



Data Quality in Web and App Diaries: A Person-Level Comparison

Stella Chatzitheochari¹ and Elena Mylona²

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

The time-use diary is a complex and burdensome data collection instrument. This can negatively affect data quality, leading to less detailed and/or inaccurate activity reporting as the surveyed time period unfolds. However, it can also be argued that data quality may actually improve over time as respondents become more familiar with the diary instrument format and more interested in the diary task. These competing hypotheses have only been partially tested on data from paper and telephone-administered diaries, which are traditionally used for large-scale data collection. Less is known about self-administered modes that make use of new technologies, despite their increasing popularity among researchers. This research note rectifies this omission by comparing diary quality in self-administered web and app diaries, drawing on data from the Millennium Cohort Study. We construct a person-level data quality typology, using information on missing data, episode changes, and reporting of key daily activity domains. Results show significant mode differences on person-level data quality, after controlling for characteristics known to influence diary mode selection and data quality. App diarists were more likely to return two diaries of inconsistent quality. Both respondent fatigue and improvement of completion over time appear more common among app diarists.

Keywords: data quality; diary mode; respondent fatigue; time diaries

<https://doi.org/10.32797/jtur-2021-2>

1 Background

Time diaries are particularly burdensome to fill in, which typically leads to low response rates in stand-alone time-use surveys (Gershuny, 2003; Abraham, Maitland and Bianchi, 2006; Ingen, Stoop and Breedveld, 2009). For example, the most recently collected UK time-use survey achieved a net response rate of approximately 33% (Morris *et al.*, 2016). There has been a lot of methodological interest on implications of non-response for time allocation estimates produced by time-use surveys, focusing on the extent to which the time allocation of non-respondents differs from that of those who complete time diaries (Gershuny 2003). This research has generally shown minor non-response effects. Burdensome data collection instruments can also compromise diary data quality. This is another reason why the majority of statistical agencies collecting time-use surveys sample two days rather than an entire week (Eurostat, 2004).

However, issues surrounding diary quality have thus far attracted less methodological interest than those related to non-response. As a result, we know little about the influence of respondent fatigue on the time-use accounts produced by respondents of time-use surveys. Diarists who experience respondent fatigue and still choose to complete their designated diaries (instead of dropping out) may be less likely to provide accurate time-use accounts as the surveyed time period unfolds. This could be the result of a higher tendency to engage in survey satisficing (Krosnick, 1991). Satisficing could be understood as the tendency to use the first available activity category and/or only a narrow set of activity categories in pre-coded “light” time diaries. Similarly, satisficing could entail a less detailed representation of the surveyed day in “heavy” open-ended time diaries by providing unfocused or inaccurate textual descriptions of activity patterns. At the same time, respondent fatigue may lead to higher levels of missing data due to decreasing engagement with the diary task. Given that respondent fatigue is positively associated with survey length (Sharp and Frankel, 1983; Galesic and Bosnjak, 2009), it would be sensible to expect diary quality to progressively drop. That is, respondents who experience respondent fatigue, will be more likely to produce worse quality diary records in their second surveyed day and onwards (if a longer time period is surveyed). Earlier methodological research on time diary quality has provided evidence on the presence of respondent fatigue, focusing on paper and/or telephone-administered diaries (Gershuny, 2003; Backor, Golde and Nie, 2007; Robinson and Godbey, 2010).

Although respondent fatigue is usually taken for granted in time-use research, it may also be hypothesized that diary quality may improve over time, as a result of a “learning effect” (Burchell and Marsh, 1992). Diarists may become more familiar with the instrument format and/or more interested in the diary task as the surveyed time period unfolds. This could lead to more detailed descriptions over time, achieved by making use of a wider range of pre-coded activities in “light” time diaries or by providing more detailed and frequent textual descriptions in “heavy” open-ended diaries. Indeed, Freedman et al (2012) have recently shown that diary quality is higher on the second rather than the first surveyed diary day, using data from the Panel Study of Income Dynamics. Similarly, analyses of the Belgium Modular Online Time Use Survey suggest that missing time is reduced significantly after the first surveyed diary day (Minnen *et al.*, 2014).

