NEPS
National Educational Panel Study

Research Data

Starting Cohort 6: Adults (SC6)
SUF-Version 3.0.1
Data Manual

Jan Skopek

SPONSORED BY THE

Federal Ministry
of Education
and Research

Data Manual

Starting Cohort 6

Adult Education and Lifelong Learning

Release 3.0.1

NEPS Data Center
Jan Skopek

August ςπ, 201σ

Research Data Papers
at the NEPS Data Center, Bamberg

This series presents documentation resources prepared to support the work with data from the National Educational Panel Study (NEPS).

This release of scientific use data from Starting Cohort 6 – "Adult Education and Lifelong Learning" was prepared by the staff of the NEPS Data Center. It represents a major collaborative effort. The contribution of the following staff members of the NEPS is gratefully acknowledged:

Data edition, coding, and scoring
Daniel Bela, Claus Carstensen, Tobias Koberg, Kathrin Lockl, Manuel Munz, Steffi Pohl, Michael Ruland, Jan Skopek, Knut Wenzig, Stefan Zimmermann.

Documentation
Lydia Kleine, Tobias Koberg, Jan Skopek, Knut Wenzig.

Data manual
Jan Skopek
Thomas Leopold, Marcel Raab (co-authoring previous version of this manual)

National Educational Panel Study (NEPS)
Data Center
Wilhelmsplatz 3
96047 Bamberg, Germany
Contact: userservice.neps@uni-bamberg.de
Web: http://www.neps-data.de

Table of Contents

1 Introduction

1.1 About this manual

This manual is intended to assist your work with the data of the NEPS Starting Cohort 6 – Adult Education and Lifelong Learning (NEPS SC6 version 1.0.11, doi:10.5157/NEPS:SC6:3.0.1). We aim at providing a detailed guide of how to use these data for your research. Therefore, our focus is on practical aspects of data usage such as the dataset structure, key variables, and examples of data retrievals. This manual significantly extends the data manual accompanying the first data release 1.0.0 (cf. Leopold et al. 2011).

Be aware that this manual is not a comprehensive documentation resource. Please consult our website

https://www.neps-data.de/datacenter

for background information on the studies, survey instruments, a structured documentation, and many more resources.

We aim at keeping this manual as short and simple as possible. At several places, we reference supplementary documents presenting additional information that we consider essential for working with our data:

- Codebook (Supplement A)
- How-to guides
  - Merging data (Supplement B)
- Technical reports
  - Weighting (Supplement C)
  - RegioInfas (infas geodaten) (Supplement D)
  - Anonymization Procedures (Supplement E)

You can download these documents here:

https://www.neps-data.de/datacenter/researchdata/startingcohortadults

We welcome feedback from our users that will help us improve the quality of this manual and our data for future releases. Please report any feedback to:

userservice.neps@uni-bamberg.de

1.2 Obtaining the data

There are three simple steps to obtain the data of this release:

- Sign the data use contract and mail it to the NEPS Data Center
- After approval, sign in as a registered NEPS user
- Access the data via one of our three access modes (see below)

Depending on which access mode(s) you choose, you will find all further instructions required to access the data on our website:

https://www.neps-data.de/datacenter/dataaccess

1.3 Three modes of data access

We offer you three modes of access to the data:

- Download from our website,
- RemoteNEPS (remote access via a virtual desktop),
- and on-site access.

These three solutions are designed to support the full range of users' interests and maximize data utility while complying with strict standards of confidentiality protection.

Sensitive data

Each access mode corresponds to a specific level of data sensitivity. Files that are offered for download include data with the highest level of anynomization. These data are available to registered users from the web portal via a secure connection. Files offered via RemoteNEPS contain more sensitive data within a controlled environment. The analysis of information in high resolution (e.g., fine-grained regional information) is only provided on-site in Bamberg where these data are available within a secure site. For details on the access modes, see our website at

https://www.neps-data.de/datacenter/dataaccess

This concept of data dissemination translates into an "onion-shaped" model of datasets: The most sensitive data ("on-site") that include weakly anonymized information in high resolution represent the outer layer, with "remote access" and "download" levels being subsets of these data. That is, any data contained within a less sensitive level is also included in the higher level(s).

An overview on which types of data are offered at each of these levels as well as detailed information on how the "on-site", "remote access" and "download" versions of the data were generated can be found in Supplement E (see 1.1).

File Format

All files are available in Stata and SPSS format with bilingual variable labels and value labels (German and English). Data stored in Stata format contain both languages within one file (see section 7). SPSS files are delivered separately in both languages.

1.4 Publications with NEPS data

If you publish with NEPS data, it is mandatory to quote the following reference:

Blossfeld, H.-P., H.-G. Roßbach, and J. von Maurice (eds.) (2011). "Education as a Lifelong Process – The German National Educational Panel Study (NEPS)", Zeitschrift für Erziehungswissenschaft: Special Issue 14.

In addition, publications using data from this release must include the following acknowledgement:

This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort 6 – Adults (Adult Education and Lifelong Learning), doi:10.5157/NEPS:SC6:3.0.1. The NEPS data collection is part of the Framework Programme for the Promotion of Empirical Educational Research, funded by the German Federal Ministry of Education and Research and supported by the Federal States.

A digital object identifier (DOI) uniquely identifies each release of NEPS data. The DOI of this release redirects to a landing page providing basic information on the data:

http://dx.doi.org/10.5157/NEPS:SC6:3.0.1

1.5 Release History

AUGUST, 2013, Release 3.0.1 *(doi: 10.5157/NEPS:SC6:3.0.1)*

Errors in files *spSchool*, *Education*, *spVocTrain*, *FurtherEducation* have been corrected. We strongly recommend researchers to use version 3.0.1 instead of version 3.0.0.

See detailed release notes on

https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC6/3-0-1/NEPS_SC6_3-0-1.txt

The data manual has been updated, too. Syntax examples for SPSS and R have been added.

JUNE 6, 2013, Release 3.0.0 *(doi: 10.5157/NEPS:SC6:3.0.0)*

Release 3.0.0 of Starting Cohort 6 contains data from three waves: the ALWA forerunner study (2007/2008) marked as wave 1 in the data files, the first NEPS wave (study B72, 2009/2010) marked as wave 2, and the second NEPS wave (study B67, 2010/2011) marked as wave 3. Note, by this release the release number (3.0.0) has been adjusted to resemble the number of data collection waves that are integrated in the data set. Additionally, compared to first data release several improvements in data edition have been implemented. Variables with occupational and educational information to targets, parents and partners have been adjusted. Spell files containing spells that went into the data revision module (also check module) at time of interview have been extended by additional generated variables providing spell times as they were corrected in the check module. Moreover, non-response to multiple response variables (like refusals and don't knows) has been re-coded in a more convenient way. Several new data files are included: *spChildCohab* (comprising cohabitation with children in a convenient long spell format previously stored in a cumbersome wide format in spChild), *xTargetCompetencies* (comprising competence data that has been collected in study B67), and *xCompMethods* (containing para data collected by the CAPI module guiding the competence assessment). Please consult the Appendix for a detailed list of changes that are included in release 3.0.0. Be aware that some of the changes possibly imply minimal syntax adjustments of already existing code.

See detailed release notes on

https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC6/3-0-0/NEPS_SC6_3_0_0_0.txt

DECEMBER 22, 2011, Release 1.0.0 *(doi: 10.5157/NEPS:SC6:1.0.0)*

First official data release of the NEPS Starting Cohort 6. Major improvements have been included after several months of beta phase.

See detailed release notes on

https://www.neps-data.de/Portals/0/Neps/Datenzentrum/Forschungsdaten/SC6/1-0-0/SC6_1_0_0.txt

AUGUST 2011, Beta-Release *(not available anymore)*

Data from the NEPS Starting Cohort 6 has been released in a beta version.

2 Conventions

Names of data files as well as names of variables in this data release follow certain conventions. These conventions are designed to improve usability of the data since they provide information on type, format, accessibility and content of dataset and variables. Additionally, this section documents different kinds of missing codes.

2.1 File names

The names of the datasets included in this release were defined by a number of conventions which are displayed in Table 1.

Table 1: Naming conventions of file names

Element	Definition
SC[1-6]	**Indicator of starting cohort** 1 = Infants 2 = Kindergarten 3 = 5th grade students 4 = 9th grade students 5 = First-year undergraduate students 6 = Adults
[filename]	**Filename conventions** Prefix: sp = spell file; p = panel file Keyword/mnemonic: A keyword or mnemonic indicates the content of the corresponding file (e.g., spEmp contains employment spells) Filenames of generated datasets do not have a prefix and always start with a capital letter (e.g., *Biography*)
[D,R,O]	**Confidentiality Level** D = Download version R = Remote access version O = On-site version
[#]-[#]-[#] (_beta)	**Version** First digit: denotes the main release number; the main release number is incremented with every wave of a starting cohort Second digit: indicates major updates; major updates affect the data structure (e.g., release of imputed datasets); updating your syntax files may be necessary Third digit: indicates minor updates; minor updates affect the content of cells but not the data structure; updating your syntax files is not necessary _beta-suffix: this suffix indicates a preliminary release which allows users to test the data in advance of the main release. The beta version is no longer available after the main release.

2.2 Names of variables

The variable naming conventions are aimed at ensuring the consistency of variable names across panel waves. They reflect the panel structure of the NEPS data and allow users to conveniently identify variables across waves.

A variable name consists of up to four elements. The principles of the naming conventions are illustrated by the following example. More detailed information is given in Table 2.

Figure 1: Elements of variable names

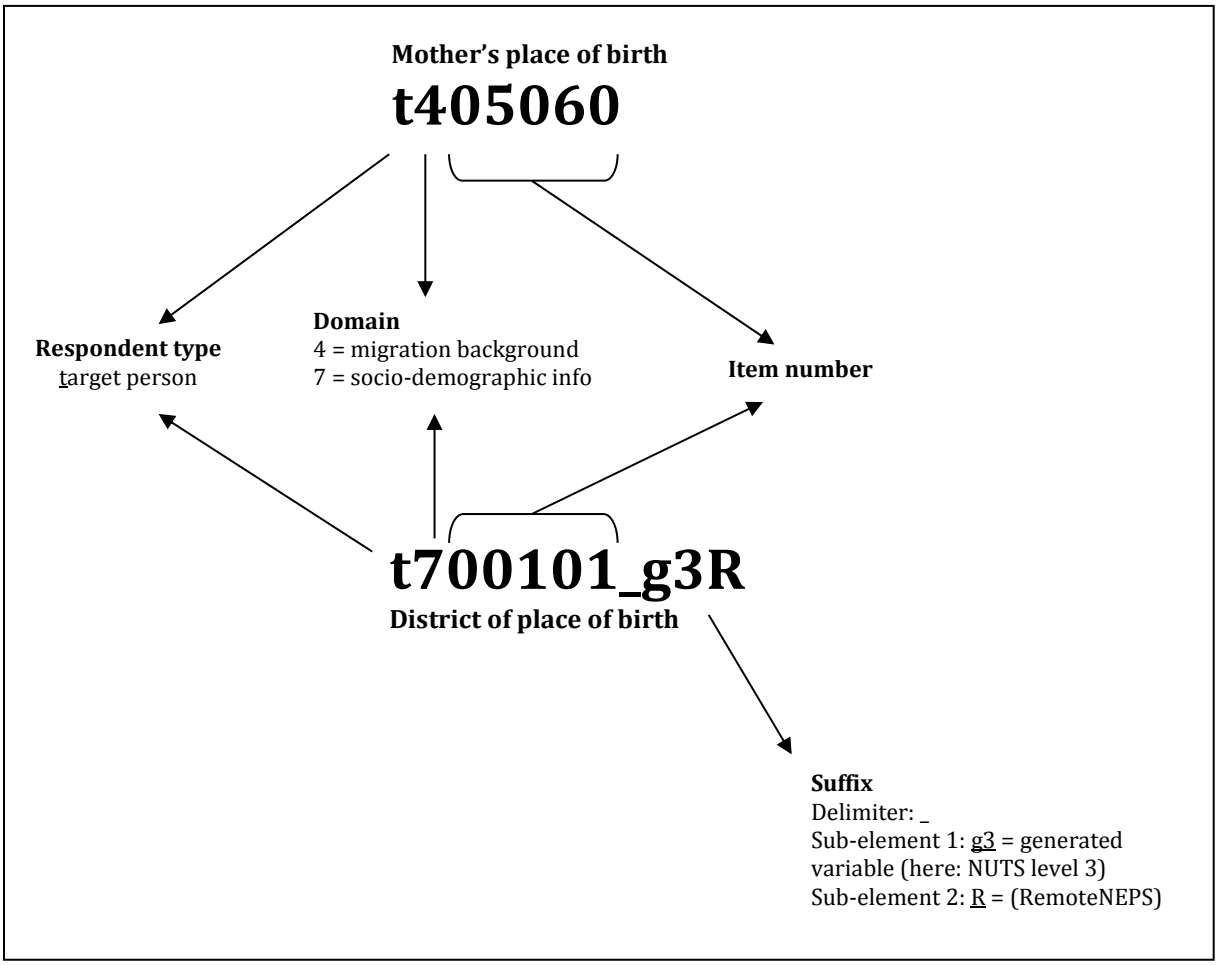


Table 2: Naming conventions for variable names

Digit	Description
1	Indicates to which **respondent type** the variable refers; in starting cohort 6, this character is always *t* (target person)
2	**Topic/domain** (according to the theoretically coordinated dimensions of the NEPS): 1 = competence development (pillar 1) 2 = learning environments (pillar 2) 3 = educational decisions (pillar 3) 4 = migration background (pillar 4) 5 = returns to education (pillar 5) 6 = working group "interest, self-concept and motivation" 7 = socio-demographic information h = adult education and life-long learning (stage 8) s = basic program; variables developed for the NEPS stages 6 and 8 x = generated variables
3–7	**Item number:** The item number typically consists of four numeric characters plus one alphanumeric character
8–11	**Suffix** (optional): Suffixes are separated from the previous characters by an underscore. There are four types of suffixes: • Version suffixes: Some questions receive minor updates or changes across panel waves. This leads to different versions of similar items. The variable name of last item version does not have a version suffix. The remaining versions are indicated by the following suffixes: o v1 = first item version (item changed for the first time) o v2 = second item version (item changed for the second time) o v3 = third item version (item changed for the third time) In most cases we were able to integrate earlier versions of variables into the updated version of the variable. If this was not possible, harmonized variables which retain the information common to both versions were generated and marked by the following suffix (see below for examples): o ha = harmonized variable • Suffixes for generated variables: Generated variables are indicated by the suffix _g# (_g1, _g2, etc.). In most cases, the running number after _g is a simple enumerator. However, there are two types of generated variables that assign meanings to these running numbers: regional and occupational variables. o Regional codes based on the Nomenclature of Territorial Units for Statistics (NUTS) ▪ g1 = NUTS level 1 (federal state/Bundesland) ▪ g2 = NUTS level 2 (government region/Regierungsbezirk) ▪ g3 = NUTS level 3 (district/Kreis) o Occupational/prestige codes ▪ g1: KldB 1988 (German Classification of Occupations 1988) ▪ g2: KldB 2010 (German Classification of Occupations 2010) ▪ g3: ISCO-88 (International Standard Classification of Occupations 1988) ▪ g4: ISCO-08 (International Standard Classification of Occupations 2008) ▪ g5: ISEI-88 (Internat. Socio-Economic Index of Occupational Status 1988) ▪ g6: SIOPS-88 (Standard International Occupational Prestige Scale 1988)

  - g7: MPS (Magnitude Prestige Scale)
  - g8: EGP (Erikson, Goldthorpe, and Portocarero's class categories)
  - g9: BLK (Blossfeld's Occupational Classification)
  - g12: Coding scheme
  - g13: KKZ (Course code / Kurskennziffer)
  - g14: ISEI-08 (Internat. Socio-Economic Index of Occupational Status 2008)
  - g15: CAMSIS (Social Interaction and Stratification Scale)
  - g16: SIOPS-08 (Standard International Occupational Prestige Scale 2008)
- Wide-format suffix:
  Wide-format variables stored in spell files are indicated by the suffix _w# (e.g., _w1, _w2, etc.).
- Confidentiality suffix:
  This suffix pertains to all variables that were anonymized (see 1.4). The suffix indicates a variable's degree of anonymization. This suffix may either stand alone (e.g., country of birth: *t405010_R*) or be combined with other suffixes (e.g., district of place of birth: *t700101_g3R*)
  - O: on site; data on this variable are only available on site
  - R: remote access; data on this variable are available on site or via RemoteNEPS
  - D: download; data on this variable are available via all three modes of access

The following examples illustrate the integration and harmonization of variables that have changed across waves. As Table 3 shows the ALWA study (2007/08) surveyed the current marital status in greater detail than the NEPS in 2009/10. To integrate these two versions, we collapsed categories 1 and 2 of the ALWA (2007/08) item into one category "married". No harmonization was necessary.

