SOEP The German Socio-Economic Panel Study

117

SOEP Survey Papers

Series D - Variable Description and Coding

SOEP – The German Socio-Economic Panel Study at DIW Berlin 2012

SOEP 2011 – Documentation on Biography and Life History Data for SOEP v28

SOEP Group

DIW SOEP

Running since 1984, the German Socio-Economic Panel Study (SOEP) is a wide-ranging representative longitudinal study of private households, located at the German Institute for Economic Research, DIW Berlin.

The aim of the SOEP Survey Papers Series is to thoroughly document the survey's data collection and data processing.

The SOEP Survey Papers is comprised of the following series:

Series A – Survey Instruments (Erhebungsinstrumente)

Series B – Survey Reports (Methodenberichte)

Series C – Data Documentation (Datendokumentationen)

Series D – Variable Descriptions and Coding

Series E – SOEPmonitors

Series F – SOEP Newsletters

Series G – General Issues and Teaching Materials

The SOEP Survey Papers are available at
http://www.diw.de/soepsurveypapers

Please cite this paper as follows:

SOEP Group. 2012. SOEP 2011 – Documentation on Biography and Life History Data for SOEP v28. SOEP Survey Papers 117: Series D. Berlin: DIW/SOEP

ISSN: 2193-5580 (online)

Contact: DIW Berlin
SOEP
Mohrenstr. 58
10117 Berlin

Email: soeppapers@diw.de

SOEP GROUP

SOEP 2011 – DOCUMENTATION ON BIOGRAPHY AND LIFE HISTORY DATA FOR SOEP V28

Berlin, 2012

DOI: *10.5684/soep.v28*

Editor: Jan Goebel

Contents

1 General Introduction

By compiling a comprehensive set of questions on the individual life history into user-friendly variables, the SOEP database provides users with a representative collection of biographical information for the entire German population. This covers information on the individual career path since the age of 15, on marital status and childhood biography, the first job, social background and immigration history. The function of these data is, on the one hand, to make important background information available for analyses (e.g. information on fertility as an explanatory variable when analyzing labor market supply of women), and, on the other hand, to support self-contained analyses (e.g. on occupational careers or intergenerational transmission of education).

In general, each respondent of the SOEP questionnaire (surveying age starts in the calendar year a person turns 17 years) will answer the biographical questions only once (retrospectively). In the beginning of the SOEP, this occurred within the framework of the first three waves (1984 to 1986). Due to the inevitable 'mortality rate' of the panel (refusal to participate, death, relocation abroad), this process unfortunately leads to missing biographical entries for persons who did not participate in all three waves. Because of this, since 1988 all biographical information (occupation, marital status, family, first job and social background) is, in principle, collected during the first interview for new respondents in existing sample households. It should be noted that - due to the costs involved and the increased response burden - the main objective of surveying the biographical information in the course of the very first interview is not applied to the first wave of new subsamples. For example, in sample C (East Germany, field work started in 1990) the biographical questionnaire was first collected in 1992. Consequently, the surveyed persons in sample C who left SOEP before 1992 or who refused to complete the biography questionnaire in 1992 have no biographical information included in the SOEP data.

Summing up, in principle most of the biographical information in the SOEP is collected by means of the so-called 'Lebenslauf' (`life history') questionnaire. Although naming conventions, positioning of questions and the scope of this questionnaire have been changed and revised several times (see below), it has been addressed once at each respondent throughout the SOEP. Since 2000, a separate youth questionnaire exists which contains youth-specific questions.[1] A whole new series of age-triggered instruments for collecting biographical data was implemented in 2003. The target of the *first* of these questionnaires is to collect information about *newborn* children. It is aimed at their mothers of children aged up

[1] Due to survey-related reasons biographical information was not asked of first-time respondents aged 16 or 17 years until 1999. For this group of persons, much of the biographical information (i.e. on marital status and family information, occupation history since the age of 16 and social background) can generally be reconstructed using variables collected by means of the Individual Questionnaire, i.e. from the yearly ongoing survey.

to 15 months. As a result, the SOEP has started to survey the development of children from the very beginning of their life and will provide users with a completely new type of data. In 2005, a follow-up questionnaire targeted at children aged 2 to 3 years was implemented. Again, the information was collected from the mothers. It contained questions on the child's individual development and the mother's specific experiences during this formative period of raising the child. There will be follow-up interviews to collect data about these children at specific ages which are typically associated with decisions relevant to their individual development. The respective questionnaire targeted at children aged 5 to 6 was implemented in 2008[2]. A questionnaire targeted at children aged 7 to 8 years is used for the first time in 2010. This questionnaire will be answered by mothers and fathers (in contrast to earlier age-triggered instruments which were answered by mothers, only) and will therefore delivered in two separate files.

A chronological listing of the various changes related to the survey of biographically relevant information for the time period 1984 to 2010 can be found below. The differences in gathering information among and between the various sub-samples are reported with respect to 'Timing' (*when* respondents were asked), 'Coverage' (*which* parts of the biographical topics and single indicators were asked), and 'Positioning' of the biographical questions in the diverse survey instruments.

1984 The focus of the survey from samples A and B was the occupational biography. This information was collected (retrospectively) with the help of a 'life-course calendar' and covered the time period from the age of 15 up to the current age (or up until and including the maximum age of 65). The 'calendar' takes the form of a matrix with one column for every year of age and up to nine specifications of occupational activities (school, apprenticeship / training, military and community service, employed full time, employed part time, unemployed, househusband/wife, retired and other; question 62 in the standard Individual Questionnaire in Wave 1).

1985 The focus for samples A and B was on collecting marital and family status information in retrospect (questions 81-88 in the standard Individual Questionnaire in Wave 2). The number of children born up to that point in time is collected in detail as well as the eventual date at which the children moved out of the parents' home (only female participants were asked those questions). In addition, residency during childhood, the date a person moved out of the parents' home, as well as the start, end and potential reasons for the termination of up to three marriages are asked of each surveyed person.

[2] **Help for (very) old friends**: Starting with data release 2006 (up to wave V) naming conventions for files containing data collected by means of age-specific questionnaires have been changed: Data from the "Mother and Child" questionnaire on newborns can be found in the file BIOAGE01 (file BIOCHILD in earlier SOEP-data releases), data collected by means of the questionnaire on 2 to 3 year old children can be found in the file BIOAGE03, and the biography information collected from 16 to 17 year old first time respondents is stored in the file BIOAGE17 (file BIOYOUTH in earlier SOEP-data releases).

1986 The focus of the life history data was on social background and entry into the workforce (questions 10-13 and 80-87 in the standard Individual Questionnaire, Wave 3). Information on every surveyed person's parents is included, i.e. their year of birth and, where appropriate, the year of death, their level of education and vocational training as well as their working status at the time the respondent was 15 years of age. Furthermore, the father's type of gainful employment was also asked. With respect to entering the workforce, information is available on the age of the surveyed person when he/she first started to work and the type of employment he/she had. When appropriate, the age at each job change was also asked.

1987 No biographical information was collected in this year.

1988 The complete collection of biographical questions was included in the blue Individual Questionnaire for first time participants. For those persons who were new additions to the SOEP since 1985 and who had missed portions or all of the biographical questions, this missing information was collected in 1988. For this reason, it is only since 1988 that complete biographical information has been available on all three biographical areas for all persons surveyed up to this point (as long as they were still included in the SOEP population). Young adults up to age 17 were excluded from this retrospective collection of biographical information due to reasons of content (it did not make sense to collect the biographical information here). However, some technical problems arose when determining the exact minimum cut-off age of the persons included in this retrospective survey.

1989 to 90 During this time period, the form of the survey on biographical data from 1988 remained unchanged.

1990 The SOEP random sample was expanded. For the first time individuals and households in East Germany (sample C) were surveyed. However, biographical data for sample C were collected later on.

Since 1991 New respondents of sample A (West Germany) and B (Foreigners / 'Guest Workers') answered the biographical questions in an independent Biography Questionnaire.

1992 Due to the differing occupational titles, educational degrees, and biographies between East and West Germany, an additional biography questionnaire was developed for sample C (East Germany) which was first used in 1992 in order to cover all respondents in this sample. This additional questionnaire is identical with the western version in its structure and the format of its questions, with just a few answer categories being modified and speech delimitations were effected (for example, on occupational position or the description of the successfully completed apprenticeship).

This extensive group of questions was also applied to everyone who had been a new addition to the survey in East Germany since 1993.

1994 An updated version of the biography questionnaire version called 'Lebenslauf' ('life history') was introduced for all the four samples A, B, C, and D1/D2.[3] The formats of some of the questions were slightly changed, and new questions were added, although some questions were included for only one of the samples (i.e. questions relevant to immigration were only directed towards sample D).

1996 The Biography Questionnaire 'Lebenslauf' ('life history') was fully integrated for all samples, for example, using appropriate filter questions the immigration relevant information was also asked from persons in samples A to C in an identical form.

1998 Introduction of the supplementary sample E.

1999 The 1996 form of the Biography Questionnaire 'Lebenslauf' ('life history') was given to members of sample E for the first time.

2000 The 1996 version of the Biography Questionnaire 'Lebenslauf' ('life history') was changed slightly. For example, information on having own children is collected for men as well, as is the information on the respondent's mother's occupation at the time that the respondent was 15 years old.

A preliminary-version of the Youth Questionnaire was designed and given to 17 year old youths (only samples A to E). Data on social background were collected from young adults with single or no parents in the household.

In the year 2000, a new supplementary sample F with over 6000 surveyed households was established.

2001 The Biography Questionnaire 'Lebenslauf" ('life history') was further expanded and now also includes more questions on school, i.e. marks, and activities during childhood.

Biographical data are collected for the first time for all persons belonging to sample F using this updated Biography Questionnaire.

The revised Youth Questionnaire, the standard version for the forthcoming years, is used in the field for all 17 year old teenagers in addition to the Individual Questionnaire.

2002 A new sample G is drawn, which is only targeted at high-income households, i.e. households with a monthly net household income of more than 7,500 DM (≈ 3,850 €). This sample was also asked retrospective information on inheritances, which was collected in 2001 for samples A through F.

[3] (A) 'West Germany', (B) 'Guest Workers / Foreigners', (C) 'East Germany', (D1/D2) 'Immigrants since 1984', persons from D2 were first surveyed in 1995.

2003 Persons from sample G answered the Biography Questionnaire for the 'first' time. The new questionnaire 'Mother and Child' was given to mothers of newborns (all samples).

2004 The Biography Questionnaire was slightly expanded with questions concerning the 'numbers of brothers and sisters' and the 'location a person lived at before reunification (East Germany, West Germany, abroad)'. The question on siblings is also asked in the Youth Questionnaire.

2005 The new questionnaire 'Mother and Child II' ("Infants") targeted at children aged 2 to 3 years was implemented (for all samples).

2006 Introduction of the supplementary sample H with valid interview information for about 2.600 individuals. As a standard procedure, these new respondents do not fill in the biography questionnaire in order to reduce response burden in wave 1.

Starting in 2006, the age for first-time respondents has been changed to be the calendar year in which the person turns 18 years of age. Those aged 17 in 2006 are asked to fill in the extended "Youth Questionnaire" (data is stored in the file $PAGE17) instead of the "Individual Questionnaire" (data stored in $P). These extended questions cover indicators on subjective well-being, health (including body measures), labor force participation and education.

2007 Members of Sample H have answered the biographical background questionnaire for the very first time.

2008 The new questionnaire 'Mother and Child III' ("Pre-School") targeted at children aged 5 to 6 years was implemented for the first time (for all samples).

2009 Introduction of the new subsample I with valid interview data on about 2.500 adults. As a standard procedure, these new respondents do not fill in the biography questionnaire in order to reduce response burden in wave 1.

2010 Members of Sample I have answered the biographical background questionnaire for the very first time. The new questionnaire 'Parents I' targeted at children aged 7 to 8 years was implemented for the first time.

Looking ahead: Members of the new subsample J (2011) are asked to fill in an integrated version of the individual questionnaire and the biographical background questionnaire. It is planned to implement a questionnaire for 9/10-year-olds in 2012.

A series of problems may emerge when combining biographical information and storing data collections spanning multiple waves. This is due to the fact that the biographical information over time both within and between the sub-samples of SOEP is not always consistent with regards to

- *Positioning* (this includes differences among the various surveying instruments, i.e. the Individual Questionnaire and the single Biography Questionnaire 'Lebenslauf' ('life history'), as well as differences in the position of several indicators in the various versions of the questionnaires),
- *Coverage* (this includes both the changes in the targeted population and the partitions of the survey asked of each person and the corresponding indicators used), and
- *Timing* (this refers to the point in time when the biographical information was collected for a person in relation to the very first survey).

The biography data sets can always be divided into time invariant (e.g. first year of immigration to Germany, first job, place a person grew up) and time dependent (e.g. marital status, number of children, occupational biography) variables. Whereas time invariant information is by definition valid at every point in time after it has been collected, the time dependent information originally collected needs to be updated whenever a change has occurred. Alternatively, the information that is still valid must be included over the entire analysis period under investigation. In other words, since for the most part identical biographical information for different individuals is collected in SOEP at various points in time, all information regarding an eventual status change or an expansion of the original information must be accounted for over the entire time period of the analysis.

A yearly update of the biographical data therefore involves the following tasks:

- *Time dependent* information must be
  - collected for persons answering the survey questions for the first time and
  - carried forward or changed for persons repeating the SOEP interview.
- *Time invariant* information must be integrated into existing data sets for persons answering the survey questions for the first time.

The goal is for all biography relevant information provided to be up-to-date, without any loss of information with respect to the original variables, and in a user friendly form within the framework of the yearly data set updates. The time dependent variables will correspond then to the status of the most recently realized personal interview. The individual steps of the complex revision of the data sets are described in the corresponding documentation.

Additional Information:

- Unless otherwise indicated, the symbol '$' in a variable name or a file name stands for a wave specific prefix or suffix: for example, the variable $KMUTTI from the file $KIND indicates the vector of the variables AKMUTTI, …, ZKMUTTI up to BAKMUTTI from the file AKIND, …, ZKIND to BAKIND. '$$' indicates the survey year (2 digits) and is used as a suffix: for example, NATION$$ stands for NATION84 to NATION10 from the files APGEN to BAPGEN.[4]
- The file BIOLELA is mentioned frequently within the framework of the following documentation of the individual steps needed for generating biographical variables. This file is not a component of the standard updates of the SOEP data sets, but encompasses all of the biographical entries collected until 1996 (in the Individual Questionnaire and the Biographical Questionnaire) from the SOEP respondents. This file is rather complex due to the differences in the surveying procedures mentioned above and is therefore one central input for nearly all of the following variables on individuals who entered the survey prior to 1996. However, BIOLELA does not contain information necessary for updates (e.g. giving birth after having answered the Biographical Questionnaire). Furthermore, identical information is distributed over a multitude of single variables. The information in BIOLELA is only suitable for very restricted analyses without additional tests and supplements. Beginning with 1997, there are wave-specific $LELA files containing the biography information as collected in the respective year. These files (i.e. BIOLELA and $LELA) can be made available on request to interested users of the SOEP data.

The following table displays in a general overview the full set of biographical information as surveyed in the Biography Questionnaire 'Lebenslauf' ('life history') in 2006 and the current version (September 2011) of the user-friendly edition of this information. The designated numbers in the Biography Questionnaire 'Lebenslauf' ('life history') refer to the 2006 version with all samples fully integrated; due to the multitude of differences in the data collection process (as mentioned above), this does not imply that all of the following named variables were collected from all respondents nor that all information is available accordingly in the final biographical files.

[4] The SOEP data set released in 2011 (SOEP V27) will include, for the first time in the survey's 27 years, a two-letter rather than a single-letter wave prefix. Since we came to the end of the Latin alphabet with the letter Z in data release V26, we decided to use the wave prefix BA for the cross-sectional data format.

Table 1: Biographical data in SOEP

Biography Sub-area	Number of Question in the 'Lebenslauf' Questionnaire (2006)	Comparable Questions in the Youth Question-naire (2006)	SOEP Target Population	Files in the SOEP Database	Analysis Unit	Update Requirements (Source File for Update)	Status: Available / Not Available (up to Wave BA)
Place of birth	2, 3	*61, 62*	All persons surveyed	PPFAD	Individual	No	Available
Year of immigration	4	*63*	For persons not born in Germany	PPFAD	Individual	No	Available
Immigration biography	5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 15a	*64, 65, 66, 67, 68, 69, 70, 71*	For persons not born in Germany	BIOIMMIG	Individual	No	Available
Living in East or West Germany in 1989	16	-	All persons surveyed	PPFAD	Individual	No	Available
Place of childhood; Life at childhood residence; grew up with parents, Living together with parents	17,17a, 19, 20	*72, 73, 75, 76*	All persons surveyed	BIOPAREN	Individual	No	Available
Number of brothers and sisters	18	*74*	All persons surveyed		Individual	Yes	Not available
Parents living region, year of birth, year of death, nationality, country of birth	21, 22, 23, 23a	*77, 78, 79*	All persons surveyed	BIOPAREN	Individual	Partly (year of death from PPFAD)	Available

Biography Sub-area	Number of Question in the 'Lebenslauf' Questionnaire (2006)	Comparable Questions in the Youth Question-naire (2006)	SOEP Target Population	Files in the SOEP Database	Analysis Unit	Update Requirements (Source File for Update)	Status: Available / Not Available (up to Wave BA)
Religious affiliation of parents	28	*85*	All persons surveyed	BIOPAREN	Individual	No	Available
Parents took care about efforts at school	29	*41*	All persons surveyed	BIOSOC	Individual	No	Available
Respondent's last school marks	30	*37*	All persons surveyed	BIOSOC	Individual	No	Available
Relationship to parents during youth	31	*13*	All persons surveyed	BIOSOC	Individual	No	Available
Sport and activities during youth	32, 33, 34, 35	*16, 21,22, 25*	All persons surveyed	BIOSOC	Individual	No	Available
Occupational biography	36	-	All persons surveyed	PBIOSPE	Spell	Yes ($P, $PKAL)	Available
Year and place of acquiring a school degree	37, 38, 41	*27*	All persons surveyed	BIOSOC	Individual	No (although possible using $P)	Available
Level of school degree	39, 40, 42	*28*	All persons surveyed	$PGEN	Individual	Yes ($P)	Available
Number of foreign classmates in last attended school class	43	*45*	All persons surveyed	BIOSOC	Individual	No	Available
Target school degree	44, 45	*29, 30*	All persons surveyed	BIOSOC	Individual	No	Available

Biography Sub-area	Number of Question in the 'Lebenslauf' Questionnaire (2006)	Comparable Questions in the Youth Question-naire (2006)	SOEP Target Population	Files in the SOEP Database	Analysis Unit	Update Requirements (Source File for Update)	Status: Available / Not Available (up to Wave BA)
Target vocational degree	53, 54	*48, 49*	All persons surveyed	BIOSOC	Individual	No	Available
First job (age, occupational position, public sector, industry)	55, 56, 57, 58, 59, 60a, 60b	-	All persons surveyed	BIOJOB	Individual	Yes, if person previously did not work ($P)	Available
Occupational changes	61	-	All persons surveyed	BIOJOB	Individual	Yes	Available
Last job (year, scope, public sector branch, occupational position)	62, 63, 64, 65, 66, 67	-	All persons surveyed	BIOJOB	Individual	Yes	
Year since living personally in current apartment; second residence	68, 69	-	All persons surveyed	BIORESID	Individual	No	Available
Births	70	-	All women surveyed; since 2000 men, too	BIOBIRTH BIOBRTHM	Individual	Yes ($P, $PBRUTTO, $KIND)	Available
Family status (marriage biography)	71, 72	-	All persons surveyed	BIOMARSY	Spell	Yes ($P, $PBRUTTO)	Available

Biography Sub-area	Number of Question in the 'Lebenslauf' Questionnaire (2006)	Comparable Questions in the Youth Question-naire (2006)	SOEP Target Population	Files in the SOEP Database	Analysis Unit	Update Requirements (Source File for Update)	Status: Available / Not Available (up to Wave BA)
Youth	Youth Questionnaire		16 and 17 year old respondents	BIOAGE17	Individual	No	Available
Newborns	Mother & Child Questionnaire		Mothers of newborns	BIOAGE01	Individual	No	Available
Infants	Questionnaire on children aged 2 to 3 years		Mothers	BIOAGE03	Individual	No	Available
Preschooler	Questionnaire on children aged 5 to 6 years		Mothers	BIOAGE06	Individual	No	Available
Elementary school	Questionnaire on children between the age of seven and eight		Parents	BIOAGE08A BIOAGE08B	Individual	No	Available
Educational history			All persons surveyed	BIOEDU	Individual	Yes	As beta release available

2 Biographical Information in the Meta File PPFAD (Month of Birth, Year of Death, Immigration Variables, Living in East or West Germany in 1989)

by Joachim R. Frick, Olaf Groh-Samberg and Christian Schmitt

The file PPFAD includes, among other more survey related variables like responding status, some most important demographical information for each person who has ever participated in SOEP in at least one wave. These are, on the one hand, longitudinally checked data on sex (variable SEX) and the date of birth (year of birth in variable GEBJAHR in 4-digits and month of birth in variable GEBMONAT), and, on the other hand, generated demographic variables on the year of death (TODJAHR and TODINFO), on the migration background of an individual (MIGBACK and MIGINFO), on the year of the first immigration to Germany (IMMIYEAR), on the country of origin (GERMBORN and CORIGIN) as well as on the geographic area a person lived in prior to German unification (LOC1989). In the following section, the construction of these generated variables will be explained briefly.

2.1 Month of birth

From wave T onwards (2003) the data set PPFAD contains – in addition to the year of birth – the month of birth (GEBMONAT). This new variable is accompanied by the supporting variable GEBMOVAL which indicates the data source for the month of birth.

GEBMONAT and GEBMOVAL can take the following characteristics:

- GEBMONAT: Month of birth;
  1 (January) to 12 (December)
- GEBMOVAL: Month of birth - data-source
  1 Generated
  2 Info as stored in PPFAD
  3 Info derived from data set $KIND & BIOAGE03
  4 Info derived from data set SP (own response)
  5 derived from data set $LELA (own response)
  6 derived from BIOAGE01 (mother-child-questionnaire) & BIOAGE03 (NEW with Wave W / Surveyyear 2006)
  7 derived from Youth Questionnaire (own response) (NEW with Wave Z / Surveyear 2009)

The month of birth was asked in wave S individual questionnaire (SP). Furthermore, the month of birth was asked in the biography data set, starting with wave T ($lela, file not available with the SOEP data distribution). Additionally the month of birth is recorded for all children within the file $KIND (starting with wave T) and BIOAGE06 (since wave BB). Since Wave W, information from the mother-child questionnaire, which is filled in by mothers of newborns (BIOAGE01; BIOAGE03 added in wave BB), is being considered, while information from the Youth Questionnaire ist considered since Wave Z. This procedure provides the relevant information for most of the current panel members. The information remains missing for persons who lack any of the above information, including temporary dropouts or people who exited in a previous wave. For some of those persons, the month of birth could be reconstructed. This reconstruction remains an approximation and might differ from the true month of birth in individual cases. The variable GEBMOVAL also displays an ordinal scaling of the level of reliability, where individual response on one's own date of birth is given preference over derived information.

2.2 Construction of variables

The month of birth is constructed in an hierarchical order from the files:

- Generated (basis: $P, $PBRUTTO $KIND)
- $KIND & BIOAGE06
- SP
- $LELA
- BIOAGE01 & BIOAGE03
- Youth Questionnaire ($page17, starting with wave Z, 2009)

whereas the latter information overrides the former.

This means the generated information will only be utilized if no further, questionnaire based information for the month of birth is available.

The generated month of birth could only be constructed for people who were born while their parents were members of the SOEP. The information was derived from two sources:

- For newborn children the month of moving into the household was used as an approximation of the real month of birth (relevant file $PBRUTTO).
- For parents who reported a birth in a certain month, a link to the child was established, assigning the month of birth to the child (relevant file $P).

Several adjustments and tests of the generated data have been done which showed that – in the cases in which the generated data was also collected by SP, $LELA or $KIND – the data

generation is almost always congruent with the collected data and therefore has proven to be reliable.

Frequency examples for the variables “month of birth” and “data source for month of birth” in the file PPFAD(Version 2011 / up to Wave BB).

GEBMONAT Month of Birth – Wave BB 2011

		Frequency	Percent	Valid Percent	Cum. Percent
Valid	-3 not valid	1	,0	,0	,0
	-1 no info	28428	38,3	38,3	38,3
	1 January	4007	5,4	5,4	43,8
	2 February	3776	5,1	5,1	48,8
	3 March	4075	5,5	5,5	54,3
	4 April	3802	5,1	5,1	59,5
	5 May	3988	5,4	5,4	64,8
	6 June	3691	5,0	5,0	69,8
	7 July	3884	5,2	5,2	75,1
	8 August	3792	5,1	5,1	80,2
	9 September	3925	5,3	5,3	85,5
	10 October	3770	5,1	5,1	90,6
	11 November	3437	4,6	4,6	95,2
	12 December	3561	4,8	4,8	100,0
	Total	74137	100,0	100,0	

GEBMOVAL Geburtsmonat, Datenquelle

	Frequency	Percent	Valid Percent	Cum. Percent
-3 not valid	1	,0	,0	,0
-1 no info	28428	38,3	38,3	38,3
1 Generated	1445	1,9	1,9	40,3
3 Parents response($kind)	5618	7,6	7,6	47,9
4 Own response (SP)	21536	29,0	29,0	76,9
5 Own response ($lela)	13343	18,0	18,0	94,9
6 Mothers response	2136	2,9	2,9	97,8
7 Own response ($page17)	1630	2,2	2,2	100,0
Total	74137	100,0	100,0	

x.x Year of birth
(not generated)

2.3 Year of death

Variable TODJAHR Year of death - 4 digits -

The Variable TODJAHRcontains the four digit year entered as the year of death.

