CENTRE FOR
TIME USE RESEARCH

UNIVERSITY OF
OXFORD

CHAPTER 1: OVERVIEW OF THE MTUS

1.1 Introduction

The origins of the Multinational Time Use Study (MTUS) go back to the 1970s following an initiative of Professor Jonathan Gershuny. The idea was to create a cross-nationally harmonised set of time use surveys composed of comparably recoded variables. A detailed discussion of the historical development of the MTUS appears in Chapter 2.

The MTUS archive, located at the Centre for Time Use Research in the Department of Sociology at the University of Oxford includes:

- **Original files and documentation**: Episode and aggregate files, not generally available for distribution except be arrangement with the data provider. In some cases, CTUR only holds an interim version of the original data or only the MTUS version of the data.
- **Harmonised core file (HCF)**: A harmonised aggregate file with all surveys from all countries except those surveys collected by the Australian Bureau of Statistics, Statistics Finland and Statistics Sweden. In this file, each row represents a 24-hour observation (diary). This file covers summary time in a simplified range of 25 time use activity categories. This dataset draws first from surveys harmonised to the current MTUS standards. We also have replaced data that was available only in the older versions of the MTUS, though surveys that have yet to be upgraded from the old to the new version of the MTUS have less complete information in this format. Full details of the conversion process are explained in the MTUS coding procedures document. Appendix 3 details variable coding issues related to files yet to be upgraded from older versions.
- **Harmonised aggregate files (HAF)**: Two harmonised aggregate files (one for adult diarists aged 18+ and a separate file for diarists aged less than 18) for a smaller sub-set of the surveys. Like the core file, these files cover summary time in a wider range of 69 activity categories, total time with a spouse or partner for diarists in couples, and a wider range of survey, household and person-level variables. Each row represents a 24-hour diary.
- **Harmonised episode file (HEF)**: Two harmonised episode files (one for adult diarists aged 18+ and a separate file for diarists aged less

than 18) for the smaller subset of the surveys included in the Harmonised aggregate files. These files include identifier variables, sex, age, and diary details: main activity, coded in the 69 category frame and an older 41 category code frame; secondary activity in the 69 category frame, location, mode of transport whether diarists used the internet or computers during the episode, and who else was present. In these files, rows represent a change in any element of the diary report.

- **Restricted files** for Australia, Finland and Sweden, are released separately and **require additional permission to access.**
- **Supplementary files** with region and ethnicity variables for a small selection of the surveys, additional variables for the UK, and special documentation for using the diaries from children, as well as a special collection of surveys from the USA (the American Heritage Time Use Study – AHTUS) also are available on the MTUS and CTUR websites.

Three variables have only harmonised column names, but original survey categories: EDUCA, INCORIG and EMPINCLM. The labels for the categories for each individual survey are included in the readme files for each survey. These variables enable users to construct their own customised variables. **NOTE THAT EDUCA, INCORIG and EMPINCLM ARE NOT SUITED TO CROSS-NATIONAL ANALYSIS UNTIL AFTER THE USER MAKES TRANSFORMATIONS REQUIRED FOR EACH ANALYTIC PURPOSE.**

We split the megafiles into child and adult files for two reasons. First, the age of the youngest diarists vary highly across the surveys. Also, as children's time use differs markedly from the activities of adults, excluding children from the main file increases the harmonisation of the adult files. Additionally, there are a number of surveys that only sampled young people. We hope in the future to include some of these surveys in the child files.

1.2 Format and structure of the datasets

The harmonised surveys are saved in **SPSS**, **STATA**, and **Flat text** formats. People who download the flat text files can access the variable and value labels in both English (Appendix 1) and Spanish (Appendix 2).

In most descriptive analyses, MTUS Users are encouraged to use all cases and disregard the fact that the total number of cases (diaries) correspond to a smaller number of respondents for some surveys. However, when carrying out analyses based on inferential statistics, MTUS users should be aware of the non-

independence of cases and should use appropriate statistical techniques for those surveys which collected multiple diaries from the same respondent.

The MTUS only includes cases where participants completed diaries. MTUS weights also permit users to analyse only good quality diary cases. Some original surveys also include information on partial participants and non-respondents, but to access this information, users would need to match back to original data. MTUS identifiers can readily be renamed to original survey identifiers to facilitate such matching.

The MTUS team has undertaken substantial data preparation and cleaning work, and produced added value variables not available in original surveys.