Overall, methodological literature is inconclusive on the effect of number of diary days on diary quality. It is worth noting that the majority of existing research provides diary-level rather than person-level analyses of diary quality. However, the latter is more appropriate for examining the presence of respondent fatigue and “learning effects”. At the same time, previous studies have mostly focused on paper or telephone administered diaries, which have long constituted the main instruments for large-scale time-use data collection. Although recent years have witnessed increasing interest in the development and use of low cost self-administered app and web diaries for data collection (Chatzitheochari *et al.*, 2018; Cornwell, Gershuny and Sullivan, 2019; Sullivan *et al.*, 2020), there have only been few attempts to analyse data quality in such new instruments (with the exception of Minnen *et al.*, 2014; Chatzitheochari *et al.*, 2018; Bonke and Christensen, 2019). It is also worth noting that there are no reports surrounding reliability testing during the development stage of new time-diary instruments. However, questions surrounding the effect of diary mode and diary format on data quality are of utmost importance for the new era of time-use data collection.

In this research note, we provide a person-level analysis of diary quality in web and app diaries. We draw on data from the time-use element of Wave 6 of the Millennium Cohort Study (University of London, Institute of Education, and Centre for Longitudinal Studies, 2020), which made use of a mixed-mode diary design to measure the daily time allocation of 14 year olds in the UK. In the next section, we describe our data and measures. We also outline the main differences between the Millennium Cohort Study

web and app time diary, and reflect on how instrument characteristics may be associated with person-level diary quality.

2 Data and Methods

2.1 The Millennium Cohort Study Time-Use Record

The Millennium Cohort Study (MCS) is the youngest of UK's renowned cohort studies, following approximately 19,000 individuals born between 2000 and 2002 (Connelly and Platt, 2014). The MCS sample was designed to be nationally-representative at the baseline, and has included oversampling of certain sub-groups to ensure adequate representation in the study. The study has thus far completed 7 waves of data collection: at age 9 months, 3, 5, 7, 11, 14, and 17 years.

MCS cohort members became the main informants of the study at Wave 6 of data collection (age 14). Wave 6 also included a substantial time-use element, which invited a sub-sample of cohort members to fill in 24-hour pre-coded time diaries and wear accelerometers for two randomly selected days, a weekday and a weekend day of the same week. The time-diary collection employed a mixed-mode design: cohort members were offered a smartphone app or a web-based diary. A paper diary was available for those unable or unwilling to use the app or web mode. A random subsample of 81% MCS households in England were selected for the time-use element, whereas all households in Wales, Scotland, and Northern Ireland were included. Among cohort members who consented to the diary task and completed time diaries, over 90% did so using the web or the app mode.

Diary differences across MCS diary modes have been discussed in detail in Chatzitheochari et al (2018). Readers interested in the development stage and layout of the three diary modes can also refer to Chatzitheochari et al. (2015). For the purposes of this research note, we outline key similarities and differences between the web and app mode that our analysis focuses on. Both modes made use of soft and hard checks, which can potentially improve data quality by minimizing missing data (Chatzitheochari et al., 2018). The smartphone app employed a question-based approach, while the web mode employed a conventional time-grid approach. The web instrument is thus similar to a standard paper-administered diary, which is still recognized as the "gold standard" for large-scale time-use data collection (Bauman, Bittman and Gershuny, 2019). Question-based measurement approaches are typically seen as more challenging, as they do not aid recall in the same way the time grid does (Robinson and Godbey, 2010). Each time an app diarist recorded an activity for a specific period of time, they had to respond to 3 contextual questions about location, co-presence, and enjoyment. This was not the case for web diarists who could simply complete their record in a different manner and entirely neglect contextual questions if they wanted to. The question-based approach of the app diary allowed respondents to specify the start and end times of their activities, while the web mode made use of 10-minute blocks. Combined together, these aspects suggest that the app diary was more cognitively demanding than the web diary. It is therefore more likely to observe fluctuations in person-level data quality among app diarists as opposed to web diarists, after controlling for individual characteristics that influence selection into diary mode. At the same time, completing such a demanding task in a relatively small size screen may further influence diary quality (Couper and Peterson, 2017). We hasten to note that the preliminary analysis presented in this research note is not able to ascertain which aspects of the diary format and administration may be driving mode effects on person-level diary quality. This would require more detailed data from respondents that were not available for this analysis.