Codes

$$$$ effective year entered for persons whose year of death could be determined

(1) from the drop-out file PBR_EXIT[5], that is, the outcome of the yearly field work
(2) within the scope of the Infratest-Verbleibstudie (Study conducted by Infratest to follow-up on drop-outs) carried out in 1992
(3) within the scope of the Infratest-Verbleibstudie (Study conducted by Infratest to follow-up on drop-outs) carried out in 2001

(4) within the scope of the Infratest-Verbleibstudie (Study conducted by Infratest to follow-up on drop-outs) carried out in 2007

(5) within the scope of the Infratest-Verbleibstudie (Study conducted by Infratest to follow-up on drop-outs) carried out in 2008

Missing codes

(-2) Persons currently living or no longer existing in the sample

Essentially, the deaths of SOEP respondents are reported in the course of the yearly household interview during which the status of the currently living members of the household, as well as the changes due to births and deaths since the last year are surveyed. Furthermore, within the framework of, up to now, three subsequent address investigations of SOEP drop-outs ("Infratest-Verbleibstudie"), demographical drop outs due to mortality or move abroad have been identified. The mortality information is used in generating the variable TODJAHR.

In the first "Verbleibstudie" conducted from April to June in 1992 a total of 53 persons could be identified as deceased. In incorporating this information into the variable TODJAHR

[5] Help for old friends: The file PBR_EXIT includes all observations that exited from survey households since the previous wave for demographic reasons (death, emigration). Together with the file PBR_HHCH (covering observations who changed household from one wave to the next) these two files replace the file YPBRUTTO used in former releases of SOEP data.

attention was given to the fact that an exact year of death could be determined for only 35 of these persons. An exact date was missing for 16 persons, that is, only the qualitative information on their death was available. As a substitute for these cases, the year of the Wave in which the person dropped out of SOEP was used. For 2 persons implausible entries were corrected.

Within the scope of the second Infratest-"Verbleibstudie" conducted in 2001, over 700 persons were identified as deceased. Included in this number are multiple identifications, i.e., persons who were already determined to be deceased through the standard follow up process or in course of the first "Verbleibstudie 1992" mentioned above. This displays essentially a very high correspondence of results from the standard follow up and the ex-post determination of the time of death. For 10 persons the missing information on the year of death was imputed with the help of the year in which they dropped out of the SOEP sample.

In the few cases in which there were conflicting information between the first two follow-up studies and the information from PBR_EXIT (formerly YPBRUTTO), in principle the information from the "Verbleibstudie" was used.

In the third of those studies, another 21 individuals were identified as deceased between 2001 and 2005. For 18 of those persons a valid year of death could be investigated, the remaining three observations are set to the standard missing code "-1".

Again, some of these deaths have also been registered in the most recent of those follow-up studies which was carried out in 2008. In this study a total of 982 individuals were identified as deceased some of which date back to the late 1980s.

Variable TODINFO Year of death – source of information

Codes

1 'from continued surveying (PBR_EXIT / YPBRUTTO)'
2 'Infratest-Verbleibstudie (Follow-up Study) 1992'
3 'Infratest-Verbleibstudie (Follow-up Study) 2001'
4 'Infratest-Verbleibstudie (Follow-up Study) 2007'
5 'Infratest-Verbleibstudie (Follow-up Study) 2008'

For all of the persons who could be identified as deceased, the variable TODINFO contains the corresponding source of information.

2.4 Immigration information

Variables MIGBACK and MIGINFO (Olaf Groh-Samberg, Florian R. Hertel, Ingrid Tucci)

The objective of the variable MIGBACK is to identify if an individual has a migration background according to the following characteristics: (1) being an immigrant, (2) having or having had a non-German citizenship, (3) having at least one parent who is either non-German or who was born outside Germany. Based on all persons who ever have been part of the SOEP (i.e., the population in the file PPFAD), MIGBACK consistently defines the migration background as either direct or indirect or, due to missing values on GERMBORN, as not further differentiable.

Please notice that in the first release of the MIGBACK variable in 2008 persons (the large majority of them have never been SOEP respondents) were assigned the value 4, but are now assigned "-1" due to the lack of direct information. As a consequence, the current distribution of the MIGBACK variable differs from time it was first available.

Variable MIGBACK Individual Migration Background

Code	Label	Frequency	Percent
-1	No information available	4,236	6.3
1	No migration background	48,198	72.1
2	Direct migration background	7,946	11.9
3	Indirect migration background	5,284	7.9
4	Migration background, not differentiated	1,149	1.7
	Total	66,813	100,0

Sample I respondents answered the biography questionnaire for the first time in 2010 so that it was possible to assign them a migration background.

Besides the variable GERMBORN which indicates whether the person has a direct migration background (i.e. if he/she has immigrated) the time-variant information on nationality is important too. In order to build the indicator of the migration background, the first step consists in catching all available information on citizenship and citizenship change(s). Therefore the NATION$$ variables in the $PGEN files and additionally the information on the acquisition of German citizenship ($P files) were used. Accordingly, people who answered that they acquired German citizenship after birth are recoded as having a migration background. In order to identify also the citizenship of children who have completed neither

the personal nor the biography questionnaire, proxy information gathered from the head of the household in 1984 and 1988 (variables AK07A and EK03A) were used. Finally, by using the field agency's (TNS Infratest) proxy information on nationality, it is possible to determine if children or persons who are part of the SOEP gross sample ("Brutto") but never answered the person questionnaire or the biography questionnaire, have a migration background. But the personal information – whether proxy or direct - is not sufficient to know if a person has a migration background because the change of nationality for some persons might have occurred before they took part in the SOEP-survey and they might have never answered the question on the time of becoming German citizen.

In a second step information on mothers and fathers are used. In particular for those parents who were also SOEP-respondents we use the variable constructed accordingly to step one. Additionally the VNAT and MNAT (German/Foreigner) variables from BIOPAREN are used for those parents who were not SOEP-respondents. As respondents might have mixed parents, a respondent is given a migration background as soon as at least one of his or her parents is of migrant origin. Finally, the variable using information concerning the immigrant group (BIIMGRP) located in the BIOIMMIG file is also used as an additional indicator.

MIGINFO informs on the sources used to construct MIGBACK. There are four different possible codes indicating whether the information stemmed from direct response or proxy personal information and if parental information (nationality, country of birth, etc.) is available or not.

Variable MIGINFO Source of Information to define Migration Background

Code	Label	Frequency	Percent
-1	No information available	4,235	6.3
1	Direct information without parental information	33,732	50.5
2	Proxy information without parental information	463	0.7
3	Direct information with parental information	17,451	26.1
4	Proxy information with parental information	10,932	16.4
	Total	66,813	100,0

The 10,932 persons coded 4 are mostly children whose information on the migration background are derived from parental information and interviewer report. A part of the population identified as "without a migration background" (MIGBACK coded 1) and for which we have no parental information (code 1 and 2 on MIGINFO) however might have a migration background. Unfortunately this cannot be tracked due to the lack of available data

(parental information and information on nationality before the respondents entered the SOEP survey).

Variables IMMIYEAR, GERMBORN and CORIGIN (Joachim R. Frick)

The objective of these variables, based on all persons who ever have been a part of SOEP (i.e., the population from PPFAD), is to identify individuals who have immigrated to Germany from any other country since 1949 (the founding year of the Federal Republic of Germany). For this group of persons the variable IMMIYEAR specifies the calendar year (4 digits) in which the first immigration to territories of the Federal Republic of Germany occurred. Persons who had immigrated to Germany up until and including 1948 are included in the group of those identified as "born in Germany" (see variable GERMBORN). The variable CORIGIN specifies the country of origin.

The "Immigrant Sample" D classifies everybody who moved from abroad to West Germany as "immigrant", including persons who came from the GDR ('*Übersiedler'*). However, this latter group does not have an immigrant status as a result of the definition used here and as such, *Overview 1* may show a surprisingly low share of immigrants for Sample D. A specific problem regarding immigration to Germany arises from the group of persons who were born in Eastern Europe at a time when these regions were considered German territory and who later immigrated to the Federal Republic of Germany. For this group, according to the formulation of the corresponding SOEP questions regarding the country of origin, inconsistent answers can be expected. If at all possible, these persons are identified as immigrants in the variables described here (cumulative up to Wave R this affects merely 115 persons).

Overview 1 illustrates that across all SOEP sub-samples, persons have been identified as being immigrants. Therefore, an analysis on immigration questions clearly should *not* be limited to samples B and D. It is relevant to note that – almost by definition – the share of *non*-immigrants in Sample B (the "foreigner" sample which was started in 1984) increases permanently due to children born to these persons after their migration to Germany ("second-generation"). The large proportion of individuals with a "No Answer" Code in Sample G is mainly due to a change in the selection scheme for the second Wave of the High Income Sample. Therefore biographical information is not available for sample G respondents which were not followed up in 2003.

Overview 1: Distribution of the Immigration Information According to SOEP Samples A to I (up to Wave BA, 2010)

GERMBORN	Sample									Total
	A	B	C	D	E	F	G	H	I	
No Answer (Codes -1/-3)	162	96	11	7	21	12	710	76	77	1,172
	1.2	*1.8*	*0.2*	*0.5*	*0.9*	*0.1*	*22.7*	*2.7*	*2.9*	*2.3*
Born in Germany or immigrated before 1949 (Code 1)	13,595	1,413	6,707	551	2,169	11,660	2,301	2,541	2,344	43,281
	95.8	*26.3*	*98.5*	*35.6*	*90.8*	*88.8*	*73,5*	*90,0*	*87,07*	*83.1*
Immigrated since 1949 (Code 2)	438	3,864	89	992	198	1,448	119	205	271	7,624
	3.1	*71.9*	*1.3*	*64.0*	*8.3*	*11.1*	*3.8*	*7.3*	*10.1*	*14.6*
Total	14,195	5,373	6,807	1,550	2,388	13,120	3,130	2,822	2,692	52,077
	100	*100*	*100*	*100*	*100*	*100*	*100*	*100*	*100*	*100*

Source: All survey participants with at least one SOEP interview from 1984 to 2010 (n=52,077).

The objective of generating immigration information in PPFAD is to fill the variables GERMBORN, CORIGIN and IMMIYEAR for all of the survey participants who have had at least one SOEP interview, as well as for the children in all households who have realized interviews; the informational base for permanent non-responders is not sufficient. The elements involved in generating the immigration information are those listed in the *Overview 2* taken from the wave specific individual questionnaire or from the course of the SOEP installed variations of the "Biography/Life history" questionnaires:

- $P (wave-specific survey data from individual questionnaire),
- $PAUSL (wave specific survey data collected only for sample B from the personal data questionnaire up until Wave M),
- $PBRUTTO (wave specific information collected by the interviewer on all household members),
- BIOLELA (integrated biographical file for the years 1984 to 1995 = Wave A to Wave L)
- $LELA (Curriculum vitae information for those interviewed for the first time since 1996 = wave M)
- $JUGEND (youth questionnaire for 16-17 year olds since 2000, wave Q) and $PAGE17 (extended youth questionnaire for first-time respondents since 2006, wave W).

Difficulties in generating the immigration information arise in part for those persons for whom none of the mentioned information was originally surveyed (for example, in the "*Lebenslauf-Fragebogen*" Biographical Questionnaire): This affects in particular sub-samples A and C and in most cases is caused by the fact that these persons did not take part in or no longer took part in the SOEP survey at the time that the survey with the immigration relevant questions was implemented (see Introduction). Questions regarding immigration were first asked of survey participants in sample A in 1990 and the Biographical Questionnaire was first included in the eastern sample C in the third wave (1992) out in the field. Moreover, Item-Non-Response is also significant, i.e., not answering a question. In order to minimize the number of missing entries for immigration relevant variables, for persons for whom corresponding information is missing, information from other variables or from the household context are used to the extent that these permit inferences on the immigration biography of the individual.

Overview 2: Input-Variables for Generating Immigration Information

Variable	File	Variable Label	Sample
AP62A	APAUSL	Country of birth	B
BP98A	BPAUSL	Country of birth	B
CP98AB	CPAUSL	Country of birth	B
DP95A	DPAUSL	Country of birth	B
EP88A	EPAUSL	Country of birth	B
FP105A	FPAUSL	Country of birth	B
GP105A	GPAUSL	Country of birth	B
HP105A	HPAUSL	Country of birth	B
IP105A	IPAUSL	Country of birth	B
JP105A	JPAUSL	Country of birth	B
AP63A	APAUSL	Year moved to Germany	B
BP99A01	BPAUSL	Year moved to Germany	B
CP99AB01	CPAUSL	Year moved to Germany	B
DP96A01	DPAUSL	Year moved to Germany	B
EP89A01	EPAUSL	Year moved to Germany	B
FP106A01	FPAUSL	Year moved to Germany	B
GP10801	GP	Year moved to East, West Germany (Filter)	A B
GP10802	GP	Year moved to East, West Germany	A B
GP10803	GP	Area of origin	A B
GP106A01	GPAUSL	Year moved to Germany	B
HP108B01	HP	Year moved to East, West Germany (Filter)	A B
HP108B02	HP	Year moved to East, West Germany	A B
HP108B03	HP	Area of origin	A B
HP106A01	HPAUSL	Year moved to Germany	B
IP10801	IP	Year moved to East/West Germany (Filter)	A B C
IP10802	IP	Year moved to East/West Germany	A B C
IP10803	IP	Area of origin	A B C
IP106A01	IPAUSL	Year immigrated to Germany	B
JP108B01	JP	Year moved to East/West Germany (Filter)	A B C
JP108B02	JP	Year moved to East/West Germany	A B C
JP108B03	JP	Area of origin	A B C
JP106A01	JPAUSL	Year immigrated to Germany	B
LPGRUPPE	LPBRUTTO	Immigration group	D
LPHERKFT	LPBRUTTO	Country born in	D
MP90A01	MP	German	A B C D
MP90A02	MP	Emigrant of German descent from Eastern Europe	A B C D
MP90A03	MP	Has already lived in Germany in 1984	A B C D
NP101	NP	Born in Germany	A B C D
NP102	NP	Has already lived in Germany in 1984	A B C D
OP103	NP	Born in Germany	A B C D E
OP104	NP	Has already lived in Germany in 1984	A B C D E
PP117	NP	Born in Germany	A B C D E
PP118	NP	Has already lived in Germany in 1984	A B C D E
QP12101	QP	Born in Germany?	A B C D E F
QP12102	QP	Country of birth	A B C D E F
QP122	QP	Has already lived in Germany in 1984	A B C D E F
QP124	QP	Born in Germany?	A B C D E F
RP11701	RP	Born in Germany?	A B C D E F
RP11702	RP	Other country of birth	A B C D E F
RP118	RP	Has already lived in Germany in 1984	A B C D E F
RP120	RP	Foreign citizenship: Born in Germany?	A B C D E F
SP120	SP	Foreign citizenship: Born in Germany?	A B C D E F G
TP121	TP	Place of residence before reunification 1989	A B C D E F G
TP126	TP	Foreign citizenship: Born in Germany?	A B C D E F G
UP130 …	UP	Foreign citizenship: Born in Germany?	A B C D E F G
VP139 …	VP	Foreign citizenship: Born in Germany?	A B C D E F G

Variable	File	Variable Label	Sample
WP131 …	WP	Foreign citizenship: Born in Germany?	A B C D E F G
XP142 …	XP	Foreign citizenship: Born in Germany?	A B C D E F G H
YP136 …	YP	Foreign citizenship: Born in Germany?	A B C D E F G H
ZP141 …	ZP	Foreign citizenship: Born in Germany?	A B C D E F G H I
BAP134	BAP	Foreign citizenship: Born in Germany?	A B C D E F G H I
P031Z	BIOLELA	Year of immigration to Germany	D
P051Z	BIOLELA	Year of immigration to Germany	D
P060Z	BIOLELA	Has always lived in Germany since immigration	D
P04Z	BIOLELA	Country of birth	D
P021Z	BIOLELA	Lived where in 1984?	D
B34	BIOLELA	Area of origin	A B C
B36	BIOLELA	Country of birth	A B C
B37	BIOLELA	Year of immigration to Federal Republic of Germany	A B C
$B03	$LELA	Country of birth	A B C D E F G H I
$B02	$LELA	Born in Germany?	A B C D E F G H I
$B04	$LELA	Year of immigration to Federal Republic of Germany	A B C D E F G H I
QJ54	QJUGEND	Born in Germany?	A B C D E
QJ55	QJUGEND	Country of birth	A B C D E
QJ56	QJUGEND	Year of immigration to Federal Republic of Germany	A B C D E
$J61	$JUGEND	Born in Germany?	A B C D E F G H I
$J62	$JUGEND	Country of birth	A B C D E F G H I
$J63	$JUGEND	Year of immigration to Federal Republic of Germany	A B C D E F G H I
ZZJAHR	Generated on the basis of the variables $PZUG from the files $PBRUTTO	Immigrations to Germany which were documented through the interview within the scope of the field work. This information is used only if no other valid information is available.	A B C D E F G H I

In the following sections, the PPFAD variables are described in detail; special attention is given to the central filtering function of the variable GERMBORN in reference to the variables IMMIYEAR and CORIGIN.

Variable GERMBORN Born in Germany?

Codes

1 Persons who were born in Germany (including immigrants before 1949)

2 Persons who have immigrated to Germany since 1949

Missing codes

-2 For permanent non-respondents

-1 Provided that no entry could be derived from all of the variables

The following persons have been identified as immigrants despite the fact that information is missing for the variables relevant to GERMBORN (The process follows the sequence stated here, i.e., if the first condition does not hold, then the second condition is tested and so on):

1.1 Persons from sample D for whom a foreign country as the country of origin is entered in the address protocol in the first survey wave (variable LPHERKFT). The variable CORIGIN is also given the respective value.

1.2 Persons with valid entries for the variable IMMIYEAR who over the course of their survey in SOEP have at least once indicated a non-German nationality. This nationality is alternatively used instead to generate the variable CORIGIN.

1.3 Persons whose mothers have immigrated after having given birth to this individual in case the mother's immigration occurred less than 18 years after the birth of the respective person. For CORIGIN and IMMIYEAR, in this case, information from the mother's immigration history is used, assuming that mother and child have always lived together.

1.4 For persons who have acquired an educational degree abroad, who at any time have indicated a non-German nationality and in whose households (given by HHNR) immigrants reside, CORIGIN is given the corresponding value of the nationality; as a rule IMMIYEAR remains "missing" ("-1").

On the other hand, for the following persons it is assumed that they were born in Germany:

2.1 Persons from sample D for whom in the first survey wave Germany as the country of origin is entered in the address protocol.

2.2 Persons whose mothers were born in Germany or had immigrated before giving birth.

2.3 Minor children who live in a household with no immigrants.

2.4 Persons from samples A, C, E, F, G and H who live in a household with no immigrants and who have never indicated a non-German nationality.

Furthermore, generating immigration variables proves to be difficult for persons who in the course of the survey years submit conflicting answers on whether or not they were born in Germany. The number of these inconsistencies has increased with the inclusion of the yearly repetition of the corresponding question in the individual data questionnaire (2000). In such cases, as a rule decisions can only be based on plausibility and in comparison with further answers. If, for example, concrete information on the country of origin or the immigration year exists, then it is assumed that the person is an immigrant. For such cases where the individual information given by the person itself is not strong enough to base a decision on, then additional information from the household context is applied. If no final decision on the basis of all available information can be made, then the person with an inconsistent answer is given the code for missing values; thereafter the affected person's data file is run through the generation process described above.

In assessing the cases with inconsistent answers it must be considered that the questions relevant for GERMBORN in the SOEP surveying instruments are based on different definitions of "Germany". The questions from 1990 to 1993 that related to the regions of the Federal Republic of Germany and the German Democratic Republic (DDR) since 1949, respectively have been switched in the Biographical Questionnaire since 1996 to "the Federal Republic of Germany (West Germany), the German Democratic Republic (East Germany) or Germany as defined at the time of your birth". From this, answers which appear to be in contrast to each other could emerge for persons who were born prior to 1949 in "formerly German regions" in Eastern Europe which today are no longer part of the Federal Republic. Persons belonging to this group are considered to be, in the manner described here, immigrants if they immigrated after 1949.

For persons who, according to GERMBORN, are not born in Germany, the variables IMMIYEAR and CORIGIN should designate the year of the initial immigration to Germany, respectively the country of origin.

Variable IMMIYEAR Year of the initial immigration to Germany after 1948 (4 digits)

Codes

1949 ... 2009 Immigration year

Missing codes

-3 for successfully surveyed persons without a valid entry for the immigration variables (GERMBORN could not be filled with a valid answer)
-2 if born in Germany or immigrated before 1949, respectively and survey participants without an interview
-1 Immigrants for whom no valid answer can be derived from all the original variables

Persons who have been identified as immigrants and for whom it was not possible to determine the immigration year from the original variables are assigned the following values:

1. The year of entry into SOEP, in the event that $PZUG from the file $PBRUTTO indicates the code "moved into household from abroad" (see the variable ZZJAHR in *Overview 2*)
2. The year of the mother's immigration, in the event that the mother had immigrated prior to the year the person turned 18 years of age.

Variable CORIGIN Country of origin

Codes

1	Germany
2 to 155	Turkey, (Ex-)Yugoslavia, Greece, Italy, Spain, …. Turkmenistan
222	unspecified Eastern Europe
333	other unspecified foreign country
444	unspecified countries within EU

Missing codes

-3 for successfully surveyed persons without a valid entry for the immigration variables (GERMBORN could not be filled with a valid answer).
-2 all survey participants without an interview
-1 Immigrants for whom no valid answer is derivable from all of the original variables

The variable CORIGIN represents the country of origin, respectively of birth. Every person born in (West and East) Germany is assigned code "1" for the variable CORIGIN (see also the variable GERMBORN). Starting with the January 2004 release of SOEP-data (including wave S (survey year 2002)), the information on the country of origin is also contained in the 95% Scientific-Use-Version, which had not been the case before (the same is true for the variables on nationality in $PGEN).

Persons, who have been identified as immigrants and from whose original variables no country of origin could be determined, are assigned as a resource the following codes:

1. The code of the country which corresponds to their non-German nationality.
2. The code "222", in the event that it is evident from the original variables (code "2" in GP10803 to JP108B03) that the person in question immigrated from Eastern Europe, respectively from the former German territories in Eastern Europe. This also includes, under certain circumstances, a small number of persons from sample D if they were identified in P070Z as persons of German descent from Eastern Europe.
3. The code „333", in the event that from the original variables (code "3" in GP10803 to JP108B03) it is evident that the person comes from a region other than Eastern Europe.
4. The country of origin given in the address protocol from sample D (variable LPHERKFT).
5. The mother's country of origin in the event that the mother has immigrated prior to the year the person turned 18 years of age.

People who immigrated from ‘Croatia’, ‘Serbia’, ‘Slovenia’, ‘Bosnia Herzegovina’, ‘Macedonia’ and ‘Kosovo’ before 1991 to Germany assigned to the corigin code 3 ‘Ex Yugoslavia’ as ‘Croatia’, ‘Serbia’, ‘Slovenia’, ‘Bosnia Herzegovina’, ‘Macedonia’ and ‘Kosovo’ did not exist those days.

2.5 Living in East or West Germany in 1989

The variable LOC1989 in the meta-file PPFAD provides information about the geographic area a person lived in *prior to* the German reunification, differentiating “East Germany (DDR incl. East Berlin”, “West Germany (Bundesrepublik Deutschland incl. West Berlin)”, and “abroad (Ausland)”. This information has been generated for all individuals in SOEP with at least one successful interview since 1984 as well as for children (i.e., $NETTO>= 10 & <30).

Variable LOC1989 “Where did you live in 1989?”

Codes

1 East Germany (German Democratic Republic [DDR] including East Berlin)
2 West Germany (Federal Republic of Germany [BRD] including West Berlin)
3 Abroad (Ausland)

Missing Codes

-2 does not apply; born after 1989
-1 not available

After asking this information from all respondents in 2003 (variable TP121 in file TP), a corresponding question has been included in the biography questionnaire since wave U (2004) [Question 16 / variable UB16 in file ULELA] which will collect this time-independent information from all future first time respondents. For all respondents interviewed up until 2006, the following information was used as input to generate LOC1989:

- Information on place and date of last school attendance [variables BSSCHEND and BSSCHWO in file BIOSOC / variables $B38 and $B3701 in file $LELA with $ starting in wave U, 2004],
- Sample affiliation [variable PSAMPLE in file PPFAD],
- year moved in at current address [variable BRMOVEIN in file BIORESID / variable $B68 in file $LELA with $ starting in wave U, 2004],
- sample region [variables $SAMPREG in file PPFAD],
- year of first immigration to Germany [variable IMMIYEAR in file PPFAD]

- In case of inconsistent information from these various sources, the data collected in 2003 via variable TP121 and the information from the biography questionnaire collected since 2004 is considered superior. Persons without any individual information and aged less than 18 years in 1989 were assigned parental information, if available.
- Since biographical information for members of sample I has been collected in survey year 2010 for the first time the variable LOC1989 is completed for SOEP samples A through I.

