parent. We also consider whether the young person lives in a two-parent or a lone-parent household, or a workless household (i.e., a household where all adults are economically inactive or unemployed). Considering the role of cognitive ability on mode effects (Nandi and Platt, 2017) we also control for cohort members' achieved score on a naming vocabulary test administered at age 14, derived from a shortened version of the APU Vocabulary Test (Closs and Hutchings, 1976). Finally, we use a variable on personal computer ownership (yes/no) as this is likely associated with diary mode choice as well as time allocation[3].

3.4 Analytical Approach

Poisson regression models examine diary episode changes and logit models estimate the probability of returning a good quality record. We then examine total time spent in all broad diary activities with OLS regression, which is appropriate for modelling diary data (Stewart, 2013). Models have been tested for multicollinearity by examination of variance inflation factors (VIFs). In order to circumvent the "reference-category" problem, we calculate quasi-variance comparison intervals to allow comparisons between the three diary modes (Firth and De Menezes, 2004).

---

[2] These categories represent different levels of secondary and post-secondary educational attainment in the United Kingdom.

[3] MCS did not collect information on ownership of other electronic devices such as smartphones and video game consoles.

Analyses are presented separately for school day and non-school days, considering distinctive differences in time-use patterns. School day diaries were identified as those who reported one of the 4 following activity codes: In class, School break, School club, Detention. Because our dataset includes cohort members that completed two non-school or two school day diaries, we use robust cluster standard errors to adjust for multiple observations per person.

We exclude 1,189 deposited diaries from those with missing information on independent variables and those whose parents had “overseas educational qualifications” and “other ethnicity” (these categories were too small for separate analyses). Data quality analyses use all remaining diaries (n=7,416). For analyses of broad time-use categories, we only focus on good quality diaries (n=5,841), given that diaries of insufficient quality are routinely excluded from time-use analyses. Results from multivariate analyses are presented graphically. For reasons of parsimony, we solely concentrate on the relationships of interest (i.e., mode effects). Full models are available in the online supplement.

3.5 Limitations

We acknowledge that diary mode effects can only be identified with the use of experimental designs. Additionally, our approach to capturing selection effects consists of controlling for observable socio-demographic differences. Although this approach remains the most popular in mode effects literature, it can be seen as partial (Vannieuwenhuyze and Loosveldt, 2013). Results from our study should therefore not be seen as providing causal evidence. Finally, we note that our investigation is inevitably constrained by the small number of young people who completed paper diaries, which resulted in relatively large comparison intervals.

4 Results

Table 2 presents sample characteristics by diary mode. We note that our sample is positively selected in terms of parental education, which was confirmed by analyses of diary non-response patterns (not presented in this paper). Approximately 69% of cohort members in our sample filled in app diaries, 25% web diaries, and 6% paper diaries. This is in line with statistics surrounding ownership of smartphones and tablets among young people from this age group in the UK (Ofcom, 2015). Table 2 shows that girls were more likely than boys to choose and complete app diaries, while the opposite was the case for web diaries. There were also national and seasonal variations in mode choice. Black and Pakistani/Bangladeshi young people were more likely to fill in paper diaries, and White young people more likely to fill in web and app diaries. The relationship between socio-economic status and diary mode is evidenced by cross-tabulations with family structure, parental worklessness, and parental educational attainment that show that young people from an advantaged socio-economic background were generally more likely to fill in web or app instruments.

As expected, app and web diarists were more likely to own a PC compared to paper diarists. Finally, in contrast to paper and app diarists, web diarists scored higher than average in the word activity score administered in the study. Many of these differences remained significant in a multinomial logit model predicting time diary mode choice (see Table A1 in the online supplement).

Table 2: Sample Characteristics by Time Diary Mode, column %; person-level analysis

Measures	%	Mode			
	All	Web (24.9%)	App (69.4%)	Paper (5.8%)	χ^2 p-value
Sex					p < 0.001
Male	45.4	54.8	41.8	48.5	
Female	54.6	45.2	58.3	51.5	
Ethnicity					p < 0.001
White	86.8	83.6	88.5	80.8	
Mixed	4.1	4.8	3.8	4.4	
Indian	2.7	3.7	2.4	2.2	
Pakistani/Bangladeshi	4.1	4.9	3.6	7.9	
Black	2.2	2.9	1.7	4.8	
Country					p < 0.001
England	60.0	63.7	58.4	62.9	
Wales	15.7	13.6	16.1	20.1	
Scotland	13.6	13.4	14.2	6.1	
Northern Ireland	10.8	9.3	11.3	10.9	
Season					p < 0.05
Autumn	18.9	18.6	19.6	12.2	
Winter	14.9	15.3	15.0	13.1	
Spring	38.6	41.5	37.1	44.5	
Summer	27.5	24.6	28.3	30.1	
Parental education					p < 0.001
No qualifications	2.7	1.9	2.7	6.6	
NVQ 1 (Level 1/CSE)	2.9	2.9	2.8	3.9	
NVQ2 (O levels)	15.3	13.3	15.7	19.7	
NVQ3 (A Levels)	13.7	11.4	14.3	16.6	
Degree or higher	65.3	70.4	64.5	53.3	
Family structure					p < 0.01
Two-parent household	80.1	82.6	79.6	74.7	
Lone-parent household	19.9	17.4	20.4	25.3	
Subjective health					p > 0.05
Excellent/very good	53.1	55.9	52.2	52.4	
Good	36.4	34.8	37.0	36.2	
Fair/Poor	10.5	9.3	10.9	11.4	
Workless Household					p < 0.001
No	92.4	92.7	92.9	84.7	
Yes	7.6	7.3	7.1	15.3	
Owns their own PC					p < 0.001
Yes	86.1	85.7	87.1	75.6	
No	13.9	14.3	12.9	24.5	
Word activity score - standardised (SE)		0.16 (1.06)	-0.06 (0.96)	-0.05 (1.10)	

