

CHAPTER 3: DEMOGRAPHIC AND SOCIO-ECONOMIC VARIABLES

3.1 List of variables

3.2 Coding of the demographic variables

General information about surveys and cases

Basic demographic information

Household structure and presence of children

Employment status, income and education

Other harmonised information

In addition to the time-use variables, the MTUS datasets contain a set of harmonized demographic and socio-economic variables. Several of the original datasets contain a larger number of such variables. In the context of MTUS, priority had to be given to variables that were available in all (or most of) the datasets. The list of these variables appears below, while details on how each of these variables was constructed may be found in the survey-specific documentation (the README documents).

3.1 LIST OF VARIABLES

Demographic and socio-economic variables included in World 5.5

- COUNTRY: Country where the study was conducted
- PERIOD: Time survey period
- SURVEY: Survey year
- SWAVE: Longitudinal study survey marker
- MSAMP: Multiple samples using same diary instrument
- ID: Case Identifier
- PERSID: Person level identifier
- HLDID: Household identifier
- DAY: Day of interview
- MONTH: Month the diary was kept
- YEAR: Year diary kept
- DIARY: Diary day
- SEX: Sex
- AGE2: Age
- AGEGR5Y: Five-year age groups
- CIVSTAT: Civic status
- COHAB: Respondent is cohabiting
- FAMSTAT: Individual level family status

- CPHOME: Unmarried child living in parental home
- HHTYPE: Household type
- SINGPAR: Whether diarist is a single parent
- HHLDSIZE: Number of people in household
- NCHILD: Number of children under 18 in household
- AGEKID: Age of the youngest child in household
- RELREFFP: Relation to household reference person
- EMPSTAT3: New employment status
- EMP: In paid work
- UNEMP: Unemployed
- STUDENT: Student status
- RETIRED: Retirement status
- EMPSP: Employment status of the spouse/partner
- WORKHRS: Hours of paid work last week including overtime
- INCORIG: Original household income
- INCOME: Total household income - grouped
- EMPINCLM: Original monthly income from employment or self-employment
- EDUCA: Educational level - original study code
- EDTRY: Harmonised level of education
- DISAB: Diarist has a disability or long-term limiting health condition
- URBAN: Urban or rural household

Note that a series of weights are also included. They are described in Chapter 4.

3.2 CODING OF THE DEMOGRAPHIC AND SOCIO-ECONOMIC VARIABLES

Information is given below on the coding of each demographic and socio-economic variable. Information on how the variables were originally coded in the surveys and how they were recoded in these harmonized variables may be found in the survey-specific documentation (the README documents).

The harmonised background variables cluster into the following sets:

- General information about the surveys and cases
- Basic demographic information
- Household structure, presence of children, and marital status
- Employment status, income, and education
- Other harmonised information

Note that the actual order of the variables in World5.5 varies from the one presented here.

General information about the surveys and cases

- COUNTRY
- PERIOD
- SURVEY
- ID
- PERSID
- HLDID
- DAY
- MONTH
- YEAR
- DIARY

COUNTRY: Country where the study was conducted

This variable records the country where the survey was carried out. Note that there are separate codes for West Germany, East Germany, and reunified Germany.

Value	Label	Value	Label
1	Canada	18	Italy
2	Denmark	19	Australia
3	France	20	Israel
4	Netherlands	21	Sweden
5	Norway	22	Germany
6	UK	23	Austria
7	USA	24	South Africa
8	Hungary	25	Brazil
9	West Germany	26	Estonia
10	Poland	27	India
11	Belgium	28	Japan
12	Bulgaria	29	Korea (South)
13	Czechoslovakia	30	Mexico
14	East Germany	31	New Zealand
15	Peru	32	Portugal
16	Yugoslavia	33	Romania
17	Finland	34	Slovenia

PERIOD: Time survey period

This variable records the period during which the survey was carried out. The values range from 1961-69 (value '1') to 2000-04 (value '7'). Precise information on when the survey was carried out is recorded in the variable 'Survey'.