Where did you live in 1989?	Freq.	Percent	Cum.
n/a; t.n.z (GEBJAHR>=1990)	9,290	14.63	14.63
n/s; k.A.	6,509	10.25	24.88
GDR; DDR (including East Berlin)	10,820	17.04	41.91
FRG; BRD (including West Berlin)	34,886	54.93	96.85
Abroad; Ausland	2,003	3.15	100.00
Total	63,508	100.00	

Source: POPULATION of PPFAD as of wave BA (2010) with at least one interview since 1984 or living as child in a responding household ($NETTO-codes >=10 and <30)

3 Activity Biography in the Files PBIOSPE and ARTKALEN

by Henning Lohmann and Marco Giesselmann (based on earlier work by Rainer Pischner)

The spell file PBIOSPE is based on the information on activity status over the life course, which are collected as a matrix from every respondent using the Biography Questionnaire (Question 37 in 2011).[6] The observations start at the age of 15 and end at the current age (up to age 65). This information on activity status covers only the period up to the time the biography is collected. To update the ongoing occupational career in PBIOSPE, information from the yearly Individual Questionnaire is also used. In this questionnaire, respondents are always asked their occupational status for every month of the previous year (Question 91 in 2011).[7] Therefore, the information on activity status collected on a monthly basis in the yearly personal questionaire and stored in the file ARTKALEN in spell format is aggregated into yearly values and combined with the information gathered from the Biography Questionnaire.[8]

In the following, the method of combining the data is described. There have been no changes how the data is generated since the previous version, distributed in 2010. But if you have been working with older versions of the dataset (versions distributed in 2008 and earlier) you should check the section at the end of the chapter, where you will find information on previous changes. But before we move on to the details, we provide a brief overview of the contents of PBIOSPE. Table 1 contains a list of all the variables in the dataset. The variables BEGIN and END indicate the beginning and the end of a spell. These variables are age entries. There are also variables that refer to calendar years: BEGINY and ENDY (Y stands for Year). The variable SPELLTYP contains information on the activity status during the spell, e.g., employed full-time or unemployed. The SPELLNR is a serial identifier of spells of a given person. Missing information on the beginning or end of a spell causes what are known as censoring problems. There are two types of missing data. First, data can be missing on periods outside the observation window (before the age of 15 and after the age of 65). Second, data can be missing on years within the observation window due to item non-response in particular years or due to temporary drop-outs (the latter applies to calendar information only). In this case, we speak of "gaps." There are nine different patterns (variable ZENSOR)[9]:

1. uncensored: beginning observed, end observed
2. right-censored: beginning observed, end not observed
3. right-censored (gap): beginning observed, end not observed because of gap
4. left-censored: beginning not observed, end observed

[6] See Chapter 1 for general information on the collection of biography information.
[7] For persons who were temporarily unavailable for interviewing, it is sometimes possible to fill in the gaps in their occupational status. If these persons fill out the additional questionnaire for temporary drop-outs later on, we can use the information collected there (see files $PLUECKE).
[8] For more information, see Haisken-DeNew, John and Joachim R. Frick (2005): *DTC - Desktop Companion to the German Socio-Economic Panel Study (SOEP)*, Chapter 3.
[9] The same goes for the ZENSOR-variable in ARTKALEN.

5. left- and right-censored: beginning not observed, end not observed
6. left-censored and right-censored (gap): beginning not observed, end not observed because of gap
7. left-censored (gap): beginning not observed because of gap, end observed
8. left-censored (gap) and right-censored: beginning not observed because of gap, end not observed
9. left-censored (gap) and right-censored (gap): beginning not observed because of gap, end not observed because of gap

Table 1: Contents of PBIOSPE (variables)

Variable	Description
HHNR	Original Household Number
PERSNR	Never Changing Person ID
SPELLNR	Serial Number Of The Spell Per Person
SPELLTYP	Type Of Spell
BEGIN	Age Spell Begins
END	Age Spell Ends
BEGINY	Year Spell Begins
ENDY	Year Spell Ends
ZENSOR	Censor Variable
SPELLINF	Spell Construction Information
ERHEBJ	Survey Year Biography Data
KALYEAR	First Observation Year Calendar
BEGINB1	Age Spell Begins, 1st Initial Biography Spell
ENDB1	Age Spell Ends, 1st Initial Biography Spell
BEGINK1	Age Spell Begins, 1st Initial Calendar Spell
ENDK1	Age Spell Ends, 1st Initial Calendar Spell
BEGINYB1	Year Spell Begins, 1st Initial Biography Spell
ENDYB1	Year Spell Ends, 1st Initial Biography Spell
BEGINYK1	Year Spell Begins, 1st Initial Calendar Spell
ENDYK1	Year Spell Ends, 1st Initial Calendar Spell
BEGINB2	Age Spell Begins, 2nd Initial Biography Spell
ENDB2	Age Spell Ends, 2nd Initial Biography Spell
BEGINK2	Age Spell Begins, 2nd Initial Calendar Spell
ENDK2	Age Spell Ends, 2nd Initial Calendar Spell
BEGINYB2	Year Spell Begins, 2nd Initial Biography Spell
ENDYB2	Year Spell Ends, 2nd Initial Biography Spell
BEGINYK2	Year Spell Begins, 2nd Initial Calendar Spell
ENDYK2	Year Spell Ends, 2nd Initial Calendar Spell
BEGINB3	Age Spell Begins, 3rd Initial Biography Spell
ENDB3	Age Spell Ends, 3rd Initial Biography Spell
BEGINK3	Age Spell Begins, 3rd Initial Calendar Spell
ENDK3	Age Spell Ends, 3rd Initial Calendar Spell
BEGINYB3	Year Spell Begins, 3rd Initial Biography Spell
ENDYB3	Year Spell Ends, 3rd Initial Biography Spell
BEGINYK3	Year Spell Begins, 3rd Initial Calendar Spell
ENDYK3	Year Spell Ends, 3rd Initial Calendar Spell
BEGINK4	Year Spell Begins, 4th Initial Biography Spell
ENDK4	Year Spell Ends, 4th Initial Biography Spell
BEGINYK4	Year Spell Begins, 4th Initial Calendar Spell
ENDYK4	Year Spell Ends, 4th Initial Calendar Spell

As mentioned above, PBIOSPE combines information collected in the biography questionnaire and the calendar matrix of the individual questionnaire. The two types of

information are merged into PBIOSPE following a number of rules. First of all, it is important to acknowledge that the Biography Questionnaire Matrix as well as the Individual Questionnaire Matrix allow for multiple activity statuses for a given year or month. No concept of main activity is used. A common combination is, for instance, "housewife/-husband" and "working part-time". There are a number of other plausible combinations, but also combinations that are less plausible. However, a list of valid combinations of activity statuses defined according to legal or similar constructs would need to be based on very strong assumptions. In addition—in particular in case of the yearly matrix in the Biography Questionnaire—activities are reported that took place in a calendar year in consecutive months, which makes it impossible to exclude combinations of activities. Therefore, no data cleaning is performed at this stage. As a consequence, the data may contain information on more than one activity for a given point in time.

This also defines the rules for aggregating the monthly ARTKALEN data into yearly values. Take, for example, a person who was in full-time employment from January to November 2007, and unemployed in December 2007. The exact months are recorded in the dataset ARTKALEN. In the aggregated data, which is merged with the yearly data from the Biography Questionnaire, you find the information that the person worked full-time and was also unemployed in the year 2007. There is a second level of aggregation of ARTKALEN information as the data on type of activity, which is recorded in the variable SPELLTYP is more detailed than in PBIOSPE. The respective information is aggregated as described in Table 2.

Table 2: Aggregation of ARTKALEN spell information into PBIOSPE

	PBIOSPE	ARTKALEN
1	School/University	School, College (1)
2	Apprenticeship/Training	Vocational Training (4), First Job Training, Apprenticeship (13), Continuing Education, Retraining (14)
3	Military/Civilian service	Military, Community Service (9)
4	Full-time employed	Full-Time Employment (1), Short Work Hrs (2)
5	Part-time employed	Part-Time Employment (3), Second Job (11), Mini-job (up to 400 euros) (15)
6	Unemployed	Unemployed (5)
7	House-Husband/Wife	Housewife, Husband (10)
8	Retired	Retired (6)
9	Other	Maternity Leave (7), Other (12)
99	Gap	Information on gaps in ARTKALEN is not used. Gaps are calculated on the basis of the merged dataset.

As stated above, the calendar information is used to update the biography information. However, there is also a certain overlap of the periods covered by the two types of data. This is shown in Table 3. It indicates, for persons included in PBIOSPE, the year in which the biography information was collected (variable ERHEBJ). This year is usually also the last

year for which biography information is available.[10] The table also shows the first year recorded in the calendar data (variable KALYEAR).

Table 3: Overlap between biography and calendar information

	First observation in ARTKALEN (compared to erhebj*)					
		earlier				
erhebj*	same year or later %	1 year %	2 years %	3 years %	4+ years %	Total n
1984	0.1	100.0	0.0	0.0	0.0	11,001
1987	0.0	36.4	33.5	30.1	0.0	505
1988	0.0	100.0	0.0	0.0	0.0	164
1989	0.5	99.5	0.0	0.0	0.0	193
1990	0.0	100.0	0.0	0.0	0.0	180
1991	0.0	100.0	0.0	0.0	0.0	157
1992	0.0	8.4	3.6	88.0	0.0	3,930
1993	0.0	76.6	0.3	2.3	20.7	304
1994	0.2	98.3	0.3	0.2	1.0	918
1995	0.2	99.1	0.0	0.1	0.6	1,037
1996	0.2	97.9	0.0	0.0	1.9	480
1997	0.0	98.5	0.0	0.0	1.5	478
1998	0.7	98.1	0.0	0.2	1.0	415
1999	0.1	26.6	72.8	0.0	0.5	1,821
2000	0.0	90.2	0.9	7.7	1.3	235
2001	0.0	6.3	93.6	0.0	0.0	7,529
2002	0.2	48.1	0.4	39.0	12.4	526
2003	0.1	16.9	81.3	0.1	1.6	2,193
2004	0.0	68.8	4.2	20.1	6.9	432
2005	0.0	89.0	3.4	0.7	6.9	292
2006	0.0	92.2	4.2	0.0	3.7	217
2007	0.0	16.2	83.4	0.1	0.3	1,858
2008	0.0	68.9	2.9	26.9	1.3	309
2009	0.0	89.5	2.1	0.5	7.9	190
2010	0.0	20.6	79.3	0.1	0.1	1,553
2011	0.0	97.1	0.7	2.2	0.1	4,118
Total	0.1	56.7	32.5	10.0	0.7	41,035

Notes: *) Year of biography data collection (variable erhebj).
Source: SOEP v28 (PBIOSPE).

In the majority of cases (56.7 percent), the earliest calendar information is available for the year before the biography interview. This is the case for persons who answered the Biography Questionnaire in their first year as survey respondents. The calendar in the Individual Questionnaire refers to the year before the survey. There are, however, changes over time. In 1998, it was decided that first-time respondents from new samples would not be given the Biography Questionnaire in the first wave but in the second in order to reduce the entry threshold for these new respondents. Consequently, for the majority of persons in years after

[10] Please note that some biographies were collected in 2011 although they are part of Wave 27. This results from the fact that some members of Sample I were interviewed in early 2011 instead of 2010.

new samples were integrated (1999, 2001, 2003, 2007, 2010/11 – Samples E to I), the earliest calendar information is available two years before the biography information was collected. However, first-time respondents who are members of an old sample (e.g., persons who moved into a panel household) still answer the Biography Questionnaire at the time of their first interview. The pattern is quite stable for most years before 1999. A notable exception is the year 1992. This is explained by the integration of East Germany into the SOEP in 1990 (Sample C). The majority of the respondents in this sample answered the Biography Questionnaire in 1992 (and some in 1993). Another exception is the year 1987. In the years 1985 to 1987, the life course matrix was not part of any of the questionnaires. Therefore the respective biography information was only available for persons who were interviewed in 1984. In 1988, biographic information was also collected for persons who became respondents in 1985, 1986, and 1987 (for all years ERHEBJ=1987). While there are only very few cases where the calendar information starts in the same year as the Biography Questionnaire or later (0.1 percent = 21 cases), there are some more cases (0.7 percent = 301 cases) where the biography information was collected a long time after the person started to respond to the Individual Questionnaire (up to 20 years). These are respondents who failed to answer to the Biography Questionnaire at a given time and therefore the biography information was collected later. In these—albeit very rare—cases, there is substantial overlap between the periods covered by the calendar and biography information.

Table 4: Sources of PBIOSPE spells

	n	%	% cum.
biography only	141,029	48.6	48.6
calendar only	101,238	34.9	83.6
1 biography, 1 calendar spell	46,433	16.0	99.6
2+ biography, 1 calendar spell(s)	402	0.1	99.7
1 biography, 2+ calendar spell(s)	796	0.3	100.0
2 biography, 2+ calendar spell(s)	23	0.0	100.0
Total	289,966	100.0	

Source: SOEP v28 (PBIOSPE).

After merging the information from the Biography Questionnaire and ARTKALEN, the data is transformed into spells, whereby each spell is defined by the duration of a given status. A question that arises when merging the data is how to handle overlapping pieces of information. The basic principle is to assign a value of a given status in a given year if the status is recorded in the calendar or in the biography information or both. An example might help to illustrate this: the calendar records full-time employment for the years 2005 and 2007 while the biography records full-time employment for the period from 2000 up to 2006. The

merged data from PBIOSPE contains a spell that begins in 2000 and ends in 2007. However, the initial information is restored by including additional variables, which allows for alternative ways of merging the data (see below). The variables SPELLINF, ERHEBJ, and KALYEAR contain general information on the sources of the information captured in a given spell. Table 4 shows that the majority of spells are based on biography information only (48.6 percent). Slightly more than one-third of all spells (34.9 percent) are not observed in the Biography Questionnaire but only in the calendar data. The remainder of spells contain information from biography as well as calendar data. Usually these spells combine one period observed in the Biography Questionnaire with a period observed in the calendar. Only 0.4 percent of the spells combine more than one period in any of the two sources (SPELLINF=4, 5 or 6).

The variables BEGINB1-ENDYK4 document the initial information from the two different sources and are probably not of interest to the majority of users. However, on the basis of these variables, users are able to fully separate the Biography data from the aggregated ARTKALEN data. This is advisable if you want to use the more detailed ARTKALEN information and combine it with the yearly information from PBIOSPE for earlier years only. The variable names indicate the "source" of the original information utilized (B: Biography - Questionnaire or K: calendar information from the yearly survey). As an example, we discuss one of the spells that combines information on more than one period from any of the two sources. The spell number 4 of person 9205 starts in 1983 and ends in 1994 (SPELLTYP=4: full-time employment). As the variable SPELLINF (=5) shows, this a spell that combines one period from the biography data with two periods from the calendar data. According to the biography data, the person worked full-time from 1983 (BEGINYB1) until 1992 (ENDYB1). There is overlapping information from the calendar data available from 1986 onwards (KALYEAR). According to these data, the person worked full-time from 1986 (BEGINYK1) to 1990 (ENDYK1) and from 1993 (BEGINYK2) to 1994 (ENDYK2). During the years 1991 and 1992, no full-time employment is recorded in the calendar data, which contradicts the information from the biography data.

Table 5: Example of combined spell

persnr	spellnr	spelltyp	beginy	endy	spellinf	erhebj	kalyear	beginyb1	endyb1	beginyk1	endyk1	beginyk2	endyk2
9205	4	4	1983	1994	5	1998	1986	1983	1992	1986	1990	1993	1994

Source: SOEP v28 (PBIOSPE).

In PBIOSPE, no attempt is made to "resolve" such contradictions, as this would require rather strong assumptions. More important, such assumptions would differ according to the research question, which makes it even more difficult to provide a standard solution. Therefore, in such cases, we generate spells in the same manner as in less difficult cases, namely by combining

the information from the calendar and the biography data. In the given example, this results in a full-time employment spell that starts in 1983 and ends in 1994. As mentioned above, there are very few spells that combine information on two or more periods (SPELLINF=4, 5, 6, less than 0.5 percent of all spells). There are even fewer such spells where the period of overlap is as long as in this example, where the biography data was collected many years after the persons joined the survey (ERHEBJ=1998, KALYEAR=1986). However, users who are interested in combining biography and calendar data in a different manner can use the variables BEGINB1-ENDYK4 to fully separate the two types of data and to recombine the data on the basis of different rules of aggregation.

Changes in the previous version of PBIOSPE (release 2009):

The description in this chapter refers to the version of PBIOSPE released in 2012 (waves 1-28). There have been no changes how the data is generated since the previous version, distributed in 2009. But users who are only familiar with older versions of PBIOSPE (releases 2008 and earlier) will observe some differences. In 2009, the data generation has been updated completely, but without changing the basic principles. Therefore, there are only a few barely discernible deviations in the main variables (due to slight changes in the consistency checks of the data). But there are a number of visible changes in the form of additional variables or additional values in already existing variables:

- documentation of censoring:
    - gaps in the data are recorded as spells (SPELLTYP=99)
    - the variable ZENSOR is more detailed and informs about the type of censoring (end of observation window, gap due to missing data)
- documentation of set-up of single spells:
    - new variable KALYEAR: contains the first year for which calendar information is available
    - new variables BEGINB1-3, ENDB1-3, BEGINYB1-3, ENDYB1-3, BEGINK1-4, ENDK1-4, BEGINYK1-4, ENDYK1-4 (these variables replace BEGINBIO, ENDBIO, BEGINYB, ENDYB, BEGINKAL, ENDKAL, BEGINYK, ENDYK): Like the replaced variables, these variables document the original calendar and biography data. The new variables have been added to have a full documentation also for spells in which three or more initial spells are merged (spells with SPELLINF>=4). For the large majority of spells (SPELLINF<=3) only the first of each set of variables is filled. The new variables can be used to separate biography and calendar data, e.g., if you want to combine on your own biography data with data from ARTKALEN.

- additional value in variable SPELLINF: the value 6 indicates that a spell has been constructed out of 2 or more biography and 2 or more calendar spells

- additional changes:
  - variable ERHEBJ: value -2 if no biography information for a person is available (old version: value 0)
  - The variable FEHLCODE is no longer provided, as its values appeared to be more confusing than helpful. It contained information on data problems in the biographies collected in 1984 only. Information on gaps and overlaps is now documented for all years but not in a single variable.

4 BIOJOB: Detailed Information on First and Last Job

by Tanja Schmidt
(Update waves X/Y/Z/BA/BB (2007/2008/2009/2010/2011) by Hansjoerg Haas, based on work of Anita Kottwitz, Daniel Wachtlin, Mathis Schroeder & Thorsten Schneider)

4.1 Overview

Biographical data in the GSOEP stem from various sources. All information for the waves 1984 to 1995 is compiled in the BIOLELA-file of the SIR-GSOEP-database. Since 1996 a standardised version for all samples has been provided, and new biographical data is stored in wave-specific files ($LELA). To have a general phrasing, all biographical files are referred to as LELA-files. (LELA stems from the German 'LEbensLAuf', curriculum vitae.)

The LELA-data relevant for BIOJOB consists of

- the age at entry into the working force
- the type of occupation at entry (blue/white collar worker, self-employed, civil servant)
- detailed occupational information at entry
- changes of occupation
- intended educational degree or vocational/professional training
- the year of the last employment
- the type of occupation in the last job.

Since 2000 a new questionnaire (in the following referred to as Youth Questionnaire) has been provided for respondents who are 16 or 17 years old. The youth respondents answer the Youth Questionnaire instead of the biographical one. The Youth Questionnaire provides less detailed information about the job biography because respondents usually have not entered the labour market at the age of 16 or 17.

In 2001 members of the F sample became part of the biojob population. They had to answer the biography questionnaire if their year of birth was prior to 1982. Members of the F sample with a birth year in the range from 1982 to 1984 answered the Youth Questionnaire.

Members of sample G (2002) answered the biography questionnaire in 2003, Persons who where born between 1986 and 1987 answered the Youth Questionnaire.

Members of sample H (2006) answered the biography questionnaire first time in 2007 and therefore are part of the BIOJOB population.

Sample I has now been moved to the SOEP-Innovation study and, since 2011 (wave 28/BB), are no longer part of the core SOEP population. Members of sample I are still part of of the BIOJOB population until 2010.

Since 2006 respondents who are 16 or 17 years old filled in a youth questionnaire instead of the standard Individual Questionnaire, which provides less detailed information about the current job.

The purpose of BIOJOB is to provide a file, that offers the user convenient access to biographical information on past job activities. Up to now all but two variables of BIOJOB are time-invariant. Information on occupational changes and on the age at the most recent change of occupation refer to the date of the respondent's biography interview.

4.2 Structure and Contents of BIOJOB

BIOJOB consists of generated variables as well as plain questionnaire information. In this section the generated variables are explained and their coding is illustrated.

Concerning different sources of information, the following priority scheme is applied: First the plain information stemming directly from questions on the relevant topic in the latest valid LELA-file is used. In case of inconsistencies, which will be explained later on, the latest valid information stemming from the PBIOSPE file is also used. The PBIOSPE file consists of spell data concerning the retrospective question 'what did you do since the age of 15' in the Biography Questionnaire as well as the question on activities in the last year in the Individual Questionnaire (for detailed information see chapter 3).

4.2.1 Contents of BIOJOB

Population: All persons with an entry in any LELA-/YOUTH-file up to 2011, even if information on employment is missing.

number of cases: 53,063 *waves*: A(1984) - BB(2011) *samples*: A, B, C, D, E, F, G, H, I, J

variables:

HHNR	original household identifier
PERSNR	unique individual identifier
BIOYEAR	year of biography / youth interview
AGEFJOB	age at first job
AGEINFO	information source AGEFJOB
NOJOB	never worked before the time of the interview
STILLFJ	still employed in first job
OCCFJOB	occupational position first job
FULLTIME	first job was a full-time or part-time job
FJBLUE	first job blue collar worker
FJSELFE	first job self-employed

FJSEFSIZ	number of employees FJSELFE
FJWHITE	first job white collar worker
FJCIVS	first job civil servant
ISCO88	International Standard Classification of Occupation 1988, first job
STBA	classification of career according to the Federal Statistical Office, Germany, (Statistisches Bundesamt), version 1992, first job
EGP	Erikson and Goldthorpe's Class Category (EGP), first job
ISEI	International Socio-Economic Index of Occupational Status after Ganzeboom (ISEI), first job
MPS	Magnitude Prestige Scale after Wegener, first job
SIOPS	Treiman Standard Int. Occ. Prestige Scale, first job
REQEDUC	required education for first job
CIVILSFJ	first job was in civil service
NACEFJ	NACE branch code first job
OCCMOVE	number of occupational changes
AGEATMV	age at most recent occupational change
INTEDUC1 to INTEDUC4	intended educational degree
CURREMPL	employed at time of biography interview
YEARLAST	year of last employment
SCOPELJ	last job was a full-time or part-time job
CIVILSLJ	last job was in civil service
NACELJ	NACE branch code last job
OCCLJOB	occupational position last job
LJBLUE	last job blue collar worker
LJSELFE	last job self-employed
LJSEFSIZ	number of employees LJSELFE
LJWHITE	last job white collar worker
LJCIVS	last job civil servant

If data are missing, we use the SOEP missing value definition:

-1	no answer / don't know: item nonresponse
-2	does not apply
-3	after intensive checks a given value was found to be implausible and was fnally deleted (to be interpreted like -1)

4.2.1.1 Description of variables

AGEFJOB/AGEINFO

The variable AGEFJOB provides the age at entry into the working force. AGEINFO is a pointer variable indicating the source of the age information.

In the Biography Questionnaire people either have to give information on their age at entry into the working force or have to state that they have never worked before the time of the interview. The latter information is used in the variable NOJOB.

In the Youth Questionnaire people have to answer whether they are currently working in a regular occupation. They are not asked about the age at their first occupation, but since people answering the Youth Questionnaire are normally at the age of 16 or 17, in most cases we can assume that a full-time job at this age is their first regular employment.

Information on the coding procedure of AGEFJOB is provided in the following subsections where (a) to (i) refer to LELA respondents, (j) to (p) to youth respondents respectively.

LELA-respondents

a) For people who are or have ever been employed at the time of answering the biographical questions their age at the time of entry into the working force is taken from the LELA-files.

b) When we observe, that the person has not been in the working force at the time of responding, but starts to work later on, data of the PBIOSPE-file is used. Using the spell information in PBIOSPE, we are able to collect the age at the first job.

c) A replacement of the LELA-data takes place, when respondents state that they have worked before the age of fifteen, but have a spell entry later than the age of fifteen. This rule is not applied when the spell starts at the age of fifteen, since this is the minimum value for spell data in the questionnaires.

d) The same procedure is applied, when people answer, that they have never worked at the time of the interview, but have a spell which starts before the first interview.

e) In some cases the AGEFJOB value is higher than the start of the corresponding working spell in PBIOSPE. In general, the AGEFJOB value is maintained. Only when the value is greater than 27, is it replaced by the PBIOSPE data. (95% of these cases have an AGEFJOB below 27.)

f) If we observe item non response concerning AGEFJOB and NOJOB, but spell information is available, the missing value is replaced by the corresponding PBIOSPE spell data.

g) If even the 'What did you do since you were 15' question had not been answered, there still was a chance to extract similar information out of the PBIOSPE-file by considering the question 'What did you do every month last year'.

h) If we still had no valid information, the value of AGEFJOB was left out of the dataset.

i) Due to the fact that PBIOSPE information are collected only until the end of the year preceding the actual wave (in this version of BIOJOB: December 2010), for respondents without first job information from both the biography questionnaire and PBIOSPE we further look for a first job using information from the current wave individual questionnaire.

YOUTH-respondents

j) For respondents who are regularly employed, information is taken from the Youth Questionnaire; AGEFJOB is coded as year of questioning minus year of birth minus one (only if the respondent does not state that he/she is still in school, etc.).

k) If we additionally observe a spell starting before the respondent answers the Youth Questionnaire, information from PBIOSPE is used if the respondent does not state in the current questionnaire that he/she is still in school, etc.

l) If respondents answer that they have no regular employment but provide an employment spell starting after the time of the first interview, information from $P (for details see m) is taken if available (only if the respondent does not state that he/she is still in school, etc.).

m) For respondents with inconsistent first job information (simultaneous employment and school attendance/apprenticeship, differing job info in Youth Questionnaire and PBIOSPE) the question 'Are you currently engaged in paid employment?' asked in the Individual Questionnaire turned out to be the most reliable source of information. If a respondent states to be full- or part-time employed in a wave subsequent to the youth interview, AGEFJOB info is derived from the latest information of that kind.

n) If people do not answer at least one of the questions 'Do you currently earn money?" and 'Do you earn money as an apprentice, full-time worker or part-time-worker?" but have an employment spell, like in m) the earliest $P information is taken if available (only if the respondent does not state that he/she is still in school, etc.).

o) If information from the Youth and the Individual Questionnaire (including PBIOSPE) are inconsistent concerning AGEFJOB, then the variable is set to missing.

p) Due to the fact that PBIOSPE information are collected only until the end of the year preceding the actual wave (in this version of BIOJOB: December 2010), for respondents

without first job information from both the Youth Questionnaire and PBIOSPE we further look for a first job using information from the current Individual Questionnaire.