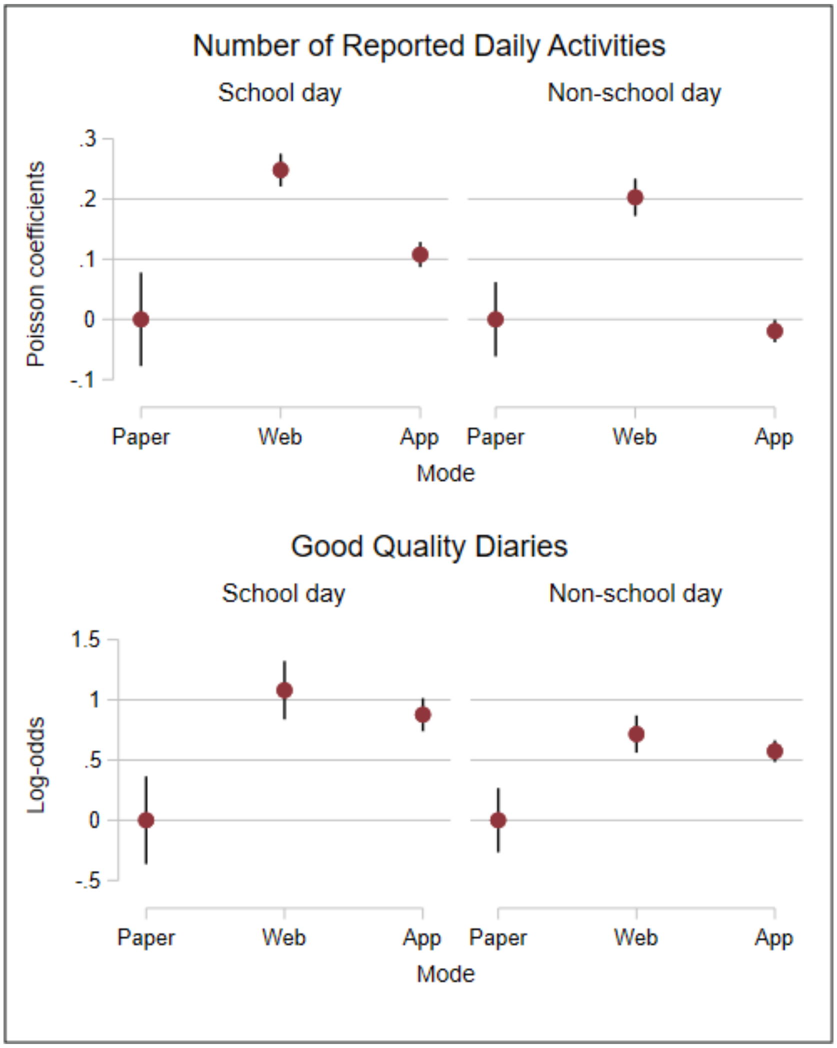
Source: *Millennium Cohort Study*
Note: N=3,982.

Figure 1 presents diary quality results, focusing on number of episode changes and good quality diaries by mode (full models shown in Tables A2 and A3 of the online supplement). Results for episode changes clearly demonstrate the strength of the web diary in eliciting detailed diary completion: Web diaries have a higher average of episode changes compared to paper and app diaries on both school days and non-school days after controlling for observable socio-economic differences. This finding refutes our hypothesis surrounding “unfolding” activity categories negatively influencing diary completion. Rather, it is indicative of the benefits of combining the grid measurement approach with the use of soft and hard checks for diary data collection. Figure 1 also shows that app diaries report a slightly higher number of episodes than paper diaries on school days. However, there are no statistically significant differences between these two modes on non-school days, as shown by the overlapping comparison intervals. App and web modes are more likely to produce good quality diaries than the paper mode. Web diaries are marginally more likely to yield good quality records than app diaries on non-school days, but differences between these two modes are not statistically significant on school days. Noting that report of non-substantive “other” activities is higher among app records (see Tables 3a and 3b), we excluded these codes from our episode measure as a robustness check. These models yielded similar results (see Tables A4 and A5 in the online supplement). Alternative definitions of good quality diaries (e.g., only focusing on a minimum number of episode changes rather than type of reported activities) did not alter the main results either (see Table A6).

Tables 3a and 3b presents descriptive statistics on participation rates and mean time spent on different broad time-use activities at the sample level, focusing on good quality diaries. We can see that some of the obtained time-use estimates are remarkably similar across time diary modes. For example, there are only negligible differences in sleeping/personal care, eating/drinking, as well as work estimates. In contrast, there are time-use domains where larger mode differences arise such as TV, Internet, and Digital Media: Web and paper diarists spend approximately 20 minutes longer on these activities than app diarists do during school days, while web diarists also report a substantially higher time expenditure on this domain during non-school days (a mean of 302 minutes, as opposed to 273 and 254 minutes for paper and app diarists respectively). Table 3b also shows that app diarists used the “other” activity category more often than web and paper diarists (e.g., a mean of 49 minutes, as opposed to 33 and 28 minutes for web and paper diarists on non-school days).