Note that the length of each period is not equal. Cut-off points were chosen to maximise the number of countries in each period and to ensure that there was only 1 survey per period for any specific country. In cases for which multiple surveys were available for a country during a specific period, only one of these surveys has been included in the MTUS dataset.

Value	Label
1	1961 - 1969
2	1970 - 1975
3	1976 - 1984
4	1985 - 1989
5	1990 - 1994
6	1995 - 1999
7	2000 - 2004
8	2005 – 2009
9	2010 – 2014

SURVEY: Year the survey began

This variable records the year during which the survey was carried out. Note that in an earlier version of the harmonised dataset, the survey year was recorded using a 2-digit code (e.g. 66 instead of 1966). We now use 4-digit codes.

Note that the data collection in some surveys spanned more than one year. In such cases, the variable SURVEY takes the value of the beginning year (with exceptions described in the relevant README documents).

SWAVE: Longitudinal study wave marker

This variable is relevant only for surveys that are longitudinal.

Value	Label
0	Not longitudinal
1	Wave 1
2	Wave 2
3	Wave 3
4	Wave 4

Note that in the case of Denmark 1987/2001 (with multiple samples), the code '1' indicates a longitudinal case, while the code '0' indicates that that it is not a longitudinal case.

MSAMP: Multiple samples using the same diary instrument

Value	Label
0	One sample
1	Szalai USA 1965 sample
2	National USA 1965 sample
3	Original NHAPS
4	1995 NHAPS supplement

ID: Case identifier

This variable records the diary case ID. This variable is not harmonised across surveys. In most surveys, we kept the original ID assigned by the institution having carried out the survey. In some surveys, we computed values from various survey and demographic characteristics of the respondent in order to obtain one unique ID per case. Note that in multiple-day diaries, the ID (being a case identifier) takes a different value than PERSID. In contrast, in one-day diaries, PERSID = ID.

PERSID: Person level identifier

This variable records the respondents' identifier.

HLDID: Household identifier

This variable uniquely identifies households for those studies where more than one household member completed a diary. For surveys in which only one person per household completed a diary and no household identifier is included in the original data, HLDID=0. For surveys in which only one person per household completed a diary but a household identifier is included, HLDID takes the original value for the corresponding variable.

DAY: Day of interview

This variable records the day of the week when the diary was kept. Note that in the current version of World5.5, we still have some surveys that appear as 7-day averages rather than as single-day cases. We intend to change the data to individual day data in future versions of the MTUS.

Value	Label
1	Sunday
2	Monday
3	Tuesday
4	Wednesday
5	Thursday
6	Friday
7	Saturday
8	Whole week average
9	Weekday

MONTH: Month the diary was kept

This variable records the month when the diary was kept. Some surveys only record the quarter or season when the survey was carried out. In such cases, we assumed that the survey was conducted during the first month of the quarter even though the data collection was actually spread throughout the quarter. Such cases are documented in the relevant README documents.

Value	Label
1	January
2	February
3	March
4	April
5	May
6	June
7	July
8	August
9	September
10	October
11	November
12	December

YEAR: Year diary kept

This variable is a supplement to the variable SURVEY and is especially useful for surveys whose data collection spanned more than one year. The variable YEAR records the year when the diary was kept.

DIARY: Diary day

This variable records the number of the diary (e.g. 1st completed diary, 2nd completed, 3rd completed). In the case of single-day diaries, this variable takes the value 1. In the case of multiple-day diaries, this variable can take the value 1, 2 or 3.

Value	Label
1	First diary day
2	Second diary day
3	Third diary day
4	Fourth diary day
5	Fifth diary day
6	Sixth diary day
7	Seventh diary day
8	Weekly average

Basic demographic information

- SEX
- AGE2
- AGEGR5Y
- CIVSTAT
- COHAB

SEX: Sex

Value	Label
1	Man
2	Woman

AGE2: Age

This variable records the age of respondents (up to 3 digits). For surveys in which age was recorded in categories, we recoded age into a continuous variable by assigning the mid-point of each age group (e.g. 17 for age group 15-19). When surveys only included the year of birth of respondents, we computed AGE2 by subtracting the year of birth from the year of the survey.