The pointer variable AGEINFO provides the coding information described above. Value labels of AGEINFO indicating the source of information are:

(1) LELA-files (case (a) above)

(2) PBIOSPE if AGEFJOB<15, but spell begin > 15 (c)

(3) PBIOSPE if 'not worked' at interview but later spell begin (b)

(4) PBIOSPE if 'not worked' at interview but earlier spell begin (d)

(5) PBIOSPE if AGEFJOB>27 and earlier spell begin (e)

(6) implausible information therefore set missing (h)

(7) PBIOSPE if 'not worked'-question and AGEFJOB not answered, but 'what done at 15'-question answered (f)

(8) PBIOSPE if 'not worked'-question, AGEFJOB and 'what done at 15'-question not answered, but 'what done last year'-question answered (g)

(9) completely missing

(10) SP if no info from bio interview and PBIOSPE but employment in current Individual Questionnaire (i)

(11) info drawn from Youth Questionnaire(j)

(12) info drawn from PBIOSPE for persons who state in the Youth Questionnaire to be regularly employed and additionally have an employment spell starting earlier (k)

(13) info drawn from $P for persons who state in the Youth Questionnaire not to earn money relating to an employment/job or to earn money but relating to a part-time job or a practical training, and have a subsequent employment spell (l)

(14) info drawn from $P for persons with inconsistent first job information from the Youth Questionnaire or PBIOSPE, but valid employment information from an Individual Questionnaire subsequent to the biography interview (m)

(15) info drawn from $P for persons with item non response in one of the questions 'Do you already earn money from jobs?' or 'Do you earn that money as a trainee, full-time or part-time employee?' and with info in PBIOSPE (n)

(16) completely missing

(17) set to missing because of inconsistent information (o)

(18) info drawn out of UP, the last wave of the SOEP (p)

For more than 50% of the cases with AGEINFO = 3, 7, or 8 (AGEINFO=7 or 8 only if information collected after biography interview) it is possible to extract information from the regular questionnaires.

For respondents with AGEINFO=10 or 11, information referring to the variables OCCFJOB, FJBLUE, FJWHITE, FJSELFE, FJSEFSIZ, FJCIVS, REQEDUC and CIVILSFJ are taken from the Individual Questionnaire (same year as of youth interview). While for respondents having AGEINFO=10 this approach is intuitive, for the persons having AGEINFO=11 we act on the assumption that the job declared in the respective Individual Questionnaire is still the first job of that person. This assumption seems plausible due to the low age of all persons responding to the YOUTH Questionnaire.

In the YOUTH Questionnaire there is no question on the first job. But we can follow up their professional career by the statements given in the activity calendar in the subsequent waves. This can lead to problems if these youths report student jobs. For that reason we decided to take information from the question "Are you currently engaged in paid employment?" asked in the Individual Questionnaires of subsequent waves as the relevant source of information for this group of respondents. The earliest information of that kind determines the variable AGEFJOB.

Some respondents have very low values with respect to AGEFJOB. Most of these jobs turn out to be low-skilled and starting before 1970. The respective persons are either blue collar workers (mostly unskilled) or self-employed (mostly helping in family business). We think these characteristics suggest that these specifications are valid.

NOJOB

The underlying question for the variable NOJOB is 'I have never been employed up to this date'. This variable has the label 'never been employed until the date of the interview" (1).

If NOJOB has a missing value, in general there should exist AGEFJOB information, for special cases, see above. Due to the lack of a comparable question in the Youth Questionnaire, respondents of this questionnaire are given the value (1) as long as no consistent AGEFJOB information is available.

STILLFJ

This variable is based on the question ‘Are you still employed in the same job and at the same place?’. It applies only to LELA respondents who do not state ‘I have never been gainfully employed’ and whose biography interview was after 2000.

Value labels:

(1) Yes
(2) No

FULLTIME

The FULLTIME-variable is used to indicate, whether the first job of a person was a full-time or a part-time job. The value labels are

(0) part-time job or marginal employment

(1) full-time job.

This variable is generated out of the file PBIOSPE for all respondents. For persons with first job information stemming from the Biography Questionnaires, FULLTIME possibly does not refer to the declared first job if PBIOSPE does not contain the respective job spell (i.e. due to item non response or incomplete answering of the activity biography within the Biography Questionnaire).

OCCFJOB

The variable OCCFJOB provides information on the occupational position at the first job. Due to different versions of the questionnaires in the GSOEP´s different samples we face some difficulties. Table 1 gives an overview.

Table 1: Number of Possible Values for Occupational Classifications in the First Job

	Farmers (not self-employed)	Blue Collar Workers	Self-employed	White Collar Workers	Civil Servants
Sample A, B (84-95)	-	5	5	5	4
Sample C (90-95)	4	5	5	4	4
Sample D (94/95)	4	5	5	4	4
Sample A,B,C,D (96)	-	3	4	3	4
Sample A,B,C,D (97-99), E (99)	-	3	4	4	4
Sample A,B,C,D,E (00)	-	3	6	4	4
Sample A,B,C,D,E,F (01)	-	3	10	4	4
Sample A,B,C,D, E,F (02)	-	5	10	6	4
Sample A,B,C,D, E,F,G(06),H(06), I(10),J(11)	-	5	10	6	4

Facing these differences we decided to standardise the occupational classification. Only four types of occupational status were taken into account: blue collar workers, white collar workers, civil servants, and self-employed. The group 'Farmers' is included in the blue collar worker group.

The potential value labels for OCCFJOB are:

(1) blue collar worker
(2) self-employed
(3) white collar worker
(4) civil servant

Further details are provided by the variables FJBLUE (for blue collar workers), FJSELFE (self-employed), FJWHITE (white collar workers), and FJCIVS (civil servants). Table 2 shows the number of possible values.

Table 2: Number of Possible Values for the subcategories of the variable OCCFJOB

	FJBLUE	FJSELFE	FJWHITE	FJCIVS
Sample A,B,C,D, E,F,G,H,I,J(84-11)	9	4	7	4

Due to the fact that the PBIOSPE-file is used for the coding of AGEFJOB in certain cases (see above) there is less information on OCCFJOB than on AGEFJOB.

FJBLUE

The FJBLUE variable provides detailed information on the first occupational status if the person was a blue collar worker. Certain value labels are only given for certain samples, because of the already mentioned differences in the questionnaires.

The following value labels are assigned:

(10) un- and semiskilled farmers (sample C/D)
(11) unskilled worker
(12) semiskilled worker
(20) skilled worker
(30) farmers (sample C/D) being foreman or master craftsman
(31) foreman (sample A/B)
(32) foreman (sample C/D)
(40) master craftsman
(41) farmers (sample C/D) in middle and higher management

FJSELFE/FJSEFSIZ

The FJSELFE variable provides detailed information on the first occupational status if the person was self-employed. FJSEFSIZ gives the number of employees in the respondent's firm. Again there are differences due to the different versions of questionnaires.

The following value labels are assigned:

(10) independent farmer
(20) free lances, self employed academics
(30) other self employed workers
(40) helping within family business

FJSEFSIZ has the following value labels:

(10) number of employees ≤ 9 (all subsamples (see exceptions for samples C/D), up until wave M)
(11) no co-workers (all subsamples, from wave R on)
(12) number of co-workers 1-9 (all subsamples, from wave N on)
(20) number of employees > 9 (all subsamples (see exceptions for samples C/D))
(30) number of employees ≤ 10 (sample C (waves I to L) / D (waves K to L), only if info drawn from biography questionnaire)
(40) number of employees > 10 (sample C (waves I to L) / D (waves K to L) , only if info drawn from biography questionnaire)

FJWHITE

FJWHITE gives detailed information on persons, who were first employed as white collar workers. The subvalues of unskilled labour without degree (21), or with degree (22) are, due to uncomparable values in the LELA-files, only drawn from the $P-Files. (Beginning with BIOJOB 2004).

Potential value labels:

(10) industrial foreman
(20) employee / unskilled labour
(21) same as (20), but without degree
(22) same as (20), but with degree
(30) employee / skilled labour
(40) employee / professional labour
(50) employee / managerial labour

FJCIVS

FJCIVS provides detailed information on first employment as a public servant.

The following value labels occur:

(10) low level civil servant
(20) middle level civil servant
(30) high level civil servant
(40) executive civil servant

ISCO88, STBA EGP, ISEI, MPS, SIOPS

These variables – job classifications and different prestige scores – concerning in each case the first job but are not generated within this file and therefore they are not described within this documentation.

REQEDUC

REQEDUC provides information about the required education for the first job. This information has been asked in the Biography Questionnaire for the first time in the year 2001, but comparable information are gathered by the Individual Questionnaire in all waves.

For all respondents having their first job subsequent to their biography interview, information is drawn out of the generated file $PGEN. Neither respective variables in $P nor those in $PGEN provide full information for all waves. In both data sources no differentiation is made between vocational college degree and university degree. As $PGEN info is equally coded in all waves, it is preferred to $P info.

Potential value labels:

(10) no training
(20) completed vocational training
(30) vocational college or university degree
(31) vocational college degree
(32) university degree

CIVILSFJ

CIVILSFJ indicates if the first job was assigned to the civil service or not. This information has been asked in the 2001 Biography Questionnaire for the first time

For respondents having their first job subsequent to their biography interview, information is drawn out of the generated file $PGEN where this information is provided since the first wave in 1984.

The following value labels occur:

(1) Yes
(2) No

NACEFJ

NACEFJ provides information about the industrial sector of the first job according to the branch classification NACE. This variable is not generated within this file. The description of its value labels is therefore not part of this documentation.

OCCMOVE

The variable OCCMOVE is based on the question 'Did you change your occupation and if you did, more than once?'. Information stems from the year of the biography interview. For respondents of the Youth Questionnaire as well as persons having their first job after the biography interview no information is available.

Labels of OCCMOVE:

(1) never changed occupation
(2) changed once
(3) changed more than once

AGEATMV

This variable is based on the question 'If you changed your occupation, how old were you at the most recent change?'. Information stems from the year of the biography interview. For respondents of the Youth Questionnaire as well as persons having their first job after the biography interview no information is available.

CURREMPL

This variable is based on the question ' Are you gainfully employed at the current time?'. The question applies only to LELA respondents who do not state 'I have never been gainfully employed' or 'Still employed in the first job'. This question has been asked in 1994 for the first time.

Value labels:

(1) Yes
(2) No

YEARLAST

This variable is based on the question ‘When was the last time you were gainfully employed?’. The question applies only to LELA respondents who do not make at least one of the following statements in their biography interview:

‘I have never been gainfully employed.’

‘Still employed in the first job’

‘Gainfully employed at the current time’.

This question has been asked in 1994 for the first time.

SCOPELJ

SCOPELJ indicates if the last job was a full time or part time job.

Information is only provided for respondents who answer the respective question within the Biography Questionnaires. The respective question applies only to respondents who do not make at least one of the following statements:

‘I have never been gainfully employed.’

‘Still employed in the first job’

‘Gainfully employed at the current time’.

This question has been asked in 1994 for the first time.

For youth respondents no information is available.

Value labels:

(1) full-time employed
(2) part-time employment
(3) marginal / irregular employment

CIVILSLJ

CIVILSLJ indicates if the last job was assigned to the civil service or not.

Information is only provided for respondents who answer the respective question within the Biography Questionnaires. The respective question applies only to respondents who do not make at least one of the following statements:

‘I have never been gainfully employed.’

‘Still employed in the first job’

‘Gainfully employed at the current time’.

This question has been asked in 1994 for the first time.

For youth respondents no information is available.

The following value labels occur:

(1) Yes
(2) No

NACELJ

NACELJ provides information about the industrial sector of the last job according to the branch classification NACE. The respective question applies only to respondents who do not make at least one of the following statements in their biography interview:

‘I have never been gainfully employed.’

‘Still employed in the first job’

‘Gainfully employed at the current time’.

This question has been asked in 1994 for the first time.

This variable is not generated within this file. The description of its value labels is therefore not part of this documentation.

OCCLJOB

The variable OCCLJOB provides information on the occupational position at the last job. The respective question applies only to respondents who do not make at least one of the following statements in their biography interview:

‘I have never been gainfully employed.’

‘Still employed in the first job’

‘Gainfully employed at the current time’.

This question has been asked in 1994 for the first time.

Due to different versions of the questionnaires in the GSOEP´s different samples we face some difficulties. Table 3 gives an overview:

Table 3: Number of Possible Values for Occupational Classifications in the Last Job

	Farmers (not self-employed)	Blue Collar Workers	Self-employed	White Collar Workers	Civil Servants
Sample A,B (94/95)	-	5	5	5	4
Sample C,D (94/95)	4	5	5	4	4
Sample A,B,C,D (96-99), E (99)	-	5	5	6	4
Sample A,B,C,D,E (00)	-	5	6	6	4
Sample A,B,C,D, E,F (01/02)	-	5	10	6	4
Sample A,B,C,D, E,F,G(06),H(06),I(10),J(11)	-	5	10	6	4

Facing these differences we decided to standardise the occupational classification. Only four types of occupational status were taken into account: blue collar workers, white collar workers, civil servants, and self-employed. The group 'Farmers' is included in the blue collar worker group.

The potential value labels for OCCLJOB are:

(1) blue collar worker
(2) self-employed
(3) white collar worker
(4) civil servant

Further details are provided by the variables LJBLUE (for blue collar workers), LJSELFE (self-employed), LJWHITE (white collar workers), and LJCIVS (civil servants). Table 4 shows the number of possible values.

Table 4: Number of possible values for the subcategories of the variable OCCLJOB

	LJBLUE	LJSELFE	LJWHITE	LJCIVS
Sample A,B,C,D, E,F,G,H,I,J(84-11)	9	4	7	4

LJBLUE

The LJBLUE variable provides detailed information on the last occupational status if the person was a blue collar worker. Certain value labels are only given for certain samples, because of already mentioned differences in the questionnaires.

The following value labels are assigned:

(10) un- and semiskilled farmers (sample C/D)
(11) unskilled worker
(12) semiskilled worker
(20) skilled worker
(30) farmers (sample C/D) being foreman or master craftsman
(31) foreman (sample A/B)
(32) foreman (sample C/D)
(40) master craftsman
(41) farmers (sample C/D) in middle and higher management

LJSELFE/LJSEFSIZ

The LJSELFE variable provides detailed information on the last occupational status if the person was self-employed. LJSEFSIZ gives the number of employees in the respondent's firm. Again there are differences due to different versions of questionnaires.

The following value labels are assigned:

(10) independent farmer
(20) free lances, self employed academics
(30) other self employed workers
(40) helping within family business

LJSEFSIZ has the following value labels:

(10) number of employees ≤ 9 (all subsamples (see exceptions for samples C/D), until wave M)
(11) number of co-workers = 0 (all subsamples, from wave N on)
(12) number of co-workers 1-9 (all subsamples, from wave N on)
(20) number of employees > 9 (all subsamples (see exceptions for samples C/D))
(30) number of employees ≤ 10 (sample C (waves I to L) / D (waves K to L), only if info drawn from biography questionnaire)

(40) number of employees > 10 (sample C (waves I to L) / D (waves K to L) , only if info drawn from biography questionnaire)

LJWHITE

LJWHITE gives detailed information on persons, who were last employed as white collar workers. The values (21) and (22) are drawn from the BIOLELA-File and from the $P-files.

Potential value labels:

(10) industrial foreman
(20) employee / unskilled labour
(21) same as (20), but without degree
(22) same as (20), but with degree
(30) employee / skilled labour
(40) employee / professional labour
(50) employee / managerial labour

LJCIVS

LJCIVS provides detailed information on last employment as a public servant.

The following value labels occur:

(10) low level civil servant
(20) middle level civil servant
(30) high level civil servant
(40) executive civil servant

INTEDUC1 to INTEDUC4

The variables INTEDUC1, INTEDUC2, INTEDUC3, and INTEDUC4 provide information on the educational degree or the vocational/professional training a respondent intends to complete in the future, asked at the time of the biography interview. We create these four variables since multiple answers are explicitly allowed in the questionnaire. The intended education is stored with respect to the hierarchy given by the questionnaire, i.e., the highest degree is placed in INTEDUC1. For example, a person intending to finish an apprenticeship (1) and university (7) would have INTEDUC1 = 7 and INTEDUC2 = 1. Since this question has been asked for the first time in 1996, we do observe a large number of missing values for INTEDUC1 to INTEDUC4.

(1) apprenticeship
(2) full-time vocational school
(3) technical school
(4) education as a civil servant
(5) accredited professional school
(6) technical or professional college
(7) university

General remark:

Some persons answered more than once the Biography Questionnaire (but this occurs very rarely). The data-set BIOJOB contains only information from one Biography Questionnaire, in most cases the earlier one.

4.3 Steps of Coding

1. Creating a dataset using the data concerning all aspects of the job biography (working force entry, position, etc.) drawn from BIOLELA, MLELA, NLELA, OLELA, PLELA, QLELA, RLELA, SLELA, TLELA, ULELA, VLELA, WLELA, XLELA, YLELA, ZLELA, BALELA, BBLELA (internal DIW files with biographical information up to wave BB), QJUGEND, RJUGEND, SJUGEND, TJUGEND, UJUGEND, VJUGEND, WJUGEND, XJUGEND, YJUGEND, ZJUGEND, BAJUGEND, BBJUGEND (internal DIW youth biography files), QP, RP, SP, TP, UP, VP, WP, XP, YP, ZP, BAP, BBP (needed for consistency checks with respect to the youth biography files).
2. Using the PBIOSPE-data to retrieve spell information during the first occupation.
3. Using PPFAD for personal data (year of birth, sex, sample).
4. Using several files containing generated information about job classification (ISCO), prestige scores and industry sector classification (NACE) concerning the first job.
5. Combining all data concerning the employment biography into a new data file BIOJOB, where priority is set as mentioned above.
6. Coding of AGEFJOB. (for details, see above)
7. Setting the pointer variable AGEINFO indicating the source of the information of AGEFJOB. (for details, see above)

8. Excluding one value for respondents, who stated to have two occupational positions in their first job. Exclusion based on consistency checks.
9. Assignment of the variable OCCFJOB, with respect to the different versions of the questionnaire. Possible value labels: FJBLUE, FJSELFE, FJWHITE, FJCIVS.
10. Definition and assignment of new value-labels for the sub-category FJBLUE, nine labels possible, for details see above.
11. Definition and assignment of new value-labels for the sub-category FJSELFE, four labels possible, for details see above.
12. Definition of the variable FJSEFSIZ, indicating the numbers of employees.
13. Definition and assignment of new value-labels for the sub-category FWHITE, seven labels possible, for details see above.
14. Definition and assignment of new value-labels for the sub-category FJCIVS, four labels possible, for details see above.
15. Coding of the variables REQEDUC and CIVILSFJ.
16. Coding of the variables INTEDUC1 to INTEDUC4.
17. Computing the age at the most recent change of occupation if necessary.
18. Check of consistency: Does information about the age at the most recent change of occupation make sense? If inconsistencies appear, the value is set to a missing value.
19. Assignment of value labels for the variables specifying the last job:
20. Definition and assignment of value labels of the variable CURREMPL indicating if a respondent is gainfully employed at the time of the biography interview.
21. Specification of the year of last employment (YEARLAST).
22. Coding of the variables SCOPELJ and CIVILSLJ.
23. Excluding one value for respondents, who stated to have two occupational positions in their last job. Exclusion based on consistency checks.
24. Assignment of the variable OCCLJOB, with respect to the different versions of the questionnaire. Possible value labels: LJBLUE, LJSELFE, LJWHITE, LJCIVS.
25. Definition and assignment of new value-labels for the sub-category LJBLUE, nine labels possible, for details see above.
26. Definition and assignment of new value-labels for the sub-category LJSELFE, four labels possible, for details see above.
27. Definition of the variable LJSEFSIZ, indicating the numbers of employees.

28. Definition and assignment of new value-labels for the sub-category LJWHITE, seven labels possible, for details see above.
29. Definition and assignment of new value-labels for the sub-category LJCIVS, four labels possible, for details see above.
30. Collecting of job information for people with AGEINFO = 3, 7 or 8, if possible.
31. Collecting of job information for people with AGEINFO = 12, 14 or 16, if possible.
32. Coding of the variable FULLTIME.
33. Definition of missing values for all variables.
34. Hand-editing of inconsistencies between different variables.
35. Final listing
36. Definition and assignment of new value-labels for the sub-category LJCIVS, four labels possible, for details see above.
37. Collecting of job information for people with AGEINFO = 3, 7 or 8, if possible.
38. Collecting of job information for people with AGEINFO = 12, 14 or 16, if possible.
39. Coding of the variable FULLTIME.
40. Definition of missing values for all variables.
41. Hand-editing of inconsistencies between different variables.
42. Final listing

- o

5 The couple history files BIOCOUPLM and BIOCOUPLY, and marital history files BIOMARSM and BIOMARSY

by Juliana Werneburg

(This documentation is based on the comparable FiD documentation[11] and the older SOEP documentation on BIOMARSY and BIOMARSM[12] and has benefited from the work by Olaf Groh-Samberg and Florian R. Hertel. For readability reasons, we do not specifically cite and specify text that has been used directly from the older SOEP document.)

SOEP provides individual partnership and marital histories in four data files. BIOCOUPLY is a spell dataset containing annual spells of the individuals' partnership status, BIOCOUPLM contains the same information as monthly spells. BIOMARSY provides annual spells on the marital status, whereas in BIOMARSM this information is measured on a monthly basis accordingly. All files comprise data on marital and couple biographies of respondents with a personal interview and additionally of adults living in an interviewed household without their own interview if other information is available.

BIOCOUPLY and BIOMARSY contain whole relationship biographies starting at the year of birth. Thus, they as well include retrospective information. Because until wave 27 no questions were asked on a respondents' couple history, BIOCOUPLY includes only those respondents who have answered the biography questionnaire in wave 28 or who were observed in the SOEP at least since the age of 17. In addition, retrospective data in both files are extended by information given in every subsequent personal interview. Note that the marital status in the $PGEN data files, stored as $FAMSTD, is derived from BIOMARSM for the time of the interview. Moreover, the partner indicator PARTZ$$ supplied in the $PGEN data files might not match the information provided in BIOCOUPLM, BIOMARSM or $FAMSTD in its entirety.

This documentation proceeds with a brief description of the four data files. Users interested in more details may read further on how information on couple histories was collected in SOEP and on the editing process of constructing logically consistent marital and couple histories.

[11] Juliana Werneburg (2012): BIOMARSY and BIOCOUPLY. The Marital and Couple History. Chapter 8 in: FiD - Familien in Deutschland - Data Documentation (FiD 2.0, 2010-2011), DIW Berlin.

[12] Olaf Groh-Samberg and Florian R. Hertel (2010): The Marital History Files BIOMARSM and BIOMARSY. Chapter 5 in: Joachim R. Frick / Henning Lohmann (Eds.): Biography and Life History Data in the German Socio Economic Panel (SOEP, v26, 1984-2009), DIW Berlin.

What's new?

BIOCOUPLM and BIOCOUPLY are newly provided datasets that contain couple histories of the respondents. Associated with that innovation the construction process of the two known datasets BIOMARSM and BIOMARSY changed as well. Note that BIOMARSM and BIOMARSY are derived directly from BIOCOUPLM/Y datasets. Because of its specific rules, some spells might have changed or have been added to the histories that were not included with previous data distributions.

Note that in BIOMARSM/Y we renamed the known SOEP code "1" 'single' to 'not married'. This is to indicate that it is possible the respondent might have a partner anyway. If you are interested in this information, we recommend using BIOCOUPLM/Y instead of BIOMARSM/Y.

Please take into account that for the current data distribution v28 months of BEGIN and END in BIOCOUPLM and BIOMARSM are not imputed when the exact date is not known from the data. Instead, the time of the interview is left as the BEGIN or END date when a change is observed in the data. Thus, ties might occur frequently and might have to be treated accordingly.

5.1 BIOCOUPLM: A monthly couple biography

The spell data in file BIOCOUPLM contain prospectively collected information on marital biographies starting with the month of the first personal interview. The data file comprises thirteen variables: the household, couple and individual identifiers HHNR, COUPID and PERSNR as well as ten spell specific variables.

Variable SPELLNR is a chronological index number for each individual's spells during the observation period. Due to the fact that the spells' duration is measured in months, it is important to note that an individual may encounter several events in the same year. In this case the variable SPELLNR allows the user to order spells with respect to the respondent's life course. Variables BEGIN and END indicate the month in which a marital spell starts and ends. Monthly histories start with a value of 1 in January 1983 and ranges until the current margin in month 348, i.e., December 2011. For compatibility reasons with the other monthly spell data in the SOEP, BIOCOUPLM starts in January 1983 although there are no observations prior to January 1984 (i.e., month 13). In principle, the month in which a spell starts is the very same month the previous spell ends in.

Please take into account that for the current data distribution v28 months of BEGIN and END are not imputed when the exact date is not known from the data. Instead, the time of the interview is left as the BEGIN or END date when a change is observed in the data. Thus, ties might occur frequently and might have to be treated accordingly.