We now proceed to multivariate analyses to ascertain whether the differences shown in Tables 3a and 3b remain after accounting for differences in sample composition across the three modes. Figure 2a and 2b presents results from 22 OLS regression models predicting time spent in different time-use domains on school days and non-school days (for full models see Tables A7-A17 of the online supplement). We can see that, after controlling for observable characteristics associated with mode selection and time-use, there are no significant differences between web and paper time-use measures during school days, as shown by the overlapping comparison intervals. Similarly, there are no significant differences between paper and app instruments. In contrast, there are some significant differences between web and app diary estimates. For example, web diarists spent 10 minutes longer on hobbies and 17 minutes longer on Internet, TV, and Digital Media compared to app diarists (Figure 2b). However, the majority of these differences are negligible, amounting to less than 15 minutes, which corresponds to less than 1% in the 24-hour period.

Figure 1: Diary Quality Indicators by Time Diary Mode: a) Poisson Regression Estimates and 95% Comparison Intervals for Number of Episode Changes b) Log Odds and 95% Comparison Intervals from Logistic Regression Models Predicting Good Quality Diaries



Source: *Millennium Cohort Study*
Notes: N=2,584 (school days); N=5,163 (non-school days). Good quality diaries report six activities or more, at least one episode of personal care (including sleep) and one episode of eating or drinking, and less than 90 minutes of missing data in the activity diary column. Models control for sex, ethnicity, country of residence, season the time diary was completed, parental education, family structure, subjective health status, PC ownership, cognitive ability, and parental worklessness. Robust cluster standard errors adjust for multiple observations per person.

Table 3a: Participation and Mean Time Spent in Daily Activities (Minutes per Day) by Time Diary Mode; Good Quality Diaries: School Days

	Web (26.5%)		**App (69.0%)**		**Paper (4.6%)**	
	Mean (SE)	%	Mean (SE)	%	Mean (SE)	%
Episode Changes	20.5 (6.2)		18.2 (6.4)		17.3 (5.0)	
Activities						
Sleep and personal care	572.4 (80.5)	100.0	582.6 (93.8)	100.0	575.5 (101.5)	100.0
Eating and Drinking	67.2 (35.7)	100.0	66.6 (57.4)	100.0	69.6 (38.8)	100.0
School, etc.	397.5 (98.1)	100.0	403.3 (101.4)	100.0	386.7 (96.6)	100.0
Paid or unpaid work	15.3 (38.8)	30.3	17.1 (40.8)	30.9	12.1 (30.3)	28.7
Physical exercise	41.7 (60.8)	46.0	45.4 (69.0)	46.3	45.3 (72.7)	47.9
Travelling (incl. active transport)	73.3 (52.0)	92.5	68.4 (57.1)	90.4	61.9 (56.3)	83.0
Social and family time	38.5 (61.4)	50.2	47.7 (77.6)	53.4	40.9 (69.9)	51.1
TV, Internet, Digital Media	172.9 (121.7)	91.0	153.0 (116.5)	88.4	172.7 (139.3)	87.2
Volunteering	3.8 (20.7)	5.3	3.3 (27.9)	3.2	2.9 (10.5)	9.6
Hobbies, etc.	42.3 (72.9)	46.9	30.7 (59.7)	37.0	37.2 (62.3)	39.8
Any other activity	12.9 (38.7)	18.8	21.6 (55.1)	30.7	15.6 (44.6)	21.3

Source: *Millennium Cohort Study*

Notes: N=2,055. Good quality diaries report six activities or more, at least one episode of personal care (including sleep) and one episode of eating or drinking, and less than 90 minutes of missing data in the activity diary column.

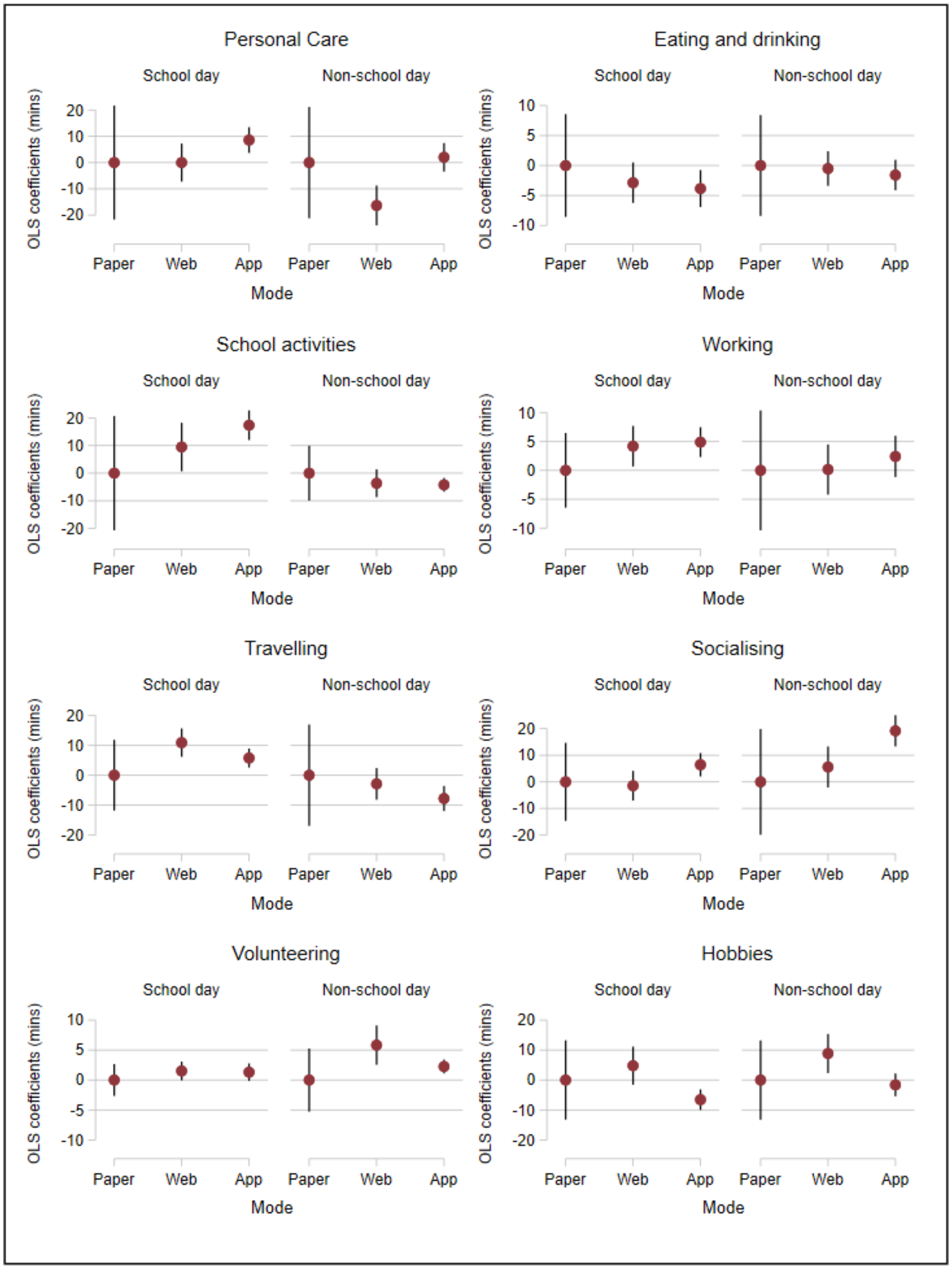
Table 3b: Participation and Mean Time Spent in Daily Activities (Minutes per Day) by Time Diary Mode; Good Quality Diaries: Non-school Days