AGEGR5Y: Five-year age groups

This variable records the age of respondent in 5-year age group categories. This variable is derived from AGE2.

Value	Label
1	0-4
2	5-9
3	10-14
4	15-19
5	20-24
6	25-29
7	30-34
8	35-39
9	40-44
10	45-49
11	50-54
12	55-59
13	60-64
14	65-69
15	70-74
16	75-79
17	80+

CIVSTAT: Civic status

This variable records the diarist's marital status.

Value	Label
1	Married or cohabiting
2	Not living with a spouse/partner

This variable is highly comparable across countries apart from the fact that most of the earlier surveys did not include a separate category for 'cohabiting' or 'common-law'. It is not possible to know how people living in such unions declared their marital status. They could have declared themselves as being married or as being single therefore introducing some elements of non-comparability.

COHAB: Respondent is cohabiting

This variable indicates whether or not the diarist is cohabiting with (but not legally married to) a partner.

Value	Label
0	Not cohabiting
1	Cohabiting

Household structure and presence of children

- FAMSTAT
- CPHOME
- HHTYPE
- SINGPAR
- HHLDSIZE
- NCHILD
- AGEKID
- RELREFP

FAMSTAT: Individual level family status

This variable is an individual characteristic, which means that not every member of a household would be coded the same way (in the case of multi-member surveys). It records the presence of children in the household, but does not always allow us to identify whether the children were the diarist's own children, children of the spouse/partner, or children of another household member.

Value	Label
0	Adult aged 18 to 39 with no co-resident children <18
1	Adult 18+ living with 1+ co-resident children aged <5
2	Adult 18+ living with 1+ co-resident children 5-17, none <5
3	Adult aged 40+ with no co-resident children <18

4	Respondent aged <18 and living with parent(s)/guardian(s)
5	Respondent aged <18, living arrangement other or unknown

This variable is highly comparable across countries, though the information about the age of children varies by survey. Discrepancies are documented with the variable AGEKID. Note that it was not possible to identify category '4' for some surveys.

CPHOME: Unmarried child living in parental home

This variable indicates whether or not diarists who are not married or cohabiting live with their parents, regardless of the diarists' age. This variable could not be computed in a large number of surveys.

Value	Label
0	Not a child in parental home
1	Child in parental home

HHTYPE: Household type

This variable records the type of household in which the diarist lived at the time of the survey. Note that the value '1' (one person household) includes never-married, separated, divorced, and widowed people living alone. This variable was computed from a household type variable when available, and from a combination of marital status and household size when no household type classification was available in the original survey.

Value	Label
1	One person household
2	married couple alone
3	married couple + others
4	other household types

In some surveys, we cannot identify cohabiting couples, and these people may be miscoded as HHTYPE =4.

In contrast to FAMSTAT, this variable is a household characteristic and all household members should be coded the same way.

SINGPAR: Whether diarist is a single parent

This variable records whether or not the diarist is a single-parent.

Value	Label
0	No
1	Yes

HHLDSIZE: Number of people in household

This variable records the total number of household members. This variable is highly comparable across surveys.

In some surveys, the size of large households is capped, with the value 'n' meaning 'n or more members'. Such cases are documented in the README documents.

NCHILD: Number of children under 18 in household

This variable records the total number of children aged under 18 in the household. The children are not necessarily the diarist's own children. They can be a spouse/partner's children, children of another household member, younger siblings, or the respondent him/herself if he/she is under 18 years old.

Although this variable is highly comparable across countries, it is affected by four problems. First, in numerous surveys, the original variable used to create NCHILD referred to the number of children under 15 instead of 18. Second, in some surveys the original variables used to create AGEKID and NCHILD referred to children of different age groups (for example under 15 for NCHILD, and 13-17 for AGEKID). Third, in most surveys NCHILD refers to the number of children residing in the household, while in others it refers to the number of children of the diarist (without specifying whether or not they reside in the diarist's household). Finally, in some surveys the original question about the number of children was only asked to diarists who were over 18 years old therefore leaving diarists less than 18 years old with a missing value. In all these cases, we made adjustments and corrections when possible. Users are asked to consult the README documents for more detailed explanations.