Variables of BIOCOUPLM

HHNR	Identifier of original sample household
PERSNR	Personal identifier
COUPID	Couple identifier
SPELLNR	Consecutive spell number (chronological order) [1 to 21]
SPELLTYP	Partnership status (1) Married, spouse in household (2) Married, spouse not in household (3) Coupled, partner in household (4) Coupled, partner not in household (5) Single (6) Married, separated (7) Registered same-sex partnership, living together (8) Registered same-sex partnership, living separately (98) Unknown (99) Unit nonresponse
BEGIN	Month when spell begins [1=Jan 1983 to 348=Dec 2011]
END	End month of spell [1=Jan 1983 to 348=Dec 2011]
BEGINY	Year spell begins [1984 to 2011; -3=implausible; -1=missing]
ENDY	Year spell ends [1984 to 2011; -3=implausible; -1=missing]
PDEATH	Death indicator: spell ends with death of partner?
DIVORCE	Divorce indicator: spell ends with divorce?
CENSOR	Censoring information [0 to 14] (see explanation below, Table 1)
REMARK	Error code (1) Original spell (2) Edited spell (3) Gap spell (4) First spell

Variable SPELLTYP documents partnership status with the possible categories ‘married, spouse in household’, ‘married, spouse not in household’, ‘coupled, spouse in household’, ‘coupled, spouse not in household’, ‘single’, ‘married, separated’, ‘registered same-sex partnership, living together’ and ‘registered same-sex partnership, living separately’. Note that it was only asked for same-sex partnerships since the current wave 28, thus, this information cannot be reconstructed retrospectively. The state ‘single’ does not equal the legal state ‘never married’, but refers only to a person’s current state of having a partner or not. ‘Separated’ spells apply only to former couples who are still married, i.e. who are separated but not yet divorced. Note that ‘separated’ is a redundant spell in terms of completeness of couple histories: it always overlaps with other spells that contain the actual couple status(es) over the entire separation episode. Hence, by deletion of all separation spells the seamless couple history is preserved. The additionally assigned codes ‘unknown’ and ‘unit nonresponse’ indicate a lack of information for the respective period.

In addition, the indicator variables PDEATH and DIVORCE are provided. PDEATH indicates whether a respective spell ends with the death of a person’s partner. Single and gap

spells are assigned a "(-2)" (does not apply). Please note that indicator PDEATH is not restricted to married persons, thus does not only refer to widowhood. By summing up PDEATH over the person's life course, the number of subsequent states of widowhood can be retrieved easily (be careful to change missings into system missing values if you plan to sum up these indicators column-wise). DIVORCE works in a similar fashion. It indicates whether the last marriage spell, that is the separated spell, ended in divorce. Hence, if it did not end in divorce it is coded as a still ongoing marriage. In this case ENDY is updated by the year of the last interview.

It is important to note that a 'first spell' in BIOCOUPLM is not the very first spell of a person, but the first observed partnership status since the person is taking part in the SOEP. Accordingly, the first spell in BIOCOUPLM is (almost) always left-censored. Variable CENSOR informs about whether a spell is left- or right-censored and if so, why. Most spells in BIOCOUPLM are in fact censored. In order to provide the user with detailed information on the nature of censorship we distinguished 'left', 'right', and combined 'left- and right-censored spells' with respect to the reason for censoring: 'first spell' or 'last spell', 'spell ends with death', spell 'precedes' or 'succeeds a gap' (see Table 1). Of course, 'death' and 'last spell' are not mutually exclusive, thus we overwrite the latter with the former reason for being right-censored if the last interview is in the year of death or precedes it.

Table 1: Coding of the variable CENSOR in BIOCOUPLM

Right: Left:	not censored	censored missing	censored before gap	censored last spell	censored death
not censored	0	3	4	5	6
censored first spell	1	7	8	9	10
censored after gap	2	11	12	13	14

Variable REMARK provides information on whether we had to edit or supplement original information provided by respondents in order to construct consistent couple biographies. Spells in BIOCOUPLM are marked as 'edited' (in contrast to 'original') if the editing process involved substitution of or additions to original information as reported in the questionnaire. This happened, for example, if a respondent failed to report a relationship but valid information from a partner was available. Similarly, we inserted a divorce between two marriages, even if it was not specified, because two marriages with separate persons at the same time are not legal (see later section for more details on editing). Furthermore, 'first spells' and 'gap spells' are marked separately. Note that inserted 'single' episodes between two consecutive reported relationships are edited as 'original spell'.

5.2 BIOCOUPLY: A yearly couple biography

The spells in the data file BIOCOUPLY contain retrospectively collected information on couple history since a respondent's year of birth on an annual basis. Since until wave 27 no questions on a respondents' couple history were asked retrospectively, BIOCOUPLY includes only those respondents who have answered the biography questionnaire in wave 28 or who were observed in the SOEP at least since the age of 17. Thus, BIOCOUPLY comprises a much smaller sample than BIOCOUPLM and the BIOMARSM/Y datasets. The newly developed retrospective part of the person questionnaire covers up to four relationships in addition to the current status. Persons aged 17 and interviewed only via Youth Questionnaire are not included in any of the datasets.

Variables of BIOCOUPLY

HHNR	Identifier of original sample household
PERSNR	Personal identifier
COUPID	Couple identifier
SPELLNR	Consecutive spell number (chronological order) [1 to 21]
SPELLTYP	Partnership status (1) Married, spouse in household (2) Married, spouse not in household (3) Coupled, partner in household (4) Coupled, partner not in household (5) Single (6) Married, separated (7) Registered same-sex partnership, living together (8) Registered same-sex partnership, living separately (98) Unknown (99) Unit nonresponse
BEGIN	Month when spell begins [1=Jan 1983 to 348=Dec 2011]
END	End month of spell [1=Jan 1983 to 348=Dec 2011]
BEGINY	Year spell begins [1914 to 2011; -5=not asked for; -3=implausible; -2=does not apply; -1=missing]
ENDY	Year spell ends [1932 to 2011; -5=not asked for; -3=implausible; -2=does not apply; -1=missing]
PDEATH	Death indicator: spell ends with death of partner?
DIVORCE	Divorce indicator: spell ends with divorce?
CENSOR	Censoring information [0 to 14] (see explanation above, Table 1)
REMARK	Error code (1) Original spell (2) Edited spell (3) Gap spell (4) First spell

The following provides an overview of the variables in BIOCOUPLY. The data file contains thirteen variables as does BIOCOUPLM: the household, couple and individual identifiers HHNR, COUPID and PERSNR as well as ten spell specific variables. Variable SPELLTYP

documents the couple status with the possible categories ‘married, spouse in household’, ‘married, spouse not in household’, ‘coupled, partner in household’, ‘coupled, partner not in household’, ‘single‘, ‘separated’, ‘registered same-sex partnership, living together’, ‘registered same-sex partnership, living separately’. The state ‘single’ does not equal the legal state ‘never married’, but refers only to a person’s current state of having a partner or not. Note as before that same-sex partnerships were only asked for since since current wave 28, thus, this information cannot be reconstructed retrospectively. ‘Separated’ spells apply only to former couples who are still married, i.e. who are separated but not yet divorced. Note that ‘separated’ is a redundant spell in terms of completeness of couple histories: it always overlaps with other spells that contain the actual couple status(es) over the entire separation episode. Hence, by deletion of all separation spells the seamless couple history is preserved. The additionally assigned codes ‘unknown’ and ‘unit nonresponse’ indicate a lack of information for the respective period.

Variable SPELLNR is a chronological index number for each individual’s spell during the observation period. The variables BEGINY and ENDY provide the years in which a spell begins and ends, whereas the variables BEGIN and END indicate respondent’s age for users’ convenience. It is important to note that an individual may encounter several events in the same year. In this case the variable SPELLNR allows the user to order spells with respect to the respondent’s life course. In BIOCOUPLY, spell systems for each individual always start with the respondent’s birth. The SPELLTYP of the first spell per definition is ‘single’.

There are some missing values (-1), (-2), (-3) or (-5) in BEGINY as well as in ENDY (BEGIN and END) indicating that we do not know the exact year of change in the couple status. Missing dates indicate that the year was either not reported (-1), does not apply (-2), is implausible (-3), i.e. contradictory to other information, or was not asked for (-5). A gap might also have occurred: a) due to unit nonresponse or b) lack of substantial detail on an announced relationship or c) because it was not asked for. In order to differentiate the reasons for missing information the user can utilize variables REMARK and CENSOR.

In addition, the indicator variables PDEATH and DIVORCE are provided. PDEATH indicates whether a respective spell ends with the death of a person’s partner. Single and gap spells are assigned a “(-2)” (does not apply). Please note that indicator PDEATH is not restricted to married persons, thus does not only refer to widowhood. By summing up PDEATH over the person’s life course, the number of subsequent states of widowhood can be retrieved easily (be careful to change missings into system missing values if you plan to sum up these indicators column-wise). DIVORCE works in a similar fashion. It indicates whether the last marriage spell, that is the separated spell, ended in divorce. Hence, if it did not end in divorce it is coded as a still ongoing marriage. In this case ENDY is updated by the year of the last interview.

Variable REMARK provides information on whether we had to edit or supplement original information provided by respondents in order to construct consistent couple biographies. Spells in BIOCOUPLY are marked as 'edited' (in contrast to 'original') if the editing process involved substitution of or additions to original information as reported in the questionnaire. This happened, for example, if a respondent failed to report a relationship but valid information from a partner was available. Similarly, we inserted a divorce between two marriages, even if it was not specified, because two marriages with separate persons at the same time are not legal (see later section for more details on editing). Furthermore, 'first spells' and 'gap spells' are marked separately. Note that inserted 'single' episodes between two consecutive reported relationships are edited as 'original spell'.

Variable CENSOR indicates whether a spell is left-censored, right-censored or censored on both sides (see Table 1 above within the explanations on BIOCOUPLM). Furthermore, there is information included for the reasons of censoring. In principle, spells might be censored if they precede or follow a gap spell or if BEGIN or END is missing. The last spell for each person is marked as right-censored if a person is still in the SOEP and the current marital status is open ('last spell').

5.3 BIOMARSM: A monthly marital history

Spells in data file BIOMARSM contain prospectively collected information on marital biographies starting with the month of the first personal interview. The data file comprises ten variables: the case and individual identifiers HHNR and PERSNR as well as eight spell specific variables. Variable SPELLTYP documents marital status with the possible categories 'unmarried', 'married', 'divorced', 'widowed' and 'divorced or widowed'. Once married, a later spell 'not married' is not assigned anymore. Note that we renamed the known SOEP code "1" 'single' to 'not married'. This is to indicate that it is possible the respondent might have a partner anyway. If you are interested in this information, we recommend using BIOCOUPLM instead of BIOMARSM. SPELLTYP has one additional category 'divorced or widowed' which indicates that a marriage definitely ended though we do not know whether via divorce or death of the spouse. This may be due to missing information from the biographical questionnaires or due to a respondent's frequent shifts between both categories without ever reporting the death of the partner or divorce as an event. A sixth state is 'gap' indicating a lack of reliable data for this period.

Variable SPELLNR is a chronological index number for each individual's spells during the observation period. Variables BEGIN and END indicate the month in which a marital spell starts and ends. Monthly histories start with a value of 1 in January 1983 and ranges until the current margin in month 348, i.e., December 2011. For compatibility reasons with the other monthly spell data in the SOEP, BIOMARSM starts in January 1983 although there are no

observations prior to January 1984 (i.e., month 13). In principle, the month in which a spell starts is the very same month the previous spell ends in.

Variables of BIOMARSM

HHNR	Identifier of original sample household
PERSNR	Personal identifier
SPELLNR	Consecutive spell number (chronological order) [1 to 9]
SPELLTYP	Marital status (1) unmarried (2) married (3) divorced (4) widowed (5) divorced or widowed (9) gap
BEGIN	Month when spell begins [1=Jan 1983 to 348=Dec 2011]
END	End month of spell [1=Jan 1983 to 348=Dec 2011]
BEGINY	Year spell begins [1984 to 2011; -5=not asked for; -1=missing]
ENDY	Year spell ends [1984 to 2011; -5=not asked for; -1=missing]
CENSOR	Censoring information [0 to 14] (see explanation above, Table 1)
REMARK	Error code (1) Original spell (2) Edited spell (3) Gap spell (4) First spell

It is important to note that a 'first spell' in BIOMARSM is not the very first spell of a person, but the first observed marital status since the person is taking part in the SOEP. Accordingly, the first spell in BIOMARSM is (almost) always left-censored. Variable CENSOR informs about whether a spell is left- or right-censored and if so, why. Most spells in BIOMARSM are in fact censored. In order to provide the user with detailed information on the nature of censorship we distinguished 'left', 'right', and combined 'left- and right-censored spells' with respect to the reason for censoring: 'first spell' or 'last spell', 'spell ends with death', spell 'precedes' or 'succeeds a gap' (see Table 1 above within the explanations on BIOCOUPLM). Of course, 'death' and 'last spell' are not mutually exclusive, thus we overwrite the latter with the former reason for being right-censored if the last interview is in the year of death or precedes it.

Variable REMARK provides information on whether we had to edit or supplement original information provided by respondents in order to construct consistent couple biographies. Spells are marked as 'edited' (in contrast to 'original') if the editing process involved substitution of or additions to original information as reported in the questionnaire. This happened, for example, if a respondent failed to report a relationship but valid information from a partner was available. Similarly, we inserted a divorce between two marriages, even if

it was not specified, because two marriages with separate persons at the same time are not legal (see later section for more details on editing). Furthermore, ‘first spells’ and ‘gap spells’ are marked separately. Note that inserted ‘single’ episodes between two consecutive reported relationships are edited as ‘original spell’.

Please take into account that for the current data distribution v28 months of BEGIN and END are not imputed when the exact date is not known from the data. Instead, the time of the interview is left as the BEGIN or END date when a change is observed in the data. Thus, ties might occur frequently and might have to be treated accordingly.

Note that the generated variable $FAMSTD of the marital status at the time of the interview stored in file $PGEN is derived from BIOMARSM. Thus, $FAMSTD takes on the value derived from the spell system BIOMARSM for the month of an interview. Anyway, the partner indicator PARTZ$$ supplied in the $PGEN data files might not match the information provided in BIOCOUPLM, BIOMARSM or $FAMSTD in its entirety.

5.4 BIOMARSY: A yearly marital biography

Data file BIOMARSY supplements BIOMARSM with retrospectively collected information on the marital history since a respondent’s year of birth. Whereas the marital history in BIOMARSM is measured in months, BIOMARSY depicts the marital biography on an annual basis. BIOMARSY is derived by collapsing the dataset BIOCOUPLY while applying only minor changes to the format. Thus, both datasets are mostly consistent with each other. In fact, BIOMARSY is a subset of BIOCOUPLY, focusing only on marriage, widowhood and divorce. Like BIOCOUPLY, BIOMARSY includes both respondents and non-responding adults living in an interviewed household.

The BIOMARSY file comprises ten variables. The individual and household identifiers HHNR and PERSNR as well as SPELLTYP are basically the same in the data sets. Once married, a later spell ‘not married’ is not assigned anymore. Note again that we renamed the known SOEP code “1” ‘single’ to ‘not married’. This is to indicate that it is possible the respondent might have a partner anyway. If you are interested in this information, we recommend using BIOCOUPLY instead of BIOMARSY. Like BIOMARSM, it is important to notice that SPELLTYP has one additional category ‘divorced or widowed’ which indicates that a marriage definitely ended though we do not know whether via divorce or death of the spouse. This may be due to missing information from the biographical questionnaires or due to a respondent’s frequent shifts between both categories without ever reporting the death of the partner or divorce as an event.

Due to the fact that duration of spells is measured in years it is important to notice that an individual may encounter several (up to three) events in the same year. In this case the variable SPELLNR allows the user to order the spells with respect to a respondent’s life

course. The variables BEGINY and ENDY provide the years in which a spell begins and ends, while the variables BEGIN and END indicate for users' convenience the respective age of the respondent. The spell system for each individual in BIOMARSY always starts with the birth of the respondents. We thus created a first spell for each individual ever interviewed in the SOEP starting in the year of birth and continuing at least until the year in which a person turns age 15. The SPELLTYP of the first spell per definition is 'unmarried'. Even if a respondent reported an earlier marriage in the biography questionnaire we restricted its beginning to age 15, however, we marked this spell as 'left censored'.

Variables of BIOMARSY

HHNR	Identifier of original sample household
PERSNR	Personal identifier
SPELLNR	Consecutive spell number (chronological order) [1 to 11]
SPELLTYP	Marital status (1) unmarried (2) married (3) divorced (4) widowed (5) divorced or widowed (9) gap
BEGINY	Year spell begins [1882 to 2011; -5=not asked for; -3=implausible; -1=missing]
ENDY	Year spell ends [1897 to 2011; -5=not asked for; -3=implausible; -1=missing]
BEGIN	Age spell begins [0 to 102; -5=not asked for; -3=implausible; -1=missing]
END	Age spell ends [15 to 102; -5=not asked for; -3=implausible; -1=missing]
REMARK	Error code (1) original spell (2) edited spell (3) gap spell (4) first spell
CENSOR	Censoring information [0 to 14] (see explanation above, Table 1)

There are some missing values (-1, -3 or -5) in BEGINY as well as in ENDY (resp. BEGIN and END) indicating that we do not know the exact year of a change in the marital status.[13] This can have two reasons. First, it may simply indicate that the respondent did not report the year in which a marriage began or ended (item non-response). Second, a gap might be included because there is an unobserved period of at least two years due to repeated item or partial unit non-response. In order to differentiate the reasons for missing information the user may utilize the variables REMARK and CENSOR. Third, within single case corrections some

[13] There are only two persons in the SOEP who never reported their year of birth, thus having in all but the first spell a missing in BEGIN and END and only in the first spell a missing BEGINY and ENDY.

dates are set implausible (-3) if huge contradictions between partners or overlappings of marriages appear unsolvable.

REMARK indicates whether a spell was 'edited' or 'inserted' (other than 'original') in the same way as in BIOMARSM (see above). Variable CENSOR indicates if a spell is left or right censored or censored on both tails. Furthermore, there is information included for the reasons of censoring. In principle, spells might be censored if they precede or follow a gap spell or if BEGIN or END is missing (-1, -2, -3, -5). In addition to what was said before about gap spells in BIOMARSM, gaps or missing values in BEGIN or END may appear in BIOMARSY if a respondent reported a terminated first marriage and the beginning of a second marriage, but did not report the reason for and/or the year of the end of the first marriage (for more details see below). The last spell of each person is marked as right censored, be it that the person died, quit the SOEP or is still married currently. Some 'marriages' are marked as left censored as well if the respondents report the marriage to began before age 15.

For the construction of BIOCOUPLM, BIOCOUPLY, BIOMARSM and BIOMARSY we first collapsed all the information on current couple and marital status, family events and marital biography into a single data set. In a second step, this information was extensively checked and edited in order to obtain logically consistent couple and marital histories. In the third and last step we put these marital histories into a user friendly spell data format. These steps will be explained in more detail in the following sections.

5.5 Sources of the couple and marital history

For the construction of individual marital histories we gathered information (1.) on the biography from the biographical questionnaire $LELA, (2.) on current marital status conducted in the personal questionnaire $P (or $PAUSL or $PLUECKE or BIOIMMIG) together with (3.) monthly information on the events 'moved in together', 'marriage', 'divorce', 'separation' and 'death of partner' that may have occurred since the last personal interview, also collected in the personal interviews and stored in the files $P. Since 2011, 'starting a new relationship' since the last interview is another event the questionnaire is asking for and that was taken into account. In addition, the couple identifier COUPID and the generated partner pointer PARTZ$$ of the generated dataset $PGEN were used to link current partners living in the same household, to compare their answers, and supplement them if necessary. But mostly the original information is preserved.

Figures 1 to 7 show those parts of the Biography and Person Questionnaire for the survey year 2001 or 2011 which aimed at collecting respondents' couple history. Note again that not all information collected in 2011 is available from the very beginning of the SOEP.

Figure 1: Extraction of the Biography Questionnaire in 2001

69. Are you married or have you ever been married?

Yes, I am or was married ____

No, I was never married ____ → Skip to question 71.

70. When did you get married or when were you married?

If you have been married more than once, please provide information on your earlier marriages, too.

	First Marriage	**Second Marriage**	**Third Marriage**
Married in the year	19__	19__	19__
Still married			
Marriage ended in the year	19__	19__	19__
by means of divorce			
due to the death of your partner			

Figure 2: Introduction to marriage part of the Biography Questionnaire in 2011 (translated)

The following part deals with relationships and marriage

All questions are related to relationships no matter whether you are married or not.

Whether you married in this relationship will be asked later.

These questions ask about both your current and your previous relationships.

We start with your current relationship. Afterwards, we ask retrospectively for the long-term relationship you had previously to the last-mentioned.

Concerning previous relationships, we call them long-term, if they lasted at least six months or longer.
When we ask about the previous long-term relationship, please always consider the relationship you had before the last mentioned, which lasted at least six months.

Figure 3: Questions on previous relationships of the Biography Questionnaire in 2011 (translated)

Now we cover your previous relationship – that is the one prior to the current / the last mentioned relationship.
(Analogously: 2nd to 4th relationship)

203. Did you have a previous long-term relationship?

Yes...☐ (→ 204) No…☐ (→ Go to question 217)

204. When did that relationship start?

Year ☐☐☐☐

205. When and in which way did that relationship end?

Year ☐☐☐☐

Through separation……☐

Through death……☐

206. Did you live with this partner?

Yes...☐ (→ 207) No…☐ (→ Go to question 209)

207. When did you move in with this partner?

Year ☐☐☐☐

208. And when did you give up your common accommodation or when did you or your partner move out?

Year ☐☐☐☐

The common domicile was not vacated……☐

209. Did you marry this partner?

Yes...☐ (→ 210) No…☐ (→ Go to question 217)

210. When did you marry?

Year ☐☐☐☐

211. Did you get a divorce?

Yes...☐ (→ 212) No…☐ (→ Go to question 217)

212. When were you divorced?

Year ☐☐☐☐

Figure 4: Questions on other marriage of the Biography Questionnaire in 2011 (translated)

(only if respondent was not married in the previous three / four relationships)

213. Have you ever been married in another relationship that was not yet mentioned?

Yes...☐ (→ 214) No…☐ (→ Go to question 217)

214. When did you marry in this relationship?

Year ☐☐☐☐

215. Is this marriage still persisting?

Yes...☐ (→ 216) No…☐ (→ Go to question 217)

216. When and in which way did that relationship end?

Year ☐☐☐☐

Through separation……☐

Through death……☐

Until 2010, the Biography Questionnaire was asking only for the last three marriages (see Figure 1). In 2011, design of the retrospective questionnaire collecting information on couples changed markedly. It contains information on up to three previous marriages or relationships (see Figure 3) or three relationships plus one marriage (see Figure 4) that took place prior to the interview. Hence, up to five relationships are possible to record. There, the amount of information collected about the different relationships – the current, the second to fourth previous relationship, and a former marriage if applicable – is not identical. Only long-term relationships, defined as lasting for at least six months, should be mentioned in the questionnaire (see Figure 2).

The personal questionnaire comprises a question on the marital status at the month of interview (see Figure 5), whereby information on registered partnerships is included only since 2011. For immigrants we also used information on the marital status derived from the foreigner questionnaire stored until 1995 in $PAUSL. Additionally, information on immigrants' spouses, their whereabouts, and foreign respondent's marital status were taken from BIOIMMIG. For temporary drop outs we replaced missing information with data from $PLUECKE.

Figure 5: Extraction of the personal questionnaire in 2011 – marital status

133. What is your marital status?

Married, living together with my spouse	
Registered partnership, living together	
Married, living (permanently) separated from my spouse	
Registered partnership, living separately	
Single	
Divorced / registered parntership annulled	
Widowed / life partner from registered partnership deceased	

Figure 6: Questions on the current relationship (translated, Biography Questionnaire 2011)

194. What about the present: Do you currently have a long-term relationship?

Yes...☐ (→ 195) No…☐ (→ Go to question 203)

195. When did the relationship with this partner start?

Year ☐☐☐☐

196. Does your partner live in this household?

Yes...☐ (→ 197) No…☐ (→ Go to question 199)

197. When did you move in with your partner?

Year ☐☐☐☐

198. Please tell the first name of your partner.

_____

199. Did you live together with that partner in the past?

Yes...☐ (→ 200) No…☐ (→ Go to question 201)

200. When did you give up your common accommodation or when did you or your partner move out?

Year ☐☐☐☐

201. Are your married to this partner?

Yes...☐ (→ 202) No…☐ (→ Go to question 203)

202. When did you marry?

Year ☐☐☐☐

Figure 7: Questions on changes during last year (translated, Person Questionnaire 2011)

151. Has your family situation changed since December 31, 2010?
Please indicate if any of the following apply to you and if so, when this change occurred.

	Yes	**2011** in month	**2010** in month
Started a new relationship.....	☐	…☐☐	…☐☐
Got married…….....	☐	…☐☐	…☐☐
Moved in with my partner…….....	☐	…☐☐	…☐☐
I separated from my spouse / partner ...	☐	…☐☐	…☐☐
I got divorced.....	☐	…☐☐	…☐☐
My spouse / partner died……...	☐	…☐☐	…☐☐

The current couple status entails whether someone has a partner and whether they are living together currently. For persons who were interviewed for the first time in 2011, their current partner and marital status were asked within one block (Figure 6). The way the couple status was retrieved in the waves before varied between waves and is not shown here, hence.

For those interviewed again it was asked for any changes that occurred during the last year (see Figure 7). Of those items we only considered information on the month of a marriage, divorce or death of a partner and moving together or separation in order to construct the couple history. Since 2011, the beginning of a new relationship is another event that was included into the questionnaire and that was taken into account. Due to the fact that events were collected retrospectively from first of January of the last calendar year until the month of interview, events in the beginning of a year could have been reported twice. For temporary drop outs we collect the information on marital events from $PLUECKE.