This analysis draws on the *harmonised* MCS time-use data, whereby app records have been rounded to 10-minute blocks to allow comparisons with web and paper records. A raw file is also available for users through the UK Data Service. We focus on 6418 diaries from 3209 cohort members. We excluded 1624 diaries of cohort members who only returned one time diary and/or were missing data on our independent variables (introduced in section 2.2).

2.2 Outcome Variable

Our analysis draws on a *person-level* typology of diary quality *across two diary days*: 1) 2 good quality diaries, 2) 2 bad quality diaries, 3) good quality diary, followed by a bad quality diary, 4) bad quality diary, followed by good quality diary. We define good quality diaries as those that report at least six diary episodes, 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. This definition draws on the conventional good quality diary measure employed in time-use research (Chatzitheochari et al., 2018). However, given that we are analyzing adolescent rather than adult time diaries, we used a less stringent measure, excluding travelling from our list of basic daily activity domains. We also conducted sensitivity analyses, focusing on less strict missing data thresholds (less than 120 minutes and less than 180 minutes missing activity data). Results remained the same. These models are shown in the Appendix.

2.3 Independent Variables

Aside from diary mode (web/app), our analysis draws on a set of variables known to influence diary mode selection and data quality: sex (male, female), ethnicity (White, Mixed, Indian, Pakistani/Bangladeshi, Black), parental educational attainment (no qualifications, NVQ1, NVQ2, NVQ3, Degree or higher), family structure (two-parent households, lone-parent household), and score on a naming vocabulary test administered at age 14, which was derived from a shortened version of the APU Vocabulary Test (Closs & Hutchings, 1976). The latter was included as a proxy of cognitive ability, which has been shown to be influential on mode effects in questionnaire-based social surveys (Nandi & Platt 2017).

Lack of data prevented us from calculating the distance between the two surveyed days, which could also potentially influence diary quality. We also repeated our analysis including conventional diary quality measures (whether the diary days were unusual and whether the young person experienced difficulties when filling in the diaries). These variables did not change our substantive results. We do not present these in this paper due to sample size considerations.

2.4 Sample Characteristics

Table 1 presents descriptive information on our person-level diary quality typology and on independent variables by diary mode. All estimates have been adjusted for the complex sampling design of the study and survey non-response. As shown in Table 1, approximately 74% of cohort members in our sample opted to fill in the time diary in a smartphone app, demonstrating the widespread diffusion of internet-enabled devices in this age group. Table 1 shows that web diarists were more likely than app diarists to produce two good quality diaries (76% as opposed to 72%). It also shows a small difference between app and web diarists who returned two bad quality diaries (10% and 12% respectively). Results for the two other categories seem to confirm our expectation of higher fluctuation in the data quality provided by app diarists. Notably, app diarists are more likely than web diarists to return a good quality diary followed by a bad quality diary. This can be understood as tentative evidence of a higher likelihood of respondent fatigue among app diarists. We see a smaller mode difference among the group who returned one bad quality diary, followed by a good quality diary.

Table 1 also shows that girls were more likely than boys to fill in the smartphone diary app. The same holds for White cohort members. At the same time, Table 1 shows that web diarists were more likely to have parents with higher educational attainment and that they also scored higher in the naming vocabulary test. As these characteristics are known to influence data quality, our analysis examines whether differences in person-level diary quality are driven by variation in other factors.