AGEKID: Age of youngest child in household

This variable records information on the age of the youngest child in the household. If there are no children under 18 in the household, this variable takes the value -7 (even if the original survey gives a valid value for such cases).

Value	Label
1	Youngest child between 0-4
2	Youngest child between 5-12
3	Youngest child between 13-17

This variable is highly comparable across surveys. However, the cut-off point for the age of the child varies across surveys. Also, in some surveys the data correspond to the diarist's children rather than children residing in the diarist's household.

RELREFP: Relation to household reference person

This variable indicates the relationship of the diarist to the household reference person. In the MTUS, the reference person may be any one of the following:

- either spouse in any married couple living in the dwelling;
- either partner in a common-law relationship;
- the parent, where one parent only lives with his or her never-married son(s) or daughter(s) of any age;
- the designated head of household in some older studies.

This variable tends to be collected mainly in surveys in which more than one household member was surveyed.

Value	Label
1	Person 1
2	Spouse/ Common-law partner
3	Child
4	Parent
5	Sibling
6	Son/Daughter-in-law
7	Father/Mother-in-law
8	Brother/Sister-in-law
9	Other Relative
10	Not related

Employment status, income, and education

- EMPSTAT3

- EMP
- UNEMP
- STUDENT
- RETIRED
- EMPSP
- WORKHRS
- INCORIG
- INCOME
- EMPINCLM
- EDUCA
- EDTRY

EMPSTAT3: New employment status

This variable replaces the variables EMPSTAT and EMPSTAT2 that were part of earlier versions of MTUS. The new variable EMPSTAT3 is identical to EMPSTAT with three exceptions.

- First, in EMPSTAT diarists who said that they were in the military were coded as '3' (other, not employed). In EMPSTAT3 they are coded as '3' (employed, unknown status). This recoding affected very few surveys and very few cases.
- Secondly, diarists who said that they were employed but did not declare their full-time or part-time status are coded as '3' in EMPSTAT3.
- Thirdly, while in the previous version of MTUS full-time was defined as 30 or more hours per week, in Release 2 preference is given to self-declared full-time or part-time status.

Value	Value Label	Description
1	Employed Full Time	Employed/self-employed (including military service), full-time hours
2	Employed Part Time	Employed/self-employed (including military service), part-time hours
3	Employed, unknown status	Employed/self-employed (including military service), hours of work unknown
4	Not in paid work	Other Unemployed, looking for work Retired Homemaker Currently attending school Currently on maternity leave Disability retirement/leave

EMP: In paid work

This variable indicates whether or not the diarist was employed or self-employed (i.e. had a paid job) during the week prior to the survey (or whatever the period of reference was in the original questionnaire).

Value	Label
0	Not in paid work
1	In paid work

UNEMP: Unemployed

This variable indicates whether or not the diarist was unemployed during the week prior to the survey (or whatever the period of reference was in the original questionnaire). This variable does not differentiate between respondents who were registered as unemployed, who were not working but available for work and actively seeking work, and who reported themselves to be unemployed.

Value	Label
0	Not-unemployed
1	Unemployed

STUDENT: Student status

This variable indicates whether or not the diarist was a student during the week prior to the survey (or whatever the period of reference was in the original questionnaire). This variable was coded from a question about whether or not the diarist was a student (or was enrolled in school). If such a question was not available in the original survey, a general economic activity status variable (e.g. what was your main activity during the previous week: 1- employed; 2- looking for work; 3- retired; 4- student; 5- at home) was used.

Value	Label
0	Not a student
1	Student

In surveys where STUDENT is derived from a question about the main activity during the week prior to the survey, students may be miscoded if the survey took place during summer months. For example, a student who is working full-time during summer months and is interviewed during such a month would declare his/her main activity during the week prior to the survey as 'employed' as opposed to 'student'. In such cases, it is recommended to restrict the analysis to school months when carrying out analyses of the patterns of time use of students.