Table 3: Panel integration of the variable "current marital status"

t733001	**t733001_v1 (ALWA)**	
1: married	1: married, living together	2: married, separated
2: in a registered partnership		6: in a registered partnership
3: divorced	3: divorced	
4: widowed	4: widowed	
5: single	5: single	

The second example illustrates the harmonization of the variable "mother's place of birth" (Table 4). Here, the NEPS collected more detailed information than the forerunner study ALWA. In the upcoming panel waves, each new respondent will answer the NEPS version of the question (*t405060*). Integrating the ALWA variable (*t405060_v1*) into the NEPS variable was not possible because the response categories varied considerably. Therefore, we generated a harmonized variable (*t405060_ha*) containing the information that both versions have in common.

Table 4: Panel harmonization of the variable "mother's place of birth"

t405060	t405060_v1 (ALWA)	t405060_ha
1: in Germany (after 1949) 2: in the area that is present-day Germany (before 1950) 3: in Germany's former eastern territories (before 1950)	1: FRG/West Germany 2: GDR/East Germany 3: FRG Germany (after reunification)	1: in Germany
4: abroad (after 1949) 5: in another country (before 1950)	4: abroad	2: abroad

2.3 Special conventions for naming variables in competence data

Naming of variables corresponding to test items (usually found in competence data files) follow an alternative nomenclature. Variable names consist of three parts and additional suffixes. The first part defines the test instrument (two characters, e. g. *vo* for vocabulary), the second part defines the target group (two characters, e. g. *a2* for adults in the second NEPS wave, i. e., 2010/2011), and the third part defines the item number.

Table 5 (p. 13) gives an overview to the logic of parts. The first two characters identify competence domains. An overview of the different competence domains is given in the first part of Table 5. The target group indicates the cohort or testing wave in which the item was first used. The different target groups are listed in the second part. In some tests, (e. g., mathematic competence tests) items are implemented in different testing waves. In these cases, the variable name contains the target group for which the item was first used. The variable name of the item is then fixed and does not change when the item is used again in later waves or other cohorts (e. g., if the item is first used in grade 5, the second part of the variable name will be G5, even when the item is reused in grade 7). Thus, the target group identification in the variable name does not necessarily indicate the cohort or testing wave. However, this labeling rule assures items being used in different studies to have the same variable name. Some competence tests are not designed for specific age groups, but are implemented unmodified in different cohorts and testing waves. The target group of these tests is indicated by ci (cohort invariant). The item number is defined differently for different competence domains. For most competence domains they only indicate the different items.

The competence data files contain item variables (responses to the test items) as well as overall competence scores. There are two versions of item variables in competence data: scored items named *[varname]_c* and scored partial credit-items named *[varname]s_c*. For example, *mag9q071_c* is a scored variable measuring that the respective math item—targeted at grade 9 students—was "solved" (value 1) or "not solved" (value 0) by the respondent. Note that the item variable does not necessarily indicate that the students' mathematics skills are measured in grade 5. It could also be that the measurement was done in grade 7 and that an item was used that has already been implemented in grade 5. Additionally to the item responses, overall measures of the competence score are given. Suffix *_sc[number]* is used for several aggregated scores and

the meaning of the suffixed number is fixed as follows: 1=WLE (Weighted Maximum Likelihood estimates[1]), 2=standard error of WLEs, 3=sum, 4=mean, 5=difference.

For example, variable *grk1_sc3* represents the sum score of the grammar test of children being tested in the first wave (2010) in the Kindergarten Cohort or variable *rea2_sc1* represents the WLE of the math test implemented in the second NEPS wave of the adult cohort. Detailed descriptions on how competence scores are estimated can be found in the respective reports for the different competence domains. If there are several aggregated scores (e. g. different sum scores), letters are appended additionally (e. g. *dgg9_sc3a* is of the sum score for perceptual speed, while *dgg9_sc3b* is the sum score for reasoning – both are measures of domain general cognitive functioning).

[1] WLEs are estimated in tests that are scaled based on models of item response theory (cf. Pohl and Carstensen 2012).

Table 5: Names of variables in competence test data

Part I (2 chars): Instrument Meaning			
re	Read	mp	Meta procedural
ma	Math	md	Meta declarative
sc	Science	rs	Reading Speed
ic	ICT	at	Attention
li	Listening	nr	Native Language Russian
vo	Vocabulary	nt	Native language Turkish
or	Orthography	vi	Verbal Intelligence
gr	Grammar	ni	Nonverbal Intelligence
dg	DGCF	fa	FAIR
ef	English Foreign		
Part II (2 chars): Target Group Meaning			
n0	Newborn 0	v1	Vocation 1
...	...	...	…
n3	Newborn 3	v3	Vocation 3
k1	KiGa 1	s1	University students 1
k2	KiGa 2	...	...
g1	Grade 1	s5	University students 5
g5	Grade 5	a1	Adults 1
g9	Grade 9	...	...
ga	Grade 10	a4	Adults 4
gb	Grade 11		
gc	Grade 12		
gd	Grade 13		
ci	Cohort invariant (for instruments administered unchanged in all cohorts)		
Part III (4 chars): Item number			

Part IV: Suffix			
_c	scored item variable (0=not solved, 1=solved)		
_sc1	WLE		
_sc2	Standard error of WLE		
_sc3	sum		
_sc4	mean		
_sc5	deviation score (procedural metacognition)		
_sc6	proportion of correct items (procedural metacognition)		

2.4 Missing values

Table 6: Overview of missing codes

Code	Missing
Item nonresponse	
–97	Refused
–98	don't know
–94	not reached (competence assessment)
–5/–6/–20	item-specific missing
Not applicable	
–54	not included in sample-specific instrument of this wave
–93	does not apply
.	filtered / system missing
Edition missings (recoded into missing)	
–52/–95	implausible value removed
–53	Anonymized
–55	not determinable

We distinguish between three types of missing values (see Table 6):

- *Item nonresponse* occurs if a person did not respond to a question.
  - The most common instances of item nonresponse are refusals (–97) and don't knows (–98).
  - In the context of competence data –94 marks test items that have not been reached in the test session.
  - Additional missing codes (–5/–6/–20) pertain to specific nonresponse categories (e.g., –5 "never graduated" for father's school leaving certificate (*t731351*)).
- *Not applicable* denotes missing data that occur because the item does not apply to a person. This category comprises two kinds of missings.
  - The first concerns samples: If a question is not included in a sample-specific questionnaire, the code –54 is assigned to all respondents from this sample.
  - The second concerns individuals: If a question does not apply to a person, it is coded "not applicable" either by the respondent's or the interviewer's remark (–93) or automatically by the survey instrument (. = filtered).
- *Edition missings* are defined in the process of data editing.
  - Implausible values are recoded into missing (–52 or –95).
  - Sensitive information which is only available via RemoteNEPS and/or on site access is anonymized (–53).

- Coding schemes are used to generate variables (e.g., occupational coding). If the information from the original data is not sufficient to generate a value, we assign the missing code “not determinable” (–55).

In addition, there are special missing values (-22 and -21) for spell time variables that have been generated in the check module and are included in the spell files since release version 3.0.0.

nepsmiss: Recoding missing values in Stata

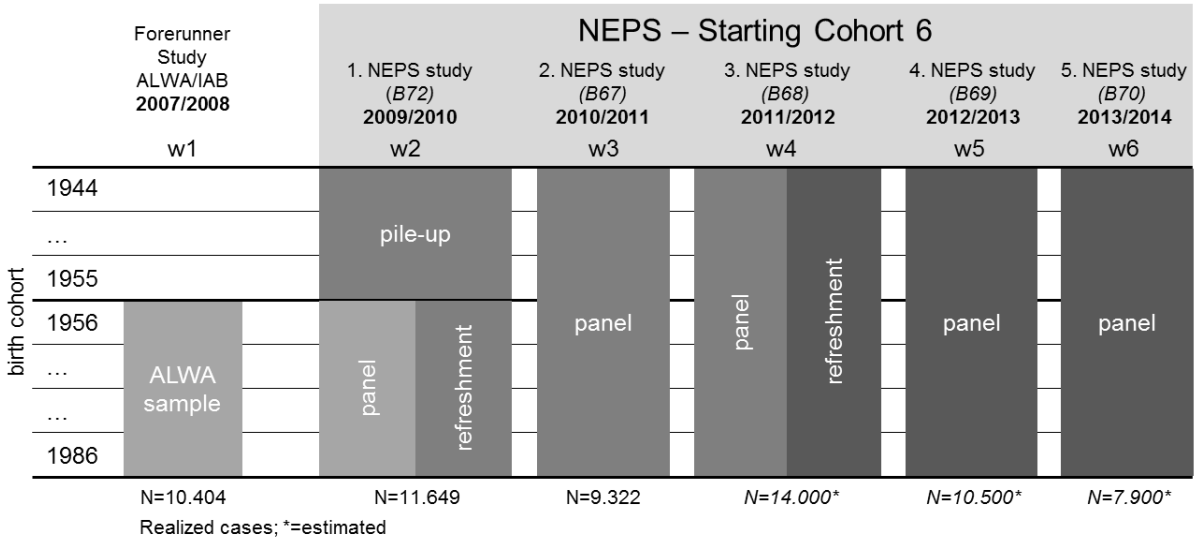
We offer a Stata ado file on our web site which automatically recodes all missing values into extended missing values (.a, .b, etc.), and vice versa, while preserving value labels. See section 7 for further information.

3 Surveys and Sampling

The study population of NEPS Starting Cohort 6 (SC6) consists of individuals born between 1944 and 1986. The data of NEPS SC6 offer detailed retrospective information on the histories of education, employment, and family of an adult population. In addition, extensive information was collected on adult education, learning environments, and decision-making processes as well as on subjective well-being and health. Furthermore, competences of respondents are assessed at every second wave of NEPS SC6.

The current release of NEPS SC6 comprises data from three data collection waves that took place in 2007/2008, 2009/2010, and 2010/2011. Note that the first wave represents a forerunner study called *Working and Learning in a Changing World (ALWA)* that had been conducted by the *Institute for Employment Research (IAB)* in Nuremberg. The National Educational Panel Study (NEPS) started at 2009/2010 with data collection (study *B72*) by taking over targets from the ALWA study (those with panel willingness) and by refreshing and piling up the original ALWA sample to older birth cohorts. Hence, the first NEPS data collection is marking the second wave effectively for a significant part of the sample. The third wave had been conducted in 2010/2011 by NEPS (study *B67*). Some cases from ALWA who temporally dropped out at wave 2 have been successfully interviewed at wave 3.[2] Figure 2 depicts the longitudinal sampling design of NEPS SC6.

Figure 2 NEPS Starting Cohort 6 - Longitudinal Sampling Design



The released data (wave 1-3) comprises two subsamples:

- ALWA (2007/08) sample: 6,855 respondents from the birth cohorts 1956 to 1986 who were recruited in 2007 by the forerunner study *ALWA*. Note that 77 non-German speaking persons were recruited for the ALWA sample but only interviewed

[2] This increases the sample size compared to the first data release.

in the second wave (2009/2010) since there was no foreign language questionnaire in ALWA. From the ALWA subsample 283 persons temporarily dropped out at the second wave (2009/2010) but were re-interviewed at wave three (2010/2011).

- NEPS (2009/10) sample: 5,077 respondents recruited for the first main study of the NEPS. Strictly speaking, the NEPS sample consists of two further subsamples:
  - Refreshment sample (N=1,971): drawn from birth cohorts 1956 to 1986
  - Pile-up sample (N=3,106): drawn from birth cohorts 1944 to 1955

The development of the sample is illustrated in Table 7. As you can see the full sample of NEPS SC6 comprises 11,932 persons of whom 6,778 were interviewed already in the ALWA study. There are 11,649 interviews completed in wave 2 and 9,322 interviews completed in wave 3.

Table 7: Sample development (cf. file *Methods*)

	Participation status			**Total**
Wave / Sample	participated (realized interview)	temporary drop out	final drop out	
Wave 1 (2007/2008)				
ALWA	6,778	–	–	
NEPS	–	–	–	
Total	**6,778**	–	–	**6,778**
Wave 2 (2009/2010)				
ALWA	6,572*	283	–	**6,855**
NEPS	5,077	–	–	**5,077**
Total	**11,649**	**283**	–	**11,932**
Wave 3 (2010/2011)				
ALWA	5,638	1,074	143	
NEPS	3,684	984	409	
Total	**9,322**	**2,058**	**552**	**11,932**

* includes 77 foreign language interviews (Russian/Turkish) for non-German speaking person who could not be interviewed at wave 1 and hence have been interviewed for the first time in wave 2.

For detailed information on the studies and sampling strategies, see Allmendinger et al. (2011), Antoni et al. (2010), and Aßmann et al. (2011).

Data were collected by computer-assisted personal and telephone interviewing (CAPI & CATI) that distinguishes between first and panel respondents. Figure 4 illustrates the basic structure of the interview from NEPS wave 2009/2010. Table 8 shows which survey instruments were used for which sample in which wave. At second wave (B72) first respondent and panel respondent interviews were conducted. Individuals from the NEPS sample as well as the non-German speaking persons, who were recruited but not interviewed at wave 1 (ALWA study), got the first respondent questionnaire. Those who

already were interviewed in the ALWA study (wave 1) got a panel questionnaire at wave 2. At wave 3 (second wave of NEPS, 2010/2011) only panel questionnaires were administered since only panel respondents were interviewed.

Table 8: Overview of samples and survey instruments (data collection from 2007 to 2011)

Wave	ALWA sample (2007)	NEPS sample (2009)
2007/2008 (ALWA)	ALWA questionnaire *(6,495 respondents)*	
2009/2010 (NEPS – study B72)	NEPS panel questionnaire *(6,495 respondents)* NEPS first questionnaire *(foreign language version) (77 respondents)*	NEPS first questionnaire *(5,077 respondents)*
2010/2011 (NEPS – study B67)	NEPS panel questionnaire *(9,322 respondents)*	

Course of Interview

The questionnaire began and ended with cross-sectional modules (see Figure 3 and Figure 4). Between these modules, the questionnaire's main part was devoted to the comprehensive collection of retrospective information on the respondents' life courses. These longitudinal data were collected within separate longitudinal modules, most of which were complemented by brief cross-sectional sub-modules. A check module identified and corrected inconsistencies in the sequence of episodes, ensuring the integrity of the life course data (see Figure 3).

There was a dynamic course module ("Courses (last year)", depicted in Figure 4), which was plugged into the interview during collection of recent military service, employment, unemployment, parental leaves, and gap episodes. Additionally, further private courses have been in collected in a further education module. After that, a course collector selected up to 2 courses randomly from the list of reported courses. Subsequently, for those random courses, further details have been asked.

Figure 3: Stylized course of interview - collection of life course data and data revision module[3]

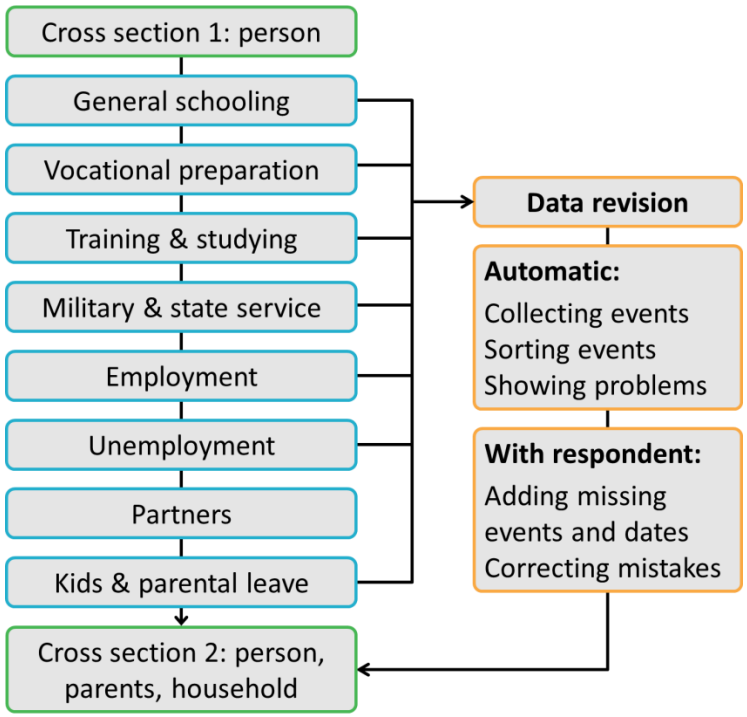
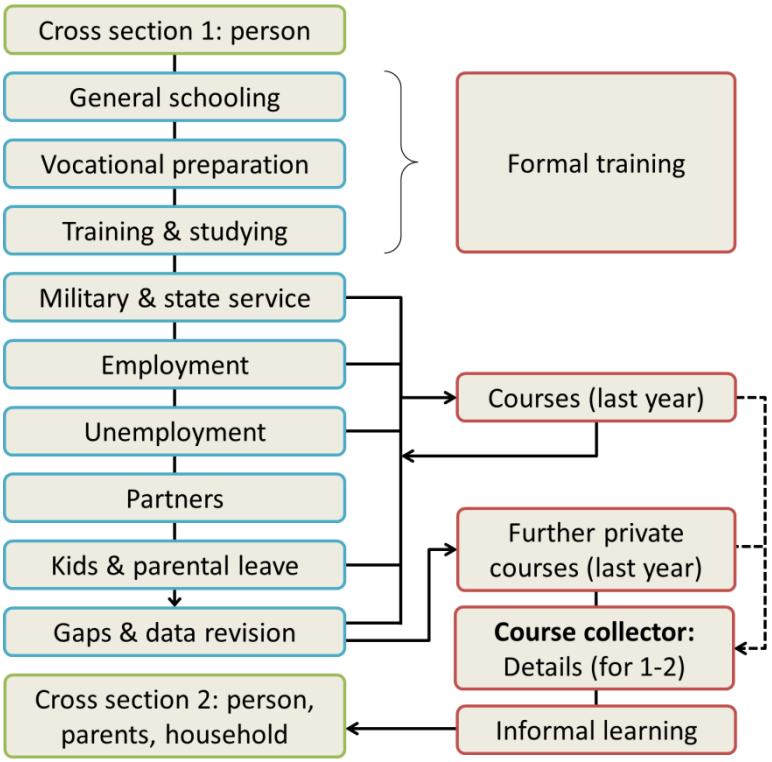


Figure 4: Stylized course of interview - collection of further education courses[4]



[3] We thank Dr. Corinna Kleinert (NEPS Stage 8) for providing this illustration.
[4] We thank Dr. Corinna Kleinert (NEPS Stage 8) for providing this illustration.