Table 1: Sample Characteristics by Time Diary Mode, column %

Measures	App		Web	
	N	%	N	%
Diary Quality				
Good quality day 1 and 2	1705	72.1	640	75.9
Good quality day 1, bad quality day 2	210	8.9	42	5.0
Bad quality day 1, good quality day 2	217	9.2	62	7.4
Bad quality day 1 and 2	234	9.9	99	11.7
Sex				
Male	984	41.6	473	56.1
Female	1382	58.4	370	43.9
Ethnicity				
White	2104	88.9	715	84.8
Mixed	91	3.9	36	4.3
Indian	57	2.4	33	3.9
Pakistani/Bangladeshi	77	3.3	37	4.4
Black	37	1.6	22	2.6
Parental education				
No qualifications	55	2.3	16	1.9
NVQ 1 (Level 1/CSE)	64	2.7	25	3.0
NVQ2 (O levels)	365	15.4	116	13.8
NVQ3 (A Levels)	341	14.4	99	11.7
Degree or higher	1541	65.1	587	69.6
Family structure				
Two-parent household	1907	80.6	701	83.2
Lone-parent household	459	19.4	142	16.8
Word activity score mean	2366	7.4	843	8.0

Notes: $N = 3,209$. Millennium Cohort Study Wave 6. Adjusted for complex survey design and survey non-response.

3 Analysis

We now present results from multinomial regression models predicting person-level diary quality. We show coefficients from unadjusted and fully-adjusted models, specifically focusing on results for mode effects. Full models are shown in the Appendix (Table A1). The reference group is two good quality diaries, which is the most common outcome in our sample. Log odds for the web mode refer to estimates for different person-level quality outcomes relative to good quality diaries, when other variables in the model are held constant. Results highlight statistically significant differences for two outcomes: 1 good quality diary, followed by 1 bad quality diary, as well as 1 bad quality diary, followed by a good quality diary. After controlling factors associated with mode selection and data quality, web diarists are less likely than app diarists to fall under these two data quality categories. These results remained robust to different definitions

of good quality diaries (see Tables A2 and A3 of the Appendix). We also examined interactions of diary mode with other socio-demographic characteristics. These interactions were not statistically significant so we retained the more parsimonious specification.

Table 2: Log-odds Estimates from Multinomial Logistic Regression Model Predicting Person-level Diary Quality

		Unadjusted	With controls⁺
		Coefficient (SE)	Coefficient (SE)
Good day 1 and 2	Web (ref.=App)	0 (-)	0 (-)
Good day 1, Bad day 2	Web (ref.=App)	-0.63 *** (-0.18)	-0.63 *** (-0.18)
Bad day 1, Good day 2	Web (ref.=App)	-0.27 (-0.15)	-0.31 * (-0.15)
Bad day 1 and 2	Web (ref.=App)	0.12 (-0.13)	0.11 (-0.13)

Notes: $N = 3,209$. Millennium Cohort Study Wave 6; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

⁺Controls: sex, ethnicity, highest parental education, lone/two-parent household, vocabulary test score. Adjusted for complex sampling design and non-response.

4 Discussion and Concluding Remarks

Self-administered time diaries that make use of new technologies can lower fieldwork costs and simplify data cleaning procedures. However, very few studies have thus far examined the quality of data produced by these time diary modes. In this research note, we provided an initial analysis of person-level data quality in web and app diary modes, drawing on data from Wave 6 of the Millennium Cohort Study, which sought to understand time allocation of cohort members in early adolescence by surveying two days of the same week at age 14. Our finding that over two thirds of cohort members across web and app modes provided consistently good quality data is particularly encouraging. However, these results are specific to adolescents who may be more accustomed to using computers and smartphones in their everyday life. Further research examining data quality in adult populations is also needed.

In line with our expectations, we found that app diarists are more likely than web diarists to complete one good quality diary, followed by a bad quality diary, as well as a bad quality diary, followed by a good quality diary. These results are suggestive of respondent fatigue as well as of a “learning effect” during diary completion. We note that results surrounding respondent fatigue should be interpreted in light of our analytic sample: Given that we only focus on those participants who returned two time diaries respondent fatigue is in fact underestimated in our analysis, as we do not take into account those who dropped out after the first surveyed day. Likewise, our reliance on a composite measure of data quality means that we are not taking into account respondent fatigue as well improvement over time among those in the two good quality diaries category. Further analyses of paradata (i.e. auxiliary variables that describe the data collection process) on diary non-response can produce a more detailed picture of respondent fatigue. Additionally, information on cohort members’ response quality on other MCS components (e.g. individual questionnaire) could shed further light on our understanding of determinants of diary data quality. However, it is important to also note that a better understanding of mode effects on diary quality also requires direct input from respondents through interviews and usability testing.