5.6 Construction of couple histories

Information on couple history mainly stems from respondents' retrospective reports on their own history. Thus, no other benchmark on the substance of these reports exists. This led to inconsistencies – for example, overlaps of two relationships occurred, or the ordering of becoming a couple, moving together and marrying was reported differently than expected. Even though many different patterns of the reported histories occurred, sometimes with unusual appearances, they are still possible for the most part. Hence, no verified decision between measurement error and uncommon reality can be made. For that reason, whenver possible, couple histories are left as they were reported. Only in rare cases, restrictions,

corrections on orderings of events or changes of reported years are conducted (read the following carefully for details). Little original information was changed (note: to retrieve original data, see the data file originating from the Biography Questionnaire, $LELA). Further corrections to smooth out irregularities are thus left to the user. Note that consistency checks between waves are done as well. That way, changes between data distributions are possible for former waves.

A clear ordering of spells needs to be sustained if information on timing is missing. Besides, information was asked on an annual basis. Thus, ordering of those spells cannot be done empirically, but has to be decided in advance. For that reason the following default rules were obeyed to obtain logically consistent histories if no other information forced to do otherwise:

1. Every individual couple history starts with the state 'single'. In general it is assumed to last at least until the age of 15. However, we did not restrict age within a relationship, that is, unmarried relationships are allowed to start anytime between the age of 0 and 15. We restricted the age of marriage to be at least 15, though.
2. Every spell set for a certain couple starts with the state 'coupled, partner not in household'. One exception exists: if respondents report a year of moving together that lies before the start of their relationship, this specific couple history starts with 'coupled, partner in household'. Note that in this case the information when this couple moved together is not available in BIOCOUPLM/Y anymore. You would have to look it up in the original source stored in the data file $LELA.
3. If there is no evidence to the contrary, it is assumed that married couples live together and moved together before marriage. That is: for married couples their specific couple history starts with 'coupled, partner not in household' and is followed by a spell 'coupled, partner in household' before there marriage spell 'married, spouse in household' starts. Thus, if a couple moved together in the same year they married, a spell 'coupled, partner in household' is included anyway. Additionally, this assumption applies to a marriage reported as the fourth/fifth reported relationship in 2011 (see Figure 4). Note that dates of becoming a couple and moving together or whether they moved together at all are not known in that case.
4. To ensure that any information possible is included into the spell dataset, a spell 'married, spouse not in household' is created even if the date of moving out and the end of a relationship fall into the same year or if either of the two dates is not known. Likewise, if a partner moves out of the joint household in the year of marriage, the spell 'married, spouse in household' is included anyway, even though it may be redundant. Note that the date of moving out is lost if it was reported to come after the end of a relationship. (It can still be retrieved from the original source stored in the data file $LELA.)

5. Any (formerly) married couple that is not an active relationship anymore, i.e. married who are separated but not yet divorced, ends with a spell ‘separated’. As long as they are not divorced yet, the end date of those separated spells is the same as their last interview year. Again, one exception exists: the separation spell is not added if marriage clearly ended with the death of the respondent’s partner. If not known, a ‘separated’ spell is added. Note that ‘separated’ is a redundant spell in terms of couple history: it always and fully overlaps with other spells that contain the actual couple status(es) over the entire separation episode. Hence, by deletion of all separation spells, the seamless couple history is preserved. It should be noted that information on how a marriage ended – via death or divorce – is saved in the separation spells in PDEATH and DIVORCE, hence care should be applied when deleting those spells.
6. Because BIOCOUPLM/Y documents couple statuses and not marital statuses, it is possible to become single after marriage (in BIOMARSM/Y the only possible change from ‘divorced’ or ‘widowed’ is to ‘married’).
7. As long as it was possible to mention another relationship in the questionnaire (up to four were allowed, see above), it is assumed that periods between those relationships mentioned were ‘single’ states and thus filled accordingly. Note that this relates to long-term relationships only.
8. Information on the death of partner or a divorce is stored in variables PDEATH and DIVORCE. If not applicable, that is if a person is currently single, PDEATH and DIVORCE are set to “-2”, does not apply. If a respective couple is not married, DIVORCE is set to “-2”, does not apply, as well. (Be careful to change missings into system missing values if you plan to sum up these indicators column-wise.)

For many non-interviewed persons’ information on the current relationship can be reconstructed, because the respective partner is available from his or her personal interview. Thus, those non-interviewed adults remain in the BIOCOUPL datasets, even though information is not directly retrieved from them. If information on there current couple status was given by their partner, it is fully copied to these non-respondents as well. Concerning their prior history, the previously mentioned rules are followed: from age 0 to 15 they are stated ‘single’. From age 15 until the current relationship a gap – a spell with SPELLTYP ‘unit nonresponse’ (99) – is included. Non-interviewed persons, for whom no information is available get a gap – a spell with SPELLTYP ‘unit nonresponse’ (99) – from age 15 onwards.

Consistency checks are possible in several ways: a) contradictions between information given by a respondent and her relationship to the head of household (see $PSTELL in the dataset $PBRUTTP). This information was given by the head of household in the Household

Questionnaire and processed in partner pointer PARTZ$$. b) for the current relationship answers of both partners can be compared if a possible partner was identified via COUPID and both were interviewed; c) dates within a personal couple history might appear plausible or not. a) and b) are accomplished and single-case corrections done if reasonable. Concerning c) it is tried to leave as much of the original information as it is given by the respondent to allow for analyses of uncommon patterns as well. Some overlaps and inconsistencies might still appear in the dataset, which you may want to remove. Generally, inconsistencies were dealt with the following way:

1. If a respondent reported an early marriage in the Biography Questionnaire, we restricted its beginning to age 15. But we did not restrict age within a relationship, that is, unmarried relationships are allowed to start anytime between the age of 0 and 15.
2. Within a persons' reported history, dates are not edited or overwritten (such as dates of marriage or begin of relationships). Thus, it is possible that a person has multiple, overlapping relationships in a certain period. That is, relationships do not need to be consecutive. The variable SPELLNR provides the sorting order, whereby the most recent spell comes last, that is, the relationship that ended most recently (no matter which are the starting dates). If no clear dates were available, the sorting in SPELLNR reflects the order of reporting by the respondent.
3. Contradictions in both partners' answers are left as they were, even if both were interviewed in the current wave. E.g. information on starting date or the date of marriage of their joint relationship often differs. As before, neither date is edited or overwritten between both partners of a current couple. To check for contradictions you can link them via couple identifier COUPID or PARTNR$$.
4. For very few cases, there were contradictions about the actual partnership between to individuals in the sample. E.g., if one part of a couple or the position to the head of household suggested a couple in a household, but one or both answered to be single or coupled with someone else (outside the household), it was tried to link them by COUPID in their respective former relationship. Note that those single case corrections are not flagged if they just concern the couple identifier COUPID.

Some respondents refused to answer some or all questions in the relationship part. To some extent, missing values and contradiction are similar in the generating sense. Hence, some of the following rules applied to missing values, may seem familiar:

1. In some cases it is possible that a respondent might have had earlier relationships, but they could not be named anymore due to the above mentioned questionnaire restriction. Thus, it is not known whether there were other partners before the last

mentioned. In these cases, gaps were introduced, that is a spell ‘unknown’ (98) was inserted.

2. If a respondent stated to be married in a relationship, but it is not known whether the couple lived together as well, the above mentioned default course is assumed. In some rare cases details on the relationship are not known, as just a relationship was reported. There thus exists no information whether the couple moved together or whether they were married. Here, gaps, (98) ‘unknown’, are introduced as well.
3. Some interviewed persons did not report a current but – given other information – likely relationship. This contradiction is apparent either by the information of the related partner, by the relation to the head of household or seen in case-to-case checks, e.g. via common children identified through the parent questionnaires. For these cases, the relationship is included into the person’s relationship record with a default record fitting to the identified marital status, but all dates are set to (-1), ‘no answer’. The decision to copy the dates from the corresponding reporting partner is left to the user.

5.7 Construction of marital histories

Marital histories are derived directly from BIOCOUPLM/Y. Hence, the information used is identical to BIOCOUPLM/Y, although shorter and sometimes in a different format. Because of that procedure, spell sets might have changed compared to the previous data distribution. Especially, additional ‘unmarried’ spell might occur before the first marriage The following rules hold:

1. Every individual marital history has to start with the state ‘unmarried’. We did not allow a person to be married before age 15.
2. From ‘unmarried’, one can only change to ‘married’.
3. There is no possible return to ‘unmarried’ once a person was ever ‘married’ (except for the annulations of a marriage. However, given that this event is extremely rare – in particular compared to the many returns to ‘unmarried’ that we find in the data – we did not consider the possibility of annulations [as we did in the former version of BIOMARSM/Y]). The only possible change from ‘married’ is to ‘divorced’, ‘widowed’ or ‘divorced or widowed’.
4. The only possible change from ‘divorced’ or ‘widowed’ is to ‘married’.

Algorithms rearrange and correct original data to produce logically consistent marital histories following these four rules. Thus, we inserted an obligatory first ‘unmarried’ spell starting with birth and ending with the first marriage. Possible contradictions in the data were checked and

edited as well, i.e.: a) contradictions between responses given in the same year in the different, previously described sources; b) illogical sequences between years; c) contradictions between both partners on an identified marriage.

In principle, the notion of "romantic wedding" that was introduced in 2010 is further exerted. This idea assumes that respondents often define their marital status in subjective or even affective terms rather than referring to the legal marital status. For instance, we frequently observe unmarried couples both reporting to be married for one year, and then returning to define themselves as singles (this phenomenon we call "romantic marriage"). Another rather emotional defined marital status often occurs when married couples start to perceive themselves as singles prior to a divorce some years later ("second spring"). A third example refers to unmarried respondents who report to be widowed after their long-term partner died.

In general, indications of being 'unmarried' between two marriages were changed into 'divorced'. If responses on marital status alternate between 'divorced' and 'widowed' without reporting another marriage in between, the first information reported was used to correct these contradictions. Sequences alternating short-termed between 'divorced'/'unmarried' and 'married' are replaced by 'married', as long as a new partner, a longer consistent sequence or the report of an event does not confirm this possibly new marriage. Marriages, that were only reported for one year and not confirmed by the following sequence, are mostly interpreted as "romantic marriage" and thus replaced by the previous and following marital status. If the very end of a reported sequence indicates an 'unmarried' spell after being married and without further information, the marital history ends with a gap instead. Likewise, reported weddings during still existing marriages are ignored mostly.

Note as well, that separation events are taken into account for the first time in the construction process. The difference between two consecutive marriages with different partners and one marriage, where separation and reunion took place, might have effected the number of reported marriages in the datasets.

In cases where information of both partners on their joint marriage is available, dates were counterchecked. Contradictions in both reports were either validated or corrected via single case inspections if possible. If contradictions in dates were not possible to dissolve but neither decisive, i.e. dates differ for example by five years but have no effect on both respondents' sequences of marital history, original information is not replaced. Clearly implausible information is either replaced by the more likely information of one partner or dates were set implausible (-3) if no decision could be made.

Completeness is a fifth criterion for construction of marital histories in the sense that the spell system is a closed system of spells starting from birth to the last year of sample membership. Due to item as well as partial unit non-response (i.e., a person of a SOEP households refuses

to give a personal interview) and due to inconsistent information, ‘gap’ spells are introduced as another category of SPELLTYP on its own. Gap spells can occur at any place in the spell system, i.e., there are no restriction rules like the ones above. If information on marital status is missing for more than two years we inserted gap spells indicating that we have no knowledge of what happened during these periods. Likewise, missing retrospective biography information is indicated by an inserted gap spell. Also, if three finished marriages were reported in the biography questionnaire and a fourth current marriage via the next marital status, another gap is inserted.

Missing information due to item non-response in the life course questionnaire may also affect the dates of the beginning or end of a spell. In these cases the date is declared as missing (-1) if there is no other reliable source like a spouse reporting the beginning of a marriage with the individual in question. The value of -1 also indicates that the first marriage spell and the following divorced or widowed spell are right or left censored, respectively.

As mentioned above, in many cases we observed a change in the marital status whereas the event that caused the change was not reported. In these cases the exact month of the change is missing and was not imputed for the current data distribution v28.

6 BIOBIRTH – A Data Set on the Birth Biography of Female Respondents

by Joachim Frick and Christian Schmitt

6.1 Population and purpose of the data set BIOBIRTH

The file BIOBIRTH is based on every woman who has ever had at least one successful SOEP interview. For each of these women the data set BIOBIRTH documents the birth biography. The annual update focuses on including new information on giving-birth collected in the individual questionnaire or in the biographical questionnaire. Furthermore women who have been interviewed for the first time but who have no information on giving-birth yet are included. The latter are either new female household members or female teenagers who have reached the required minimum for a SOEP participation (16 years). For that reason BIOBIRTH can be described as an accumulative data set, in which the entire birth biography of all female SOEP respondents is presented. BIOBIRTH covers the following information:

(1) sum, birth year and sex of the biological children of a woman up to the last date of interview[14]

(2) Person identifier (PERSNR) of the children – provided the child could be identified within the SOEP.

6.2 Structure of the data set

BIOBIRTH contains the following variables for all women:

- HHNR Invariable number of the original household
- PERSNR Invariable personal number of the woman
- BIOVALID Status of the birth biography:

  (Attention! The variable BIOVALID has been altered in SOEP wave T (2003) containing 2 digit-information).

  10: no birth biographical entries
  (was code "0" in SOEP distribution 2002 and before).

  20: youth biography questionnaire completed, no children in biography (new code).

14 While the wave specific files $KIND present the social, thus time-dependent, mother-child relationships for children aged 16 or younger in the household, BIOBIRTH documents only biological mother-child-relationships.

30: birth biography questionnaire completed, no children in biography
(was code “2” in SOEP distribution 2002 and before).

31: birth biography questionnaire completed, one or more children in biography (was code “1” in SOEP distribution 2002 and before).

- BIOYEAR Year of the survey of the birth biography (1985ff.), respectively “-2” for women without information steming from this special survey instrument.
- BIOAGE Age of the woman at the time of the birth biography survey. If no birth biographical information is available yet, the age at the very first survey is indicated.
- SUMKIDS Total number of children born (more precisely: total number of children identifiable within SOEP by merging all available data up to the time of the last observation (SUMKIDS=BIOKIDS+NEWKIDS).
- BIOKIDS Total number of children identified through the birth biography. For women who haven’t filled in the birth biographical questionnaire yet, the code “-2” applies.
- NEWKIDS Total number of children identified through $PBRUTTO or $KIND.
- KIDGEB[n] Year of birth of the children (for the first child up to the fifteenth child).
- KIDSEX[n] Sex of the children (for the first child up to the fifteenth child).
- KIDPNR[n] Personal number of the children (for the first child up to the fifteenth child), in so far as it is identifiable in the SOEP.
- KIDMON[n] Month of birth of a child (for the first child up to the fifteenth child).

  For the variables KIDGEB[n], KIDSEX[n], KIDPNR[n], and KIDMON[n] identical missing codes apply: The code “-2“ is assigned if there’s no [n]th child found for this mother. The code “-1” applies if information about the [n]th child is found but information about the birth year or the sex is missing or the child could not be identified by a personal identifier (“persnr”) within the SOEP.

For every woman a maximum of 15 entries for children is provided, although the biography questionnaire enables only eight possible entries regarding birth information. If there have been additional births up to the time the biography questionnaire is collected, they are recorded separately by the interviewer and are included in BIOBIRTH. The sequence of children within BIOBIRTH is recorded with regards to the age of the children. The oldest

child is recorded under KIDPNR01 the second oldest under KIDPNR02 and so on. If the age is missing the lowest personal identifier applies.

6.3 Information basis of the birth biography

The main basis of the individual birth biography in BIOBIRTH is normally the information collected by the biography questionnaire[15], in which the number, birth year and sex of the biological children for every woman are collected. For women with information on children steming from the biography questionnaire the BIOVALID code "31" is assigned. Women who completed this questionnaire but did not report on any births receive the code "30". In addition, the variable BIOAGE contains the age at the time of the collection of the life history. Apart from this one-time collection within the scope of the first SOEP interview, there is no other possibility in SOEP to collect information on the number, the residence status, or if appropriate, the year of death of children who were born before the first SOEP interview.

A minority of women have no information from the biographical questionnaire due to several reasons[16]. In this cases the variable BIOVALID has the code "10" and the variable BIOAGE contains the age at the first time of SOEP interview. The group can be divided into different sub-populations and is in principle affected by the risk of underestimating the total number of births:

- Woman who were at the time of the first interview only 16 years old. In most cases these women participate at a later date in the biography survey. Thus, the mother-child relationship recorded earlier in BIOBIRTH can be checked later with the birth biography.
- Women who are 30 years old or younger at the first interview. In this sub-population, children are not yet adults and live in most cases in the parents' household. Since information from the biographical questionnaire is missing, a final distinction in social and biological children is not possible.
- Women who are over 30 years old at the time of the first interview. Some of the children don't live any more in the parents' household at the time of the first interview and therefore they are not part of the survey population. For that reason the number of biological children might be underestimated in this group of women (over 30 years) to a larger extent compared to younger women.

[15] The information collected over the course of the biography survey for every woman on the number, on the year of birth, on the sex, on the residence status within the household, and, if necessary, on the year of death of the biological children was stored up until 1995 in a biography data set which spanned the various waves (BIOLELA). Since 1996, this biographical information has been stored in wave specific files ($LELA) Both BIOLELA as well as $LELA belong to the files which have never been distributed to the SOEP-user community.

[16] Beside the reason 'refusal', the collection date of the life history biographies differ among SOEP sub-samples.

6.4 A new source of biographical information – the youth questionnaire

From wave T onwards the data within BIOBIRTH includes information from a further biographical instrument: the youth biography. The youth-questionnaire has been in ciculation since the year 2000 (wave Q) for all young adults, one year after they have reached the required age for completing the individual-questionnaire. Apart from exceptions described in table 1, this means the age of 17. What is important for the BIOBIRTH data-set is that these individuals who fill in the youth-questionnaire complete this questionnaire *instead* of the biographical questionnaire. The age groups which instead fill in the youth-questionnaire of the biographical module differ slightly among the SOEP-subsamples (table 1):

Table 1: Target population of the Youth Questionnaire by year, sample and age

Sample	2000	2001	2002	2003 and later
A-E	17 years	17 years	17 years	17 years
F		17-19 years	17 years	17 years
G and beyond				17 years

The youth-biography does not contain any birth-biographical modules. Assuming that only very few women give birth before the age of 17 and that these few can be identified in the household context (as long as they remain within the SOEP) this does not pose any problems for compiling the birth-biography of the respondents. Nevertheless, a few changes to the BIOBIRTH data-set have to be outlined:

- In the variable BIOVALID a new code ("20": "youth biography questionnaire completed") is added. As the youth questionnaire doesn't contain any information about own children the addendum "no children in biography" is always added to the code "20".
- While calculating the age at the time of the biographical questionnaire (BIOAGE), the age upon completion of the youth questionnaire is applied.
- The variable BIOKIDS always remains at zero as no biographical information on parenthood can be derived from the youth-biography (in this cases no missing code is applied in BIOKIDS).

6.5 Identification process of the children in the SOEP data base

The starting point for the process of identifying children is the relationship of a household member to the head of the household (HH) (variable $STELL in the file $PBRUTTO). The variable $STELL has the following codes:

Code	Label
0	head of the household (HH)
1	spouse of HH

2 "life companion" of HH
3 daughter / son (including adopted/step-children) of HH
4 foster child of HH
5 daughter in law / son in law of HH
6 father / mother of HH
7 father in law / mother in law of HH
8 brother / sister / brother in law / sister in law of HH
9 grandchild of HH
10 other relation to HH
11 not related to HH
12 child of "life companion" of HH (included since 1999)

However, there are only certain combinations among household members in which a biological mother-child relationship among a female adult and another person can be assumed.

Potential mother-child relationships as a combination of the variable $STELL

$STELL of the		**Potential mother-child relationship**
woman	**another person**	**In this case the person is the…**
0	3	Child of reference person (reference person = head of he household)
1	3	Child of the wife of reference person
1	11	Child of the wife of reference person, but not child of reference person
1	12	
2	3	Child of "life companion" of reference person and of reference person
2	11	Child of "life companion" of reference person but not of reference person
2	12	
3	9	Child of daughter of reference person
4	9	Child of foster child of reference person
5	9	Child of daughter in law of reference person (3 generation household)
6	0	Child is reference person, lives with his mother in the same household
6	8	Child is the sister / brother of reference person, the siblings live with their mother in the same household
7	1	Child is spouse of reference person and lives together with spouse and mother in the same household
7	8	Child is daughter / son of the mother in law of reference person, but not the spouse of the reference person rather the sister in law / brother in law of reference person
8	10	Child is niece / nephew of reference person, mother is sister / sister in law of reference person
9	10	Child is another relation to reference person, great grandchild of reference person
10	10	Mother and child have another relation to reference person
11	11	Child and mother are in no way related to reference person

For the remaining unassigned children within a household a thorough check is performed to scan for mother-child ties that can be derived from more complex household relationships. It should be noted here that the larger the number of persons living in a household, the more complicated the relationships become among the individual household members. For that reason only the combination of the information from the biography questionnaire with the information from $PBRUTTO (and in most cases the mother indicator $KMUTTI from $KIND) provides the most reliable definition of a biological mother-child relationship.

6.6 Identification of the children for women with biographical data

If one woman mentioned in the biography questionnaire the birth, the sex and the birth year of a child, the identification process has been started. In the first step, the program identified the woman's relationship to the reference person and looked on the basis of the mother-child combination - as illustrated above- for a potential child in the household. If the birth year and the sex of this person are the same as named in the life history interview of the assumed mother, the person has been identified as the child of this woman. Since the majority of the households with children present small nuclear families including one potential mother, this kind of identification process was completely sufficient. In other, rather complex households a careful hand editing has been examined, in order to identify the 'right' child to the 'right' mother. The same has been done, if the sex or the important information – the year of birth - of a child mentioned in the biography questionnaire of a woman was missing.

In the case of a successful identification the variable KIDPNR[n] has been filled with the person identifier of this child. Children, for whom the woman in the biography questionnaire has reported that they were deceased or had moved out, were assigned the personal number (KIDPNR[n]) "-1", for missing information, in BIOBIRTH.

6.7 Identification of the children for women who have no biography data/ not completed the biography questionnaire

To get as close as possible to the definition of a biological child, for this group of women only, specific relationships among household members were allowed. Since the main important information – from the biography questionnaire - is missing, a careful analysis of the composition and the history of the household in which the children live has been examined, in order to assign the 'right' child to the 'right' mother.

Potential mother-child relationships as combination of the variable $STELL

$STELL of the		Potential mother-child relationship
woman	another person	In this case the person is the…
0	3	Child of reference person
1	3	Child of the wife of reference person
2	3	Child of "life companion" of reference person and of reference person
3	9	Child of daughter of reference person

6.8 Last step of the identification process

If a child is identified through the above described process and it was not yet old enough to be surveyed itself (under 16 years old) and lived at least in a partially realized household, an additional source of information is made available along with the data set $KIND to check the relationship. For each person in this group (under 16 years old), there is an indicator for the mother ($KMUTTI) containing the person identifier of the mother. Although this information does not necessarily indicate a biological mother-child relationship, at least the identification process for a large portion of the women could be checked or compared.

6.9 Updating BIOBIRTH

As mentioned in section 6.1 the annual update of the data set BIOBIRTH is examined with respect to two dimensions. First, updating the birth biography of the BIOBIRTH population and second, extending BIOBIRTH by new persons. The latter are either new female household members or female teenagers who have reached the required age for giving a first interview (16 years). Since the extension of BIOBIRTH follows the generation rules as described above, the following only summarizes the updating of the birth biography of the BIOBIRTH population.

New born children in the SOEP study are documented in the variable $PZUG in the data set $PBRUTTO:

Code	Label
11	Born since the last survey
17	Born before the last survey, but only now first mentioned
31	Born two years ago

For this group of new born persons the identification process starts this time in the reverse direction: While we have looked before for the children, now we are looking for the mothers.

Potential mother-child relationships as combination of the variable $STELL

$STELL of the		**Potential mother-child relationship**
woman	**another person**	**In this case the person is the…**
0	3 /4	Child / foster child of reference person
1	3	Child / foster child of the wife of reference person
2	3	Child / foster child of the reference person's life's companion and of the reference person
2	11	Child / foster child of the reference person's life's companion, but not
2	12	child of reference person
3	9	Child of daughter of reference person
4	9	Child of foster child of reference person
5	9	Child of daughter in law of reference person
6	8	Child is sister / brother of reference person, the brothers and sisters live with their mother in the same household
7	8	Child is daughter / son of the mother in law of reference person, but not the spouse of the reference person rather the sister in law / brother in law of reference person
8	10	Child is niece / nephew of reference person, mother is sister / sister in law
9	10	of reference person Child is great grandchild of reference person, mother is granddaughter of reference person

Since new born children per definition could not be a reference person or a partner of the reference person, several $STELL codes are excluded. Further combinations are excluded, in which the basis of a mother-child relationship could not have been identified with sufficient significance. Again, the majority of the households with children, in particular with small children, are nuclear households or lone parent households. For that reason this kind of identification process often yields in a successful identification of the new born children. Nevertheless, the mother indicator information ($KMUTTI) from $KIND was again drawn upon as a check.