Table 9: Questionnaire modules, availability over waves, and data files

Module (code)	Module	Wave 1 (ALWA)	Wave 2 (1. NEPS, study B72)	Wave 3 (2. NEPS, study B67)	Data files
20QS1	Cross-Section 1	X	X	X	pTarget
22AS	Schooling	X	X	X	spSchool (spells), pTarget (cross-section)
23BV	Vocational Preparation	X	X	X	spVocPrep (spells), pTarget (cross-section)
24AB	Vocational Training	X	X	X	spVocTrain (spells), pTarget (cross-section)
25WD	Military Service	X	X	X	spMilitary (spells), pTarget (cross-section)
26ET	Employment	X	X	X	spEmp (spells), pTarget (cross-section)
27AL	Unemployment	X	X	X	spUnemp (spells), pTarget (cross-section)
28PA	Partners	X	X	X	spPartner (spells), pTarget (cross-section)
29KI	Children + Cohabitation & Parental leaves	X	X	X	spChild (children), spChildCohab (cohabitation spells), spParLeave (parental leave spells), pTarget (cross-section)
30X	Check-Module / Gap-Episodes	X	X	X	spGap (spells)
31WB	Further Education	n.a.	X	X	spFurtherEdu1 (further courses), spFurtherEdu2 (course details for 2 randomly selected course), spFurtherEdu3 (German courses), pTarget (cross-section)
32QS2	Cross section 2	X	X	X	pTarget
35KU	Dynamic Course Module	n.a.	X	X	spCourses (up to three courses per spell)
36T1	Competence assessment: CAPI-module + test booklets	n.a.	n.a.	X	xCompMethods (para data collected in CAPI mode) xTargetCompetencies (scored test items)

Notes: Spell modules like employment or unemployment typically collect episodes in questionnaire loops. However, for most spell modules cross-sectional data not related directly to spells is collected too. These spell-type specific cross-sectional data is usually integrated into the pTarget file. The course module (35KU) is called dynamically in the context of reporting on episodes of military/civil service, employment, unemployment, parental leave, and gaps if episode times fall into the last 12 months and respondent reports participation at further education during these episodes.

4 Data Structure and Datasets

Aims and scope of the NEPS surveys inevitably create complex data. We structured these data in a user-friendly way and generated a number of additional datasets from one or more of the original files to ease the preparation and analysis of life course data. Basically, there are so called *survey files* and *generated files*. Survey files contain data as it was collected in the interviews. Contrary, generated files contain data that was not directly interviewed. Usually generated files represent derivations of survey files and data that are already prepared to some extent (like *Biography* or *Education*). Moreover, method data (para data from the interview, sample and participation data) and regional data are available in generated files. In addition to that, we distinguish between four types of files according to their data structure: cross-sectional files, panel files, spell files, and event time files.

- *Cross-sectional files*: Cross-sectional files contain data from a specific wave. Each respondent has only one entry. Hence, variable *ID_t* identifies rows uniquely in these files.
- *Panel files*: Panel files include panel data in a long format. Usually, each row corresponds to an interview of an individual at one wave. Rows in panel files are usually identified via variables *ID_t* and *wave*.
- *Spell files:* Spell files include episode data in a long format. Usually, each row corresponds to an episode reported by an individual at one wave. Rows are identified by ID_t and spell identifiers.
- *Event time files:* In contrast to spells, event times do not specify the time spent in a certain state (e.g., unemployed) but the point in time at which a transition between two states (i.e., an event) occurred.

4.1 Identifiers

All files contain the *ID_t* variable that identifies a target person uniquely over waves and over starting cohorts of NEPS. Most of the files additionally have a *wave* variable, identifying the *wave* in which the data row was collected. Note, that wave is just a counter variable starting from the ALWA forerunner study conducted in 2007/2008. Hence, wave=1 marks the ALWA study and not the first wave of NEPS. The first NEPS wave is coded by 2, the second NEPS wave by 3, the third NEPS wave by 4 and so on. In most of the spell files there is a variable called *splink* that identifies a spell (e. g., an employment spell) of a target person. Note that *splink* identifies a spell uniquely only within a person identified by *ID_t*. There are other identifiers for reported children (*child*), reported partners (*partner*), and reported further education courses (*course*). In the following we will discuss the data structure and relevant identifiers will be presented. Additionally, check section 4.5 and section 6 for further information and examples on how to use the identifiers.

4.2 Survey files

Panel and cross-sectional files

Main panel file: pTarget

We merged all cross-sectional information collected at each panel wave into one single dataset *(pTarget)*. This dataset is composed of the two main cross-sectional modules as well as panel data from all cross-sectional sub-modules. These data are stored in long format. That is, one record represents one respondent at one wave. The variables in are organized by questionnaire modules (see Table 9) and there are so called delimiter variables that group variables accordingly. For example, the delimiter variable *DELIM_20QS1* indicates that the following variables relate to questions that have been asked in the 20QS1 module of the interviews.

The file *pTarget* includes basic socio-demographic information and (repeated) cross-sectional measurements. Figure 5 shows an exemplary data snapshot. The first respondent (*ID_t* = 8000215) participated in all three waves – ALWA (coded as 1), first NEPS wave (coded as 2), and second NEPS wave (coded as 3) – and therefore has three records, whereas the second respondent (*ID_t* = 8000334) was recruited in the second wave (*wave* = 2) and therefore has no record for the first wave. The example shows panel data on three variables in long format: gender (*t700001*) , date of birth (*t70000y*), and life satisfaction (*t514001*). As one can see, life satisfaction stays stable for both respondents. Since life satisfaction was not asked in wave 1, there is a missing value coded with -54 ("missing by design", see section 2.4).

Figure 5: Data snapshot from *pTarget*

ID_t	wave	t700001	t70000y	t514001
8000215	1	2	1960	-54
8000215	2	2	1960	7
8000215	3	2	1960	7
8000334	2	2	1959	5
8000334	3	2	1959	5

Competence data: xTargetCompetencies

File *xTargetCompetencies* contains data from competence assessments conducted at the second NEPS wave (marked as wave 3 in the data). Scored item variables as well as scale variables are available in a cross-sectional format. Note, that not all respondents took part at the assessment. Since assessments were conducted in CAPI mode, those persons who were interviewed in CATI-mode have been excluded from testing. Additionally, those who had severe debilities of sight or were even blind were excluded from the assessment.

Interview control data of competence assessment: xCompMethods

In the CAPI interview situation test booklets have been administered by the interviewer. The whole test situation was moderated by an assessment module implemented in the CAPI software. The file *xCompMethods* contains data from that module. Variable *teststat* indicates who was eligible for the competence assessment. A further variable *splitgr* reveals who got which kind of randomized test booklets.

Spell files

Three types of spells: Duration, entity and event spells

This study collected several types of life history data, such as episodes of general education, employment, unemployment, and parental leave. Each of these spell types is stored in a separate dataset. These files always include longitudinal data with each row representing one spell. There are three types of spells:

- *Duration spells* specify a duration spent in a state or episode, such as "employed"
- *Entity spells* pertain to specific entities, such as partners, children, or courses
- *Event spells* defining event times

The variables *spell*, *child, partner,* and *course* are enumerators for each spell of a person:

- *spell* always refers to duration spells
- *child* and *partner* always denote entity spells
- *course* identifies courses in *spFurtherEdu2, spVocTrain, spCourses*, and *spFurtherEdu1* (see examples in section 6 for details); in *spFurtherEdu3* (courses in German), the corresponding identifier is *gcourse*.

Panel updates of spells

Spells are either complete or right-censored. Right-censoring occurs if a spell continues until the time of the interview. The design of the NEPS allows updating right-censored spells prospectively at every panel wave. The updating of these spells was executed as follows:

- Spells that were right-censored at the preceding wave (2007/2008 ALWA) were divided into three subspells which are represented by the variable *subspell*:
  - original right-censored episodes from the preceding wave (*subspell* = 1)
  - continued episodes from the panel interview with updated information (*subspell* > 2)
  - harmonized episodes (*subspell* = 0 & *spgen* = 1; see below); in most cases, these edited episodes include the latest information from the panel subspell (*subspell* = 2). However, if this information was either "filtered / system missing", these missing values from the panel subspell were replaced by data from the previous subspell (*subspell* = 1). For a few selected variables, there were exceptions to these rules which were guided by plausibility criteria (see below for an example).

The main advantage of this procedure is that it retains all information from the original spells while at the same time offering a convenient way of obtaining a harmonized spell data structure. The variable *subspell* is coded 0 both for completed and harmonized spells. Therefore, you can easily obtain a harmonized spell structure by selecting all observations that satisfy the condition

$$subspell = 0.$$

We generally recommend executing this selection at the start of your data preparation unless you are specifically interested in subspell information. However, be aware that data of harmonized spells may come from different waves because these spells always include the latest valid information available. There is another caveat: Do not use this selection if you work with information stored in wide format (like interruption episodes of vocational training spells stored in a wide format in spVocTrain).

The following example in Figure 6 illustrates the identification and selection of subspells as well as the logic of harmonized episodes. Employment spells (identified by *splink* or *spell*) of two persons (identified by *ID_t*) are displayed. Subspells are identified via *subspell* and harmonized spells are marked by *spgen=1*. The example shows 5 variables characterizing the spells: *ts2311y* (start year), *ts2312y* (end year or interview date for right censored spells), *ts23410* (net job income, monthly), *ts23223* (effective weekly hours of work), *ts23228* (required training for this job).

The first respondent (*ID_t* = 8001123) has three job spells. The first spell was reported in the ALWA study (in 2007/2008) and finished at time of interview. The second spell (*splink* = 260002) was reported for the first time in wave 1 and updated prospectively in wave. Hence, this employment spell is divided into three subspells: *subspell* = 1 represents the right-censored spell from the preceding wave (2007/2008, ALWA); *subspell* = 2 denotes the continued spell from the second wave (2009/2010, NEPS); *subspell = 0* denotes the harmonized version of both subspells. Since, the job was reported to be finished at date of the second interview there was no additional panel update in wave 3 (2010/2011, NEPS). However, in wave 2 the respondent reported a third job (*splink* = 260003) as current employment. Consequently, this job was preloaded and re-asked in the interview at wave 3.

The second respondent (ID_t = 8001204) reported two jobs in sum. While the first job lies completely in the past (ended in 1997), the second job started in 2006 and reported in wave 1 has been followed up over three waves.

As a result not only respondents but also their episodes are followed up in a panel fashion. That means, that spell data provides time-varying information on spells. In the example of the employment spells you see that within the second job of respondent 8001204 reported net income increased over panel waves. Similarly, weekly hours of work decreased between waves from 40 to 36 in the second job of respondent 8001138.

How does the harmonization of spells work? First, for every panel-extended spell a harmonized subspell is provided by NEPS data edition. Those harmonized spells can be identified via the *spgen* variable very easily. Second, for most variables the most recent non-system missing value is provided in the harmonized spell version. Usually this

pertains to the value of the last interview. If in the current interview a question is filtered, the value from last waves is put into the harmonized spell. For example, the question of which kind of training is needed for the job (variable *ts23228*) is asked only once. In follow up waves this variable is skipped by filtering. According to the rules of harmonization the last non-system missing value is coded in the harmonized spell (e.g. see second job of respondent 8001204). As mentioned earlier there are some exceptions to this rule. For example, the rule of taking the last valid information is not applied for the net job income, as one can see in the second job data of the first respondent. For this variable, it was considered not plausible to replace the missing income from the second subspell by data from the first subspell. Consequently, data on income is missing in the harmonized episode.

Figure 6: Data snapshot from duration spell file *spEmp*: spells and subspells

ID_t	wave	splink	spell	subspell	spgen	ts2311y	ts2312y	ts23410	ts23223	ts23228
8001138	1	260001	1	0	0	2006	2007	.	25	3
8001138	2	260002	2	0	1	2007	2009	.	36	3
8001138	1	260002	2	1	0	2007	2007	3800	40	3
8001138	2	260002	2	2	0	2007	2009	.	36	.
8001138	3	260003	3	0	1	2009	2010	.	10	3
8001138	2	260003	3	1	0	2009	2010	400	4	3
8001138	3	260003	3	2	0	2009	2010	.	10	.
8001204	1	260001	1	0	0	1984	1997	.	0	3
8001204	3	260002	2	0	1	2006	2011	400	10	2
8001204	1	260002	2	1	0	2006	2008	363	10	2
8001204	2	260002	2	2	0	2006	2010	400	10	.
8001204	3	260002	2	3	0	2006	2011	400	10	.

Harmonized spells are generated spells and thus can be easily distinguished from complete spells by the indicator variable *spgen*. This variable is coded 1 for all generated spells.

For most analyses it is reasonable to delete the subspells and keep only the generated harmonized episodes (*subspell* = 0 & *spgen* = 1) as well as the complete episodes (*subspell* = 0 & *spgen* = 0). Keeping all observations that satisfy the condition *subspell* = 0 returns a convenient harmonized spell structure with each row representing one episode.

Figure 7: Data snapshot from duration spell file *spEmp*, restricting to subspell=0

ID_t	wave	splink	spell	subspell	spgen	ts2311y	ts2312y	ts23410	ts23223	ts23228
8001138	1	260001	1	0	0	2006	2007	.	25	3
8001138	2	260002	2	0	1	2007	2009	.	36	3
8001138	3	260003	3	0	1	2009	2010	.	10	3
8001204	1	260001	1	0	0	1984	1997	.	0	3
8001204	3	260002	2	0	1	2006	2011	400	10	2

It should be noted, that the harmonized spells provided in the duration files represent just a simple approach for harmonizing time-varying spell data. It enables you to work with harmonized spell data out of the box. However, you might consider alternative methods of harmonizing which may be better suited for your substantial analysis. For instance, with regard to job net income you could consider taking the mean income over subspells rather than using the last reported value. Anyhow, you can easily implement your own way of harmonizing subspells.

Duration spell files

General education history: spSchool

This module covers each respondent's general education history from school entry until the date of (anticipated) completion, including

- episodes of elementary schooling,
- completed episodes of secondary schooling that led to a school leaving certificate, and
- incomplete episodes of schooling that would have led to a school leaving certificate if they had been completed.

A new episode is generated only if the school type changes. That is, a change from one Gymnasium to another is not recorded. As a result, a single schooling episode may take place at more than one location. In such cases, only information on the last location is included.

A new episode is generated at each school type change even if both schools offer the same certificate. Below you find an example for a person who took four schooling spells to obtain a secondary degree. During the first spell (April 1967 until July 1971), the person was enrolled in elementary school (*ts11204_ha* = 1) which does not award a certificate. Therefore, data on the variables for aspired (*ts11214*) and obtained (*ts11209*) certificates are missing. In the second spell, this person attended a comprehensive school (*ts11204_ha* = 3), aspiring the Abitur (*ts11214* = 5) but not attaining any certificate within this spell (*ts11209* = -5). The third spell represents a (futile) attempt at the Gymnasium (*ts11204_ha* = 5) which lasted from April 1976 until July 1977. Because the school type had changed, a new episode was generated although neither the aspired nor the attainable degree had changed. Back in comprehensive school, the person was finally awarded the Abitur in July 1980. Note that the aspired degree is set to missing because a degree was actually obtained within this episode.

Figure 8: Data snapshot *spSchool*

ID_t	wave	splink	spell	ts11204_ha	ts11214	ts11209	ts1111m	ts1111y	ts1112m	ts1112y
8002268	1	220001	1	1	.	.	4	1967	7	1971
8002268	1	220002	2	3	5	-5	8	1971	7	1976
8002268	1	220003	3	5	5	-5	4	1976	7	1977
8002268	1	220004	4	3	.	5	8	1977	7	1980

Vocational preparation schemes: spVocPrep

This module comprises episodes of vocational preparation after general education, including

- pre-training courses,
- basic vocational training years, and
- work preparation courses of the employment agency.

Data were collected on the duration from taking up until completing a vocational preparation scheme, including possible intermissions.

Vocational education history: spVocTrain

This module covers all further training, vocational and/or academic, that a respondent ever attended:

- Vocational training and retraining
- Training at technical schools, such as schools of public health, full-time vocational schools (excluding basic vocational training years), other vocational schools, and master craftsmen's colleges.
- Training in specialized fields of medicine
- Accredited training courses to receive licenses
- Conferral of a doctorate or postdoctoral thesis
- Tertiary education at universities, specialized colleges for higher education, colleges of advanced vocational studies, and colleges of advanced administrative and commercial studies.

