

Introduction to quantitative time diary analysis

2. Structure and design of time diary data

Pierre Walthéry

10 October 2023

UK Data Service, University of Manchester

Centre for Time Use Research, University College London



Some basic vocabulary

Activity Action recorded by respondents in the time diary
ie what the respondent was doing during an episode

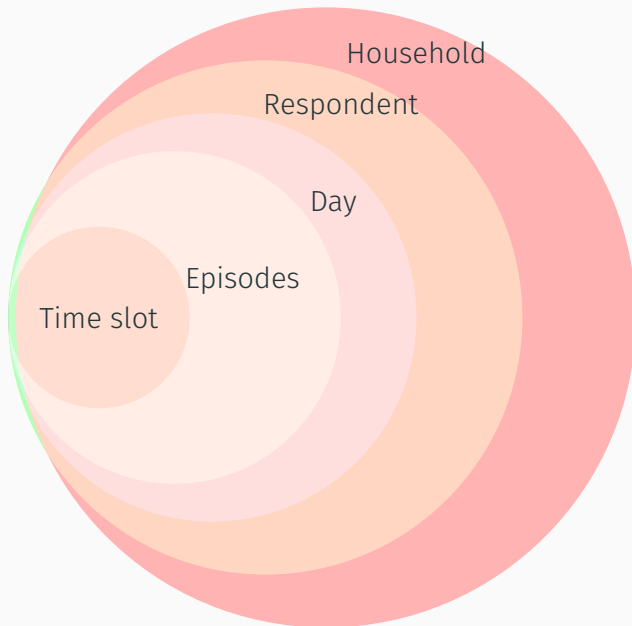
- Example: I am teaching time diary analysis
- Multitasking: Primary vs secondary

Episode Any unique combination of primary, secondary activity, copresence and location

- Episode 1. I watch TV whilst eating crisps alone at home;
- Episode 2. I watch TV whilst eating crisps with my son at home
- Episode 3. I play video games whilst eating crisps with my son at home

Time slot Minimum duration of an episode (ie resolution of the time diary). Usually 10 or 15 mins

Typical data structure in time use surveys



Data structure and files

1. Individual files ie person-level data
 - Socio-demographic characteristics ie age, education...
2. Day-level files (aka 'aggregate' files)
 - Each line in the dataset records a day
 - Often, 2 or more lines per respondent
 - Aggregate variables (ie time spent on activities)
 - Also sometimes include day level variables: *Is this a rushed day?*
 - May also include time diary data in wide format

Data structure in the *long* format – 1

3. Episode-level files in long format

- Each line in the dataset records an *episode*
- On average 15 episodes per day per person
- Episodes are embedded within days within persons
- One of the most common time diary format
- Intuitive, but requires more storage space and computing power than wide format
- Requires episode number, duration, start and end time

Ep nr	Person nr	Day nr	Duration	Activity	Start	End
1	1	1	360	Sleep	10	370
2	1	1	20	Shower	370	390
3	1	1	30	Breakfast	390	420

Data structure in the *long* format – 2

4. Slot-level data

- Each line records the time unit of the diary
- Largest datasets
- 144 observations (ie 10 minute slots) per day
- Requires a slot identifier variables

Slot nr	Person nr	Day nr	Activity
1	1	1	Sleep
...
42	1	1	Shower
43	1	1	Shower
44	1	1	Breakfast
...
144	1	1	Sleep

- In many time use surveys, days begin at 4 AM and end at 3.59 AM

Data structure in wide format

- During this course, we will be working with episode files in the long format
- However, episode or time slots datasets are also available in wide format:
- Example of a time slots dataset in the wide format:

Day	Pers.	Activity – TS 1	...	Act. – TS 42	...	Act. – TS 144
1	1	Sleep	...	Sleep	...	Watch TV
2	1	Sleep	...	Commute	...	Sleep
1	2	Sleep	...	Shower	...	Sleep

- Each line represents a day, each time slot and its attributes (activities, location, copresence are recorded in one variable)
- Lots of variable, less intuitive but more parsimonious than the long format

Common time diary variables

- Most time diary datasets include similar variables:
 - Primary and secondary activities
 - If episode files: duration
 - Incremental time/clock time
 - Copresence, ie 'with who were you?'
 - Location
 - Day of the week, month, calendar date
 - Diary number (ie whether this diary was the first one collected)
- Less common variables/recent innovations
 - Whether feels rushed on the day
 - Enjoyment
 - Device use
 - Work schedule

Nomenclatures of activities

- In many time use surveys, respondents still fill in their time diary by hand
- Activities are written down literally (ie *I walked the dog around the block*) and need to be recoded
- Indispensable to standardise activities
- National and international norms and guidelines for coding activities
- Three main international 'Western' nomenclatures: MTUS, HETUS, ICATUS (UN-based)
- National nomenclatures: US (ie ATUS), Japanese, Korean, Indian Time Use surveys have their own