Overall, this initial exploration suggests that the traditional “gold-standard” time-grid format is more likely to provide good quality data at the person level. MCS data do not allow us to gain a deeper understanding of the specific app diary format aspects that may be implicated in lower engagement with the

diary task or difficulties of completion. However, given the increasing interest and opportunities offered by this mobile time-use data collection, we contend that future methodological analyses of diary quality and non-response for this mode are urgently needed.

Acknowledgments

We are grateful to the Centre for Longitudinal Studies for the use of MCS data and to the UK Data Service for making the data available. These organisations bear no responsibility for the analysis and data interpretations presented in this article. Elena Mylona's contribution was supported by an Economic and Social Research Council studentship from the Midlands Graduate School Doctoral Training Partnership. We also thank Killian Mullan for useful comments and suggestions.

References

- Abraham, K. G., Maitland, A. and 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(5), pp. 676–703.
- Backor, K., Golde, S. and Nie, N. (2007) 'Estimating survey fatigue in time use study', in *international association for time use research conference. Washington, DC*. Citeseer.
- Bauman, A., Bittman, M. and Gershuny, J. (2019) 'A short history of time use research; implications for public health', *BMC public health*, 19(2), p. 607.
- Bonke, J. and Christensen, A. E. W. (2019) *Monitoring daily time-use by an electronic time use app (ETUS)*. Copenhagen: Rockwool Foundation. Copenhagen: Rockwool Foundation.
- Burchell, B. and Marsh, C. (1992) 'The effect of questionnaire length on survey response', *Quality and quantity*, 26(3), pp. 233–244.
- Chatzitheochari, S. *et al.* (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(1), pp. 379–390.
- Chatzitheochari, S., *et al.* (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>
- Connelly, R. and Platt, L. (2014) 'Cohort profile: UK millennium Cohort study (MCS)', *International journal of epidemiology*, 43(6), pp. 1719–1725.
- Cornwell, B., Gershuny, J. and Sullivan, O. (2019) 'The social structure of time: Emerging trends and new directions', *Annual Review of Sociology*, 45, pp. 301–320.
- Couper, M. P. and Peterson, G. J. (2017) 'Why do web surveys take longer on smartphones?', *Social Science Computer Review*, 35(3), pp. 357–377.
- Eurostat, E. C. (2004) *Guidelines on harmonised European time use surveys*. Luxembourg: Luxembourg.
- Freedman, V. A. *et al.* (2012) 'Assessing time diary quality for older couples: An analysis of the panel study of income dynamics' disability and use of time (DUST) supplement', *Annals of economics and statistics*, 105, p. 271.
- Galesic, M. and Bosnjak, M. (2009) 'Effects of questionnaire length on participation and indicators of response quality in a web survey', *Public opinion quarterly*, 73(2), pp. 349–360.
- Gershuny, J. (2003) *Changing times: Work and leisure in postindustrial society*. Oxford University Press on Demand.

- Ingen, E. V., Stoop, I. and Breedveld, K. (2009) ‘Nonresponse in the Dutch time use survey: Strategies for response enhancement and bias reduction’, *Field Methods*, 21(1), pp. 69–90.
- Krosnick, J. A. (1991) ‘Response strategies for coping with the cognitive demands of attitude measures in surveys’, *Applied cognitive psychology*, 5(3), pp. 213–236.
- Minnen, J. *et al.* (2014) ‘Modular Online Time Use Survey (MOTUS)-Translating an existing method in the 21 st century.’, *Electronic International Journal of Time Use Research*, 11(1).
- Morris, S. *et al.* (2016) ‘The UK Time Diary Study 2014-2015’.
- Robinson, J. and Godbey, G. (2010) *Time for life: The surprising ways Americans use their time*. Penn State Press.
- Sharp, L. M. and Frankel, J. (1983) ‘Respondent burden: A test of some common assumptions’, *Public Opinion Quarterly*, 47(1), pp. 36–53.
- Sullivan, O. *et al.* (2020) ‘Time use diary design for our times-an overview, presenting a ‘click-and-drag’ diary instrument for online application’, *J. Time Use Research*.
- University of London, Institute of Education, Centre for Longitudinal Studies. (2020). *Millennium Cohort Study: Sixth Survey, 2015*. [data collection]. 7th Edition. UK Data Service.