  Note: Only the main subjects are surveyed. New episodes are generated if
  - a main subject changes over the course of studies, or
  - the attainable degree changes over the course of studies (e.g., from MA to teaching certification).

  Episodes are continued in case of location changes unless the main subjects change as well.

Training courses for licenses are comparable to courses in the *spCourses*, *spFurtherEdu1*, and *spFurtherEdu2* modules and can therefore be identified by the spell indicator *course*. This enumerator allows linking information about the few courses included in this module to the courses in the modules described below (also see below). Course IDs in spVocTrain are only available in wave 2.

Interruptions of vocational training spells, so-called vocational interruption episodes, are stored in wide format (be aware of this when working with harmonized spell data!).

Courses: spCourses

This module comprises courses and trainings attended within the past 12 months during episodes of employment, unemployment, parental leave, military or civilian service as well as episodes from the *spGap* module. The starting and end dates of the spells in this module represent the original episodes (from *spEmp*, *spUnemp*, etc.) in which a course was taken. For each of these episodes, information on up to three courses is included in wide format (variables with the suffix *_w* – see section 2.2); *spCourses* comprises all spells from the past 12 months which were recorded in the modules mentioned above. Spells may also be included if no course was taken during this episode. The only criterion for inclusion in the module is that a person provided information on at least one course.

The following example illustrates the data structure of this module. Data from two persons are displayed. The first person (*ID_t* = 8000523) reported 8 courses in the context of two employment spells (*sptype* = 26). In the past 12 months of wave 2, s/he attended courses within both employment spells (260007 and 260008), amounting to a total of five courses taken. At wave 3 interview s/he reported 3 further course attendances related again to job 260007. The second person (*ID_t* = 8000645) is represented by one employment spell and one unemployment spell. Note that both spells are included although this person only attended courses during the unemployment spell (*sptype* = 27).

Figure 9: Data snapshot *spCourses*

ID_t	wave	splink	sptype	t27800a	t27800b	t27800c	t27800d	course_w1	course_w2	course_w3
8000523	2	260007	26	4	1990	11	2009	1	2	3
8000523	2	260008	26	4	2001	11	2009	4	5	.
8000523	3	260007	26	11	2009	4	2011	301	302	303
8000645	2	260008	26	9	1999	9	2009	.	.	.
8000645	2	270001	27	10	2009	11	2009	1	2	3

Note that in *spCourses*, the course enumerator is stored in wide format (*course_w1*, *course_w2*, and *course_w3*), whereas in the other course modules (*spFurtherEdu1* and *spFurtherEdu2*) there is only a single enumerator (*course*) (also see the examples in section 6). You can merge *spCourses* with spell data using *ID_t*, *splink*, and *wave* as identifiers.

Additional courses: spFurtherEdu1

This module contains information on further courses (also private courses) attended in addition to courses reported in *spCourses* or in *spVocTrain*. These include both professional trainings (similar to those from *spCourses*) and courses attended for private purposes (e.g., cookery course, yoga course, fortune telling, NLP coaching) within the past 12 months. In contrast to *spCourses*, this module's spells refer to the actual starting and end dates of the courses.

German courses: spFurtherEdu3

Information on courses in German as a foreign language is collected only for immigrants. These data are stored in this module. The course enumerator for the German courses is *gcourse*. German courses have not been collected in wave 3 (second NEPS wave , 2010/11).

Military / civilian service and voluntary gap years: spMilitary

This module includes episodes of military or civilian service as well as gap years taken to do voluntary work in the social or environmental sector. Regular or professional soldiers are considered employed and are therefore included in the employment module.

Employment history: spEmp

This extensive module covers all spells of regular employment, including traineeships. Information on second jobs is only collected for activities that continue to the interview date. Vacation jobs, volunteering, and internships are not included.

New episodes are created at the following events:

- Change of employer
- Change of occupation
- Interruption of employment (e.g., unemployment or military service)

Unemployment history: spUnemp

This module includes all episodes of unemployment irrespective of whether a person was registered as unemployed or not. Questions on registration of unemployment and receipt of benefits refer both to the beginning and to the end of an unemployment spell.

Parental leave: spParLeave

For each child in *spChild* (except for deceased children), information is collected on whether the respondent took a parental leave. Each parental leave episode contributes one record to *spParLeave*.

Parental leaves do not include maternity protection. These periods are added to the corresponding employment episode. As a result, an employment spell is not interrupted if the mother only takes the maternity leave without an additional parental leave.

Cohabitation with children: spChildCohab

With release of NEPS SC6 3.0.0, we introduced the new file *spChildCohab*. It contains respondents' cohabitations with their children in a regular long spell format equivalent to duration spells like employments or parental leaves. Formerly, child cohabitation episodes have been provided in a wide format included in the *spChild* file. In the figure below, you find a small data snapshot from the *spChildCohab* file. Cohabitation spells are related to children by the *child* number. If cohabitation episodes are extended over panel waves, harmonized spells are generated (*subspell=0 & spgen=1*). Similar to the other spell files, you can easily work with the file by restricting observations to *subspell* having

value *0*. The data example below shows entries for three respondents. The first respondent (*ID_t = 8000335*) reported finished cohabitation episodes for two children. Respondent 8000338 reported cohabitation spells for three children in wave 2, all of which have been extended to wave 3. Consequently, for all three spells harmonized subspells are provided. Respondent 8001203 reported two cohabitation spells with two children. Both spells have been extended over three waves.

Figure 10: Data snapshot *spChildCohab*

ID_t	wave	child	spell	subspell	spgen	ts3331y	ts3332y
8000335	2	1	101	0	0	1978	2007
8000335	2	2	201	0	0	1980	2005
8000338	3	1	101	0	1	1994	2011
8000338	2	1	101	1	0	1994	2009
8000338	3	1	101	2	0	1994	2011
8000338	3	2	201	0	1	1996	2011
8000338	2	2	201	1	0	1996	2009
8000338	3	2	201	2	0	1996	2011
8000338	3	3	301	0	1	2000	2011
8000338	2	3	301	1	0	2000	2009
8000338	3	3	301	2	0	2000	2011
8001203	3	1	101	0	1	2001	2011
8001203	1	1	101	1	0	2001	2007
8001203	2	1	101	2	0	2001	2010
8001203	3	1	101	3	0	2001	2011
8001203	3	2	201	0	1	2005	2011
8001203	1	2	201	1	0	2005	2007
8001203	2	2	201	2	0	2005	2010
8001203	3	2	201	3	0	2005	2011

Gaps: spGap

Gaps in individual life courses are identified by a check module. Such gap episodes are included in the *spGap* module. The spells in this file refer to different types of gaps which can be distinguished by the variable *ts29101*. The most common gap episodes are (extended) holidays and looking after home or family.

Entity spell files

History of partners: spPartner

This module covers the partnership history of the respondent. Respondents' subjective reports define whether they live in a relationship and whether they cohabit or not. A comprehensive set of additional questions refers to the present partner. For earlier partners, only information on the year of birth and education is available. Information on the current partner is collected regardless of the cohabitation status, whereas previous partners are only included if they cohabitated with the respondent. The enumerator variable *partner* identifies partners "within" respondents. This variable is coded 1 for the first partner and counts upward until the last (current) partner.

The following example illustrates the data structure of this module. The respondent reported on four partners. Partners 1, 2,and 3 are previous partners. Remember that previous partners are only included in this module if they cohabited with the respondent. Cohabitation with partner 1 lasted from 1979 until 1985 (variables *ts3131y* and *ts3152y*). The respondent married this partner in 1981 (variable *ts3141y*) and divorced in 1985 (variable *ts3153y*). The second relationship was a consensual union which began in 1997 and ended in 2003. No data on cohabitation is available for the third and fourth partner. That is, information on this partner is recorded although s/he never cohabited with the respondent. This is only possible for current partners at date of interview who are included regardless of the cohabitation status. Compared to previous partners, information on current partners is available in greater detail. This example shows one of these variables (*ts31211*) which indicates whether the partner is German or not. This information was not collected retrospectively for previous partners. As a result, data on this variable is missing for partners 1 and 2. However, basic information like date of partner's birth (variable *ts3120y*) is recorded for all partners.

Figure 11: Data snapshot from *spPartner*

ID_t	wave	partner	ts3131y	ts3141y	ts3152y	ts3153y	ts31211	ts3120y
8000334	2	1	1979	1981	1985	1985	.	1954
8000334	2	2	1997	.	2003	.	.	1963
8000334	2	3	.	.	.	.	1	1964
8000334	3	4	.	.	.	.	1	-97

Children: spChild

This module contains information on

- all biological, foster, and adopted children of the respondent, and
- any other child that currently lives or has ever lived together with the respondent (e.g., children of former and current partners).

In cases of twins and higher orders of multiple births, separate episodes are generated for each child. Episodes generally refer to the periods in which the respondent and the child shared a household. The enumerator variable *child* identifies children within respondents. Note that a child episode was skipped in the interview if the respondent reported that the child was deceased. Spell data on cohabitation with children is stored in file *spChildCohab* and spell data on parental leaves relating to children is stored in *spParLeave*.

Selected courses: spFurtherEdu2

The survey instrument randomly selected two courses from the *spVocTrain*, *spCourses* and *spFurtherEdu1* modules, collecting additional information on these courses (e.g., costs, motivation, and certificates). These data are included in *spFurtherEdu2*.

4.3 Generated files

Method dataset: *Methods*

This dataset offers a variety of information on the data collection (e.g., age, gender and education of the interviewer; interview date; interview duration; incentives), individual survey participation, sampling design (e.g., strata variables), and weighting (design weights and calibrated design weights). Detailed information on the calculation and the use of weights is available in Supplement C (see 1.1). You will find data examples in section 6.6.

Basic information: *Basics*

This file contains the most recent basic information on each respondent (e.g., socio-demographic variables, current job and household characteristics). These data are generated from the file *pTarget* and a number of spell files (see below). The *Basics* file is updated prospectively. That is, the file is cross-sectional (i.e., one row per person) and always includes updated information from the latest panel wave (if available).

This simplified data structure can help to gain a first insight in the data. However, it should be handled with care, as it may not feature the "best" information about the respondent.

Integrated life course data: *Biography*

The *Biography* file is designed to facilitate the analysis of complex life course data that were collected both retro- and prospectively. This dataset pulls together episodes from the following duration spell files: *spSchool*, *spVocPrep*, *spMilitary*, *spVocTrain*, *spEmp*, *spUnemp*, *spGap*, and *spParLeave*.

In contrast to the "raw" life course data from each of these modules, the Biography file offers more consistent life course data that are thoroughly checked and edited. During the interview, inconsistencies in individual life course data were identified and corrected by the data revision module (also "check module" or "Prüfmodul" in German). Those corrected times can be found in the duration spell files as **_g1* variables (e.g. variable *ts2311y_g1* in spEmp contains the starting date of an employment spell as corrected by the check module). Those corrected times are the starting point for further corrections that have been implemented in the data editing process for *Biography*. Overall, the following measures were taken to ensure the integrity of the life course data in the *Biography* file:

- All subspells were removed; *Biography* includes only completed, harmonized, or right-censored episodes (i.e., *subspell* = 0).
- Episodes revoked by the respondents during the interview (i.e., disagreement in the introductory question for episode updating in the panel questionnaire) were deleted. Note that the revoked episodes are included in the original spell files and can be identified using the variable *spstat* (91 = spell missing in *Biography*).
- Starting and end dates of episodes were smoothed and corrected:

- One-month overlaps between adjacent episodes were resolved.
- Gaps between adjacent episodes which did not exceed two months were closed; gaps of more than two months were defined as specific gap episodes (edition gaps) within the *Biography* file.

The linking variables *ID_t* and *splink* allow matching information from the following duration spell files to the *Biography* file (see section 6 for examples): *spSchool*, *spVocPrep*, *spMilitary*, *spVocTrain*, *spEmp*, *spUnemp*, *spParLeave*, and *spGap*.

Therefore, we recommend using *Biography* as a starting point for life course analyses.

The example displayed below illustrates two respondents' life courses. Episodes follow a clear chronological order: The first respondent (*ID_t* = 8000342) records two school spells (*sptype* = 22) prior to a vocational training episode (*sptype* = 24). There is an edition gap (i.e., a generated spell that bridges a gap in the reported episodes of more than one month; *sptype* = 99) before the first employment spell (*sptype* = 26). This employment spell is right-censored (i.e., it continued until the interview date) and overlaps with a second vocational training episode (e.g., a course) which took place in 2002 (September – December). The second respondent (*ID_t* = 8000357) provided a complete educational and occupational biography without any gaps.

Figure 12: Data snapshot from integrated life course file *Biography*

ID_t	splink	sptype	startm	starty	endm	endy
8000342	220001	22	9	1987	8	1991
8000342	220002	22	9	1991	8	1997
8000342	240001	24	9	1997	5	2000
8000342	990001	99	6	2000	3	2001
8000342	260001	26	4	2001	6	2011
8000342	240002	24	9	2002	12	2002
8000357	220001	22	4	1959	3	1963
8000357	220002	22	4	1963	3	1965
8000357	220003	22	4	1965	7	1967
8000357	240001	24	8	1967	1	1970
8000357	260001	26	2	1970	3	1971
8000357	260002	26	4	1971	3	1973
8000357	260003	26	4	1973	6	1992
8000357	260004	26	7	1992	11	1996
8000357	260005	26	12	1996	8	1997
8000357	270001	27	9	1997	9	1998
8000357	260006	26	10	1998	3	2000
8000357	260007	26	4	2000	2	2010
8000357	300001	30	3	2010	12	2010

Many users may want to restrict their analyses to one life course domain such as the employment career. You can do this by selecting the corresponding spell type. In our example, this spell type is employment (*sptype = 26*).

Figure 13: Data snapshot from integrated life course file *Biography*, restricted to employment spells

ID_t	splink	sptype	startm	starty	endm	endy
8000342	260001	26	4	2001	6	2011
8000357	260001	26	2	1970	3	1971
8000357	260002	26	4	1971	3	1973
8000357	260003	26	4	1973	6	1992
8000357	260004	26	7	1992	11	1996
8000357	260005	26	12	1996	8	1997
8000357	260006	26	10	1998	3	2000
8000357	260007	26	4	2000	2	2010

The next screenshot presents the original dates included in *spEmp*. This allows a comparison between smoothed start and end dates in the *Biography* file (*startm*, *starty, endm, endy*) and the original dates of the complete and harmonized episodes (*subspell* = 0) in *spEmp* (*ts2311m, ts2311y, ts2312m, ts2312y*). Three corrections were executed for the second respondent (*ID_t* = 8000357). Information on the end month of the third employment spell and the start month of the fourth employment spell was not precise in the original data (*ts2312m* and *ts2311m*= 27, "middle of the year"). The upper screenshot shows that this value was replaced by 7 (July) for the start month of the fourth spell. To avoid an overlap, the end date of the previous spell was set to 6 (June). Another overlap in the original data occurred between the end month of the sixth and the starting month of the seventh spell (both *ts2311m* and *ts2312m* have the value 4, "April"). Again, the end date of the previous month was adjusted (*endm* = 3, "March").

Figure 14: Data snapshot from *spEmp* - originally reported spell times

ID_t	splink	subspell	ts2311m	ts2311y	ts2312m	ts2312y
8000342	260001	0	4	2001	6	2011
8000357	260001	0	2	1970	3	1971
8000357	260002	0	4	1971	3	1973
8000357	260003	0	4	1973	27	1992
8000357	260004	0	27	1992	12	1996
8000357	260005	0	12	1996	8	1997
8000357	260006	0	10	1998	4	2000
8000357	260007	0	4	2000	2	2010

Basic information on children: *Children*

This entity spell file was generated from the *spChild* module, offering basic information (e.g., current cohabitation state, cohabitation history) about all biological, step, foster, and adopted children as well as other children with whom the respondent has ever cohabited.

Transitions in educational careers: *Education*

This generated file provides longitudinal information on transitions in respondents' educational careers. It contains only persons who have an educational degree at a lower secondary level or higher. We used all information on educational attainment from

spSchool (lower, intermediate, and upper secondary school degrees – Hauptschule, Realschule, (Fach-)Abitur), *spVocPrep* (participation in vocational preparation schemes), and *spVocTrain* (all successfully completed trainings). Three measures of educational attainment are available: CASMIN (variable *tx28101*), ISCED-97 (*tx28103*), and years of education (*tx28102*; derived from CASMIN). You can easily merge data from the original spells to *Education* using the variable *splink*. Note that the ISCED-classification is a little more fine-grained in wave 2 and wave 3 (NEPS) compared to wave 1 (ALWA) because the measures of educational attainment were more differentiated in the NEPS survey instrument.

The file stores transitions in a long event time format. That is, each row represents a transition in at least one classification (CASMIN and/or ISCED-97). Variables on month and year of the transition (*datem* and *datey*) specify the event time. We considered only upward educational transitions in CASMIN levels and upward as well as lateral transitions in ISCED-97 levels (CASMIN is ordinal, whereas ISCED-97 has some nominal elements). Because ISCED-97 and CASMIN follow different concepts, some educational transitions (approximately 6-7 % in these data) are captured by only one of these classifications.