The time use diary is a narrative account and not a series of quantitative questionnaire answers. People can give a full account of their day without necessarily completing all columns of the diary for all potential time slots. For example, when offered a main activity and a mode of transport column, some diarists opt to make a time-saving minimal entry of writing “drove car” or “train” in the mode of transport column while not recording anything in the main activity column when their activity is transport. We consider entries across the whole diary to define the main activity, not just the information recorded in the main activity column.

We use full diary reports to add detail to some incomplete reports in three specific instances. In cases where the main activity is missing at the beginning or end of the diary day, where the diarist records being at home or at another person’s home, where the subsequent activities after the initial gap at the beginning or the preceding activities before the final gap at the end of the diary day are the sorts of activities which typically follow or precede sleep (have a drink, low-activity leisure like watch TV or read, dress/undress, personal care) we presume the missing activity to be sleep. (We make this assumption both on account of the pattern of activity and as time-diary surveys tend to start the observation window as a point of the day when most people in the population are asleep). Second, in circumstances of a missing period of 25 or fewer minutes that precedes leaving home to travel somewhere else or following returning home after travel from activities away from home, we assume the missing activity to include a combination of personal and household care. The documentation specific to each survey (conversion files) includes information on the number of cases amended by these procedures. Finally, where the diarist gives location information indicating that they are not at home but records no activity, we mark the case with the category unknown activity away from home on the newer 69 category activity code frame (though these cases are treated as missing time on the older 41 category activity list).

The MTUS team does not alter participant reports. Where an activity seems unusual, such as walking at home (which may occur on a treadmill), or food preparation while travelling (which a passenger might undertake in limited circumstances), we do not overwrite and change priority of the original account. We also **do not impute** data based on estimations from what similar people do on similar days. All adjustments to diaries work from information that the participant supplies.

After this data preparation work, we define any diary to be low quality, when:

- continues to have 91 or more minutes of missing time,
- it has fewer than 7 episodes,
- it is missing two or more of four basic activities, or
- it was filled in by a diarist whose age or sex is not known.

We define the four basic activities as those necessary for basic functioning on a day-to-day basis:

1) eating or drinking (measured by time in these activities, or time recorded working with food (set or clear table, food preparation, cooking and the like), or the diarist being in a location where they are likely to be around food and drink, that is attending a feast or being at a pub or in a restaurant);
2) sleep or rest (including do nothing, think, time out, or take a work break)
3) personal care (including assumed self care preceding or following travel and receiving personal services, such as at hair dresser or doctor)
4) exercise and/or travel (including leisure excursions, gardening, walk dogs, imputed travel where no activity is recorded but the diarist records a change of location or records a mode of transport).

Diaries with large volumes of missing time do not account for enough of the day to allow imputation of what is likely to have taken place in the missing periods.

Very low episode diaries and diaries missing basic activities do not give complete accounts of the day, and both lead to over-estimates of the activities the diarist did record and under-estimates of the activities the diarist did not record. Age and sex are two variables highly associated with specific time use patterns, and these variables are required for the creation of the recommended weight.

We make **five exceptions in relation to these quality rules**.

- Diarists may not record any travel when their travel episodes are very short, but may record a pattern in the diary that lets us know that they did travel and where in the day the travel took place. In such diaries, you will find patterns of continuous reports of activity, and a change of

location (such as eating breakfast at home then doing paid work at the office) with no report of travel in between the change of location. We handle these cases by making a flag variable for unreported travel present (0=no such missed travel, 1=missed travel). If the diary includes 2 of the 4 basic activities, one of the 2 missing activities is exercise or travel and the diary is flagged as including missing travel, then we count this diary as a good diary. We do not alter the diary record in such cases, and users of the MTUS would have to make their own adjustments to the entries made by the diarists if they wish to account for such travel.

- Diarists may not record any personal care when their episodes of personal care are short. In such diaries, you will find patterns of continuous reports of activity, and transitions where personal care is highly likely to have occurred (sleep for 2+ hours at home followed by other activities with no care, eating meals at home where the meal consumption lasts at least 10 minutes followed by other activities with no record of personal care). If such patterns are present, we make a flag variable for unreported personal care (0=no, 1=yes). If the diary is missing 2 basic activities, and one of these missing activities is personal care and we can flag this diary as having unreported personal care patterns, then we count this diary as a good diary. We do not alter the diary record in such cases, and users of the MTUS would have to make their own adjustments to the entries made by the diarists if they wish to account for such personal care.
- Diaries of carers (either the variable “carer” flagging cases of people who look after an adult needing assistance =1 for yes, or the diary includes any time in any form of adult or childcare) who otherwise meet the other 4 good diary criteria count as good diaries.
- Diaries including only 2 of the basic activities but that have at least 12 episodes where the diarist reports being at home all day (defined as no travel but eloc=1 – own home, or eloc=2 – other’s home for at least 1000 minutes), but otherwise meet the other 4 good diary criteria count as good diaries.
- Other diaries including only 2 of the basic activities and 15 or more episodes count as good diaries.