Appendix

Table A1: Log Odds from Multinomial Regression Models Predicting Person Level Diary Quality

	Unadjusted		Adjusted	
	Log-odds	S.E.	Log-odds	S.E.
Good quality day 1 and 2	-		-	
Good quality day 1, bad quality day 2				
App	0.00	(.)	0.00	(.)
Web	-0.63***	-0.18	-0.63***	-0.18
Male			0.00	(.)
Female			-0.29*	-0.14
White			0.00	(.)
Mixed			0.15	-0.35
Indian			-0.20	-0.47
Pakistani/Bangladeshi			-0.42	-0.44
Black			0.92*	-0.39
No qualifications			0.00	(.)
NVQ 1 (Level 1/CSE)			0.81	-0.57
NVQ2 (O levels)			0.09	-0.51
NVQ3 (A Levels)			0.06	-0.52
Degree or higher			-0.09	-0.5
Two-parent household			0.00	(.)
Lone-parent household			0.21	-0.17
Word activity score <i>mean</i>			-0.08**	-0.03
Constant	-2.09***	-0.07	-1.39**	-0.53

(continued)

Table A1 (continued): Log Odds from Multinomial Regression Models Predicting Person Level Diary Quality

	Unadjusted		Adjusted	
	Log-odds	S.E.	Log-odds	S.E.
Good quality day 1 and 2		-	-	
Bad quality day 1, good quality day 2				
App	0.00	(.)	0.00	(.)
Web	-0.27	-0.15	-0.31*	-0.15
Male			0.00	(.)
Female			-0.45***	-0.13
White			0.00	(.)
Mixed			0.33	-0.31
Indian			0.22	-0.37
Pakistani/Bangladeshi			0.19	-0.33
Black			0.44	-0.45
No qualifications			0.00	(.)
NVQ 1 (Level 1/CSE)			0.46	-0.53
NVQ2 (O levels)			-0.14	-0.45
NVQ3 (A Levels)			-0.24	-0.46
Degree or higher			-0.23	-0.44
Two-parent household			0.00	(.)
Lone-parent household			0.01	-0.17
Word activity score <i>mean</i>			-0.08**	-0.03
Constant	-2.06***	-0.07	-1.08*	-0.47

(continued)

Table A1 (continued): Log Odds from Multinomial Regression Models Predicting Person Level Diary Quality

	Unadjusted		Adjusted	
	Log-odds	S.E.	Log-odds	S.E.
Good quality day 1 and 2	-		-	
Bad quality day 1 and 2				
App	0.00	(.)	0.00	(.)
Web	0.12	-0.13	0.11	-0.13
Male			0.00	(.)
Female			-0.56***	-0.12
White			0.00	(.)
Mixed			0.78**	-0.25
Indian			0.32	-0.34
Pakistani/Bangladeshi			0.22	-0.29
Black			0.86*	-0.37
No qualifications			0.00	(.)
NVQ 1 (Level 1/CSE)			-0.26	-0.43
NVQ2 (O levels)			-0.40	-0.33
NVQ3 (A Levels)			-0.55	-0.34
Degree or higher			-1.11***	-0.32
Two-parent household			0.00	(.)
Lone-parent household			0.14	-0.15
Word activity score <i>mean</i>			-0.13***	-0.03
Constant	-1.99***	-0.07	-0.0004	-0.36
Pseudo R-squared	0.003		0.032	