An overview over central variables in the file BIOBIRTH (Version 2011 / up to Wave BB)

BIOVALID Status of the birth biography					
		Frequency	Percent	Val. Percent	Cum. Percent
Valid	10 No Birthbio. - No Kids from Bio.	2311	7,8	7,8	7,8
	20 Youthbio - No Kids from Bio.	1936	6,6	6,6	14,4
	30 Birthbio - No Kids from Bio.	9375	31,7	31,7	46,1
	31 Birthbio - Kids according to Bio.	15921	53,9	53,9	100,0
	Total	29543	100,0	100,0	

BIOYEAR Year of the survey of the birth biography

	Frequency	Percent	Val. Percent	Cum. Percent
-2 No Birth-Biography	2311	7,8	7,8	7,8
1985	6619	22,4	22,4	30,2
1986	83	,3	,3	30,5
1987	104	,4	,4	30,9
1988	220	,7	,7	31,6
1989	211	,7	,7	32,3
1990	202	,7	,7	33,0
1991	161	,5	,5	33,5
1992	2635	8,9	8,9	42,5
1993	230	,8	,8	43,2
1994	592	2,0	2,0	45,2
1995	528	1,8	1,8	47,0
1996	255	,9	,9	47,9
1997	231	,8	,8	48,7
1998	203	,7	,7	49,4
1999	1033	3,5	3,5	52,9
2000	243	,8	,8	53,7
2001	4928	16,7	16,7	70,4
2002	467	1,6	1,6	71,9
2003	1321	4,5	4,5	76,4
2004	410	1,4	1,4	77,8
2005	329	1,1	1,1	78,9
2006	268	,9	,9	79,8
2007	1333	4,5	4,5	84,3
2008	290	1,0	1,0	85,3
2009	213	,7	,7	86,0
2010	1165	3,9	3,9	90,0
2011	2958	10,0	10,0	100,0
Total	29543	100,0	100,0	

SUMKIDS Sum of kids in BIOBIRTH

		Frequency	Percent	Val. Percent	Cum. Percent
Valid	,00	10734	36,3	36,3	36,3
	1,00	5708	19,3	19,3	55,7
	2,00	8151	27,6	27,6	83,2
	3,00	3190	10,8	10,8	94,0
	4,00	1084	3,7	3,7	97,7
	5,00	380	1,3	1,3	99,0
	6,00	156	,5	,5	99,5
	7,00	64	,2	,2	99,7
	8,00	44	,1	,1	99,9
	9,00	15	,1	,1	99,9
	10,00	11	,0	,0	100,0
	11,00	2	,0	,0	100,0
	12,00	4	,0	,0	100,0
	Total	29543	100,0	100,0	

BIOKIDS Sum of kids derived from birth biography

		Frequency	Percent	Val. Percent	Cum. Percent
Valid	-2	2311	7,8	7,8	7,8
	0	11311	38,3	38,3	46,1
	1	4946	16,7	16,7	62,9
	2	6759	22,9	22,9	85,7
	3	2688	9,1	9,1	94,8
	4	929	3,1	3,1	98,0
	5	328	1,1	1,1	99,1
	6	139	,5	,5	99,6
	7	60	,2	,2	99,8
	8	43	,1	,1	99,9
	9	16	,1	,1	100,0
	10	9	,0	,0	100,0
	11	2	,0	,0	100,0
	12	2	,0	,0	100,0
	Total	29543	100,0	100,0	

NEWKIDS Sum of kids identified through $PBRUTTO

		Frequency	Percent	Val. Percent	Cum. Percent
Valid	0	25684	86,9	86,9	86,9
	1	2125	7,2	7,2	94,1
	2	1329	4,5	4,5	98,6
	3	322	1,1	1,1	99,7
	4	67	,2	,2	99,9
	5	12	,0	,0	100,0
	6	3	,0	,0	100,0
	10	1	,0	,0	100,0
	Total	29543	100,0	100,0	

7 BIOBRTHM – The Birth Biography of Male Respondents in the SOEP

by Christian Schmitt

7.1 Contents of the BIOBRTHM data-set

In the year 2001 a new biographical module was implemented in the SOEP. The birth-biography – up to then only available for women – was also collected from the men in the SOEP. The information is included in the file BIOBRTHM that was introduced to the family of SOEP biography data-sets in wave T (2003). This documentation deals with important features and limitations that are unique to BIOBRTHM data-set. The data-set BIOBRTHM – the birth biography of male respondents – contains the same set of variables as the initial BIOBIRTH file – the birth biography of female respondents. Refer to the documentation of the latter file for basic information, concerning the SOEP birth biography.

The variables are:

- HHNR Invariable number of the original household.
- PERSNR Invariable personal number of the man.
- BIOVALID Status of the birth biography:

  10: no birth biographical entries.

  20: youth biography questionnaire completed, no children in biography.

  30: birth biography questionnaire completed, no children in biography.

  31: birth biography questionnaire completed, one or more children in biography.
- BIOYEAR Year of the birth biography survey (2001ff.), respectively "-2" for men without information steming from this special survey instrument.
- BIOAGE Age of the man at the time of the birth biography survey. If no birth biographical information is available yet, the age at the very first survey is indicated.
- SUMKIDS Total number of children born (more precisely: total number of children identifiable within SOEP by merging all available data up to the time of the last observation (SUMKIDS=BIOKIDS+NEWKIDS).

- BIOKIDS Total number of children identified through the birth biography. For men who haven't filled in the birth biographical questionnaire yet, the code "-2" applies.
- NEWKIDS Total number of children identified through $PBRUTTO or $KIND.
- KIDGEB[n] Children's year of birth (for the first child up to the fifteenth child).
- KIDSEX[n] Sex of the children (for the first child up to the fifteenth child).
- KIDPNR[n] Personal number of the children (for the first child up to the fifteenth child), as far as this person is identifiable in the SOEP.
- KIDMON[n] Month of birth for the children (for the first child up to the fifteenth child).

  For the variables KIDGEB[n], KIDSEX[n], KIDPNR[n], and KIDMON[n] identical missing codes apply: The code "-2" is assigned if there's no [n]th child found for this mother. The code "-1" applies if information about the [n]th child is found, but information about the birth year or the sex is missing or the child could not be identified by a personal identifier ("persnr") within the SOEP.

For a complete overview of contents and construction of variables refer to the document "BIOBIRTH – A Data Set on the Birth Biography of Female Respondents"

7.2 Population of BIOBRTHM

Like the female birth-biography, the population of BIOBRTHM consists of all respondents of the relevant sex (here, all *male* respondents) with at least one personal interview up to the current wave. However, two features differentiate the male from the female birth-biography data:

- First: only information about men with at least one completed questionnaire *in 2001 or later* is contained in the BIOBIRTHM file.
- Second: information from the birth-biography will only be added for *new* Panel members in 2000 and later, as only these persons fill in a new biography interview (usually one wave after the first participation in the SOEP which in our case means in 2001 or later). Most of the members who have completed a questionnaire before 2000 have also already completed the biographical modules that are only collected once for every person.

The module collecting information about (non-) fatherhood was introduced in 2001. Therefore, most men in subsample "F" (which started in 2000) have completed the birth-biography as most of the biographical questionnaires are usually completed one wave after the starting wave. For all the other men within BIOBRTHM, the information about fatherhood is

underestimated as only the context of the household can be taken into account when considering the number of children (see the documentation of the file BIOBIRTH, 6.3 for the reasons of this underestimation and further details). Furthermore, this underestimation is more severe for men who did not complete the biographical questionnaire than it is for women as children remain more often with the mother than with the father after the split of a relationship.

7.3 Construction of variables - Particularities

The construction of variables for the data-set BIOBIRTHM mostly resembles the generation of variables within the BIOBIRTH file. The major difference affects information from the data-set $KIND from which the mother-child-pointer ($KMUTTI) is used when generating the birth-biography of women to link the right child to a given mother. For men such a pointer is not available. Therefore, the focus of variable construction remains on utilizing information from $PBRUTTO where the household context is taken into account. For details of BIOBRTHM variable construction based on $PBRUTTO please refer to the documentation of the file BIOBIRTH as the construction remains – besides the mentioned differences – identical.

An overview over central information in the file BIOTWIN (Version 2011 / Wave BB)

BIOVALID Status of the birth biography

		Frequency	Percent	Val. Percent	Cum. Percent
Valid	10 No Birthbio. - No Kids from Bio.	7285	36,2	36,2	36,2
	20 Youthbio - No Kids from Bio.	1890	9,4	9,4	45,5
	30 Birthbio - No Kids from Bio.	3673	18,2	18,2	63,8
	31 Birthbio - Kids according to Bio.	7298	36,2	36,2	100,0
	Total	20146	100,0	100,0	

BIOYEAR Year of the survey of the birth biography

		Frequency	Percent	Val. Percent	Cum. Percent
Gültig	-2 No Birth-Biography	7285	36,2	36,2	36,2
	2000	118	,6	,6	36,7
	2001	4501	22,3	22,3	59,1
	2002	432	2,1	2,1	61,2
	2003	1370	6,8	6,8	68,0
	2004	412	2,0	2,0	70,1
	2005	338	1,7	1,7	71,8
	2006	262	1,3	1,3	73,1
	2007	1243	6,2	6,2	79,2
	2008	304	1,5	1,5	80,7
	2009	225	1,1	1,1	81,9
	2010	1084	5,4	5,4	87,2
	2011	2572	12,8	12,8	100,0
	Gesamt	20146	100,0	100,0	

SUMKIDS Sum of kids in BIOBRTHM

		Frequency	Percent	Val. Percent	Cum. Percent
Gültig	,00	9070	45,0	45,0	45,0
	1,00	3652	18,1	18,1	63,1
	2,00	4947	24,6	24,6	87,7
	3,00	1769	8,8	8,8	96,5
	4,00	493	2,4	2,4	98,9
	5,00	144	,7	,7	99,6
	6,00	44	,2	,2	99,9
	7,00	13	,1	,1	99,9
	8,00	9	,0	,0	100,0
	9,00	2	,0	,0	100,0
	10,00	1	,0	,0	100,0
	11,00	1	,0	,0	100,0
	12,00	1	,0	,0	100,0
	Gesamt	20146	100,0	100,0	

BIOKIDS Sum of kids derived from birth biography

		Frequency	Percent	Val. Percent	Cum. Percent
Gültig	-2	7285	36,2	36,2	36,2
	0	5563	27,6	27,6	63,8
	1	2160	10,7	10,7	74,5
	2	3333	16,5	16,5	91,0
	3	1268	6,3	6,3	97,3
	4	364	1,8	1,8	99,1
	5	114	,6	,6	99,7
	6	37	,2	,2	99,9
	7	11	,1	,1	99,9
	8	9	,0	,0	100,0
	9	2	,0	,0	100,0
	Gesamt	20146	100,0	100,0	

NEWKIDS Sum of kids identified through $PBRUTTO

		Frequency	Percent	Val. Percent	Cum. Percent
Gültig	0	16083	79,8	79,8	79,8
	1	1900	9,4	9,4	89,3
	2	1595	7,9	7,9	97,2
	3	433	2,1	2,1	99,3
	4	105	,5	,5	99,9
	5	22	,1	,1	100,0
	6	4	,0	,0	100,0
	7	2	,0	,0	100,0
	10	1	,0	,0	100,0
	11	1	,0	,0	100,0
	Gesamt	20146	100,0	100,0	

8 BIOTWIN – Information on TWINS in the SOEP

by Christian Schmitt

8.1 Population and contents of the data set BIOTWIN

The file BIOTWIN contains all twins that were ever identified within the SOEP. To be classified as a twin, a person has to:

- have exactly the same age as his or her sibling,
- have a relationship to the head of the household that indicates that he or her and a second persons are siblings and
- has to have the identical mother (as far as a pointer to the mother is available).

Furthermore, it is not only twins that are recorded in the BIOTWIN data set, but also triplets or quadruple siblings. The following variables are stored within the BIOTWIN data set:

- HHNR Invariable number of the original household.
- PERSNR Invariable personal identifier of the first sibling .
- PNRTWIN Invariable personal identifier of the second sibling, the twin.
- PNRTRIP Invariable personal identifier of the third sibling.
- PNRQUAD Invariable personal identifier of the fourth sibling.
- PNRMOTH Pointer to the personal identifier of the mother of the twin-group.
- BIOMONOZ Monozygotic group? Information if the group is monozygotic.
- INFSOURC Source of information from which the status of being a twin is derived

The central variable PERSNR is assigned to the sibling with the lowest personal identifier in the twin group. The PNRTWIN and – in rare cases if available – PNRTRIP or PNRQUAD contain the personal identifier of second, and third or fourth sibling in the group. This means that every case in the data set consists of a *group* of twins (or triplets or quadruplets). The code "-2" is assigned to PNRTRIP and/or PNRQUAD if a third or fourth twin sibling doesn't exist. PERSNR and PNRTWIN however should always contain valid codes.

The variable PNRMOTH provides the link to the mother of the group and is derived from the data sets $KIND and/or BIOBIRTH.

8.2 The twin survey of 2006

In 2006, a questionnaire was distributed among all households with potential twin groups, identified up till then. The aim was to validate that none of these twins had been identified by mistake. The variables INFOTWIN and BIOMONOZ contain new information which was derived from this survey.

The result of the survey could widely validate the selection of the twin population, contained in the BIOTWIN data set of the SOEP. More than 80% of households with twins could be contacted and were interviewed in the twin survey. Among these only 3 groups of twins turned out to be identified erroneously (those false positives were removed from the BIOTWIN data set). Thus the algorithms of identifiying twins within the SOEP could proove to be widely reliable. Additional information that was collected with the twin survey contributed to identifying a number of mothers of twins, for whom the mother-child-link was missing previously. Furthermore the twin survey provided additional information on monozygotic respectively dizygotic twins. The variable BIOMONOZ was extended, in order to reflect this additional information (see below for more details).

8.3 Construction of variables in the data set BIOTWIN

The variable BIOMONOZ[17] indicates if the group is monozygotic. If the information could be validated in the twin-survey in 2006 the code is set to 1 for monozygotic twins and 2 for dizygotic twins. If the information on being mono- or dizygotic twins could *not* be validated in the twin survey, which was carried out in 2006, the code is set to 0 if the sex of all the siblings is identical, and this group thus *might* be monozygotic. Please pay attention to the fact that the labels and values of the variable BIOMONOZ from wave W onwards are not consistent with values and labels from previous waves.

The variable INFOTWIN is introduced with wave W and provides information on the source from which the status of being a member of a twin group is derived from and whether this information could be validated in the twin-survey in 2006.

INFOTWIN can take the following characteristics:

1 Generated from identical number of household and month of birth – not validated by the twin survey from 2006

2 Possible Twin or Triplet – Information not revisable in twin survey 2006

3 Possible Twin or Triplet –Answer refused in twin survey 2006

[17] This variable existed before wave W but was restructured to reflect the addtional information which became available with the 2006 twin questionnaire.

4 Twin or Triplet – Information validated by twin survey 2006

5 Twin or Triplet – New since 2007
(congruent years & months of birth)

6 Twin or Triplet – New since 2007
(congruent years / missing info on months of birth)

The selection of twins within the SOEP, which compiles the data set BIOTWIN, is based on the either the month of birth, or an identical year of birth. Priority is given to congruent months of birth, as a woman might – in rare casaes – give birth at two different times in a year. Hence the month of birth plays a central role in identifiying potential twin-groups. According to that logic people with a) valid month of birth information or b) identical month of birth, or c) with an identical year of birth *and* missing data on the month of birth among both siblings are classified as twins.

In a second step, the relationship of these potential twins to the head of household is scanned ($STELL). If the relationship of both persons assures that they are siblings, then they are assumed to be twins.

In a third step the pointer to the mother is checked for both siblings with focus on the files $kind / BIOBIRTH. If this maternal link is identical for both siblings, it is transferred into the variable PNRMOTH.

An overview over central information in the file BIOTWIN (Version 2011 / Wave BB)

Table 1 Sibling groups in the BIOTWIN

group size		valid mother pointers
twins	267	251
triple	5	5
quadruple	-	-

BIOMONOZ Monozygotic or dizygotic siblings?

		Frequency	Percent	Val. Percent	Cum. Percent
Valid	0 Possible Monozygotic	120	44,1	44,1	44,1
	1 Definite Monozygotic	23	8,5	8,5	52,6
	2 Definite dizygotic	129	47,4	47,4	100,0
	Complete	272	100,0	100,0	

INFOTWIN Status twin - Source of information?

		Frequency	Percent	Val. Percent	Cum. Percent
Valid	1 Generated from identical household and mont of birth – not validated	7	2,6	2,6	2,6
	2 Possible Twin or Triplet – Information in survey 2006 not revisable	36	13,2	13,2	15,8
	3 Possible Twin or Triplet – In survey 2006 answer refused	1	,4	,4	16,2
	4 Twin or Triplet – Information in survey 2006 validated	97	35,7	35,7	51,8
	5 Twin or Triplet – New since 2007 (month & year)	19	7,0	7,0	58,8
	6 Twin or Triplet – New since 2007 (year & missing month)	112	41,2	41,2	100,0
	Complete	272	100,0	100,0	

9 BIOSIB (beta version): Information on siblings in the SOEP

by Daniel D. Schnitzlein

9.1 General description of the data set

BIOSIB provides information on siblings living within the SOEP households. The data set contains the person numbers of all siblings in an observed family. It includes information on their sex, their year of birth, the number of siblings, the individual's position within the birth order, and on the relationship between the observed siblings.

In the recent wave BIOSIB is included as a beta version. Any positive or negative feedback as well as suggestions are highly welcome. Please address these to dschnitzlein@diw.de.

9.2 Sources of information on siblings in the SOEP

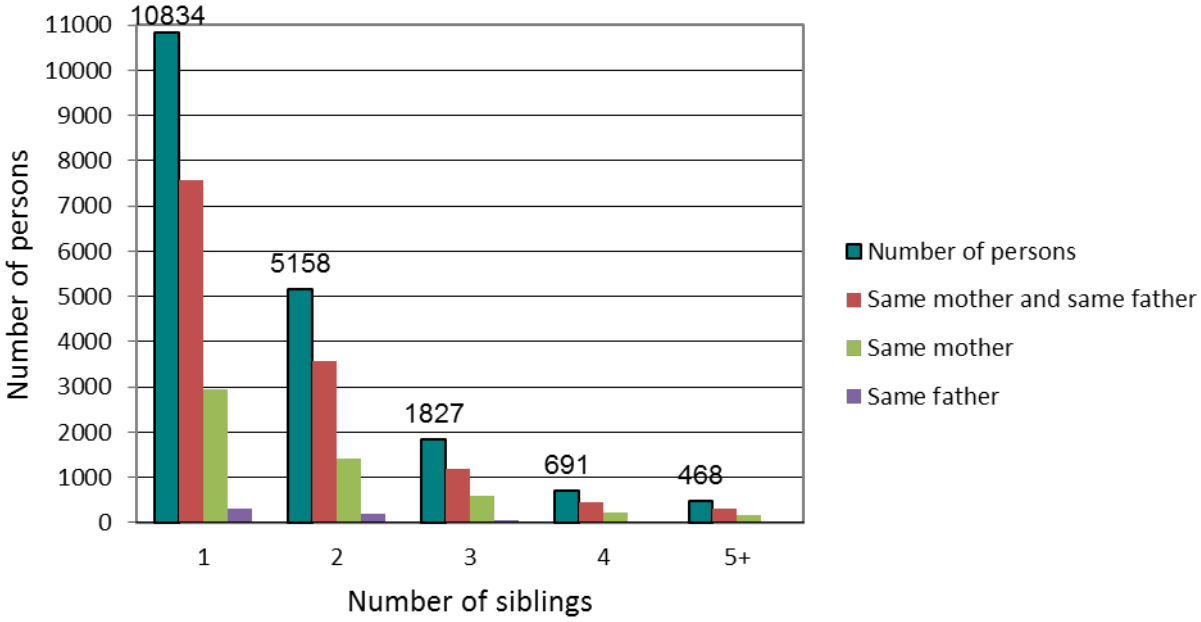
Information on siblings in the SOEP is available from three sources:

- First, the respondents are asked about their siblings in the biography questionnaire and the youth questionnaire. This information (for example *have or ever had siblings yes/no, number of sisters, number of brothers*) is stored in the file BIOPAREN (for detailed information please see the chapter on BIOPAREN).
- Second, in the years 1991, 1996, 2001, 2006 and 2011 all respondents were asked about their family relations. Among other questions on their family, the individuals were asked about their siblings (like above, *have or ever had siblings yes/no, number of sisters, number of brothers*) outside the SOEP household in 1991, 1996 and 2001. From 2006 on they were asked about all siblings within and outside the household. The information from these questions is stored in the $$p-files.
- Third, siblings within the SOEP households are observed directly. The aim of this file is to provide the never changing person ID of the siblings of each respondent as far as they can be identified in the SOEP. With this information it is possible to use the whole range of personal or household information for each sibling to carry out detailed sibling analyses. This file therefore adds to the information in BIOTWIN (for details see chapter on BIOTWIN), which provides the person IDs for all twins in the SOEP.

9.3 Overview on the number of siblings in BIOSIB

Figure 1 gives an overview on the information stored in BIOSIB. The data set contains 7,817 families with 18,978 individuals that have at least one sibling identified in the SOEP. 10,834 of them have one sibling, 5,158 have two siblings, 1,827 have three siblings, 691 have four siblings and 468 have five or more siblings identified in the data. As is apparent from Figure 1 in all cases most of the siblings are identified by having the same mother and the same father (for more details about the identification of siblings and the identification of biological siblings please see the description to SIBDEF1-SIBDEF11 on the subsequent pages).

Figure 1: Number of persons with information on their siblings, by number of siblings



Source: own calculation, based on SOEPv28.

9.4 Organization of the data in BIOSIB

Each row in the dataset represents one individual for which at least one sibling could be identified. Therefore a family with three siblings appears three times in BIOSIB, one time for each child. The person IDs of the siblings are ordered by birth order starting with the oldest sibling.

List of variables

HHNR	Original Household Number
PERSNR	Never Changing Person ID
SIBPNR1 – SIBPNR11	Person Number of 1st – 11th Sibling
SIBDEF1 – SIBDEF11	Sibling Relation to 1st – 11th Sibling
FAMCOUNT	Family Counter
POS_SIB	Position in the birth order
NUM_SIB	Number of observed siblings in the SOEP
SEX	Gender of Individual
GEBJAHR	Year of Birth of Individual
SEXSIB1 – SEXSIB11	Gender of 1st – 11th Sibling
GEBSIB1 – GEBSIB11	Year of Birth of 1st – 11th Sibling

The variables HHNR, PERSNR, SEX, GEBJAHR, SEXSIB1-SEXSIB11 and GEBSIB1-GEBSIB11 are generated from the information stored in PPFAD. The newly generated variables SIBPNR1-SIBPNR11, SIBDEF1-SIBDEF11, FAMCOUNT, POS_SIB, and NUM_SIB are described on the next pages.

Variable **SIBPNR1 – SIBPNR11**

Label: Person Number of 1st-11th Sibling

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

Description: The variables provide the never changing person IDs for the siblings of the individual identified by PERSNR. The sibling relationship is generated from the parent information in BIOBIRTH, BIOBRTHM and BIOPAREN (for detailed information on these files please see the relevant chapters above). Two persons are defined as siblings if they report both, the same mother and father, only the same mother, or only the same father. This information on the sibling relationship is stored in SIBDEF1-SIBDEF11.

In the case of inconsistent information on parents in BIOBIRTH, BIOBRTHM and BIOPAREN, BIOPAREN was assigned the lowest priority.

Please note, that BIOPAREN uses a social definition of parenthood based on cohabitation. In contrast, BIOSIB contains both biological (BIOBIRTH/BIOBRTHM) and social siblings with a higher priority on biological relations.

Variable **SIBDEF1 – SIBDEF11**

Label: Sibling Relation to 1st-11th Sibling

Values:
(-1) No answer
(-2) Does not apply
(-3) Answer improbable
(1) Same mother (B); same father (B)
(2) Same mother (B); same father (nB)
(3) Same mother (nB); same father (B)
(4) Same mother (nB); same father (nB)
(5) Same mother (B)
(6) Same mother (nB)
(7) Same father (B)
(8) Same father (nB)

Description: The variables provide the information on the sibling relationship between the individual identified by PERSNR and the respective sibling. Two siblings can have either, the same mother and the same father, only the same mother, or only the same father. The indicator further provides information if the identified parent is a biological (B) (indicator from BIOBIRTH/BIOBRTHM) or non-biological parent (nB) (indicator from BIOPAREN). So for example, variable value (1) indicates that the two individuals share the same biological mother and the same biological father.

Variable	**FAMCOUNT**
Label:	Family Counter
Values:	(-1) No answer (-2) Does not apply (-3) Answer improbable
Description:	The variable contains a non-systematic counter of families occurring in BIOSIB. All siblings (biological and non-biological), who belong to one family, are assigned the same value of FAMCOUNT. The variable can be used, for example, in multilevel analyses to define the family level. Note: In the case a family splits up, children from the new partnerships of the parents are no siblings. Children of the early partnerships are siblings to all children in the new partnership.

Variable	**POS_SIB**
Label:	Position in birth order
Values:	(-1) No answer (-2) Does not apply (-3) Answer improbable
Description:	The variable contains the individual's position in the birth order of the observed siblings (biological and non-biological).

Variable	**NUM_SIB**
Label:	Number of siblings observed in SOEP
Values:	(-1) No answer (-2) Does not apply (-3) Answer improbable
Description:	The variable contains the total number of identified siblings, including the respondent, in the SOEP (biological and non-biological).

BIOMONOZ Monozygotic or dizygotic siblings?

		Frequency	Percent	Val. Percent	Cum. Percent
Valid	0 Possible Monozygotic	108	43,0	43,0	43,0
	1 Definite Monozygotic	23	9,2	9,2	52,2
	2 Definite dizygotic	120	47,8	47,8	100,0
	Complete	147	100,0	100,0	

INFOTWIN Status twin - Source of information?