Some original surveys additionally include row cases for non-respondents who do not complete a diary. Nevertheless, most of the surveys do not include specific information on non-respondents in the data files. The MTUS format provides a suitable platform to analyse good-quality diaries as well as low-quality diaries, but users would need to take greater account of original survey information to investigate people who do not respond at all. When original

surveys include case rows for non-diarists (people with 24 hours of no reported activity), we delete the non-diary cases.

1.3 Surveys included

The following table lists all the surveys included in MTUS, as well as the surveys that we hope to include in coming years. Users should note that data included only in the harmonised corle file that are drawn from an old version of the MTUS but which have not yet been upgraded are not converted to the same standard as the rest of the MTUS and do not contain the same detail of information.

Table 1.1 – List of surveys included and versions available

Country	**Surveys and years**	**Versions available** **HEF - Harmonised episode file** **HAF - Harmonised aggregate file** **HCF - Harmonised core file**		
Australia	1974	HCF	HAF	HEF
	1987	HCF - drawn from old version		
	1992	HCF - drawn from old version		
	1997	HCF	HAF	HEF
	2006	HCF	HAF	HEF
Austria	1992	HCF	HAF	HEF
	2008-09	Hope to include in future		
Belgium	1965	Hope to include in future		
	1999	Hope to include in future		
Bulgaria	1965	Hope to include in future		
	1988	Hope to include in future		
	2001-02	Hope to include in future		
Canada	1971	HCF - drawn from old version		
	1981	HCF - drawn from old version		
	1986	HCF - drawn from old version		
	1992	HCF - drawn from old version		
	1998	HCF - drawn from old version		
	2005	Hope to include in future		
	2010	Hope to include in future		
Denmark	1964	HCF - drawn from old version		
	1975	Hope to include in future		
	1987	HCF - drawn from old version		
	2001	HCF		
	2008-09	Hope to include in future		
Estonia	1999-00	Hope to include in future		
Finland	1979	HCF	HAF	HEF
	1987-88	HCF - drawn from old version		

	1999-00	HCF - drawn from old version		
	2009-10	Hope to include in future		
France	1966	HCF - drawn from old version		
	1974-75	HCF - drawn from old version		
	1985-86	Hope to include in future		
	1998-99	HCF	HAF	HEF
	2009-10	Hope to include in future		
Germany	1965	HCF - drawn from old version		
	1991-92	HCF	HAF	HEF
	2001-02	HCF		
Hungary	1965	HCF - drawn from old version		
	1976-77	HCF - drawn from old version		
	1986-87	Hope to include in future		
	1999-00	Hope to include in future		
India	1998-99	Hope to include in future		
Ireland	2007-08	Hope to include in future		
Israel	1991-92	HCF	HAF	HEF
Italy	1979-80	HCF - drawn from old version		
	1989	HCF	HAF	HEF
	2002-03	HCF - drawn from old version		
Japan	1976	Hope to include in future		
	1981	Hope to include in future		
	1986	Hope to include in future		
	1991	Hope to include in future		
	1996	Hope to include in future		
	2001	Hope to include in future		
	2006	Hope to include in future		
Netherlands	1975	HCF	HAF	HEF
	1980	HCF	HAF	HEF
	1985	HCF	HAF	HEF
	1990	HCF	HAF	HEF
	1994	HCF	HAF	HEF
	2000	HCF	HAF	HEF
	2005	HCF	HAF	HEF
Norway	1971	HCF - drawn from old version		
	1981	HCF - drawn from old version		
	1990	HCF - drawn from old version		
	2000	HCF - drawn from old version		
Poland	1965	Hope to include in future		
	1975-76	Hope to include in future		
	1984	Hope to include in future		
	2001	Hope to include in future		