	Web (25.3%)		**App (70.1%)**		**Paper (4.6%)**	
	Mean (SE)	%	Mean (SE)	%	Mean (SE)	%
Episode Changes	17.8 (7.0)		14.3 (5.9)		14.9 (5.7)	
Activities						
Sleep and personal care	657.5 (114.2)	100.0	680.5 (129.8)	100.0	674.3 (136.5)	100.0
Eating and Drinking	89.0 (43.2)	100.0	89.0 (61.0)	100.0	88.7 (53.1)	100.0
School, etc.	29.5 (73.6)	27.2	27.7 (62.3)	26.7	30.5 (64.9)	31.6
Paid or unpaid work	41.1 (65.9)	49.2	44.7 (81.8)	47.9	41.7 (65.1)	48.3
Physical exercise	58.6 (90.0)	48.6	66.7 (104.4)	48.4	81.8 (112.6)	48.8
Travelling (incl. active transport)	59.2 (85.3)	59.0	55.7 (88.2)	55.7	61.8 (108.2)	51.2
Social and family time	91.6 (116.7)	61.0	111.9 (138.5)	64.2	86.1 (127.6)	51.2
TV, Internet, Digital Media	301.7 (192.5)	95.2	253.8 (186.2)	91.6	272.8 (197.3)	92.5
Volunteering	11.1 (45.7)	9.3	7.0 (38.3)	5.5	5.7 (33.6)	6.3
Hobbies, etc.	64.7 (97.3)	53.1	53.9 (92.8)	43.3	57.4 (82.9)	52.9
Any other activity	33.4 (80.0)	28.2	49.1 (103.5)	35.8	28.3 (81.4)	23.0

Source: *Millennium Cohort Study*

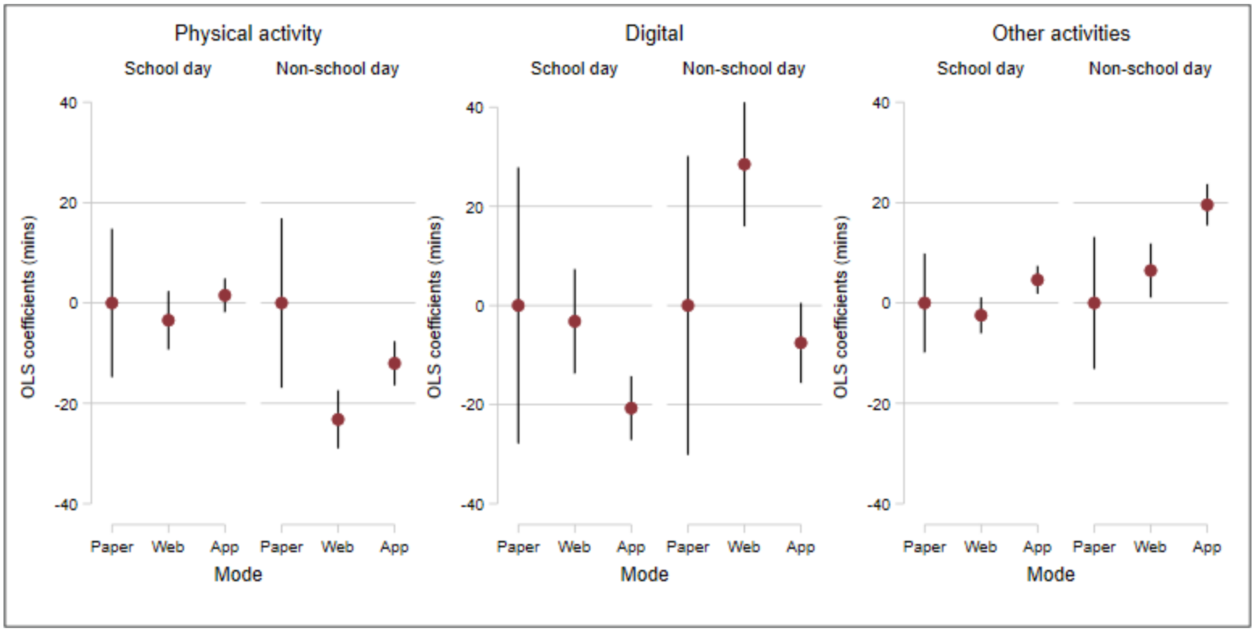
Notes: N=3,786. Good quality diaries report six activities or more, at least one episode of personal care (including sleep) and one episode of eating or drinking, and less than 90 minutes of missing data in the activity diary column.