The following example illustrates the structure of this dataset. The first respondent (*ID_t* = 8000507) obtained a lower secondary degree (Hauptschulabschluss) in March 1966. This degree is represented by the value 1 both in the CASMIN (*tx28101*) and ISCED-97 (*tx28103*) classification. The variable *tx28109* indicates that a change was recorded in both classifications (denoted by the value 3). This always applies to the first event spell of a respondent in this dataset. In September 1969 (second event spell), the respondent completes a vocational training (Lehre). Consequently, CASMIN is set to the value 2 and ISCED-97 is set to the value 4. Because this upward transition concerns both classifications, the variable *tx28109* is again 3. Three years later (September 1972), the respondent experiences a vocational upward transition (e.g., master's qualification, Meister/in). This transition is only captured by the ISCED-97 classification (*tx28103* increases from 4 to 8) but not by the CASMIN classification which remains at the value 2 (i.e., CASMIN does not differentiate between basic and advanced vocational trainings). As a result, *tx28109* is set to the value 2, indicating that only ISCED-97 changed its value. The reverse is true for the fourth (and final) event spell of this respondent in which an educational upward transition is recorded. This change is only captured by the CASMIN classification. The corresponding value of CASMIN (*tx28101*) is 6, indicating that the respondent has attained an A-levels qualification (or equivalent) in addition to the vocational training that had already been completed. Therefore, *tx28109* has the value 1, denoting a change only in CASMIN. Note that the variable *sptype* specifies the source of the information from which these event spells were generated (22 = *spSchool*; 24 = *spVocTrain*).

Figure 15: Data snapshot from generated file *Education*

ID_t	splink	datem	datey	tx28101	tx28103	tx28109
8000507	220001	3	1966	1	1	3
8000507	240001	9	1969	2	4	3
8000507	240002	9	1972	2	8	2
8000507	220002	9	1974	6	8	1
8000512	220001	8	1968	1	1	3
8000512	240002	9	1974	2	4	3
8000512	240004	8	1986	7	9	3

Table 15 (p. 45) and Table 16 (p. 46) show how the ISCED-97 and CASMIN classes are composed.

Integrated course file: *FurtherEducation* (generated file)

Information about the respondents' participation in further education is distributed across several spell files. The generated file *FurtherEducation* integrates data on all courses from *spCourses* and *spFurtherEdu1* as well as vocational courses and trainings from *spVocTrain* into one consolidated format. In *FurtherEducation*, these courses are stored as duration spells in long format. Start and end dates of courses were imputed if this information was not precise (e.g., "spring") or missing. Since wave 3 (second NEPS wave 2010/11), starting and ending dates for further courses (spFurtherEdu1) have not asked anymore. Rather, respondents have been asked whether they attended any courses since the last interview. Hence, in this case we coded the last date of interview as starting date and the current date of interview as the ending date. Be aware that for these cases the start and end dates represents only the time interval wherein the course has been attended. The indicator variable *tx28201* codes whether the course dates have been asked directly or are derived by interview or episode dates. Data on the content of courses are available as open answers and in a coded version using a classification of the *Federal Employment Agency* (Kompetenzkatalog der Bundesagentur für Arbeit).

All respondents reporting at least one participation in further education are included in *FurtherEducation*. Note that in contrast to *spCourses* and *spFurtherEdu1*, this file contains not only course participations from the last year but also from the previous life course. The latter are vocational trainings reported in *spVocTrain* that can be classified as courses and trainings related to further education. The variable *course* (course number) allows tracking courses back to the original files *spCourses*, *spFurtherEdu1*, and *spVocTrain*. For a subset of courses that have a course number, further information from *spFurtherEdu2* can be added. Furthermore, there is a second subset of courses that can be linked to spells from *spVocTrain* or *spEmp* because they have been reported in the context of these spells or (in case of spells from *spVocTrain*) are directly derived from them. The variables *ID_t*, *course*, and *splink* allow matching these original spell data to *FurtherEducation*. Table 10 provides an overview which courses are included in *FurtherEducation* and to which spells they can be linked in the original files.

Table 10: Overview of courses included in FurtherEducation

course	splink	Description
valid	missing	Further education spell reported in the further education module (stored in *spFurtherEdu1*); spell is right-censored or ended within the past 12 months.
missing	24#### (Vocational Training)	Vocational training spells related to further education and participation(stored in *spVocTrain*); spell ended more than 12 months ago.
valid	24#### (Vocational Training)	Vocational training spells of the type "further education and participation" (stored in *spVocTrain*); spell is right-censored or ended within the past 12 months
valid	25#### (Military/Civilian Service)	Further education reported in the course module; triggered by spells in *spMilitary* (courses are stored in *spCourses*); triggering spell is right-censored or ended within the past 12 months
valid	26#### (Employment)	Further education reported in the course module; triggered by spells in *spEmp* (courses are stored in *spCourses*); triggering spell is right-censored or ended within the past 12 months
valid	27#### (Unemployment)	Further education reported in the course module; triggered by spells in *spUnemp* (courses are stored in *spCourses*); triggering spell is right-censored or ended within the past 12 months
valid	29#### (Parental Leave)	Further education reported in the course module; triggered by spells in *spParLeave* (courses are stored in *spCourses*); triggering spell is right-censored or ended within the past 12 months
valid	30#### (Gap)	Further education reported in the course module; triggered by spells in *spGap* (courses are stored in *spCourses*); triggering spell is right-censored or ended within the past 12 months

= wildcard character for an integer between 0 and 9

The following example (see Figure 16) illustrates the structure of this dataset. There are two respondents. The first one (*ID_t = 8000373*) reported two courses in wave 3. The first course was collected in the context of an employment spells (as you can see from the *splink* variable having non missing value and starting with 26). Hence the starting and ending date of the course is estimated by the starting and ending date of the employment spell. Indicator *tx28201* is coded as 1 which documents that the course times represent not starting and ending times directly asked but rather an estimated time interval in which the course might have been attended. The respondent's second course was collected in the further education module (*spFurtherEdu1*) since *splink* variable is missing. Since this course was collected in the third wave, starting and ending dates have not been directly asked. Rather, the dates which are generated refer to a time interval which starts at the date of last interview and closes with the date of the current interview.

The second respondent (*ID_t = 8000410*) reported courses in wave 2 and in wave 3. In the course of the wave 2 interview, the respondent reported on a total of three courses. Each of these courses was recorded by a different module and stored in a different dataset. The variable *tx28200* identifies the source dataset of a spell (24 = *spVocTrain*; 31 = *spFurtherEdu1*; 35 = *spCourses*). While start and end dates of courses 1 and 3 have been directly surveyed, dates of courses 2 and 4 are interval estimates.

Finally, the variable *tx28202_g13* includes coded data on the content of courses based on the classification of the *Federal Employment Agency* (Kompetenzkatalog der Bundesagentur für Arbeit).

Figure 16: Data snapshot from *FurtherEducation*

ID_t	wave	number	course	splink	tx28200	tx2821y	tx2822y	tx28201	tx28202_g13
8000373	3	1	301	260006	35	2009	2010	1	172
8000373	3	2	302	.	31	2009	2010	1	226
8000410	2	1	.	240002	24	2006	2006	2	319
8000410	2	2	1	260004	35	2008	2009	1	-55
8000410	2	3	2	.	31	2009	2009	2	215
8000410	3	4	301	260004	35	2009	2011	1	117

Marital history: *MaritalStates*

This file was generated from the *spPartner* module. It contains event time data on each respondent's marital states. The variable *tx27000* distinguishes between three categories: married, divorced, and widowed. Only persons who have married are included in this file. There is a variable *problem* marking and documenting problematic cases.

Regional data: *RegioInfas*

This file has been generated from the *infas geodaten* database.[5] It comprises geographical information on four regional levels: municipality, postal code, quarters (living areas), and street sections. These data were linked to each respondent by geocoding the sample addresses. Data are available only for respondents of the first NEPS wave. Hence, this file has a cross-sectional structure. A comprehensive documentation of this dataset is available in Supplement D (see 1.1).

Please note that these data are highly sensitive and thus can only be accessed on site. (see Section 1.3).

Regional data: *pTargetMicrom*

Respondent's addresses have been recoded to geocodes by *Microm*[6], thus rendering them linkable to *Microm's* own geo database. This information, included in the panel file *pTargetMicrom*, is similar to *RegioInfas*. Three regional levels are provided: house, street section, or postal code. However, *Microm* data only provides the smallest level. If there is no data available for a geocode on this level, the next level is chosen. The *Microm* file has a long panel format and provides regional data for wave 2 (first NEPS wave) and wave 3 (second NEPS wave) of SC6.

Please note that these data are highly sensitive and thus can only be accessed on site. (see Section 1.3).

[5] This database is provided by the infas geodaten GmbH, see: http://www.infas-geodaten.de
[6] Microm Consumer Marketing, see: http://www.microm-online.de

4.4 Overview of all datasets

Table 11 presents an overview of all datasets included in this release. Note that the number of respondents contained in each file varies markedly because specific modules only apply to certain subgroups of respondents ("universes"). For example, *spPartner* only includes those who currently have a partner and/or have ever cohabited with a partner.

Note

The NEPS invested a lot to ensure the integrity of these data. However, we strongly recommend you to examine the data critically when you work with this release. Furthermore, you should always consult the questionnaire/s to obtain a precise understanding of how the data have been collected. Finally, it is important to note that each additional dataset that we created for the users' convenience was generated on the basis of certain assumptions (e.g., event times in the file *Education* were calculated from the smoothed duration times in the file *Biography*). Please keep these assumptions in mind if you use data from the generated files.

Table 11: Overview of all data files in NEPS SC6 version 3.0.1

File	Content	Data Type	Universe	Row Identifier Variables	N rows	N persons
			Surveyed Data			
pTarget	Socio-demographic information + all cross sectional data; repeated cross-sectional measurements	Panel	All respondents	ID_t, wave	27,749	11,932
xTargetCompetencies	Competence data (second NEPS wave, 2010/2011)	Cross-section	All CAPI-respondents and respondents without severe debilities of sight or being blind	ID_t	7,256	7,256
xCompMethods	Interview para data on competence assessment	Cross-section	All respondents of second NEPS wave (2010/2011)	ID_t	9,322	9,322
spSchool	General education history	Spells (duration)	Resp. who attended general school and/or received a general school certificate	ID_t, splink, subspell	27,461	11,921
spMilitary	Military / civilian service	Spells (duration)	Resp. who served in military or civilian service or completed voluntary work in the social or environmental sector	ID_t, splink, subspell	4,518	3,918
spVocPrep	Vocational preparation schemes	Spells (duration)	Resp. who attended vocational preparation schemes	ID_t, splink, subspell	1,230	1,006
spVocTrain	Vocational education history	Spells (duration)	Resp. who (at least started) vocational training	ID_t, splink, subspell	27,644	11,503
spEmp	Employment history	Spells (duration)	Resp. who reported on at least one employment or traineeship	ID_t, splink, subspell	74,650	11,833
spUnemp	Unemployment history	Spells (duration)	Resp. who were unemployed (registered or unregistered) at least once	ID_t, splink, subspell	14,958	6,875
spParLeave	History of parental leaves	Spells (duration)	See *spChild*	ID_t, splink, subspell	6,949	3,507
spGap	Gap episodes	Spells (duration)	Resp. who reported gaps between labor market and educational activities	ID_t, splink, subspell	15,405	7,156
spPartner	History of partners in the household	Spells (entity)	Resp. who currently have a partner or ever cohabitated with a partner	ID_t, partner, subspell	34,874	11,098
spChild	History of children in the household	Spells (entity)	Resp. who have children and/or ever cohabitated with children	ID_t, child, subspell	54,346	8,778

spChildCohab	Cohabitation with children spells	Spells (duration)	See *spChild*	ID_t, spell, subspell	42,630	8,685
spCourses	Courses and trainings	Spells (entity)	Resp. who attended training courses during employment, unemployment, parental leaves, military/civilian service, or gap episodes.	ID_t, splink, wave	7,498	5,756
spFurtherEdu1	Additional courses	Spells	Resp. who attended further courses	ID_t, course	4,983	3,083
spFurtherEdu2	Detailed information on two randomly selected courses	Spells	Additional information on courses in *spVocTrain, spCourses*, and *spFurtherEdu1*	ID_t, course	11,205	5,653
spFurtherEdu3	Courses in German	Spells	Resp. with migration background who ever attended a German course	ID_t, gcourse	287	240
			Generated Files: Regional and Methodological Data			
Methods	Interview, data collection, sampling design, weights	Panel	All respondents	ID_t, wave	30,642	11,932
RegioInfas	Multilevel regional information from *infas Geodaten* database	Cross-section	All respondents	ID_t, regio	46,596	11,649
pTargetMircom	Regional information from *Microm Consumer Marketing* database	Panel	All respondents	ID_t, wave	23,572	11,932
			Generated Files: Derivations			
Basics	Generated: Most recent basic information on the respondent	Cross-section	All respondents	ID_t	11,932	11,932
Biography	Generated: Integrated and edited life course data	Spells	Respondents with at least one spell in *spSchool, spVocPrep, spMilitary, spVocTrain, spEmp, spUnemp, spParLeave*, and/or *spGap*	ID_t, splink	141,257	11,930
Children	Generated: Basic information to children of target	Spells	Respondents who have children and/or ever cohabitated with children	ID_t, child	18,059	8,778
Education	Generated: upward transitions in educational careers classified by CASMIN and ISCED-97	Events	Respondents who have an educational degree at a lower secondary level or higher	ID_t, number [or splink]	27,065	11,805
FurtherEducation	Integrated course file	Spells	Respondents who reported at least one participation in further education	ID_t, number	20,553	7,465
MaritalStates	Generated: Marital biography	Events	All respondents who ever married	ID_t, number	12,074	8,929

4.5 Merging the data

A number of identifiers allow combining information from different datasets. A unique and never-changing *ID_t* (identifier of target person) is assigned to each respondent. This identifier is required for all matching procedures. In *pTarget*, the variable *wave* further indicates in which wave(s) a respondent was observed. In case of spell data, additional variables are needed to uniquely identify observations within a dataset. There are five basic matching procedures:

1. Use *ID_t* to match data from *Basics* to all other datasets.
2. Use *ID_t* and *wave* to match data from *pTarget* and *Methods* to all other datasets.
3. Use *ID_t* and *splink* to match data from all duration spells (*spSchool, spVocPrep, spMilitary, spVocTrain, spEmp, spUnemp, spCourses, spGap, Education*) to the *Biography* file.
4. Use *ID_t* and *child* to match data from *spParLeave* to data from *spChild.*
5. Use *ID_t* and *course* to match data from *spFurtherEdu2* to *spVocTrain*, *spFurtherEdu1*, *spCourses* and/or the generated file *FurtherEducation*.

See section 6 for examples on each of these matching procedures. A comprehensive overview of all matching procedures is available in Supplement B (see section 1.1).

5 Generated Variables and Weights

5.1 Coding

All string variables on occupations of respondents, their parents, and partners were coded. Table 12 presents an overview to these coded variables and the variables that are derived from them as well as the CASMIN and the ISCED-97 classification which is particularly useful if you are interested in cross-national comparisons. Table 13 and Table 14 show the coding of Blossfeld's scheme and EGP.