Note that it is possible for some respondents to be coded as employed (EMP=1) and student (STUDENT=1) if these variables were based on different questions and the respondent answered that he/she was both employed and a student. This is not the case in surveys where EMP and STUDENT are derived from the same question and where multiple answers were not allowed.

RETIRED: Retirement status

This variable indicates whether or not the diarist had retired. This variable was created from a question about retirement. If the study did not include retirement questions, the receipt of pension income was instead used. Only when this information was not available was data regarding the diarist's main activity during the week prior to the survey used to compute this variable.

Value	Label
0	Not retired
1	Retired

Note that it is possible for some diarists to be coded as employed (EMP=1) and retired (RETIRED=1) if these variables were based on different questions and the diarist answered that he/she was both employed and retired. This is not the case in surveys where EMP and RETIRED are derived from the same question and where multiple answers were not allowed.

EMPSP: Employment status of spouse/partner

This variable records the employment or self-employment status of the diarist's spouse or partner. In most surveys, this variable is derived from a question about the employment status of the diarist's spouse during the week prior to the survey (or whatever the period of reference was in the original questionnaire).

Note that if the diarist is not married (i.e. CIVSTAT=2), EMPSP was automatically coded as '-7' (i.e. not applicable because no spouse/partner present) even if the original survey gives a valid value for such cases.

Value	Label
1	Employed full-time
2	Employed part-time
3	Employed, unknown hours
4	Not in paid work

WORKHRS: Hours paid work last week including overtime

This variable records the number of hours of paid work reported during the week prior to the survey including any overtime. Note that the number of hours of paid work during the last week was given priority even if data on the number of hours 'usually worked' was available. If data on the number of paid work last week was not available, then WORKHRS was computed by using usual hours of paid work. When neither question was available, 7-day diaries or work schedules (as used in HETUS) were used to measure hours worked during the diary week. Surveys in which this variable does not represent hours worked last week should be documented in the README documents.

The variable includes reported hours of paid work for any diarist who answered the question. For example, if the number of hours of work were asked to people who reported their economic activity status as 'not working', then the diarists' answer should be included in this variable. The README file documents if all or only a subset of the diarists were asked the question.

Values of 0 mean that the diarist reported zero hours of paid work. If diarists were not asked the question, they were given a value of -9 or -7 as appropriate. If diarists did not answer the question, they were coded as -8 for this variable.

INCORIG: Original household income

This variable records the household income as originally recorded in the surveys. This variable is not harmonised (see INCOME for the harmonised variable).

Note that when merging data from different surveys, the original value labels for this variable will be lost since they are survey-specific (in fact, the labels for last survey merged will be kept and erroneously applied to all surveys). Users should be aware of this and should consult the README documents to obtain the original value labels.

INCOME: Total household income - grouped

This variable records the annual household income, recoded in quartiles.

Value	Label
1	lowest 25%

2	middle 50%
3	highest 25%

Three factors affect the quality of this variable. First, income tends to have a high percentage of cases with missing values. Users should therefore use this variable with caution. Second, in most surveys, data on income was collected and/or coded in income groups rather than interval values. As a result, the identification of the cut-off points for the first quartile (lowest 25%) and fourth quartile (highest 25%) is not precise. The example below illustrates the situation.

Table 3. Example Frequency distribution of people aged 20-59 years old by income, Canada 1998 (Variable INCMHSD = Income of the household)

Original household income		Frequency	Percent	Valid %	Cumulative%	Cut-off
1	No income	35.2	0.5	0.6	0.6	
2	Less than \$5,000	31.4	0.4	0.6	1.2	
3	\$5,000 to \$9,999	117.0	1.5	2.1	3.3	
4	\$10,000 to \$14,999	212.0	2.7	3.8	7.0	
5	\$15,000 to \$19,999	222.5	2.9	4.0	11.0	
6	\$20,000 to \$29,999	536.0	7.0	9.5	20.6	
7	\$30,000 to \$39,999	724.5	9.4	12.9	33.5	←
8	\$40,000 to \$49,999	780.0	10.1	13.9	47.4	
9	\$50,000 to \$59,999	778.2	10.1	13.9	61.2	
10	\$60,000 to \$79,999	915.4	11.9	16.3	77.5	←
11	\$80,000 to \$99,999	554.0	7.2	9.9	87.4	
12	\$100,000 or more	707.6	9.2	12.6	100.0	
97	Missing	792.8	10.3			
98	Missing	1305.3	16.9			
	Total	7711.8	100.0			
	Total valid		72.8			