Figure 2a: Ordinary Least Square Estimates and 95% Comparison Intervals for Time Spent in Daily Activities by Time Diary Mode on school days and on-school days; Good Quality Diaries



Source: *Millennium Cohort Study*
Notes: N= 2,055 (school days); N= 3,786 (non-school days). Good quality diaries report six activities or more, at least one episode of personal care (including sleep) and one episode of eating or drinking, and less than 90 minutes of missing data in the activity diary column. Models control for sex, ethnicity, country of residence, season the diary was completed, parental education, family structure, subjective health status, PC ownership, cognitive ability, and parental worklessness. Robust cluster standard errors adjust for multiple observations per person.

Figure 2b: Ordinary Least Square Estimates and 95% Comparison Intervals for Time Spent in Daily Activities by Time Diary Mode on school days and on-school days; Good Quality Diaries



Source: *Millennium Cohort Study*
Notes: N= 2,055 (school days); N= 3,786 (non-school days). Good quality diaries report six activities or more, at least one episode of personal care (including sleep) and one episode of eating or drinking, and less than 90 minutes of missing data in the activity diary column. Models control for sex, ethnicity, country of residence, season the diary was completed, parental education, family structure, subjective health status, PC ownership, cognitive ability, and parental worklessness. Robust cluster standard errors adjust for multiple observations per person.

There are somewhat stronger differences during non-school days, which are characterized by higher levels of discretionary time. Figure 2b shows that web diarists report less time on physical activity compared to paper and app diarists (23 minutes and 11 minutes respectively). In contrast, app diarists spend more time socializing than web diarists, after controlling for observable socio-demographic characteristics (Figure 2a). The most substantial mode difference highlighted by Figure 2b is on the domain of Internet, TV, and Digital Media, with web diarists reporting over half an hour longer on this activity compared app diarists on non-school days (37 minutes longer). It is worth noting that the lack of significant differences between paper-obtained time-use measures and web and app-obtained measures in these discretionary domains on non-school days is likely due to the small number of paper diaries in our analysis, which resulted in relatively large comparison intervals. In addition to this, missing data in different covariates contributed to the loss of statistical power. Alternative model specifications that did not include verbal score led to a 5% increase of sample size, yielding significant differences between paper and other modes, in line with the descriptive patterns underlined in Table 3b (for example, see Table A18 for digital media).

We have explored interactions between diary mode and verbal score as well as parental educational attainment. None of these interactions were significant. Alternative model specifications that included alternative measures of parental background (i.e., the National Statistics Socio-Economic Classification) yielded the same results.

Overall, our results highlight relatively small mode effects for a set of leisure activities, particularly during non-school days when young people have less structured time. To better understand these patterns, we further investigated reported durations in activity codes included in these particular broad time-use domains. We found that mode effects on time spent on other hobbies were largely driven by differences in time spent reading (not for school) that was higher among web diarists. Similarly, decomposition of the Internet, TV, and Digital Media category revealed that mode differences on this domain were a result of web diarists reporting longer times using the Internet and playing video games. Taken together, these results are in line with those of previous research that has shown that heavy use of personal computers has corresponding effects on other daily activities, and that it is associated with more home-centred and less mobile lifestyles (Vilhelmson et al., 2016). We consider these effects as indicative of unobserved leisure preferences and/or personality differences that are associated with diary mode choice, which cannot be fully accounted by the observable socio-economic background characteristics included in our models. To this end, we note that recent research on young people's time-use has shown that total time spent using computers and other devices (the most pronounced mode effect in our models) is not associated with SES in the UK (Mullan, 2017). We therefore interpret these mode effects as reflective of limitations of the covariates we employ to control for diary mode composition and not as evidence of genuine mode effects on measurement. With regards to the Internet, TV, and Digital Media domain, we also acknowledge the possibility that mode effects are partly an artefact of the methodology itself: young people who chose to complete time diaries would likely complete their records during non-school days, which could consequently increase time spent on other computer-related activities.

5 Conclusions

This paper sought to address the lack of research on mode effects in time-use research, specifically focusing on data quality and measurement. We argued that methodological research on time diaries needs to move beyond established indicators of diary quality, and to incorporate systematic examinations of time-use accounts provided by different time diary modes. Drawing on a large-scale longitudinal study of individuals born between 2000 and 2002 in the UK, we proceeded to examine comparability of paper, web, and app diaries administered at age 14. Our research overcomes limitations of previous diary research on mode effects that has relied on non-representative, small-scale samples (see, e.g., Chatzitheochari et al., 2018; Fernee and Sonck, 2013). At the same time, it provides timely evidence surrounding measurement of time-use in adolescence, a topic that has recently attracted considerable research attention given consistent associations with later life outcomes (Hunt and McKay, 2015).