Notes: $N=3,209$. Millennium Cohort Study Wave 6. * $p<0.05$, ** $p<0.01$, *** $p<0.001$. Adjusted for complex survey design and non-response. 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 A2: Log Odds from Multinomial Regression Models Predicting Person-Level Diary Quality; Alternative Definitions of Good Quality Diary (Less than 120 Minutes Missing Activity Time)

	Unadjusted		Adjusted	
	Log-odds	S.E.	Log-odds	S.E.
Good quality day 1 and 2	-		-	
Good quality day 1, bad quality day 2				
App	0.00	(.)	0.00	(.)
Web	-0.56**	(0.17)	-0.56**	(0.18)
Male			0.00	(.)
Female			-0.30*	(0.14)
White			0.00	(.)
Mixed			0.05	(0.36)
Indian			-0.23	(0.47)
Pakistani/Bangladeshi			-0.18	(0.39)
Black			0.59	(0.43)
No qualifications			0.00	(.)
NVQ 1 (Level 1/CSE)			0.44	(0.52)
NVQ2 (O levels)			-0.19	(0.45)
NVQ3 (A Levels)			-0.26	(0.46)
Degree or higher			-0.45	(0.44)
Two-parent household			0.00	(.)
Lone-parent household			0.17	(0.17)
Word activity score <i>mean</i>			-0.07*	(0.03)
Constant	-2.13***	(0.07)	-1.14*	(0.47)

(continued)

Table A2 (continued): Log Odds from Multinomial Regression Models Predicting Person-Level Diary Quality; Alternative Definitions of Good Quality Diary (Less than 120 Minutes Missing Activity Time)

	Unadjusted		Adjusted	
	Log-odds	S.E.	Log-odds	S.E.
Good quality day 1 and 2	-		-	
Bad quality day 1, good quality day 2				
App	0.00	(.)	0.00	(.)
Web	-0.56**	(0.17)	-0.56**	(0.18)
Male			0.00	(.)
Female			-0.30*	(0.14)
White			0.00	(.)
Mixed			0.05	(0.36)
Indian			-0.23	(0.47)
Pakistani/Bangladeshi			-0.18	(0.39)
Black			0.59	(0.43)
No qualifications			0.00	(.)
NVQ 1 (Level 1/CSE)			0.44	(0.52)
NVQ2 (O levels)			-0.19	(0.45)
NVQ3 (A Levels)			-0.26	(0.46)
Degree or higher			-0.45	(0.44)
Two-parent household			0.00	(.)
Lone-parent household			0.17	(0.17)
Word activity score <i>mean</i>			-0.07*	(0.03)
Constant	-2.13***	(0.07)	-1.14*	(0.47)

(continued)

Table A2 (continued): Log Odds from Multinomial Regression Models Predicting Person-Level Diary Quality; Alternative Definitions of Good Quality Diary (Less than 120 Minutes Missing Activity Time)

	Unadjusted		Adjusted	
	Log-odds	S.E.	Log-odds	S.E.
Good quality day 1 and 2	-		-	
Bad quality day 1 and 2				
App	0.00	(.)	0.00	(.)
Web	0.091	(0.13)	0.09	(0.14)
Male			0.00	(.)
Female			-0.56***	(0.12)
White			0.00	(.)
Mixed			0.82***	(0.25)
Indian			0.32	(0.35)
Pakistani/Bangladeshi			0.21	(0.30)
Black			0.75*	(0.37)
No qualifications			0.00	(.)
NVQ 1 (Level 1/CSE)			-0.25	(0.45)
NVQ2 (O levels)			-0.22	(0.35)
NVQ3 (A Levels)			-0.39	(0.36)
Degree or higher			-0.93**	(0.34)
Two-parent household			0.00	(.)
Lone-parent household			0.14	(0.15)
Word activity score <i>mean</i>			-0.14***	(0.03)
Constant	-2.04***	(0.07)	-0.16	(0.38)
Pseudo R-squared	0.003		0.030	

Notes: $N=3,209$. * $p<0.05$, ** $p<0.01$, *** $p<0.001$. Adjusted for complex survey design and non-response. 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 120 minutes of missing data in the activity diary column.