Portugal	1999	Hope to include in future		
Republic of Korea	1999	Hope to include in future		
	2004	Hope to include in future		
	2009	HCF		
Romania	2001	Hope to include in future		
Slovak Republic/ Czechoslovakia	1965	Hope to include in future		
	2006	Hope to include in future		
Slovenia / Yugoslavia	1965	HCF - drawn from old version		
	2000	HCF - drawn from old version		
South Africa	2000	HCF	HAF	HEF
	2010	Hope to include in future		
Spain	1992-93 (Basque)	HCF	HAF	HEF
	1997-98 (Basque)	HCF	HAF	HEF
	2002-03 (national)	HCF	HAF	HEF
	2002-03 (Basque)	HCF	HAF	HEF
	2008-09 (Basque)	HCF	HAF	HEF
	2009-10 (national)	HCF	HAF	HEF
Sweden	1991	HCF - drawn from old version		
	2000	HCF - drawn from old version		
	2010	Hope to include in future		
Turkey	2006	Hope to include in future		
United Kingdom	1961	HCF - drawn from old version		
	1974-75	HCF	HAF	HEF
	1983-84	HCF	HAF	HEF
	1987	HCF	HAF	HEF
	1995	HCF	HAF	HEF
	2000-01	HCF	HAF	HEF
	2005	HCF	HAF	HEF
USA	1965-66	HCF	HAF	HEF
	1975-76	HCF	HAF	HEF
	1985	HCF	HAF	HEF
	1992-94	HCF	HAF	HEF
	1994-95	HCF	HAF	HEF
	1998-2001	HCF	HAF	HEF
	2003-12	HCF	HAF	HEF
Totals: 23 countries	Total surveys*	HCF: **65***	HAF: **35^**	HEF:**35^**

* 74 surveys in total if one counts each of the American Time Use Survey years separately

^ 44 surveys in total if one counts each of the American Time Use Survey years separately

CENTRE FOR TIME USE RESEARCH

UNIVERSITY OF OXFORD

1.4 Technical information on the surveys

The table below contains key information on the sample size, age of respondents, response rate, etc. for each of the surveys included in the MTUS. These tables summarise the information in the top table of the readme files.

Table 1.2 Technical information on the time use surveys

Country[1]	Year	Age	Sample Size[2]	Survey Period (# months)[3]	Response rate (%)	Diary (# days)	Type of diary	Time interval	Household members[4]
AUS	1974	18+	1,491	7	67% A/W 58% Melb	1	On day	Free	No
	1987	15+	1,011	2	74.2%	2	On day	15min	Yes
	1992	15+	7,045	11	82.9%	2	On day	5 min	Yes
	1997	15+	7,246	8	72.0%	2	On day	5 min	Yes
	2006	15+	13,617	8	82.5%	2	On day	5 min	Yes
OST	1992	10+	25,233	2	47.0%	1	On day	30 min	Yes
	2008/09	10+	6,451	12	79.5%	1	On day	15 min	Yes
CAN	1971	18-64	2,141	8	72.0%	1	On day	Free	Yes
	1981	15+	2,686	3	46.0%	1	On day	Free	No
	1986	15+	9,946	3	78.9%	1	On day	Free	No
	1992	15+	9,815	12	77.0%	1	Recall	Free	No
	1998	15+	10,749	12	77.6%	1	Recall	Free	No
	2005	15+	19,957	12	59.0%	1	Recall	Free	No
	2010	15+	15,390	12	55.2%	1	Recall	Free	No
DEN	1964	15+	4,069	2	80.4%	1	Recall	30/15 min	In limited cases
	1987	16-74	3,584	3	72.7%	1	Recall	15 min	No
	2001	16-74	4,000	4	70.0%	2	On day	10 min	In limited cases
	2008/09	18-74	6,091	12	48.0%	2	On day	10 min	Yes
FIN	1979	10-64	12,038	4	81.0%	2	On day	10 min	No
	1987/88	15+	7,758	12	74.0%	2	On day	10 min	No
	1999/00	10+	10,561	12	52.0%	2	On day	10 min	Yes
	2009/10	10+		12		2	On day	10 min	Yes
FRA	1966	18-65	2,805	2	90.0%	1	On day	Free	Yes
	1974/75	18+	6,641	12	66.4%	1	On day	5 min	Yes
	1985/86	15+	16,047	12	66.9%	1	On day	5 min	Yes
	1998/99	15+	15,441	12	88.3%	1	On day	10 min	Yes
	2009/10	11+	27,903	15	88.3%	1 of 2	On day	10 min	1 + spouse
GER	1965	18-65	2,478	4	W73% E88%	1	On day	Free	Yes in West, no in East
	1991/2	12+	7,200	4	Quota	2	On day	5 min	Yes
	2001/2	10+	11,919	12	95.5%	3	On day	10 min	Yes
ISR	1991-92	14+	4,843	6	84.9%	1 (more limited cases)	Recall	15 / 30 min	In limited cases
ITA	1988/9	3+	38,110	12	70.0%	1	On day	Free	Yes
	2002/3	3+	55,773	12	91.8%	1	On day	10 min	Yes