Table 12: Overview of coded variables relating to education and occupations

Classification	Included in data file(s)	Description
KldB88	*spEmp; spVocTrain; pTarget; spPartner*	German Classification of Occupations 1988 (4-digit)
KldB2010	*spEmp; spVocTrain; pTarget; spPartner*	German Classification of Occupations 2010 (5-digit)
ISCO-88	*spEmp; spVocTrain; pTarget; spPartner*	International Standard Classification of Occupations 1988 (4-digit)
ISCO-08	*spEmp; spVocTrain; pTarget; spPartner*	International Standard Classification of Occupations 2008(4-digit)
BLK	*spEmp; spVocTrain; pTarget; spPartner*	Occupational classification by Blossfeld based on KldB-92 (Blossfeld, 1985; Schimpl-Neimanns, 2003)
ISEI-88	*spEmp; spVocTrain; pTarget; spPartner*	Metric scale to measure socio-economic status of occupations based on ISCO-08 (Ganzeboom et al., 1992; Ganzeboom & Treiman, 1996)
ISEI-08	*spEmp; spVocTrain; pTarget; spPartner*	Metric scale to measure socio-economic status of occupations based on ISCO-08 (Ganzeboom, 2010)
SIOPS-88	*spEmp; spVocTrain; pTarget; spPartner*	Metric scale to measure prestige of occupations based on ISCO-88 (Ganzeboom et al., 1992; Ganzeboom & Treiman, 1996)
SIOPS-08	*spEmp; spVocTrain; pTarget; spPartner*	Metric scale to measure prestige of occupations based on ISCO-88 and ISCO-08 respectively (Ganzeboom, 2010)
MPS	*spEmp; spVocTrain; pTarget; spPartner*	Magnitude prestige score of occupations (Wegener, 1985)
EGP	*spEmp; pTarget; spPartner*	Class scheme which assigns occupations to classes (Erikson et al., 1979)
CAMSIS	*spEmp; pTarget; spPartner*	Social Interaction and stratification scale (Prandy, 2000)
CASMIN	*pTarget; spPartner; Basics; Education*	Classification representing differentiated educational attainment and vocational training degrees
ISCED-97	*pTarget; spPartner; Basics; Education*	Classification representing differentiated educational attainment and vocational training degrees
Years of education	*pTarget; spPartner; Basics; Education*	Years of education based on the CASMIN classification

Table 13: Coding of Blossfeld's classification of occuptations

Key		English	German
1	[AGR]	Agricultural occupations	Agrarberufe
2	[EMB]	Common manual occupations	Einfache manuelle Berufe
3	[QMB]	Skilled manual occupations	Qualifizierte manuelle Berufe
4	[TEC]	Technician	Techniker
5	[ING]	Engineer	Ingenieure
6	[EDI]	Common services	Einfache Dienste
7	[QDI]	Skilled services	Qualifizierte Dienste
8	[SEMI]	Semiprofessions	Semiprofessionen
9	[PROF]	Professions	Professionen
10	[EVB]	Common commercial and adminstrative occupations	Einfache kaufmännische und Verwaltungsberufe
11	[QVB]	Skilled commercial and adminstrative occupations	Qualifizierte kaufmännische und Verwaltungsberufe
12	[MAN]	Manager	Manager

Table 14: Coding of EGP

Key		English	German
1	[I]	Higher Controllers	Obere Dienstklasse
2	[II]	Lower Controllers	Untere Dienstklasse mit hohen Qualifikationen
3	[IIIa]	Routine Non-manual	Angestellte der ausführenden nicht-manuellen Klasse mit beschränkten Entscheidungsbefugnissen
4	[IIIb]	Lower Sales-Service	Angestellte der ausführenden nicht-manuellen Klasse mit gering qualifizierten Routinetätigkeiten
5	[IVa]	Selfemployed with employees	Selbständige mit unterstellten Mitarbeitern
6	[IVb]	Selfemployed no employees	Selbständige ohne unterstellte Mitarbeiter
7	[IVc]	Selfemployed Farmer	Selbständige in der Landwirtschaft
8	[V]	Manual Supervisors	Arbeiter, Techniker, Facharbeiter
9	[VI]	Skilled Worker	Qualifizierte Arbeiter
10	[VIIa]	Unskilled Worker	Unqualifizierte Arbeiter
11	[VIIb]	Farm Labor	Landwirte

Table 15: Coding of ISCED-97

Codes in tx28103	ISCED-97		
	Key	**English**	**German**
-55		not determinable	nicht bestimmbar
0	0A/1A	Inadequately completed general education	kein Abschluss
1	2B	Lower general education	Haupt-, Volksschulabschluss, berufsvorbereitende Maßnahme
2	2A	Intermediate general education	Mittlere Reife, Realschulabschluss
3	3A	Full maturity certificates (e.g., the Abitur, A-levels)	Fachhochschulreife, Hochschulreife
4	3B	Basic vocational training, Vocational full time school, Health sector school (less than two years), civil servant of the lower grade, vocational basic skills	Lehre, Berufsfachschule, Fachschule des Gesundheitswesens (weniger als zwei Jahre), Beamter einfacher Dienst, berufliche Grundkenntnisse
5	3C	Civil servants of the medium grade	Beamter mittlerer Dienst
6	4A	Full maturity certificates (e.g., the Abitur, A-levels) (second cycle)	Fachhochschulreife, Hochschulreife (zweiter Bildungsweg)
7	4B	Basic vocational training, Vocational full time school, Health sector school (less than two years), civil servant of the lower grade, vocational basic skills (second cycle)	Lehre, Berufsfachschule, Fachschule des Gesundheitswesens (weniger als zwei Jahre), Beamter einfacher Dienst, berufliche Grundkenntnisse (zweiter Bildungsweg)
8	5B	Diploma (vocational and other specialised academies, College of public administration), Qualification of a two or three year Health-Sector School, Master's/technician's qualification	Fach- und Berufsakademische Abschluss, Verwaltungsfachhochschule, Fachschule des Gesundheitswesens (mindestens zwei Jahre), Meister/Techniker, anderer Fachschulabschluss, Beamter gehobener Dienst
9	5A	Bachelor, Master, Diploma, state examination, civil servants of the highest grade	Bachelor, Master, Diplom, Magister, Staatsexamen, Beamter höherer Dienst
10	6	Doctoral degree and postdoctoral lecture qualification	Promotion

Table 16: Coding of CASMIN

Codes in tx28101	CASMIN		
	Key	English	German
-55		not determinable	nicht bestimmbar
0	1a	Inadequatly completed general education	Kein Abschluss
1	1b	General elementary education	Hauptschulabschluss ohne berufliche Ausbildung
2	1c	Basic vocational training above and beyond compulsory schooling	Hauptschulabschluss mit beruflicher Ausbildung
3	2b	Intermediate general education	Mittlere Reife ohne berufliche Ausbildung
4	2a	Intermediate vocational qualification, or secondary programmes in which general intermediate schooling is combined by vocational training	Mittlere Reife mit beruflicher Ausbildung
5	2c_gen	General maturity: Full maturity certificates (e.g., the Abitur, A-levels)	Hochschulreife ohne berufliche Ausbildung
6	2c_voc	Vocational maturity: Full maturity certificates including vocationally specific schooling or training	Hochschulreife mit beruflicher Ausbildung
7	3a	Lower tertiary education: Lower level tertiary degrees, generally of shorter duration and with a vocational orientation	Fachhochschulabschluss
8	3b	Higher tertiary education: The completion of a traditional, academically orientated university education	Universitätsabschluss

5.2 Weights

Weight variables are included in the *Methods* file. Information on the construction of weights and how to use them can be found in the technical report on weighting (Supplement C) and the examples section (example 6). Furthermore, consult the field work reports for the first NEPS wave (study B72, Aust et al., 2011) and the second NEPS wave (study B67, Aust et al., forthcoming). Design und post-stratification weights are available for wave 2 (first NEPS wave). Wave 3 (second NEPS wave) includes probabilities of participation in wave 3 for wave 2 respondents. Note that no weights are available for wave 1, the ALWA forerunner study. The following table provides an overview to variables relevant for weighting and accounting for the sampling design.

Table 17: Variables for weighting and adjustment to sampling design

Variable	Description
psu	Primary sampling unit
stratum	Stratum identifier
weight_design_std	Standardized design weight; available for first NEPS wave (wave 2) participants
weight_design	Unstandardized design weight; available for first NEPS wave participants
weight_mc08_std	Standardized and calibrated design weight according (Mikrozensus 2008); available for first NEPS wave participants
weight_mc08	Unstandardized and calibrated design weight according (Mikrozensus 2008); available for first NEPS wave participants
weight_mc09_std	Standardized and calibrated design weight according (Mikrozensus 2009); available for first NEPS wave participants
weight_mc09	Unstandardized and calibrated design weight according (Mikrozensus 2009); available for first NEPS wave participants
prob_w3	Predicted probability of participation at second NEPS wave (wave 3) (for details see Aust et al., forthcoming). Available for all respondents that took part in first NEPS wave.[7] To calculate the longitudinal weight from first to second NEPS wave, you just have to multiply the standardized design weight from first NEPS wave with the inverse of participation probability (=1/prob_w3).
weight_isced_w3_std	Standardized ISCED calibrated longitudinal weight for second NEPS wave according to distributions of Mikrozensus 2010; account for sex and ISCED-97 level, year of birth and educational level, federal state, community size (BIK-10), and country of origin.
weight_isced_w3	Unstandardized ISCED calibrated weights for wave 3

[7] Calculated using logistic regression controlling for year of birth, sex, country of origin, mother tongue, size of household, education, parents education, income, federal state, and community size for respondents who gave a panel consent (see Aust et al, forthcoming).

6 Examples

This section gives some examples of how to merge different datasets and how to use the weights from this release. We provide you with the code to run the examples in R, SPSS, and Stata.

6.1 Example 1 – Merging *Basics* with other datasets

Variables from the cross-sectional file *Basics* can easily be merged to all other datasets of this release. In the example shown below we merge data on the respondent's gender and the father's EGP class (when the respondent was aged 15) to the employment spell file (*spEmp*).

Example 1 in R

```
# Merge information from Basics to other files

library(foreign)

read.dta("SC6_spEmp_D_3-0-1.dta", convert.factors=FALSE) -> spEmp
read.dta("SC6_Basics_D_3-0-1.dta", convert.factors=FALSE) -> Basics

Basics[,c("ID_t","t700001","t731453_g8")] -> Basics.temp

merge(spEmp,Basics.temp,by="ID_t", all.x=TRUE) -> spEmp.temp

# gender and fathers's EGP class could be merged to all 59,266 spells:
sum(table(spEmp.temp$t700001))

# gender and fathers's EGP class could be merged to all 59,266 spells:
sum(table(spEmp.temp$t731453_g8))
```

Example 1 in SPSS

```
* Procedure.
* 1. Open spEmp and Basics, sort by ID_t.
* 2. Merge variables from Basics to spEmp with Basics as key table.

GET FILE='SC6_Basics_D_3-0-1.sav'.
DATASET NAME Basics WINDOW=FRONT.

SORT CASES BY ID_t.

GET FILE='SC6_spEmp_D_3-0-1.sav'.
DATASET NAME spEmp WINDOW=FRONT.

SORT CASES BY ID_t.

DATASET ACTIVATE spEmp.

MATCH FILES /FILE=*
  /TABLE='Basics'
  /KEEP  ID_t TO ts23101 /* all variables of spEmp */
  t700001 t731453_g8 /* variables to keep from Basics */
  /BY ID_t
  /MAP.

DATASET CLOSE Basics.

* gender and fathers's EGP class could be merged to all 74,650 spells.
FREQUENCIES VARIABLES=t700001 t731453_g8.

DATASET CLOSE spEmp.
```

Example 1 in Stata

```
*Merge information from Basics to other files

/*
Procedure
1. Open spEmp
2. Merge variables from Basics to spEmp with a m:1-merge
*/

use "SC6_spEmp_D_3-0-1.dta", clear

merge m:1 ID_t using "SC6_Basics_D_3-0-1", keepusing(t700001 t731453_g8) keep(1 3)

tab _merge // gender and father's EGP class were matched to all 74,650 spells
```

6.2 Example 2 – Merging *pTarget* with other datasets

Virtually everyone who works with spell files will draw on information stored in *pTarget*, such as the respondents' gender. If you merge a spell file with *pTarget*, you should keep in mind that *pTarget* is a long-format file. If you want to merge time-constant information such as gender with a spell file (e.g., *spEmp*), you only need information from one record of each respondent in *pTarget*.

Example 2 in R

```
# Procedure
# 1. read pTarget and and select only one record for each respondent and
#    some variables
# 2. read spEmp
# 3. Merge

library(foreign)

read.dta("SC6_pTarget_D_3-0-1.dta", convert.factors=FALSE) -> pTarget

pTarget[order(pTarget$ID_t,-pTarget$wave) # sort by ID_t (asc), and wave descending
        ,c("ID_t","wave","t700001")       # only keep the required variables
        ] -> pTarget.temp

# keeps first occurrence of each ID_t with two observations in the data:
pTarget.temp[!duplicated(pTarget.temp$ID_t),] -> pTarget.temp

read.dta("SC6_spEmp_D_3-0-1.dta", convert.factors=FALSE) -> spEmp

merge(spEmp,pTarget.temp,by="ID_t",all.x=TRUE) -> spEmp.temp

sum(table(spEmp.temp$t700001)) #gender could be merged to all 74,650 spells
```

Example 2 in SPSS

```
* Procedure.
* 1. Open pTarget and select only one (the last) record for each respondent.
* 2. Open spEmp
* 3. add the gender variable from pTarget to spEmp, using pTarget as key table.


GET FILE='SC6_pTarget_D_3-0-1.sav'
  /KEEP ID_t wave t700001.
DATASET NAME pTarget WINDOW=FRONT.

SORT CASES BY ID_t (A) wave (D) /*sort by ID_t ascending and wave descending*/.

MATCH FILES
  /FILE=*
  /BY ID_t
  /FIRST=PrimaryFirst.

VARIABLE LABELS  PrimaryFirst 'Dummy for first row'.
VALUE LABELS  PrimaryFirst 0 'subsequent row' 1 'first row'.

SELECT IF PrimaryFirst = 1.
EXECUTE.

DELETE VARIABLES wave PrimaryFirst.

GET FILE='SC6_spEmp_D_3-0-1.sav'.
DATASET NAME spEmp WINDOW=FRONT.

MATCH FILES /FILE=*
  /TABLE='pTarget'
  /BY ID_t.

DATASET CLOSE pTarget.

* gender (t700001) could be matched to all 74,650 spells.
FREQUENCIES VARIABLES = t700001.

DATASET CLOSE spEmp.
```

Example 2 in Stata

```
/*
Procedure
1. Open pTarget and select only last wave record for each respondent using "duplicates
drop"
2. Save a temporary version (helpfile) of the reduced pTarget
3. Open spEmp and add the gender variable from our helpfile using a m:1-merge
*/

***

use "SC6_pTarget_D_3-0-1.dta", clear

keep ID_t wave t700001  // only keep the required variables

gsort +ID_t -wave // sort by ID_t ascending and wave descending

duplicates drop ID_t, force // drops all but the first record of each ID_t that has more
than 1 observations in the data

***

tempfile helpfile // defines the local macro "helpfile" as a temporary file
save `helpfile', replace // saves the information to be merged

***

use "SC6_spEmp_D_3-0-1.dta", clear

merge m:1 ID_t using `helpfile', keep(1 3)

tab _merge // gender was merged to all 74,650 spells
```

Note that merging time-variant panel variables (e.g., income) to a spell file is much more complicated because you have to deal with different time axes in the files you intend to merge. Whereas a row in *pTarget* represents a year in which the respondent participated in the survey, a record in a spell file corresponds to one specific episode (e.g., an employment spell) or entity (e.g., a partner).

6.3 Example 3 – Merging duration spells with Biography

This example illustrates how to merge the smoothed and corrected starting and end dates of the *Biography* file with the employment history (*spEmp*). Because the *Biography* file includes only harmonized or completed episodes, you have to delete subspells (*subspell* > 0) before merging data from duration spells with the *Biography* file.

Example 3 in R

```
# Procedure 1:
# 1. Read Biography
# 2. Read spEmp and # only keep harmonized and completed spells
# 2. Merge spEmp with Biography using ID_t and splink as key variables (1:1-merge)

library(foreign)

read.dta("SC6_Biography_D_3-0-1.dta", convert.factors=FALSE) -> Biography
read.dta("SC6_spEmp_D_3-0-1.dta", convert.factors=FALSE) -> spEmp

spEmp[spEmp$subspell==0,] -> spEmp.temp # only keep harmonized and completed spells

merge(spEmp.temp,Biography,by=c("ID_t","splink"),all.x=TRUE) -> spEmp.temp

# 50,067 episodes merged; 257 episodes are included only in master file
table(spEmp.temp$sptype, useNA="always")

#***********************

# Procedure 2:
# 1. Read spEmp
# 2. Read Biography and select the emplyment spells
# 3. Merge part of spEmp to Biography using ID_t and splink as key variables

read.dta("SC6_spEmp_D_3-0-1.dta", convert.factors=FALSE) -> spEmp

read.dta("SC6_Biography_D_3-0-1.dta", convert.factors=FALSE) -> Biography

Biography[Biography$sptype== 26,] -> Biography.temp # only keep employment spells

merge(Biography.temp,spEmp[spEmp$subspell==0,],by=c("ID_t","splink"),all.x=TRUE)
  -> Biography.temp

table(Biography.temp$sptype, useNA="always") # 50,067 episodes merged
```

Example 3 in SPSS

```
* Procedure 1: Merge Biography to spEmp.
* 1. Open Biography, select employment spells.
* 2. Open spEmp, select harmonized and completed spells.
* 3. Match files, and generate source indicator.

GET FILE='SC6_Biography_D_3-0-1.sav'.
DATASET NAME Biography WINDOW=FRONT.

* select employment spells.
SELECT IF sptype = 26.

SORT CASES  ID_t splink (A).

GET FILE='SC6_spEmp_D_3-0-1.sav'.
DATASET NAME spEmp WINDOW=FRONT.

* only keep harmonized and completed spells.
SELECT IF subspell = 0.

SORT CASES  ID_t splink (A).

MATCH FILES /FILE=*
  /FILE='Biography'
  /IN source
  /BY ID_t splink
  /DROP=wave /* wave in Biography */
  /MAP.

VARIABLE LABELS source 'Dates from Biography matched'.
VALUE LABELS  source 0 'no' 1 'yes'.

DATASET CLOSE Biography.

* 50,067 episodes merged; 257 episodes are included only in master file
FREQUENCIES VARIABLES = source.

DATASET CLOSE spEmp.

*For Procedure 2 see next page.
```

Example 3 in SPSS – continued

```
* Procedure 2: Merge spEmp to Biography
* 1. Open Biography, select employment spells.
* 2. Open spEmp, select harmonized and completed spells.
* 3. Match files using spEmp as key table.

GET FILE='SC6_Biography_D_3-0-1.sav'.
DATASET NAME Biography.

SELECT IF sptype = 26.

SORT CASES  ID_t splink (A).

GET FILE='SC6_spEmp_D_3-0-1.sav'.
DATASET NAME spEmp WINDOW=FRONT.

* only keep harmonized and completed spells.
SELECT IF subspell = 0.

SORT CASES  ID_t splink (A).

MATCH FILES /TABLE=*
  /FILE='Biography'
  /BY ID_t splink
  /DROP=wave /* wave in Biography */
  /MAP.

DATASET CLOSE Biography.

FREQUENCIES VARIABLES = subspell.

DATASET CLOSE spEmp.
```

Example 3 in Stata

```
/*
Procedure 1: Merge Biography to spEmp
1. Open spEmp and delete subspells 1 and 2
2. Merge spEmp with Biography using ID_t and splink as key variables (1:1-merge)
*/

***

use "SC6_spEmp_D_3-0-1.dta", clear

keep if subspell == 0 // only keep harmonized and completed spells

***

merge 1:1 ID_t splink using "SC6_Biography_D_3-0-1", keep(1 3)

tab _merge // 50,067 episodes merged; 257 episodes are included only in master file

************************

/*
Procedure 2: Merge spEmp to Biography
1. Open Biography and select employment spells
2. Merge spEmp to Biography using ID_t and splink as key variables (1:m-merge)
*/

***

use "SC6_Biography_D_3-0-1", clear

keep if sptype == 26 // keep employment spells

***

*Note: 1:m-merge is necessary because spEmp contains subspells 1 and 2
merge 1:m ID_t splink using "SC6_spEmp_D_3-0-1.dta", keep(1 3)

keep if subspell == 0 // keep harmonized and completed spells

tab _merge // 50,067 episodes merged
```

The example illustrated two different approaches to merging data from the *Biography* file with the *spEmp* module. Note that the first approach yields more observations after merging has been completed. This is because *spEmp* still contains episodes revoked by the respondents during the interview.