Our results demonstrate the strength of self-administered web and app instruments in obtaining good quality time-use records, characterized by low levels of missing time, adequate descriptive detail, and appropriate reporting of key daily activities. This is in line with results of previous research that has focused on mode effects on diary quality (Bonke and Fallesen, 2010; Minnen et al., 2014). However, our analyses also highlight the benefits of combining the "gold standard" grid measurement approach (Gershuny, 2003; Michelson, 2015), with the use of soft and hard checks, as evidenced by data quality comparisons of the MCS web and app instruments, as well as the higher reports of non-substantive "other" activity codes in the latter.

We only found modest mode effects for a limited number of discretionary activities. We interpreted these as suggestive of unobserved characteristics and preferences that vary by mode choice, not fully captured by our socio-economic and demographic controls for selection. Our results do not confirm negative effects from the limited visual illustration of activity codes and recorded activities in the MCS web and app diaries. However, it is important to acknowledge that our measurement analyses solely focused on broad time-use domains, which do not capture potential differences at the more granular level of detailed activity codes.

Notwithstanding results on overall app diary quality, our analyses show that app diarists were more likely to use the "any other activity" code than paper or web diarists. We argue that this is associated with the markedly different diary format of the MCS app instrument, which is more cognitively demanding than the paper and the web instrument, both of which provide a visual representation of the surveyed day and the range of broad time-use domains available to the diarist. However, our analyses cannot ascertain whether "other" reports are indicative of survey fatigue with the instrument or a result of a slower learning curve during the beginning of time-use record completion (Chatzitheochari and Mylona, 2021). Better understanding of the specificities of app instruments could be achieved with paradata (Tienda and Koffman, 2020), including timestamps for data entry, which are not made available in the MCS survey.

While our results provide valuable evidence for future time-use data collection, we hasten to note that they need to be interpreted with caution: participants of longitudinal studies are more familiar with different instruments of data collection, and typically display high levels of commitment to their study (Nandi and Platt, 2017), which is likely to have influenced their engagement with the diary task. At the same time, adolescents' familiarity with new technologies is markedly higher compared to older age groups. Future research should therefore examine diary mode effects in cross-sectional time-use studies and/or adult populations.

A question of paramount importance for time-use data collection is whether web and app diaries can replace traditional paper-administered diaries. Results from our study are not sufficient to answer this question, given that we do not examine mode effects on response rate, which is a key consideration for time-use research (Abraham et al., 2006; Gershuny, 2003). At the same time, our analysis solely focuses on main activities. While main activities constitute the most frequently used measures in time-diary research, it is worth noting that several substantive areas of social research (e.g., on housework, eating patterns, screen time) typically require information provided by contextual diary columns too (see, e.g., Bittman and Wajcman, 2000; Mullan and Chatzitheochari, 2019; Zick and Stevens, 2010). Further exploration of recording of contextual codes across different diary modes is therefore needed.

Overall, our results suggest that the majority of the MCS time-use record measures can be analysed unproblematically in a single sample. However, analyses of discretionary activities such as physical activity and digital media would benefit from including a control for diary mode in multivariate models. We acknowledge that results from our study need to be further validated by experimental research designs with random assignment to different diary modes. At the same time, future research should explore other ways of disentangling selection and measurement effects in order to provide more robust estimates of mode effects (Vannieuwenhuyze & Loosveldt, 2013).

However, we contend that out study provides a useful starting point into a topic of utmost importance for the future of time-use research.

Acknowledgments

We thank the Centre for Longitudinal Studies and the UK Data Service for making the data available. These organisations bear no responsibility for the analysis and data interpretations presented in this article. We also thank Roxanne Connelly and Richard Lampard for helpful comments during earlier stages of this research, and Killian Mullan for useful guidance. Elena Mylona's contribution was supported by an Economic and Social Research Council studentship from the Midlands Graduate School Doctoral Training Partnership.

References

Abraham, K. G., Maitland, A., & Bianchi, S. M. (2006). Nonresponse in the American time use Survey: Who is missing from the data and how much does it matter? *International Journal of Public Opinion Quarterly*, *70*, 676–703. https://doi.org/10.1093/poq/nfl037

Allum, N., Conrad, F., & Wenz, A. (2018). Consequences of mid-stream mode-switching in a panel survey. *Survey Research Methods*, *12*, 43–58. https://doi.org/10.18148/srm/2018.v12i1.6779

Bauman, A., Bittman, M., & Gershuny, J. (2019). A short history of time use research; implications for public health. *BMC Public Health*, *19*, 607. https://doi.org/10.1186/s12889-019-6760-y

Bittman, M., & Wajcman, J. (2000). The rush hour: The character of leisure time and gender equity. *Social Forces* 79, 165–189. https://doi.org/10.2307/2675568

Blair, E., & Burton, S. (1987). Cognitive processes used by survey respondents to answer behavioral frequency questions. *Journal of Consumer Research*, *14*, 280–288. https://doi.org/10.1086/209112

Bonke, J., & Christensen, A. E. W. (2019). *Monitoring daily time-use by an electronic time use app (ETUS)*. Copenhagen: Rockwool Foundation.