CENTRE FOR TIME USE RESEARCH

UNIVERSITY OF OXFORD

Country[1]	Year	Age	Sample Size[2]	Survey Period (# months)[3]	Response rate (%)	Diary (# days)	Type of diary	Time interval	Household members[4]
NET	1975	12+	1,309	1	76.0%	7	On day	15 min	No
	1980	12+	2,730	1	54.0%	7	On day	15 min	No
	1985	12+	3,263	1	54.0%	7	On day	15 min	No
	1990	12+	3,158	1	49.0%	7	On day	15 min	No
	1995	12+	3,227	1	20.0%	7	On day	15 min	No
	2000	11+	1,813	1	25.0%	7	On day	15 min	No
	2005	12+	2,204	1	37.0%	7	On day	15 min	No
NOR	1971/2	16-74	3,040	12	58.0%	2 & 3	On day	15 min	No
	1980/1	16-74	3,307	12	65.0%	2	On day	15 min	No
	1990/1	16-79	3,097	12	64.0%	2	On day	15 min	No
	2000/1	9+	3,211	12	50.0%	2	On day	10 min	Yes
SPA	1992/3 b	16+	5,040	6	73.0%	1	On day	5 min	No
	1997/8 b	16+	5,023	6	Missing	1	On day	5 min	No
	2002/3 b	10+	5,039	6	64.0%	1	On day	5 min	No
	2002/3 n	10+	46,774	12	86.0%	1	On day	10 min	Yes
	2008/9 b	10+	6,746	12	73.5%	1	On day	5 min	No
	2009/10 n	10+	19,295	12	58%	1	On day	10 min	Yes
SLO	2000/1	10+	4,500	12	52.5%	2	On day	10 min	Yes
RSA[5]	2000	10+	14,553	3	94.0%	1	Recall	30/10-15 min	Yes
SWE	1990/1	20-64	3,943	9	75.0%	2	On day	10 min	No
	2000/01	20-99	3,976	12	50%	2	On day	10 min	No
UK[6]	1961	15+	2,363	1	69.8%	7	On day	30 min	Yes
	1974/75	5+	3,583	4	60.0%	7	On day	30 min	Yes
	1983/84	14+	1,525	2	51.0%	7	On day	15 min	Yes
	1987	16+	3,035	1	70.0%	7	On day	15 min	Yes
	1995	16+	1,875	1	93.0%	1	Recall	15 min	No
	2000/1	8+	11,667	15	45.0%	2	On day	10 min	Yes
	2005	16+	4,941	10	59%	1	Recall	10 min	No
USA	1965	18-64	1,243	7	74.0%	1	On day	Free	Yes
	1975/76	18+	2,406	3	72.0%	1	On day	Free	No
	1985	12+	5,358	12	55.2%	1	On day + recall	Free	Yes
	1992/4	0+	9,386	12	63.0%	1	Recall	Free	No
	1994/5	18+	1,199	13	64.6%	1	Recall	Free	No
	1998/1	18+	1,700	12	56.0%	1	Recall	Free	No
	2003-12	15+	136,960	132	52-57%	1	Recall	Free	No

Notes:

1- More countries have carried out time use surveys. A more comprehensive list is available at the CTUR web site.

2- Unless otherwise indicated, the sample size refers to the number of individuals. The actual number of cases is larger in surveys where 2 or 3 diaries were collected.

3- 'Period' refers to different collection periods throughout the year.

4- Indicates whether or not more than 1 household member was included in the survey.

5- The South African diary includes 30 minute intervals. People could nominate multiple activities, and if they nominated more than 1 activity, were asked in the activities were simultaneous or consecutive. The original file codes multiple consecutive activities in 10 as well as 15 minute intervals.

6- The UK 1974-75 survey collected 4 rounds of 1 week diaries. Only the Monday and Tuesday diaries remain for the collection wave that took place in September 1974. The 1983-84 and 1987 surveys are treated as a single survey in the recent versions, though the variable msamp allows the user to distinguish between the two surveys.