In this example, the cut-off point for the first quartile is part of code '7' (\$30,000 to \$39,999), and the cut-off point for the third quartile is part of code '10' (\$60,000 to \$79,999). In coding the harmonised variable INCOME, we coded as '1' (lowest 25%) cases that fell below the cut-off point (i.e. excluding the cut-off point), and we coded as '3' (highest 25%) cases that were equal or fell above the cut-off point (i.e. including the cut-off point). As a result, the category identified as belonging to the lowest 25% does not represent exactly 25% of cases. On the other hand, the category identified as belonging to the highest 25% does not represent exactly 25% of cases. Note however, that this is not the case in surveys having collected the exact income figure.

Note also that in the above example, the categories '97' and '98' were recoded as -8 (unless additional information was available to show that these would be cases of 'not applicable' in which case they were recoded to -7).

Finally, because the age range covered in each survey varies substantially, and because various age groups in a population tend to occupy different places in the income distribution, we computed the income cut-off points using the population age 20-59 years old (and using weighted data). When these cut-off points are applied to the total sampled population, the percentage of respondents coded as '1' (lowest 25%) may therefore exceed 25% -- especially in surveys that included children in their sample.

EMPINCLM: Original monthly income from employment or self-employment

This variable records the monthly personal income from wages/employment/self-employment during the last month. This variable is not harmonised and is instead recorded in national currency. Note that if data is only available on the personal income from wages/employment/self-employment during the last 12 months, this was converted in months and was documented in the README file.

Note that when merging data from different surveys, the original value labels for this variable will be lost since they are survey-specific (in fact, the labels for last survey merged will be kept and erroneously applied to all surveys). Users should be aware of this and should consult the README documents to obtain the original value labels.

EDUCA: Educational level-original study code

This variable contains the diarists' education level as originally coded in the surveys. This variable is not harmonised.

Note that value labels for the variable EDUCA are excluded from the dataset merging all surveys (but not in individual World5.5 files). In surveys for which the information on the diarist's education was recorded in more than one variable, combinations of variables were used to create a single EDUCA variable.

Note that when merging data from different surveys, the original value labels for this variable will be lost since they are survey-specific (in fact, the labels for last survey merged will be kept and erroneously applied to all surveys). Users should be aware of this and should consult the README documents to obtain the original value labels.

EDTRY: Harmonised level of education

This variable contains the harmonised diarists' education level. It is based on the International Classification of Education (ISCED). This variable proved one of the most difficult to harmonise.

Value	Label	ISCED equivalent
1	uncompleted secondary or less	Not completed ISCED level 3
2	completed secondary	Completed ISCED level 3 and/or attendance at level 4
3	above secondary education	ISCED level 5 or above

This variable refers to the diarist's highest level of education completed (in the case of '1' and '2') or attended (in the case of '3').

More information regarding the ISEC classification can be found in the following document:

http://www.unesco.org/education/information/nfsunesco/doc/isced_1997.htm

Other harmonised information

- DISAB
- URBAN

DISAB: Diarist has a disability or long-term limiting health condition

This variable indicates whether or not the diarist has a disability or long-term health limiting condition.

Value	Label
0	No
1	Yes

It should be noted that the way disability is defined tends to vary across surveys, which may affect the degree of cross-survey comparability.

URBAN: Urban or rural household

This variable indicates whether or not the diarist lives in an urban area.

Value	Label
1	Urban/suburban
2	Rural/semi-rural

The survey-specific definition of 'urban' and 'rural' is included in the README file.