Table A3: Log Odds from Multinomial Regression Models Predicting Person-Level Diary Quality; Alternative Definitions of Good Quality Diary (Less than 180 Minutes Missing Activity Time)

	Unadjusted		Adjusted	
	Log-odds	S.E.	Log-odds	S.E.
Good quality day 1 and 2	-		-	
Good quality day 1, bad quality day 2				
App	0.00	(.)	0.00	(.)
Web	-0.60***	(0.17)	-0.60***	(0.18)
Male			0.00	(.)
Female			-0.29*	(0.13)
White			0.00	(.)
Mixed			0.15	(0.35)
Indian			-0.25	(0.48)
Pakistani/Bangladeshi			-0.22	(0.39)
Black			0.90*	(0.39)
No qualifications			0.00	(.)
NVQ 1 (Level 1/CSE)			0.42	(0.50)
NVQ2 (O levels)			-0.36	(0.43)
NVQ3 (A Levels)			-0.37	(0.44)
Degree or higher			-0.55	(0.42)
Two-parent household			0.00	(.)
Lone-parent household			0.21	(0.17)
Word activity score <i>mean</i>			-0.07**	(0.03)
Constant	-2.09***	(0.07)	-0.99*	(0.46)

(continued)

Table A3 (continued): Log Odds from Multinomial Regression Models Predicting Person-Level Diary Quality; Alternative Definitions of Good Quality Diary (Less than 180 Minutes Missing Activity Time)

	Unadjusted		Adjusted	
	Log-odds	S.E.	Log-odds	S.E.
Good quality day 1 and 2	-		-	
Bad quality day 1, good quality day 2				
App	0.00	(.)	0.00	(.)
Web	-0.32*	(0.15)	-0.36*	(0.16)
Male			0.00	(.)
Female			-0.46***	(0.13)
White			0.00	(.)
Mixed			0.35	(0.31)
Indian			0.25	(0.37)
Pakistani/Bangladeshi			0.21	(0.33)
Black			0.47	(0.46)
No qualifications			0.00	(.)
NVQ 1 (Level 1/CSE)			0.47	(0.53)
NVQ2 (O levels)			-0.11	(0.45)
NVQ3 (A Levels)			-0.23	(0.46)
Degree or higher			-0.26	(0.44)
Two-parent household			0.00	(.)
Lone-parent household			0.03	(0.17)
Word activity score <i>mean</i>			-0.07**	(0.03)
Constant	-2.07***	(0.07)	-1.11*	(0.47)

(continued)

Table A3 (continued): Log Odds from Multinomial Regression Models Predicting Person-Level Diary Quality; Alternative Definitions of Good Quality Diary (Less than 180 Minutes Missing Activity Time)

	Unadjusted		Adjusted	
	Log-odds	S.E.	Log-odds	S.E.
Good quality day 1 and 2	-		-	
Bad quality day 1 and 2				
App	0.00	(.)	0.00	(.)
Web	0.097	(0.13)	0.09	(0.13)
Male			0.00	(.)
Female			-0.56 ^{***}	(0.12)
White			0.00	(.)
Mixed			0.79 ^{**}	(0.25)
Indian			0.37	(0.33)
Pakistani/Bangladeshi			0.17	(0.30)
Black			0.89 [*]	(0.37)
No qualifications			0.00	(.)
NVQ 1 (Level 1/CSE)			-0.13	(0.45)
NVQ2 (O levels)			-0.21	(0.35)
NVQ3 (A Levels)			-0.37	(0.36)
Degree or higher			-0.92 ^{**}	(0.34)
Two-parent household			0.00	(.)
Lone-parent household			0.15	(0.15)
Word activity score <i>mean</i>			-0.14 ^{***}	(0.03)
Constant	-2.00 ^{***}	(0.07)	-0.17	(0.38)
Pseudo R-squared	0.003		0.031	

Notes: $N=3,209$. * $p<0.05$, ** $p<0.01$, *** $p<0.001$. Adjusted for complex survey design and non-response. 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 180 minutes of missing data in the activity diary column.