1.5 File naming conventions

We have standardised MTUS file names. The name of each file distinguishes:

- The country (2 or 3-letter code) (see table below)
- The first year in which the survey started data collection (4-digit)
- The version of the archive (HEF, HAF, HCF)
- The type of file (extensions 'sav' or 'dta' for data files, and extensions 'sps' or 'do' for programme files)

For example, Release 2 of the HEF version of Spain 2009-2010 is called 'es2009hef.sav', which should be read as:

Country: es (for Spain)
Year: 2009 (the first year in which data collection took place)
Version: hef (harmonised episode file)
Type: sav (an SPSS file)

Country	Code	Country	Code	Country	Code
Albania	AL	Hungary	HU	Poland	PL
Algeria	DZ	India	IN	Portugal	PT
Armenia	AM	Indonesia	ID	Qatar	QA
Australia	AU	Ireland	IE	Republic of Korea	KR
Austria	AT	Israel	IL	Romania	RO
Belgium	BE	Italy	IT	Russian Federation	RU
Bosnia & Herzegovina	BA	Japan	JP	Serbia	RS
Brazil	BR	Laos	LA	Slovenia	SI
Bulgaria	BG	Latvia	LV	South Africa	ZA
Canada	CA	Lithuania	LT	Spain	ES
Chile	CL	Macedonia	MK	Sweden	SE
China	CN	Mauritius	MU	Switzerland	CH
Czechoslovakia	CZ	México	MX	Tanzania	TZ
Denmark	DK	Mongolia	MN	Thailand	TH
Djibouti	DJ	Morocco	MA	Tunisia	TN
Estonia	EE	Netherlands	NL	Turkey	TR
Ethiopia	ET	New Zealand	NZ	United Kingdom	UK
Finland	FI	Norway	NO	United States	US
France	FR	Oman	OM	Uruguay	UY
Germany	DE	Pakistan	PK	Yugoslavia	YU

Ghana	GH	Palestine	PS		
Greece	GR	Peru	PE		

1.6 Missing value conventions

We use three codes to mark missing values, and a separate fourth convention for weights and identifier variables that are not present, as follows:

- "-7" refers to situations for we can create a variable for this survey, but we cannot create the variable for this diarist (or diary) as the respondent was not asked for the information on this diary or because the information is not relevant to that respondent (such as the employment status of a spouse for a person who is single and not living with a co-habiting partner). Although this missing value option potentially applies to all variables, it is mainly used for AGEKIDX, AGEKID2, WORKHRS, EMPSP, PARNTID1, PARNTID2, PARTID and EMPINCLM.
- "-8" refers to situations where we can create the harmonised variable for the study, but no information is recorded for this case (item non-response or insufficient information to create the variable for that case).
- "-9" refers to situations for which the harmonised variable could not be computed for the survey (with exceptions for weights and case identifier variables – although the identifier of spouse or of parents can have a -8 value if this could not be created for a case). Note that we use -9 with the time use activity variables to distinguish true 0s (the diarist did not record any time in this activity, though in theory they could have done so) from cases where no time is recorded in the activity because we could not create this time use category for this survey.

There are cases where an original weight is not present. In these cases, we use "0" rather than a missing value to indicate that this weight is not present in the study (and anyone attempting to use this weight would find they have no cases remaining for analysis from the survey). The conventions for the identifiers are set out in detail below.

Users also should note that we do not use missing values for the aggregated or summary time use variables, unless the category is not available for the whole survey. A value of 0 means that the diarist did not record any minutes in the activity (it is impossible to say for certain if this is because the diarist did not do any of the activity or if the diarist actually did undertake the activity but did not report doing the activity in the diary). If a category is not coded in the survey, then the summary value is set to -9 for the whole survey. Users should take notice of -9 values. If one sums time across a variable that cannot be created for

a survey without first addressing the missing categories, 9 minutes will be subtracted in error for each category that is not present.

There are no system missing cases in MTUS data files. All cases for all variables have either a valid value or a standardised missing value. The MTUS data files contain no declared missing values. MTUS users need to declare missing values if they choose to do so before running their analysis.

1.7 Data preparation

The Harmonised Aggregate Files (HAF) and the Harmonised Core File (HCF) offer **aggregate (summary)** versions of the time-use surveys, where each row case in the dataset reflects a record in one 24-hour time diary. For those studies where respondents completed more than one diary, individual diarists appear on a separate row for each diary they completed. The HAF and HCF include survey, demographic and socioeconomic information about respondents (hereafter called diarists) and their households alongside the aggregated time-use variables.

The Harmonised Episode Files (HEF) cover sequence information. In the HEF, each row case represents one episode, or change of main activity, secondary activity, location, use of computers or the internet, mode of transport, or presence of others, in a diary. Where the diarist completed more than one diary, the episodes of the subsequent diary or diaries follow the episodes of the first diary. As the HEF files are large, only the identifiers, age and sex are included in the HEF alongside the diary information. Users will need to match the HEF with the HAF file to pick up the corresponding background variables.

The process of making the HAF and HEF files takes around three to five weeks of cumbersome work, and can take longer in the case of older time use surveys where information is reported in uneven intervals and more considerable efforts are required to resolve errors in original files. For this reason, only a limited number of surveys will be coded into the HAF and HEF formats. More surveys will be included only in the core HCF format.