Bonke, J., & Fallesen, P. (2010). The impact of incentives and interview methods on response quantity and quality in diary-and booklet-based surveys. *Survey Research Methods*, *4*, 91–101. https://doi.org/10.18148/srm/2010.v4i2.3614

Chatzitheochari, S., Fisher, K., Gilbert, E., Calderwood, L., Huskinson, A., Cleary, A., & Gershuny, J. (2015). *Measuring young people's time use in the UK millennium cohort study: A mixed-mode time diary approach* (Centre for Longitudinal Studies Working Paper No. 2015-5). Retrieved from: https://cls.ucl.ac.uk/wp-content/uploads/2017/04/CLS-WP-2015-5.pdf

Chatzitheochari, S., Fisher, K., Gilbert, E., Calderwood, L., Huskinson, T., Cleary, A., & Gershuny, J. (2018). Using new technologies for time diary data collection: Instrument design and data quality findings from a mixed-mode pilot survey. *Social Indicators Research*, *137*, 379–390. https://doi.org/10.1007/s11205-017-1569-5

Chatzitheochari, S. & Mylona, E. (2021) Data quality in web and app diaries: A person-level comparison, *Journal of Time Use Research,* 2. https://doi.org/10.32797/jtur-2021-2

Closs, S. J., & Hutchings, M. J. (1976). *APU arithmetic test*. London: Hodder and Stoughton.

Connelly, R., & Platt, L. (2014). Cohort profile: UK Millennium Cohort study (MCS). *International Journal of Epidemiology*, *43*, 1719–1725. https://doi.org/10.1093/ije/dyu001

Connor, S. O., McCaffrey, N., Whyte, E., & Moran, K. (2016). The novel use of a SenseCam and accelerometer to validate training load and training information in a self-recall training diary. *Journal of Sports Sciences*, *34*, 303–310. https://doi.org/10.1080/02640414.2015.1050600

Cornwell, B., Gershuny, J., & Sullivan, O. (2019). The Social Structure of Time: Emerging Trends and New Directions. *Annual Review of Sociology*, *45*, 301–320. https://doi.org/10.1146/annurev-soc-073018-022416

Couper, M. P. (2017). New developments in survey data collection. *Annual Review of Sociology*, *43*, 121–145. https://doi.org/10.1146/annurev-soc-060116-053613

Couper, M. P., & Peterson, G. J. (2017). Why do web surveys take longer on smartphones? *Social Science Computer Review*, *35*, 357–377. https://doi.org/10.1177/0894439316629932

Daum, T., Buchwald, H., Gerlicher, A., & Birner, R. (2019). Times Have Changed: Using a Pictorial Smartphone App to Collect Time–Use Data in Rural Zambia. *Field Methods*, *31*, 3–22. https://doi.org/10.1177/1525822X18797303

Elevelt, A., Bernasco, W., Lugtig, P., Ruiter, S., & Toepoel, V. (2021). Where You at? Using GPS Locations in an Electronic Time Use Diary Study to Derive Functional Locations. *Social Science Computer Review*, 39, 509-526 https://doi.org/10.1177/0894439319877872

European Commission. (2004). *Guidelines on harmonised European time use surveys*. Luxemburg: Eurostat.

Fernee, H., & Sonck, N. (2013). Is everyone able to use a smartphone in survey research? *Survey Practice*, *6*. https://doi.org/10.29115/SP-2013-0020.

Firth, D., & De Menezes, R. X. (2004). Quasi-variances. *Biometrika*, *91*, 65–80. https://doi.org/10.1093/biomet/91.1.65

Fisher, K., & Gershuny, J. (2013). Coming full circle–Introducing the multinational time use study simple file. *Electronic International Journal of Time Use Research*, *10*, 91. https://doi.org/10.13085/eIJTUR.10.1.91-111

Frazis, H., & Stewart, J. (2012). How to Think about Time-Use Data: What Inferences Can We Make about Long- and Short-Run Time Use from Time Diaries? *Annals of Economic and Statistics, 105*, *231–245.*https://doi.org/10.2307/23646463

Gershuny, J. (2003). *Changing times: Work and leisure in postindustrial society* (2nd ed.). Oxford, England: Oxford University Press.

Gershuny, J., Harms, T., Doherty, A., Thomas, E., Milton, K., Kelly, P., & Foster, C. (2020). Testing Self-Report Time-Use Diaries against Objective Instruments in Real Time. *Sociological Methodology, 50, 318-349* https://doi.org/10.1177/0081175019884591

Glorieux, I., & Minnen, J. (2009). How many days? A comparison of the quality of time-use data from 2-day and 7-day diaries. *Electronic International Journal of Time Use Research*, *6*, 314–327. https://doi.org/ 0.13085/eIJTUR.6.2.314-327

Hox, J., De Leeuw, E., & Klausch, T. (2017). Mixed mode research: Issues in design and analysis. In: Biemer, P., P., de Leeuw, E., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L., E., Tucker, N., C., & West, B., T. (Eds.), *Total Survey Error in Practice* (pp.511–530). New York, USA: Wiley.

Hunt, E. & McKay, E. (2015). What can be learned from adolescent time diary research. *Journal of Adolescent Health, 56***,** 259-66. https://doi.org/10.1016/j.jadohealth.2014.11.007

Ipsos MORI. (2016). *Millennium cohort study sixth sweep (MCS6): Technical Report*. Ipsos MORI Social Research Institute. Retrieved from: http://doc.ukdataservice.ac.uk/doc/8156/mrdoc/pdf/mcs6_technical_report.pdf

Jäckle, A., Roberts, C., & Lynn, P. (2010). Assessing the effect of data collection mode on measurement. *International Statistical Review*, *78*, 3–20. https://doi.org/10.1111/j.1751-5823.2010.00102.x

Juster, F. T., Ono, H., & Stafford, F. P. (2003). An assessment of alternative measures of time use. *Sociological Methodology*, *33*, 19–54. https://doi.org/10.1111/j.0081-1750.2003.t01-1-00126.x