Annex IV / 1.1 Main and secondary activities

- 0** **PERSONAL CARE**
- 01** **SLEEPING**
 - 011** Sleeping
 - 012** Sick in bed
- 02** **EATING**
 - 021** Eating
- 03** **OTHER PERSONAL CARE**
 - 031** Washing and dressing
 - 032** Personal care services
 - 039** Other or unspecified personal care

Indian Time Use Survey nomenclature

1. ECONOMIC ACTIVITIES IN THE PRIMARY SECTOR (AGRICULTURE, HUNTING, FORESTRY AND FISHING, MINING AND QUARRYING)

— CROP AND ANIMAL PRODUCTION, HUNTING AND RELATED SERVICE ACTIVITIES (NIC'08 DIV-01)

-- CULTIVATION OF CROPS IN THE FIELD OTHER THAN PLANTATION, HORTICULTURE AND FLORICULTURE

Abridged Classification

Land preparation and planting in the field

- 100. Ploughing/tilling, preparing land for cultivation
- 101. Sowing, planting and transplanting, in the field
- 109. Land preparation and planting activities in the field n.e.c.

Raising crops/cultivation in the field

- 110. Applying fertilizer, manure etc in the field
- 111. Applying pesticides/insecticides/rodenticides etc. in the field
- 112. Watering of plants/irrigation operations in the field
- 113. Weeding, cutting of undergrowths/shrubs etc. in the field
- 114. Grass cutting, tree/plant pruning in the field
- 115. Plant propagation activities
- 119. Crop raising/cultivation activities n.e.c.

Harvesting and reaping in the field

- 120. Harvesting and related works
- 121. Threshing and winnowing of crops and related works, including collection of stalks and stems

American Time Use Survey nomenclature

01 Personal Care

Category definition: This category captures time the respondent spends sleeping, grooming, providing self-care, and doing sexual activities. Respondents are not asked WHO they were with or WHERE they were for sleeping (0101xx), grooming (0102xx), and personal (sexual) activities (0104), as such information is sensitive for these activities. (Modified for clarity: Apr2006)

1. Exercising for medical conditions: If it is clear from the verbatim that physical therapy exercising is done specifically for medical reasons, code as health-related self care (010301). If the purpose is not clear, code as Sports, Exercise, and Recreation/Participating in Sports, Exercise, or Recreation/specific sport] (1301xx). For example, code weightlifting exercises as weightlifting. (Updated: Feb2004)
2. Insomnia: Code as Sleeplessness (010102). **Exception**: If the respondent mentions thinking or worrying, code as Socializing, Relaxing, and Leisure/Relaxing and Leisure/Relaxing, thinking (120301).
3. Kissing hello or goodbye: Code as Socializing, Relaxing, and Leisure/Socializing and Communicating/Socializing and communicating with others (120101), not as Personal Care/Personal Activities.
4. Meditating: Code as Personal Care/Health-related Self Care/Health-related self care (010301). **Exception**: If verbatim specifically mentions Meditating for religious purposes, code as Religious and Spiritual Activities/Religious and Spiritual Practices/Participation in religious practices (140102).

MTUS nomenclature

Activity code	Description
MAIN/SEC 1	Imputed personal or household care
MAIN/SEC 2	Sleep and naps
MAIN/SEC 3	Imputed sleep
MAIN/SEC 4	Wash, dress, care for self
MAIN/SEC 5	Meals at work or school
MAIN/SEC 6	Other meals or snacks
MAIN/SEC 7	Paid work - main job (not at home)
MAIN/SEC 8	Paid work at home
MAIN/SEC 9	Second or other job not at home
MAIN/SEC 10	Unpaid work to generate household income
MAIN/SEC 11	Travel as a part of work
MAIN/SEC 12	Work breaks
MAIN/SEC 13	Other time at workplace
MAIN/SEC 14	Look for work
MAIN/SEC 15	Regular schooling, education
MAIN/SEC 16	Homework
MAIN/SEC 17	Leisure/other education or training
MAIN/SEC 18	Food preparation, cooking
MAIN/SEC 19	Set table, wash/put away dishes
MAIN/SEC 20	Cleaning
MAIN/SEC 21	Laundry, ironing, clothing repair
MAIN/SEC 22	Home/vehicle maintenance/improvement
MAIN/SEC 23	Other domestic work
MAIN/SEC 24	Purchase goods

Workflow: producing time use estimates from time diaries

- One can:
- ... work directly with aggregate variables in day-level datasets ('Aggregate files' in MTUS jargon)
- ... compute one's own, which usually entails four steps:
 1. Recoding original time use activities in the episode dataset
 2. Sum the time spent on these activities (for example total daily time) within day, person
 3. Merge the data with person-level information of interest
 4. Produce the estimate of interest (ie, mean, median, etc)