Before beginning the actual conversion, users should undertake three steps to ensure maximum data quality.

DATA PREPARATION STEP 1 – Check alternative options for the MTUS background variables to ensure that you are using the option with the cleanest profile compared with other reported results and the least missing data. If there are options and one is better than others, the choice should be documented in the conversion syntax and the Readme file. In some cases, combinations of

original variables are needed to create the MTUS variables. We also triangulate information in files to use any available information to fill gaps, not to impute data, but to make maximum use of the information coded in the file. As an example, a diarist may have no answer recorded for the question of marital status. Nevertheless, a household grid may show that this person is the spouse of another diarist, and this person may report time with this spouse in the diary, and this diarist's diary may match patterns the spouse reports being with this diarist, indicating that the person is in a couple even though the couple variable has a missing value.

DATA PREPARATION STEP 2 – Apparently missing main activity time in diaries is not necessarily missing. The point of the diary is to collect information about what people are doing at any point in time. Diarists sometimes do not fill in the main activity column – creating the appearance of missing data, but fill in other information elsewhere in the diary that nonetheless indicates their activity and allows us to properly code the time period. We should recognise that elements of the diary are not always separate. At points of overlap, diarists can record a comprehensible and clear response in the diary in one place but not in others. For example, an entry “took train to work” is both a location/mode of transport and an activity, and this dual meaning is clear even if the entry is written only once in either the main activity column or the location column. The redundancy of writing the same entry in two places is not necessary for the diary to have a full account of a participant’s activities. We recommend the following steps be undertaken where a main activity is missing before converting the data:

a) Completing a time diary can be an onerous task, and some diarists do not appreciate making redundant entries. Where diaries have a location or mode of transport column and the diarist is travelling, some diarists may write “drive my car to work” or “on the bus” in the mode of transport column and see no point in writing the same entry in the activity column. When main activity is missing but the diarist has recorded a mode of transport during this time period, we recode the missing main activity slot as unspecified travel (main=62).
b) Some diarists get confused while they complete the diary in a hurry, and may record the main activity in the secondary activity column. Another possibility is that a diarist may be undertaking a series of main activities while also doing an extended secondary activity – for instance alternating between care of pets, care of children and housework as main activities while listening to the radio. An item on the radio may be particularly interesting and attract the diarist’s full attention for 10 minutes, but the 10 minutes of main activity radio listening is more efficiently recorded by simply extending the radio listening recorded in the secondary activity column. Where main activity is missing but a valid secondary activity is recorded, we recode the main activity as the reported secondary activity, and recode the secondary activity as no

reported second activity. These are cases where the diarist has reported one valid activity.

c) For short gaps in the early hours at the beginning or end of the diary where the diarist is at home or in the same location where they report sleeping on the diary day and asleep before the gap at the end of the diary, or asleep following the gap at the beginning of the diary, we recode the gap as imputed sleep.

d) If a short gap (<25 minutes) occurs at home just before travel or at home just after travel home, set the missing time to imputed personal and household care

e) If there is other diary information that illuminates the activity in an episode where there is no recorded main activity, this should be used to identify the activity. As some examples, individual surveys in the past have recorded information as the number of cigarettes smoked during the episode, which television station the diarist watched if they watched TV during the episode, which type of material the diarist read if they read during the episode, and the like. Similar to the instance of the diarist recording a mode of transport but not recording a main activity, a diarist might record that they smoked 10 cigarettes in 15 minutes or watched a specific television station for 45 minutes but not record a main activity. Nonetheless, such records do reveal what the diarist was doing, so can be used to complete apparently missing time episodes.

All these changes are making use of information the diarist supplied about their activity, and this procedure eliminates some unnecessary wastage of diaries. All such data cleaning should be fully documented in the conversion programme.

DATA PREPARATION STEP 3 – Check to see if other diary information facilitates the coding of time use activities. Different studies code activities in different ways. Sometimes researchers need to use multiple columns from the diary to code a single activity. For instance, some surveys simply code “eating/drinking” in the main activity, and the location variable is needed to distinguish meal breaks at work (Main=5), from eating out in a restaurant (Main=39), from eating meals at home or elsewhere (Main=6). Likewise, location can distinguish paid work at home (Main=8) from paid work away from home (Main=7). Who else is present information sometimes is needed to distinguish childcare from adult care. Some cases arise peculiar to only one dataset. For instance, Denmark 1964 includes an original code for all media use, but also has a separate column where diarists indicated what media they were reading, watching, or listening to, and this second column enables the separate coding of listening to the radio (Main=58) from watching TV (Main=59) to listening to music (Main=57). All combinations of information used to code a category should be included in the documentation.