6.4 Example 4 – Merge spParLeave with spChild

If you want to link information about the respondents' children to the corresponding parental leave episodes, you have to use the key variables *ID_t* and *child*. In this example, we merge information on the child's gender (*ts33203*) and year of birth (*ts3320y*) to the parental leave file.

Example 4 in R

```
# Procedure
# 1. Read spChild, select parts of it
# 2. Read spParleave
# 3. Merge

library(foreign)

read.dta("SC6_spChild_D_3-0-1.dta", convert.factors=FALSE) -> spChild

spChild[
  spChild$subspell==0,                   # only keep harmonized and completed spells
  c("ID_t","child","ts3320y","ts33203")  # only keep the required variables
  ] -> spChild.temp

read.dta("SC6_spParLeave_D_3-0-1.dta", convert.factors=FALSE) -> spParLeave

merge(spParLeave,spChild.temp,by=c("ID_t","child"),all.x=TRUE) -> spParLeave.temp

sum(table(spParLeave.temp$child, useNA="always")) # information on 6,949 children merged
```

Example 4 in SPSS

```
* Procedure.
* 1. Open spChild and select completed and harmonized spells.
* 2. Open spParLeaveSave a temporarily version (helpfile) of the reduced spChild.
* 3. Match files using spChild as key table.

GET FILE='SC6_spChild_D_3-0-1.sav'
  /KEEP=ID_t child ts3320y ts33203 subspell /* only keep the required variables */ .

DATASET NAME spChild WINDOW=FRONT.

SELECT IF subspell = 0 /*only keep harmonized and completed spell*/ .
* spChild _now_ unique in ID_t, child.
EXECUTE.

DELETE VARIABLES subspell.

SORT CASES BY ID_t child.

GET FILE='SC6_spParLeave_D_3-0-1.sav'.
DATASET NAME spParLeave WINDOW=FRONT.

SORT CASES BY ID_t child.

MATCH FILES /FILE=*
  /TABLE='spChild'
  /BY ID_t child
  /MAP.

DATASET CLOSE spChild.

* information on 6,949 children merged.
FREQUENCIES VARIABLES=child.

DATASET CLOSE spParLeave.
```

Example 4 in Stata

```
/*
Procedure
1. Open spChild and select spells that are completed and harmonized
2. Save a temporary version (helpfile) of the reduced spChild
3. Open spParLeave and add information from the helpfile using a m:1-merge
*/

***

use "SC6_spChild_D_3-0-1", clear
keep if subspell == 0 // only keep harmonized and completed spells
keep ID_t child ts3320y ts33203 // only keep the required variables

***

tempfile helpfile // defines the local macro helpfile as a temporary file
save `helpfile', replace // saves the information to be merged

***

use "SC6_spParLeave_D_3-0-1", clear
merge m:1 ID_t child using `helpfile' , keep(1 3)

tab _merge // information on 6,949 children merged
```

6.5 Example 5 – Merge course data

Data on courses are stored in several files of this release. Some basic information on courses which the respondent attended during the 12 months before the interview can be found in *spVocTrain*, *FurtherEdu1*, and *spCourses*; *spFurtherEdu2* contains more detailed information on two randomly selected courses from these three files.

If you want to merge *spFurtherEdu2* to the other modules, remember that courses are stored in different formats across the files. In *spVocTrain* and *spFurtherEdu1*, courses are stored in spell format. Therefore, they can be easily merged with *spFurtherEdu2* using *ID_t* and course as key variables (see examples 1 and 2). However, courses in *spCourses* are stored in wide format. Here the data must be reshaped into long format before they can be merged with *spFurtherEdu2* (see example 3).

Example 5-1 in R

```
# Example 5-1:  spFutherEdu2 to spVocTrain

# Procedure
# 1. Read spFurtherEdu2#
# 2. Read spVocTrain and select the course spells
# 2. Merge and add detailed information on the courses from spFurtherEdu2

library(foreign)

read.dta("SC6_spFurtherEdu2_D_3-0-1.dta", convert.factors=FALSE) -> spFurtherEdu2
read.dta("SC6_spVocTrain_D_3-0-1.dta", convert.factors=FALSE) -> spVocTrain

spVocTrain[!is.na(spVocTrain$course),] -> spVocTrain.temp

merge(spVocTrain.temp,spFurtherEdu2,by=c("ID_t","course"),all.x=TRUE) -> spVocTrain.temp

# details available for 25 courses
xtabs(~.,data=spVocTrain.temp[,c("wave.x","wave.y")],na.action=na.pass,exclude=NULL)
```

Example 5-2 in R

```
# Example 5-2:  spFutherEdu2 to spFutherEdu1

# Procedure:
# 1. Read spFutherEdu1 and spFurtherEdu2
# 2. Merge and add detailed information on the courses from spFurtherEdu2

library(foreign)

read.dta("SC6_spFurtherEdu1_D_3-0-1.dta", convert.factors=FALSE) -> spFurtherEdu1
read.dta("SC6_spFurtherEdu2_D_3-0-1.dta", convert.factors=FALSE) -> spFurtherEdu2

merge(spFurtherEdu1,spFurtherEdu2,by=c("ID_t","course"), all.x=TRUE) -> spFurtherEdu1.temp

xtabs(~.,data=spFurtherEdu1.temp[,c("wave.x","wave.y")],na.action=na.pass,exclude=NULL)
# details available for 2,967 courses
```

Example 5-3 in R

```
# Example 5-3:  spFutherEdu2 to reshaped spCourses

# Procedure
# 1. Read spFurtherEdu2
# 2. Read spCourses and select course specific variables
# 3. Reshape spCourses from wide to long format
# 4. Select only spells with courses
# 5. Add detailed information on the courses from spFurtherEdu2 where possible

library(foreign)

read.dta("SC6_spFurtherEdu2_D_3-0-1.dta", convert.factors=FALSE) -> spFurtherEdu2
read.dta("SC6_spCourses_D_3-0-1.dta", convert.factors=FALSE) -> spCourses

spCourses[,c("ID_t","sptype","splink","subspell",
             "course_w1","t271011_w1","t271012_w1","t271013_w1",
             "course_w2","t271011_w2","t271012_w2","t271013_w2",
             "course_w3","t271011_w3","t271012_w3","t271013_w3")] -> spCourses.temp

reshape(spCourses.temp,direction="long",
        idvar=c("ID_t","splink","subspell"),
        varying=5:16,
        v.names=c("course","t271011","t271012","t271013")
        ) -> spCourses.temp

spCourses.temp[!is.na(spCourses.temp$course),] -> spCourses.temp

merge(spCourses.temp,spFurtherEdu2,by=c("ID_t","course"),all.x=TRUE) -> spCourses.temp

# details available for 8,212 courses:
table(spCourses.temp$wave,useNA="always")
```

Example 5-1 in SPSS

```
* Example 5-1:  spFurtherEdu2 to spVocTrain.

* Procedure.
* 1. Open spFurtherEdu2 and add indicator variable.
* 2. Open spVocTrain, select course spells and add indicator variable.
* 3. match files with spFurtherEdu2 as key table.
* 4. inspect indicator variables.

GET FILE='SC6_spFurtherEdu2_D_3-0-1.sav'.
DATASET NAME spFurtherEdu2 WINDOW=FRONT.

SORT CASES ID_t course.

COMPUTE spFurtherEdu2=1.

GET FILE='SC6_spVocTrain_D_3-0-1.sav'.
DATASET NAME spVocTrain WINDOW=FRONT.

SORT CASES ID_t course.

SELECT IF MISSING(course) = 0.

COMPUTE spVocTrain=1.
EXECUTE.

MATCH FILES /FILE=*
  /TABLE='spFurtherEdu2'
  /BY ID_t course
  /DROP= wave
  /MAP.

DATASET CLOSE spFurtherEdu2.

* details available for 25 courses.
IF MISSING(spVocTrain) spVocTrain=0.
IF MISSING(spFurtherEdu2) spFurtherEdu2=0.
CROSSTABS spFurtherEdu2 BY spVocTrain.

DATASET CLOSE spVocTrain.
```

Example 5-2 in SPSS

```
* Example 5-2:  spFutherEdu2 to spFutherEdu1.

* Procedure.
* 1. Open spFurtherEdu2 and add indicator variable.
* 2. Open spFurtherEdu1.
* 3. Match files using spFurtherEdu2 as key table.  on the courses (spFurtherEdu2) where
possible.
* 4. Inspect indicator variable.

GET FILE='SC6_spFurtherEdu2_D_3-0-1.sav'.
DATASET NAME spFurtherEdu2 WINDOW=FRONT.

SORT CASES ID_t course.

COMPUTE FurtherEdu2 = 1.
EXECUTE.

GET FILE='SC6_spFurtherEdu1_D_3-0-1.sav'.
DATASET NAME spFurtherEdu1 WINDOW=FRONT.

SORT CASES ID_t course.

MATCH FILES /FILE=*
  /TABLE='spFurtherEdu2'
  /DROP = wave
  /BY ID_t course
  /MAP.

DATASET CLOSE spFurtherEdu2.

if MISSING(FurtherEdu2) FurtherEdu2=0.
FREQUENCIES VARIABLES = FurtherEdu2.

DATASET CLOSE spFurtherEdu1.
```

Example 5-3 in SPSS

```
* Example 5-3:  spFutherEdu2 to reshaped spCourses.

* Procedure
* 1. Open spFurtherEdu2 and add indicator variable.
* 2. Open spCourses, reshape dataset (wide to long) and drop rows without course
information.
* 3. Match files with spFurtherEdu2 as key table.
* 4. inspect indicator variable.

GET FILE='SC6_spFurtherEdu2_D_3-0-1.sav'.
DATASET NAME spFurtherEdu2 WINDOW=FRONT.

SORT CASES ID_t course.

COMPUTE FurtherEdu2 = 1.
EXECUTE.

GET FILE='SC6_spCourses_D_3-0-1.sav'.
DATASET NAME spCourses WINDOW=FRONT.

VARSTOCASES
  /MAKE course FROM course_w1 course_w2 course_w3
  /MAKE t271011 FROM t271011_w1 t271011_w2 t271011_w3
  /MAKE t271012 FROM t271012_w1 t271012_w2 t271012_w3
  /MAKE t271013 FROM t271013_w1 t271013_w2 t271013_w3
  /KEEP=ID_t wave splink
  /NULL=KEEP.

* drop generated rows which don't store any course information
SELECT IF missing(course) = 0.

SORT CASES ID_t course.

MATCH FILES /FILE=*
  /TABLE='spFurtherEdu2'
  /BY ID_t course
  /MAP.
DATASET CLOSE spFurtherEdu2.

if MISSING(FurtherEdu2) FurtherEdu2=0.
* details available for 8,212 courses.
FREQUENCIES VARIABLES = FurtherEdu2.

DATASET CLOSE spCourses.
```

Example 5-1 in Stata

```
*Example 5-1: spFurtherEdu2 to spVocTrain

/*
Procedure
1. Open spVocTrain and select the course spells
2. Add detailed information on the courses (spFurtherEdu2) where possible
*/

***

use "SC6_spVocTrain_D_3-0-1", clear

keep if !missing(course)

***

merge 1:1 ID_t course using "SC6_spFurtherEdu2_D_3-0-1", keep(1 3)

tab _merge // details available for 25 courses
```

Example 5-2 in Stata

```
*Example 5-2: spFurtherEdu2 to spFurtherEdu1

/*
Procedure:
1. Open spFutherEdu1
2. Add detailed information on the courses (spFurtherEdu2) where possible
*/

use "SC6_spFurtherEdu1_D_3-0-1", clear

merge 1:1 ID_t course using "SC6_spFurtherEdu2_D_3-0-1", keep(1 3)

tab _merge // details available for 2,967 courses
```

Example 5-3 in Stata

```
*Example 5-3: spFurtherEdu2 to reshaped spCourses

/*
Procedure
1. Open spCourses and select course-specific variables
2. Reshape dataset from wide to long format
3. Prepare the reshaped dataset for merging
4. Add detailed information on the courses (spFurtherEdu2) where possible
*/

***

use "SC6_spCourses_D_3-0-1", clear

drop t278000-t271001  // drop unimportant variables
drop t272011_w2R t272011_g13w2 t272011_w3R t272011_g13w3 t272011_w1R t272011_g13w1

***

reshape long course_w  t271011_w t271012_w t271013_w, i(ID_t wave splink)

drop if missing(course_w)     // drop generated rows which don't store any course
information

drop _j

***

*Removing the _w-suffixes

//For Stata 12 (or newer) users: use the enhanced rename command
rename (*_w) (*)

//Alternative for older Stata versions
/*
* use looped regular expression
foreach var of varlist *_w {
        local newvar = regexr("`var'","(_w[O]?)$","")
        rename `var' `newvar'
}
*/

merge 1:1 ID_t course using "SC6_spFurtherEdu2_D_3-0-1", keep(1 3)

tab _merge // details available for 8,212 courses
```

6.6 Example 6 – Accounting for sample stratification and using weights

The file *Methods* contains variables for sample stratification as well as weights. This information can be used to correctly estimate population parameters. We present two examples. In example 6-1 we will analyze first NEPS wave data while accounting for sample stratification and the using design weight. Example 6-2 shows how you can construct longitudinal weights that adjust for unit non-response in the second NEPS wave.