Kan, M. Y., & Pudney, S. (2008). Measurement Error in Stylized and Diary Data on Time Use. *Sociological Methodology*, *38*, 101–132. https://doi.org/10.1111/j.1467-9531.2008.00197.x

Kelly, P., Thomas, E., Doherty, A., Harms, T., Burke, Ó., Gershuny, J., & Foster, C. (2015). Developing a method to test the validity of 24 hour time use diaries using wearable cameras: A feasibility pilot. *PLoS One*, *10*. https://doi.org/10.1371/journal.pone.0142198

Klasnja, P., & Pratt, W. (2012). Healthcare in the pocket: Mapping the space of mobile-phone health interventions. *Journal of Biomedical Informatics*, *45*, 184–198. https://doi.org/10.1016/j.jbi.2011.08.017

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*, 213–236. https://doi.org/10.1002/acp.2350050305

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, *50*, 537–567. https://doi.org/10.1146/annurev.psych.50.1537

Link, M. W., Murphy, J., Schober, M. F., Buskirk, T. D., Hunter Childs, J., & Langer Tesfaye, C. (2014). Mobile technologies for conducting, augmenting and potentially replacing surveys: Executive summary of the AAPOR task force on emerging technologies in public opinion research. *Public Opinion Quarterly*, *78*, 779–787. https://doi.org/0.1093/poq/nfu054

Lugtig, P., Lensvelt-Mulders, G. J., Frerichs, R., & Greven, A. (2011). Estimating nonresponse bias and mode effects in a mixed-mode survey. *International Journal of Market Research*, *53*, 669–686. https://doi.org/10.2501/IJMR-53-5-669-686

Michelson, W. H. (2015). *Time use: Expanding explanation in the social sciences*. London, England: Routledge.

Minnen, J., Glorieux, I., van Tienoven, T. P., Daniels, S., Weenas, D., Deyaert, J., Van den Bogaert, S., & Rymenants, S. (2014). Modular Online Time Use Survey (MOTUS)–Translating an existing method in the 21st century. *Electronic International Journal of Time Use Research*, *11*, 73–93. dx.doi.org/10.13085/eIJTUR.11.1.73-93

Morris, S., Humphrey, A., Cabrera, Alvarez, P., & D'Lima, O., (2016). *The UK Time Diary Study 2014-2015*. Technical Report, National Centre for Social Research.

Mullan, K. & Chatzitheochari, S. (2019) Changing times together? A time diary analysis of family time in the digital age in the UK, *Journal of Marriage and Family, 81, 795-811*. https://doi.org/10.1111/jomf.12564

Mullan, K. (2017). Technology and children's screen-based activities in the UK: The story of the Millennium so far. *Child Indicators Research*, *11*, 1781–1800. https://doi.org/10.1007/s12187-017-9509-0

Nandi, A., & Platt, L. (2017). Are there differences in responses to social identity questions in face-to-face versus telephone interviews? Results of an experiment on a longitudinal survey. *International Journal of Social Research Methodology*, *20*, 151–166. https://doi.org/10.1080/13645579.2016.1165495

Ofcom. (2015). *Communications market report*. London, England: Ofcom.

Robinson, J., & Godbey, G. (2010). *Time for life: The surprising ways Americans use their time*. University Park, USA: Penn State Press.

Rydenstam, K., & Wadeskog, A. (1998). *Evaluation of the European time use pilot survey.* Luxemburg: Eurostat.

Stewart, J. (2013). Tobit or not Tobit? *Journal of Economic and Social Measurement*, *38*, 263–290. https://doi.org/10.3233/JEM-130376

Tienda, M. & Koffman, D. (2020) Using paradata to evaluate youth participation in a digital diary study, *Social Science Computer Review.* https://doi.org/10.1177/0894439320929272

University of London, Institute of Education, Centre for Longitudinal Studies (2020). *Millennium Cohort Study: Sixth Survey, 2015* [data collection]. *6th Edition.* UK Data Service. https://doi.og/10.5255.UKDA-SN-8156-7

Väisänen, P., & Finland, S. (2006). Mean of episode lengths as a quality indicator of time use diaries. Presentation at 28th IATUR Annual Conference, Copenhagen, Denmark: International Association for Time Use Research.

Vannieuwenhuyze, J. T., & Loosveldt, G. (2013). Evaluating relative mode effects in mixed-mode surveys: Three methods to disentangle selection and measurement effects. *Sociological Methods & Research*, *42*, 82–104. https://doi.org/10.1177/0049124112464868

Vilhelmson, B., Thulin, E. & Elldér, E (2017) Where does time spent on the Internet come from? Tracing the influence of information and communications technology use on daily activities, *Information, Communication and Society*, 20, 250-263. https://doi.org/10.1080/1369118X.2016.1164741

Walthery, P., & Gershuny, J. (2019). Improving Stylised Working Time Estimates with Time Diary Data: A Multi Study Assessment for the UK. *Social Indicators Research*, *144*, 1303–1321. https://doi.org/10.1007/s11205-019-02074-3

Zeni, M., Bison, I., Gauckler, B., & Giunchiglia, F. R. F. (2020). Improving time use measurement with personal big data collection–the experience of the European Big Data Hackathon 2019. *ArXiv Preprint ArXiv:2004.11940*. https://doi.org/10.48550/arXiv.2004.11940

Zick, C. & Stevens, R. (2010). Trends in Americans' food-related time-use: 1975-2006. *Public Health Nutrition, 13,* 1064-1072. https://doi.org/10.1017/S1368980009992138