We engage in a series of quality checks in the post-harmonisation process. These quality checks are detailed in Section 11 (Quality checks) of the MTUS coding procedures document available elsewhere on the MTUS web pages.

1.8 Identifying good-quality diaries

Diaries with large volumes of missing time do not account for enough of the day to allow imputation of what is likely to have taken place in the missing periods. Low episode diaries and diaries missing basic activities do not give complete accounts of the day. Low quality diaries lead to over-estimates of the activities the diarist does record and under-estimates of the activities the diarist did not record. Age, sex and day of the week are highly associated with specific time use patterns, and these variables are required for the creation of the recommended weight. We also classify diaries missing age or sex of the diarist or the day of the week on which the diary was completed as low quality diaries.

For this reason, the MTUS includes the variable BADCASE to distinguish good quality diaries from diaries lacking sufficient standards for analysis. The MTUS team defines any diary which:

- continues to have 91 or more minutes of missing time after data cleaning,
- which has fewer than 7 episodes,
- which is missing two or more of four basic activities - with 5 exceptions (defined below)

The four basic activities necessary for basic day-to-day functioning are:

- eating or drinking (measured by time in these activities, or time recorded working with food (set or clear table, food preparation, cooking and the like), or the diarist being in a location where they are likely to be around food and drink, that is attending a feast or being at a pub or in a restaurant);
- sleep or rest (including do nothing, think, time out, or take a work break)
- personal care (including assumed self care preceding or following travel and receiving personal services, such as at hair dresser or doctor)
- exercise and/or travel (including leisure excursions, gardening, walk dogs, imputed travel where no activity is recorded but the diarist records a change of location or records a mode of transport).

The five exceptions where MTUS does not count a diary as being low quality if the only problematic dimension is missing two or more of four basic domains of activities are as follows:

- Diarists may not record any travel when their travel episodes are very short, but may record a pattern in the diary that lets us know that they did travel and where in the day the travel took place. In such diaries, you will find patterns of continuous reports of activity, and a change of location (such as eating breakfast at home then doing paid work at the office) with no report of travel in between the change of location. We handle these cases by making a flag variable for unreported travel present (0=no such missed travel, 1=missed travel). If the diary includes 2 of the 4 basic activities, one of the 2 missing activities is exercise or travel and the diary is flagged as including missing travel, then we count this diary as a good diary. We do not alter the diary record in such cases, and users of the MTUS would have to make their own adjustments to the entries made by the diarists if they wish to account for such travel.
- Diarists may not record any personal care when their episodes of personal care are short. In such diaries, you will find patterns of continuous reports of activity, and transitions where personal care is highly likely to have occurred (sleep for 2+ hours at home followed by other activities with no care, eating meals at home where the meal consumption lasts at least 10 minutes followed by other activities with no record of personal care). If such patterns are present, we make a flag variable for unreported personal care (0=no, 1=yes). If the diary is missing 2 basic activities, and one of these missing activities is personal care and we can flag this diary as having unreported personal care patterns, then we count this diary as a good diary. We do not alter the diary record in such cases, and users of the MTUS would have to make their own adjustments to the entries made by the diarists if they wish to account for such personal care.
- Diaries of carers (either the variable “carer” flagging cases of people who look after an adult needing assistance =1 for yes, or the diary includes any time in any form of adult or childcare) who otherwise meet the other 4 good diary criteria count as good diaries.
- Diaries including only 2 of the basic activities but that have at least 12 episodes where the diarist reports being at home all day (defined as no travel but eloc=1 – own home, or eloc=2 – other’s home for at least 1000 minutes), but otherwise meet the other 4 good diary criteria count as good diaries.
- Other diaries including only 2 of the basic activities and 15 or more episodes count as good diaries.

- which was filled in by a diarist whose age or sex is not known,
- the day of the week on which the diary was completed is not know

to be low-quality.

Note that only good-quality diaries have positive values in PROPWT. Low-quality diaries should have 0 values on PROPWT.

Some original surveys additionally include row cases for non-respondents who do not complete a diary. Nevertheless, most of the surveys do not include specific information on non-respondents in the data files. The MTUS format provides a suitable platform to analyse good-quality diaries as well as low-quality diaries, but users would need to take greater account of original survey information to investigate people who do not respond at all. Where original surveys include case rows for non-diarists (people with 24 hours of no reported activity), we delete the non-diary cases.