Example 6-1 in R

```
# Example 6-1: Analyzing first NEPS wave while accounting for sample stratification and
# using weights
# Procedure:
# 1. Prepare Methods-File for getting sampling and weighting information
# 2. Merge methodological information to Basics-File
# 3. Recode some variables (set some NAs)
# 4. Use library survey and set options
# 5. Generate weight-version from Basics
# 6. Compute frequencies, mean and regression with survey

library(foreign)

read.dta("SC6_Methods_D_3-0-1.dta", convert.factors=FALSE) -> Methods

Methods[
  Methods$wave!=1,                                #remove wave records from 2007/2008 (ALWA)
  c("ID_t","psu","stratum","weight_design_std")  #keep relevant variables
  ] -> Methods.temp

read.dta("SC6_Basics_D_3-0-1.dta", convert.factors=FALSE) -> Basics

merge(Basics,Methods.temp,by="ID_t") -> Basics.temp

Basics.temp[Basics.temp$t510010_g1<=-5,]$t510010_g1 <- NA
Basics.temp[Basics.temp$tx28101<=-5,]$tx28101 <- NA

library(survey)
options("survey.lonely.psu")
options(survey.lonely.psu="certainty")

Basics.weight <- svydesign(ids=~psu, strata=~stratum, data=Basics.temp,
weights=~weight_design_std, nest=TRUE)

svytable(~t700001,Basics.weight)
svymean(~tx29000, Basics.weight, na.rm=T)
csregress <- svyglm(t510010_g1 ~ as.factor(t700001)+as.factor(tx28101)+as.factor(tx27000),
Basics.weight)
summary(csregress)
```

Example 6-2 in R

```
# Example 6-2: Analyzing second NEPS wave while accounting for sample stratification and
# using longitudinal weights
# Procedure:
#  1. Prepare Methods file to obtain sampling and weighting information for wave 3
#  2. Merge this information to the Basics file
library(foreign)
library(survey)

Methods <- read.dta("SC6_Methods_D_3-0-1.dta", convert.factors=FALSE)

# keep variables for primary sampling unit and stratum identifier as well as
# standardized design weight and probability of participation at second NEPS wave for
# first NEPS wave respondents
Methods.wave3   <- Methods[
  Methods$wave==3, #keep only second NEPS wave
  c("ID_t","psu","stratum", "prob_w3") #keep relevant variables
  ]

# carry forward design weight to second NEPS wave (i.e. wave=3)
Methods.weights <- Methods[ Methods$wave==2, c("ID_t","weight_design_std") ]
Methods.temp    <- merge(Methods.weights,Methods.wave3,by="ID_t")

# calculate longitudinal weight
Methods.temp$longweight <- Methods.temp$weight_design_std * (1/Methods.temp$prob_w3)
# keep cases with valid weight (cases participating in both waves)
Methods.temp <- Methods.temp[!is.na(Methods.temp$longweight),]

# merge gender and age at interview from Basics file
Basics         <- read.dta("SC6_Basics_D_3-0-1.dta", convert.factors=FALSE)
Basics.temp    <- Basics[, c("ID_t", "t700001", "tx29000")]
Methods.Basics <- merge(Methods.temp,Basics.temp,by="ID_t")

# define complex survey data structure to adjust standard errors
options(survey.lonely.psu="certainty")
Basics.weight <- svydesign(ids=~psu, strata=~stratum, data=Methods.Basics,
weights=~longweight, nest=TRUE)
svymean(~tx29000, Basics.weight, na.rm=T)
```

Example 6-1 in SPSS

```
*Example 6-1: Analyzing first NEPS wave while accounting for sample stratification and
using weights
* Procedure.
* 1: Open Methods-File and prepare for getting sampling and weighting information  .
* 2: Open Basics file.
* 3: Match files using Methods as key table.

GET FILE='SC6_Methods_D_3-0-1.sav'
  /KEEP=ID_t wave psu stratum  weight_design_std ALWAlatecomer     /* keep relevant
variables */.
DATASET NAME Methods WINDOW=FRONT.

SELECT IF wave=2 AND ALWAlatecomer NE 1.
EXECUTE.

DELETE VARIABLES wave ALWAlatecomer.

SORT CASES ID_t.

GET FILE='SC6_Basics_D_3-0-1.sav'.
DATASET NAME Basics WINDOW=FRONT.

SORT CASES ID_t.

MATCH FILES /FILE=*
  /TABLE='Methods'
  /BY ID_t.

DATASET CLOSE Methods.

* do some descriptive analysis using standardized design weights

WEIGHT BY weight_design_std.
FREQUENCIES VARIABLES=t700001.
WEIGHT OFF.

* Turn to next page...
```

Example 6-1 in SPSS – continued

```
* define complex survey data structure for adjusting standard errors

CSPLAN ANALYSIS
  /PLAN FILE='SC6.csaplan'
  /PLANVARS ANALYSISWEIGHT=weight_design_std
  /DESIGN STRATA=stratum CLUSTER=psu
  /ESTIMATOR TYPE=WR.

* estimate the mean with standard error of age at interview.

CSDESCRIPTIVES
  /PLAN FILE='SC6.csaplan'
  /SUMMARY VARIABLES=tx29000
  /MEAN
  /STATISTICS SE
  /MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.

* regress net household income on education, gender, and civil state.

* IF-statements only neccessary, if output should be comparable to Stata's calculations.
IF t700001=1 t700001=3.
IF tx28101=0 tx28101=9.
IF tx27000=1 tx27000=5.
EXECUTE.

* Linear model for complex sample.
CSGLM  t510010_g1 BY t700001 tx27000 tx28101
  /PLAN FILE='SC6.csaplan'
  /MODEL t700001 tx27000 tx28101
  /STATISTICS PARAMETER SE CINTERVAL TTEST.

DATASET CLOSE Basics.
```

Example 6-2 in SPSS

```
*Example 6-2: Analyzing second NEPS wave while accounting for sample stratification and
using longitudinal weights
* Procedure.
* 1: Open Methods-File and prepare for getting sampling and weighting information  .
* 2: Open Basics file.
* 3: Match files using Methods as key table.

GET FILE='SC6_Methods_D_3-0-1.sav'
  /KEEP=ID_t wave psu stratum  weight_design_std prob_w3   /* keep relevant variables */.
DATASET NAME Methods WINDOW=FRONT.

SELECT IF wave NE 1.
EXECUTE.

SORT CASES ID_t wave.

* carry forward design weight to second NEPS wave (i.e. wave=3).
IF (wave EQ 3) weight_design_std=LAG(weight_design_std).
EXECUTE.

* keep only second NEPS wave.
SELECT IF (wave EQ 3).

* calculate longitudinal weight.
COMPUTE longweight=weight_design_std * (1/prob_w3).
EXECUTE.

* keep cases with valid weight (cases participating in both waves).
SELECT IF (NOT(MISSING(longweight))).
EXECUTE.

DELETE VARIABLES wave.

GET FILE='SC6_Basics_D_3-0-1.sav'
 /KEEP ID_t t700001 tx29000.
DATASET NAME Basics WINDOW=FRONT.

SORT CASES ID_t.

MATCH FILES /FILE=*
  /TABLE='Methods'
  /BY ID_t.

DATASET CLOSE Methods.

* Turn to next page...
```

Example 6-2 in SPSS – continued

```
* define complex survey data structure for adjusting standard errors
CSPLAN ANALYSIS
  /PLAN FILE='SC6.csaplan'
  /PLANVARS ANALYSISWEIGHT=longweight
  /DESIGN STRATA=stratum CLUSTER=psu
  /ESTIMATOR TYPE=WR.

* estimate the mean with standard error of age at interview.
CSDESCRIPTIVES
  /PLAN FILE='SC6.csaplan'
  /SUMMARY VARIABLES=tx29000
  /MEAN
  /STATISTICS SE
  /MISSING SCOPE=ANALYSIS CLASSMISSING=EXCLUDE.

DATASET CLOSE Basics.
```

Example 6-1 in Stata

```
*Example 6-1: Analyzing first NEPS wave while accounting for sample stratification and
using weights
/*
Procedure:
1. Prepare Methods file to obtain sampling and weighting information
2. Merge this information to the Basics file
*/

use "SC6_Methods_D_3-0-1.dta", clear
keep if wave==2 & ALWAlatecomer!=1 // keep first NEPS wave
keep ID_t psu stratum weight_design_std // keep relevant variables
tempfile weights
save `weights', replace
use "SC6_Basics_D_3-0-1.dta", clear
merge 1:1 ID_t using `weights', keep(3) nogen
nepsmiss t700001 tx29000 t510010_g1 tx28101 tx27000

* do some descriptive analyses using standardized design weights
tab t700001 [aweight=weight_design_std]
* define complex survey data structure to adjust standard errors
svyset psu [pweight=weight_design_std], strata(stratum) singleunit(certainty)
// estimate the mean and standard error of age at interview
svy: mean tx29000
// regress net household income on education, gender, and civil state
svy: regress t510010_g1 i.t700001 i.tx28101 i.tx27000
```

Example 6-2 in Stata

```
*Example 6-2: Analyzing second NEPS wave while accounting for sample stratification and
using longitudinal weights
/*
Procedure:
1. Prepare Methods file to obtain sampling and weighting information for second NEPS wave
2. Merge this information to the Basics file
*/

use "SC6_Methods_D_3-0-1.dta", clear
drop if wave==1 // drop ALWA wave

* keep variables for primary sampling unit and stratum identifier as well as
* standardized design weight and probability of participation at second NEPS wave for
* first NEPS wave respondents
keep ID_t wave psu stratum weight_design_std prob_w3

* carry forward design weight to second NEPS wave (i.e. wave=3)
by ID_t (wave), sort: replace weight_design_std = weight_design_std[_n-1] if wave==3

* keep only second NEPS wave
keep if wave == 3

* calculate longitudinal weight
gen longweight = weight_design_std * (1/prob_w3)

* keep cases with valid weight (cases participating in both waves)
keep if !missing(longweight)

* merge gender and age at interview from Basics file
merge 1:1 ID_t using "SC6_Basics_D_3-0-1.dta", nogen assert(matched using) keep(matched)
keepusing(t700001 tx29000)

* define complex survey data structure to adjust standard errors
svyset psu [pweight=longweight], strata(stratum) singleunit(certainty)
svy: mean tx29000
* estimate the mean and standard error of age at interview
svy: mean tx29000
```

7 Tools for Stata Users

Thanks to Stata's quite versatile data format, NEPS Scientific Use Files provide an enriched experience for Stata users. In addition, NEPS Data Center provides a package of additional Stata programs ("ado files") to ease work with our data even more.

7.1 Multi-lingual data sets

Our Stata files offer variable labels and value labels both in German and in English. Stata users can easily switch between these languages using the `label language` command.[8]

```
. label language
Language for variable and value labels

    Available languages:
            de
            en

    Currently set is:               . label language de
    To select different language:   . label language <name>
(more output omitted)
. label language en
```

Furthermore, we have developed two Stata programs (ado files) to ease work with our data. You can obtain these ado files from our repository using the following command:

```
net install nepstools, from(http://nocrypt.neps-data.de/stata)
```

7.2 Data signatures

On dissemination, the signature of every single data file is calculated using Stata's `datasignature` procedure. The resulting digest is saved within each file. Stata users can confirm that they use unmodified NEPS SUF data by using Stata's `datasignature confirm`.

```
. datasignature confirm

  (data unchanged since 05aug2013 15:52)
```

7.3 NEPStools

Installation and updating

As mentioned before, NEPS Data Center has developed Stata programs to ease work with our data. You can freely obtain these ado files from our web repository using Stata's built-in installation routine[9]:

[8] Please note that, due to restrictions in the data release schedule, not all data releases feature the full set of translations with the first releases. This implicitly also applies to other multi-lingual metadata.

```
. net install nepstools, from(http://nocrypt.neps-data.de/stata)
checking nepstools consistency and verifying not already installed...
installing into <directory name>...
installation complete.
```

After installation, Stata can automatically check for updates using the `adoupdate` routine. NEPS' ado files are updated regularly to fix issues or adopt new functionality.

It is highly recommended to first read the included, detailed help files for each of the provided commands before usage. To do so, type "`help <command>`" at Stata's command prompt after installation.

nepsmiss – Recoding missing values

This program automatically recodes and labels all missing values into extended missing values (.a, .b, etc.). In this example, we run `nepsmiss` on the variable *t731454*, decoding all negative values (-54, -97, -98) into Stata's extended missings (.c, .b, .a). In contrast to Stata's built-in routines `mvencode` and `mvdecode`, value labels are correctly transferred as well. `nepsmiss` also has a predefined list of values to recode, representing the standard NEPS missing codes.

```
. nepsmiss t731454
Recoded 9450 values in total
```

ID_t	wave	t731454
8010851	2	-97
8012254	1	-54
8002388	2	-98
8012254	2	5
8002388	1	1

ID_t	wave	t731454
8010851	2	.b
8012254	1	.c
8002388	2	.a
8012254	2	5
8002388	1	1

The whole process canbe reverted with the option reverse:

```
. nepsmiss t731454 , reverse
Recoded 9450 values in total
```

We generally recommend running `nepsmiss` on all variables of interest (e.g. keep interesting variables, and run "`nepsmiss _all`") prior to any further data preparation.

[9] If this procedure is not available to you (e.g. due to technical restrictions at your institution), please contact NEPS Data Center (userservice.neps@uni-bamberg.de) to obtain a copy of the programs in ZIP format.

infoquery – Display additional metadata

As you may know, NEPS Data Center saves additional metadata from survey instruments to Stata data set files. `infoquery` displays these information attached to a variable. Note that `infoquery` will produce no output for some derived variables.

```
. label language en
. infoquery t514001
------------------------------------------------------------------------------
query result for variable t514001:

t514001[instname]:
zufrie1

t514001[sufname]:
t514001

t514001[questiontext_en]:
[ITEMBAT] I would like to begin by asking you a few questions about how
satisfied you are with various aspects of your life. Please answer using a
scale of 0 to 10 "0" means that you are totally and utterly dissatisfied;
"10" means that you are entirely satisfied. You can indicate your opinion
using the numbers '1' to '9'.

t514001[variablequestion_en]:
How satisfied are you currently with your life in general?
------------------------------------------------------------------------------
```

charren – Rename variables to survey names

In NEPS' data edition process, variables are renamed from instrument names to comply with NEPS SUF variable naming conventions. However, some users may be familiar to the original names, or for other reasons like to work using them. To enable this work, both old and new names are attached to each variable. The program charren easily allows switching between these versions, called "instname" and "sufname" internally.

```
. charren t514001 , to(instname) verbose
Info: will rename t514001 to zufrie1
```

Note that charren also uses reverse-search: Even if you only know an "instname" of a variable, it will reliably find it and rename the correct variable.

```
. charren zufrie1 , to(instname) verbose
Info: zufrie1 is not a variable name in current dataset; searching for
zufrie1 in specified search space
Info: will rename t514001 to zufrie1
```

8 Further Information

Please visit our web portal for further information and comprehensive documentation resources such as CATI questionnaires, how-to guides, technical reports, and the codebook.

https://portal.neps-data.de/de-de/datenzentrum/forschungsdaten/startkohorteerwachsene.aspx

For further support, please contact the NEPS data center:

Web:

https://www.neps-data.de/en-us/datacenter/contactdatacenter.aspx

E-Mail:

userservice.neps@uni-bamberg.de

Phone:

+49-(0)951-863-3511 (Mon.-Fri. 10:00-12:00 and 14:00-16:00)

Participation in the NEPS user trainings

Furthermore, the NEPS data center offers training courses on a regular basis. These courses introduce the research design of the NEPS, the structure of datasets, terms and conditions of data usage, issues of privacy and data protection, and so on. A central module of the courses consists of hands-on work with the NEPS data supervised by our staff. As skill levels, research interests, and methods vary greatly across users and disciplines, we will offer a comprehensive portfolio of seminars ranging from introductory topics on a rather general level to advanced methodological courses.

9 References

Allmendinger, J., Kleinert, C., Antoni, M., Christoph, B., Drasch, K., Janik, F., Leuze, K., Matthes, B., Pollak, R., & Ruland, M. (2011). Adult education and lifelong learning. In *H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice* (Eds.), Zeitschrift für Erziehungswissenschaft: Special Issue 14. Education as a Lifelong Process. The German National Educational Panel Study (NEPS) (pp. 283–299). Wiesbaden: VS Verlag für Sozialwissenschaften.

Antoni, M., Drasch, K., Kleinert, C., Matthes, B., Ruland, M., & Trahms, A. (2010). Arbeiten und Lernen im Wandel. Teil I: Überblick über die Studie (FDZ Methodenreport No. 05/2010). Nürnberg.

Aust, F., Gilberg, R., Hess, D., Kersting, A., Kleudgen, M., Steinwede, A. (2011). Methodenbericht NEPS Etappe 8 – Befragung von Erwachsenen. Haupterhebung 1. Welle 2009/2010 (B72). *https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Studienuebersicht/Methodenbericht_B72.pdf*

Aust, F., Gilberg, R., Hess, D., Kersting, A., Kleudgen, M., Steinwede, A. (forthcoming). Methodenbericht NEPS Startkohorte 6 – Befragung von Erwachsenen. Haupterhebung 2. Welle 2010/2011 (B67).

Aßmann, C., Steinhauer, H. W., Kiesl, H., Koch, S., Schönberger, B., Müller-Kuller, A., Rohwer, G., Rässler, S. & Blossfeld, H.P. (2011). Sampling designs of the National Educational Panel Study: challenges and solutions. In *H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice* (Eds.), Zeitschrift für Erziehungswissenschaft: Special Issue 14. Education as a Lifelong Process. The German National Educational Panel Study (NEPS) (pp. 51–65). Wiesbaden: VS Verlag für Sozialwissenschaften.

Blossfeld, Hans-Peter (1985). Bildungsexpansion und Berufschancen. Frankfurt: Campus.

Blossfeld, H.-P., Roßbach, H.-G., & von Maurice, J. (Eds.). (2011). The German National Educational Panel Study (NEPS). Zeitschrift für Erziehungswissenschaft: Special Issue 14. Education as a Lifelong Process. Wiesbaden: VS Verlag für Sozialwissenschaften.

Erikson, R., Goldthorpe J.H. & Portocarero, L. (1979). Intergenerational Class Mobility in Three Western European Societies: England, France and Sweden. The British Journal of Sociology 30(4), pp. 415-441

Ganzeboom, Harry B. G. (2010). Questions and Answers about ISEI-08. Retrieved from *http://home.fsw.vu.nl/hbg.ganzeboom/isco08/qa-isei-08.htm*.

Ganzeboom, Harry B. G., De Graaf, Paul M., Treiman, Donald J. (1992). A standard international socio-economic index of occupational status. Social Science Research, 21 (1), pp. 1-56.

Ganzeboom, Harry B. G., & Treiman, Donald J. (1996). Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations. Social Science Research, 25 (3), pp. 201-239.

Prandy, K. (2000): The social interaction approach to the measurement and analysis of social stratification. International Journal of Sociology and Social Policy 19(9/10/11), pp. 204-236.

Schimpl-Neimanns, Bernhard (2003). Mikrodaten-Tools: Umsetzung der Berufsklassifikation von Blossfeld auf die Mikrozensen 1973-1998. ZUMA-Methodenbericht 2003/10.

Wegener, Bernd (1985). Gibt es Sozialprestige? Zeitschrift für Soziologie, 14 (3) , pp. 209-235.