		Frequency	Percent	Val. Percent	Cum. Percent
Valid	1 Generated from identical household and mont of birth – not validated	7	2,8	2,8	2,8
	2 Possible Twin or Triplet – Information in survey 2006 not revisable	36	14,3	14,3	17,1
	3 Possible Twin or Triplet – In survey 2006 answer refused	1	,4	,4	17,5
	4 Twin or Triplet – Information in survey 2006 validated	97	38,6	38,6	56,2
	5 Twin or Triplet – New since 2007 (month & year)	10	4,0	4,0	60,2
	6 Twin or Triplet – New since 2007 (year & missing month)	100	39,8	39,8	100,0
	Complete	147	100,0	100,0	

10 BIOAGE01, BIOAGE03, BIOAGE06, BIOAGE08a, BIOAGE08b: Generated variables from the "Mother & Child" and "Parent" questionnaires

By David Richter and Anne Fromm

This documentation is based on earlier versions of documentation materials on bioage data sets and has benefited from previous work by Sebastian Frischholz, Stefanie Lenuweit, Katharina Mahne, Christian Schmitt, and Jürgen Schupp.

10.1 Introduction

The *bioage* data files are generated using information collected in the "Mother & Child" and "Parent" questionnaires. There are five different "Mother & Child" and "Parent" questionnaires (see Table 1) and accordingly five different *bioage* data files with the data file-specific prefix.

Table 1. Overview of bioage files, corresponding age group, and respondents

Bioage File	Classification by Age	Respondents	First Wave
Bioage01	0-1 years old	Mothers only	2003
Bioage03	2-3 years old	Mothers only	2005
Bioage06	5-6 years old	Mothers only	2008
Bioage08a	7-8 years old	Parent 1	2010
Bioage08b	7-8 years old	Parent 2	2010

The "Mother & Child" and "Parent" questionnaires aim to follow and observe future generations of the SOEP and collect all information in age-specific files, even though the data come from different survey years. As we try to make this process as comprehensive and gap-free as possible, we begin following and documenting the development of children in SOEP households from birth onwards. Since many questions overlap, all *bioage* data files are covered in this chapter, rather than dedicating a separate chapter of this document to each *bioage* data file. By doing so, we are able to provide you with an overview of all variables covered in the *bioage* data files.

10.2 Data Files and Respondents

The file-specific suffix of each *bioage* data file reflects the age of the children in it (e.g., *bioage06* includes children who turned six that survey year, producing a range in ages between 5 years and 1 month and 6 years and 11 months, depending on the birth month and

interview month). For information on the questionnaires used with each age group, please see Table 1. Except for *bioage08a* and *bioage08b*, it is usually the mother who completes the questionnaire. In the exceptional case that a mother cannot complete the questionnaire, the father does. Both the father's and the mother's personal ID numbers have the suffix PERSNRM (from 2011 on: PERSNRE). Note that while the parents are the actual respondents, the data are organized under the child's unchanging personal ID number (PERSNR).

Bioage01: "Mother and Child" Questionnaire, Children Aged 0-1 years

The questionnaire is given to all women who gave birth to a child in the current or previous survey year, and to all women whose non-biological child was born in the same period. The questionnaire contains information on pregnancy, childbirth, childcare, and child health. In the exceptional case that a mother is unable to complete the questionnaire, the father responds. As shown in Table 2, *bioage01* includes one male respondent.

Bioage03: "Your child between the ages of two and three", Children Aged 2-3 years

The questionnaire is given to all mothers whose child turns three in the current survey year. In the exceptional case that a mother cannot complete the questionnaire, the father responds. As shown in Table 2, *bioage03* includes six male respondents.

Bioage06: "Your child between the ages of five and six", Children Aged 5-6 years

The questionnaire is given to all mothers whose child turns six in the current survey year. In the exceptional case that a mother cannot complete the questionnaire, the father responds. As shown in Table 2, *bioage06* includes eight male respondents.

Table 2. Fathers in the bioage data sets

Bioage File	Personal ID number of child PERSNR	Personal ID number of parent PERSNRM / PERSNRE
Bioage01	7225405	7225401
Bioage03	1139503	80204
	1159104	1159102
	3014505	3014501
	3190908	3190901
	8743404	8743401
	8782904	8782901
Bioage06	1159103	1159102
	2524101	2524104
	2766805	2766801
	2976003	2976001
	3014505	3014501
	3190907	3190901
	7018003	1266204
	7225405	7225401

Bioage08a and Bioage08b: "Parent" Questionnaire, children aged 7-8 years

The questionnaire is given to *both* mothers and fathers of children turning eight in the current survey year. Data of parent 1, which is usually the mother, can be found in *bioage08a*. Data of parent 2, which is usually the father, can be found in *bioage08b*.

Number of Children and Twins in the *Bioage* Data Sets

The data set has grown over the years and now contains data on 2,269 children. At the moment (2012), there are 203 children for whom information has been provided through all of the "Mother & Child" and "Parent" questionnaires (*bioage01*, *bioage03*, *bioage06*, and *bioage08a/b*). Data are available for 105 children from four questionnaires, for 434 children from three questionnaires, and for 741 children from two questionnaires. For 786 children, data are available from at least one questionnaire. An overview of the number of children in each data set is given in Table 3.

Table 3. Number of respondents to the "Mother & Child" and "Parent" questionnaires

Bioage File	Survey Year									Total
	2003	2004	2005	2006	2007	2008	2009	2010	2011	
Bioage01	318	247	246	234	205	185	196	182	146	1,959
Bioage03			257	222	237	246	186	190	173	1,511
Bioage06						237	210	214	210	871
Bioage08a								221	174	395
Bioage08b								152	117	269
Total	318	247	503	456	442	668	592	959	820	5,005

Table 4 gives an overview of the number of children per family in the specific data sets. Note that there might be more children in the household than indicated here: these children are not in the age range of the "Mother & Child" and "Parent" questionnaires and therefore not covered by the *bioage* data sets.

Table 4. Number of children in the family in the specific data sets

Bioage File	Number of Children in the Family					Total
	1 Child	2 Children	3 Children	4 Children	5 Children	
Bioage01	1,091	347	46	8	1	1,493
Bioage03	944	240	25	3		1,212
Bioage06	679	90	4			773
Bioage08a	369	13				382
Bioage08b	251	9				260

The data also includes twins. *Bioage01* contains 32 pairs, *bioage03* 19 pairs, *bioage06* 10 pairs, bioage08a 4 pairs, and *bioage08b* 3 pairs of twins.

10.3 Topics and Variables

The *bioage* data sets contain information regarding

- Pregnancy and childbirth
- Child health
- Childcare situation
- Changes in living circumstances since birth of child
- Child's abilities
- Parenting experiences
- Expectations about the child's success in school
- Educational goals and aspirations of the parents

- Educational behavior of the parents
- Self-conception of the role of parents

For a detailed overview of topics and variables, see Table 5.

The rules used to generate the variables from the questionnaires are consistent over the various *bioage* data files. For the most part, the variables in the *bioage* files have the same names as in the respective "Mother & Child" or "Parent" questionnaire. This ensures consistency over the years, even if different questionnaires (in terms of question order or content) are used. The consistency in variable names also makes it possible to compare different age groups as most variable names are identical (up to their prefix) if the contents are the same (see Table 5). Note that for survey-specific reasons, variable names change occasionally. This is not the case in the new integrated *bioage long* data set. Here, the data is presented in "long" format, i.e. this dataset will contain information from *bioage01*, *bioage03*, *bioage06*, as well as *bioage08a* and *bioage08b*. You will therefore find the variables here without their wave-specific prefixes.

A few variables are generated by combining the information from two or more variables. The present documentation provides detailed information on variables generated using information from other files. An overview of the specific variables and the rules by which they were

Table 5 Variables and Variable names in the individual files

Label	Bioage 01	Bioage 03	Bioage 06	Bioage 08a	Bioage08b	Bioage long
Original HH number	HHNR					
Current wave HH Number	HHNRAKT					
Unchanging Personal ID of child	PERSNR					
Unchanging Personal ID of parent	PERSNRM/from 2011 on PERSNRE					pidresp
Survey year	SVYYEAR					
General Information						
Child's month of birth	bckgebmo	BKGEBMO	BCPGEBMO	EBAGEBMO	EBBGEBMO	birthm
Child's year of birth	bckgebja	BKGEBJA	BCPGEBJA	EBAGEBJA	EBBGEBJA	birthy
Age (in months) of the child at the time of the survey	bckalter	BKALTER	BCPALTER	EBAALTER	EBBALTER	age
Child's gender	bcksex	BKKSEX	BCPSEX	SEX_C	SEX_C	sex
Parent's gender				SEX_E	SEX_E	sexresp
Newborn is 1st, 2nd, 3rd, etc. child	bckizahl					nchild
Biological child	bckleibl			EBABIOEL	EBBBIOEL	biochild
Child's Health						
Weight of child at birth in grams	bckgew					weightb
Height of child in cm	bckgroe	BKGROE	BCPGROE			height

Label	Bioage 01	Bioage 03	Bioage 06	Bioage 08a	Bioage08b	Bioage long
Child's Weight in kg		BKGEW	BCPGEW			weight
Head circumference of the child in cm	bckkopf					circum
Last medical examination	bckletzu					lstmedex
Medical help - number of times in the first 3 months after birth	bckarzt					medaid3mb
Medical help - number of times in the last 3 months		BKARZT	BCPARZT			medaid3m
Length of hospital stay in the first 3 months after birth in days	bckkrhau					hospital3mb
Length of hospital stay in the last 12 months in days		BKKRHAU	BCPKRHAU			hospital12m
Concerns about the child's health	bckgesu1	BKGESU7				health
Child has confirmed disorders	bckstoer					disord
Symptoms of disorders	BCKSTOE1					disordno
Child is restricted in abilities			BCPSTOE0			ill0
Respiratory disease			BCPSTOE1			ill10
Asthma		BKSTOE1				ill11
Chronic bronchitis		BKSTOE2				ill12
Spastic / acute bronchitis		BKSTOE3				ill13
Pseudocroup / croup syndrome		BKSTOE4				ill14
Middle-ear inflammation		BKSTOE5	BCPSTOE2			ill2
Hayfever		BKSTOE6				ill3

Label	Bioage 01	Bioage 03	Bioage 06	Bioage 08a	Bioage08b	Bioage long
Neurodermatitis		BKSTOE7	BCPSTOE3			ill4
Vision impairment		BKSTOE8	BCPSTOE4			ill5
Hearing impairment		BKSTOE9				ill6
Nutritional disorders		BKSTOE10	BCPSTOE5			ill7
Motor impairment		BKSTOE11	BCPSTOE6			ill8
Other impairments or disorders		BKSTOE12	BCPSTOE7			ill9
No disease / disorder		BKSTOER	BCPSTOER			illno
Pregnancy and Childbirth						
Place where birth took place	bcentbin					delivpl
Pregnancy week of birth	bcssw					birthpw
Pregnancy planned/unplanned	bcssplan					pregplan
Physical condition in the final third of the pregnancy	bcbefin1					feeling1
Physical condition in the first 3 months after birth	bcbefin2					feeling2
Mental state in the final third of the pregnancy	bcbefin3					feeling3
Mental state in the first 3 months after birth	bcbefin4					feeling4
Mother: pregnant at individual interview	BCPREGY					pregy
Mother: month of pregnancy at individual interview	BCPREGMO					pregmo
Spell beginning month of pregnancy	BCPREBEG					prebeg

Label	Bioage 01	Bioage 03	Bioage 06	Bioage 08a	Bioage08b	Bioage long
Spell end of pregnancy, birth	BCPREEND					preend
Delivery by caesarean section	BCKAISER					delivcs
Breast-feeding baby	BCSTILL	BKSTILL				breastf
Breast-feeding time in months	BCSTILLM	BKSTILLM				breastfm
Change in Living Circumstances since Childbirth						
Circumstances in life have greatly changed	bcverae1					change1
Child provides happiness and joy	bcverae2					change2
Often close to running out of strength	bcverae3					change3
Very satisfied with the role of mother	bcverae4					change4
Often unable to cope with tasks/responsibilities	bcverae5					change5
Have made new contacts through the child	bcverae6					change6
Suffering from being limited to the role of mother	bcverae7					change7
Important to provide the child with much affection	bcverae8					change8
Childcare Situation[18]						
Father lives in household	bcvater					fathinhh
Supported by partner	bcuntpa					suppartn

[18] For the variables on the child care situation, different missing codes apply. Here, the code "-1" indicates that the respondent reported that the child is cared for by the respective person or institution but gave no information on the number of hours per week.

Label	Bioage 01	Bioage 03	Bioage 06	Bioage 08a	Bioage08b	Bioage long
Mother is main caregiver	bchauptb			EBAHBETR	EBBHBETR	maincare
Cared for by partner (in h/week)	bcbetre1	BKBETRE1	BCPBETR1	EBABE1	EBBBE1	care1h
Cared for by child's father (if not resident of same household) (in h/week)		BKBETRE2	BCPBETR2	EBABE2	EBBBE2	care2h
Cared for by grandparents (in h/week)	bcbetre2	BKBETRE3	BCPBETR3	EBABE3	EBBBE3	care3h
Cared for by older siblings (in h/week)	bcbetre3	BKBETRE4	BCPBETR4	EBABE4	EBBBE4	care4h
Cared for by other relatives (in h/week)	bcbetre4	BKBETRE5	BCPBETR5	EBABE5	EBBBE5	care5h
Cared for in family day care (in h/week)	bcbetre5	BKBETRE6	BCPBETR6			care6h
Cared for by babysitter (in h/week)		BKBETRE7	BCPBETR7	EBABE6	EBBBE6	care7h
Cared for in crèche/day nursery (in h/week)	bcbetre6	BKBETRE8	BCPBETR8			care8h
Cared for by school (in h/week)				EBABE8	EBBBE8	care9h
Cared for in after-school child care (in h/week)				EBABE9	EBBBE9	care10h
Cared for by social institution, center (in h/week)				EBABE10	EBBBE10	care11h
Cared for by others (in h/week)	bcbetre7	BKBETRE9	BCPBETR9	EBABE7	EBBBE7	care12h
No use of temporary care/ cared for solely by respondent	bcbetre8	BKBETRE0	BCPBETR0	EBABE11	EBBBE11	care13
Activities done with the Child						
Singing children songs to/with child (no. of times during last 14 days)		BKAKT1	BCPAKT05			activ1
Taking walks outdoors (no. of times during last 14 days)		BKAKT2	BCPAKT02			activ2
Painting or doing arts and crafts (no. of times during last 14 days)		BKAKT3	BCPAKT06			activ3

Label	Bioage 01	Bioage 03	Bioage 06	Bioage 08a	Bioage08b	Bioage long
Reading or telling stories (in German) (no. of times during last 14 days)		BKAKT4	BCPAKT11			activ4
Looking at picture books (no. of times during last 14 days)		BKAKT5				activ5
Going to playground (no. of times during last 14 days)		BKAKT6	BCPAKT01			activ6
Visiting other families with children (no. of times during last 14 days)		BKAKT7	BCPAKT03			activ7
Going shopping with child (no. of times during last 14 days)		BKAKT8	BCPAKT04			activ8
Watching television or videos with child (no. of times during last 14 days)		BKAKT9	BCPAKT08			activ9
Frequency of playing cards or dice in the last 14 days			BCPAKT07			activ10
Frequency of computer/online gaming in the last 14 days			BCPAKT09			activ11
Frequency of trips to children's theater, circus, museum, etc. in the last 14 days			BCPAKT10			activ12
Frequency of reading or story-telling in another language in the last 14 days			BCPAKT12			activ13
Child allowed to watch television or videos alone		BKTV	BCPTV			tvyn
Child watches video/tv alone (in h/week)		BKTVH	BCPTVH			tvhrs
Language spoken with child		BKDEUT				language
Child's Temper						
Child is generally happy and satisfied	BCKGESU2	BKGESU1				temp1
Child is easily irritated and often cries	BCKGESU3	BKGESU2				temp2
Child is hard to console	BCKGESU4	BKGESU3				temp3
Child is curious and active	BCKGESU5	BKGESU4				temp4

Label	Bioage 01	Bioage 03	Bioage 06	Bioage 08a	Bioage08b	Bioage long
Child more reserved	BCKGESU6					temp5
Child communicative and talkative		BKGESU5				temp6
Child shows empathy when others are sad		BKGESU6				temp7
Child's Personality						
Child tends to be shy / outgoing		BKEIG1				char1a
Child tends to be talkative / quiet			BCPEIG01			char1b
Child tends to be focused / easily distracted		BKEIG2	BCPEIG07			char2
Child tends to be obstinate / obedient		BKEIG3	BCPEIG08			char3
Child tends to be quick to learn new things / to need more time		BKEIG4	BCPEIG09			char4
Child tends to be messy / neat			BCPEIG02			char5
Child tends to be sweet-tempered / easily excited			BCPEIG03			char6
Child tends to be disinterested / curious			BCPEIG04			char7
Child tends to be self-confident / tends to lack confidence			BCPEIG05			char8
Child tends to be withdrawn / outgoing			BCPEIG06			char9
Child tends to be anxious / confident			BCPEIG10			char10
Child's Strengths and Difficulties (SDQ)						
Child is considerate			BCPVER01			behav1
Child is restless, hyperactive, can't sit still			BCPVER02			behav2

Label	Bioage 01	Bioage 03	Bioage 06	Bioage 08a	Bioage08b	Bioage long
Child shares with other children			BCPVER03			behav3
Child often has fits of anger, quick-tempered			BCPVER04			behav4
Child is a loner, usually plays alone			BCPVER05			behav5
Child is helpful when others are hurt, sick, sad			BCPVER06			behav6
Child is fidgety			BCPVER07			behav7
Child argues with or bullies others			BCPVER08			behav8
Child is often unhappy, cries easily			BCPVER09			behav9
Child is popular with peers			BCPVER10			behav10
Child is easily distracted, lacks concentration			BCPVER11			behav11
Child is nervous or clingy in new situations			BCPVER12			behav12
Child gets teased or bullied by peers			BCPVER13			behav13
Child frequently offers to help others			BCPVER14			behav14
Child gets along with adults better than with other children			BCPVER15			behav15
Child has a lot of fears, gets scared easily			BCPVER16			behav16
Child completes tasks, can concentrate for an extended period			BCPVER17			behav17
Child's Adaptive Behavior (Vineland Scale)						
Child understands simple instructions		BKSPR1				spch1
Child forms sentences with at least two words		BKSPR2				spch2

Label	Bioage 01	Bioage 03	Bioage 06	Bioage 08a	Bioage08b	Bioage long
Child speaks in full sentences (at least four words)		BKSPR3				spch3
Child listens attentively to a story for at least five minutes		BKSPR4				spch4
Child can relate simple messages		BKSPR5				spch5
Child eats with spoon without making a mess		BKALLT1				skll1
Child blows nose without assistance		BKALLT2				skll2
Child uses toilet to do number two		BKALLT3				skll3
Child puts on pants and underpants forwards		BKALLT4				skll4
Child brushes teeth without assistance		BKALLT5				skll5
Child walks forwards down the stairs		BKBEW1				mvmn1
Child uses door handle to open doors		BKBEW2				mvmn2
Child climbs jungle gyms and other high playground equipment		BKBEW3				mvmn3
Child uses scissors to cut paper		BKBEW4				mvmn4
Child paints / draws recognizable shapes on paper		BKBEW5				mvmn5
Child calls familiar people by name		BKSOZ1				sclr1
Child plays games with other children		BKSOZ2				sclr2
Child participates in role-playing games		BKSOZ3				sclr3
Child shows particular liking for certain playmates or friends		BKSOZ4				sclr4
Child calls his/her own feelings by name, e.g., sad, happy, scared		BKSOZ5				sclr5

Label	Bioage 01	Bioage 03	Bioage 06	Bioage 08a	Bioage08b	Bioage long
School Enrolment						
First month of school enrolment				EBASCHMO	EBBSCHMO	sclenrolm
First year of school enrolment				EBASCHJA	EBBSCHJA	sclenroly
Child hasn't been enrolled in school yet				EBAKSCH	EBBKSCH	sclenroln
Parental Ambitions for Child's Education						
Ideal school completion: low (Hauptschule, lower secondary school)				EBAIDHAU	EBBIDHAU	idegrad1
Ideal school completion: medium (Realschule, intermediate secondary school)				EBAIDREA	EBBIDREA	idegrad2
Ideal school completion: high (Gymnasium, upper secondary school)				EBAIDGYM	EBBIDGYM	idegrad3
Probable school completion: lower secondary school				EBAWAHAU	EBBWAHAU	probgra1
Probable school completion: intermediate secondary school				EBAWAREA	EBBWAREA	probgra2
Probable school completion: upper secondary school				EBAWAGYM	EBBWAGYM	probgra3
Parenting Goals						
Education goal. Child should: be a good student				EBAEZI1	EBBEZI1	edgoal1
Education goal. Child should: get along with other kids				EBAEZI2	EBBEZI2	edgoal2
Education goal. Child should: be interested why things happen				EBAEZI3	EBBEZI3	edgoal3
Education goal. Child should: act like normal girl/boy				EBAEZI4	EBBEZI4	edgoal4
Education goal. Child should: be honest				EBAEZI5	EBBEZI5	edgoal5
Education goal. Child should: have good manners				EBAEZI6	EBBEZI6	edgoal6

Label	Bioage 01	Bioage 03	Bioage 06	Bioage 08a	Bioage08b	Bioage long
Education goal. Child should: have good self-control				EBAEZI7	EBBEZI7	edgoal7
Education goal. Child should: be responsible				EBAEZI8	EBBEZI8	edgoal8
Education goal. Child should: treat others with respect				EBAEZI9	EBBEZI9	edgoal9
Education goal. Child should: obey parents				EBAEZI10	EBBEZI10	edgoal10
Education goal. Child should: have good ability to judge				EBAEZI11	EBBEZI11	edgoal11
Education goal. Child should: be clean and neat				EBAEZI12	EBBEZI12	edgoal12
Education goal. Child should: strive to achieve goals				EBAEZI13	EBBEZI13	edgoal13
Education goal. Child should: fit in well in groups				EBAEZI14	EBBEZI14	edgoal14
Education goal. Child should: learn to overcome obstacles				EBAEZI15	EBBEZI15	edgoal15
Education goal. Child should: be satisfied with self				EBAEZI16	EBBEZI16	edgoal16
Education goal. Child should: learn to avoid risks				EBAEZI17	EBBEZI17	edgoal17
Education goal. Child should: be lovable				EBAEZI18	EBBEZI18	edgoal18
Parenting Style						
Education behavior: ask what he/she has experienced				EBAMET3	EBBMET3	edbeh3
Education behavior: punish when disobedient				EBAMET4	EBBMET4	edbeh4
Education behavior: threaten to punish, but do not punish				EBAMET5	EBBMET5	edbeh5
Education behavior: know where child is				EBAMET6	EBBMET6	edbeh6
Education behavior: rather strict with child				EBAMET7	EBBMET7	edbeh7

Label	Bioage 01	Bioage 03	Bioage 06	Bioage 08a	Bioage08b	Bioage long
Education behavior: comfort child when child is sad				EBAMET8	EBBMET8	edbeh8
Education behavior: shout when child makes mistakes				EBAMET9	EBBMET9	edbeh9
Education behavior: a disobedient child is ungrateful				EBAMET10	EBBMET10	edbeh10
Education behavior: do not talk to child when disobedient				EBAMET11	EBBMET11	edbeh11
Education behavior: tell child not to disobey				EBAMET12	EBBMET12	edbeh12
Education behavior: praise my child				EBAMET13	EBBMET13	edbeh13
Education behavior: scold if angry				EBAMET14	EBBMET14	edbeh14
Education behavior: try to influence child's friendships				EBAMET15	EBBMET15	edbeh15
Education behavior: reduce or retract punishment				EBAMET16	EBBMET16	edbeh16
Education behavior: disappointed about bad behavior				EBAMET17	EBBMET17	edbeh17
Education behavior: hard to be consistent in parenting				EBAMET18	EBBMET18	edbeh18
Parenting Role						
Be parents: sacrifice own wishes				EBARO1	EBBRO1	bepar1
Be parents: disobedience is aimed to bother me				EBARO2	EBBRO2	bepar2
Be parents: child is my main purpose in life				EBARO3	EBBRO3	bepar3
Be parents: wish I was less overwhelmed by parental responsibilities				EBARO4	EBBRO4	bepar4
Be parents: problems in upbringing are child's fault				EBARO5	EBBRO5	bepar5
Be parents: nothing is better than being with child				EBARO6	EBBRO6	bepar6

Label	Bioage 01	Bioage 03	Bioage 06	Bioage 08a	Bioage08b	Bioage long
Be parents: would put up with everything for child				EBARO7	EBBRO7	bepar7
Be parents: child misbehavior is intentional				EBARO8	EBBRO8	bepar8
Be parents: often put everything aside to support child				EBARO9	EBBRO9	bepar9
Be parents: look forward to spending time with child				EBARO10	EBBRO10	bepar10

10.4 Generated Variables

This section provides additional information on variables that have been generated from combinations of variables from other datasets than the parent questionnaires.

BCKALTER, BKALTER, BCPALTER, EBAALTER, EBBALTER (*bioage long*: age)

Variable label	**"Child's age in months"**
Variable format	2-digit integer
Comment:	This variable provides the child's age in months as a combination of month of birth and interview month. As the exact day of birth remains unknown, information is only an approximation and may vary by one month. Note that the information concerning year and month of birth from the "Mother & Child" and "Parent" questionnaires proved to be partially inconsistent. Therefore, as of the beginning of 2012 the child's age in months is computed using birth information from *ppfad*. For further information, refer to the documentation on *ppfad*.

BCPREGY (*bioage long*: pregy)

Variable label	**"Mother: pregnant at interview in survey year X"**
Variable format	4-digit integer
Bioage File	01
Comment	This variable is based on information from the previous year's individual questionnaire provided by the mother on her pregnancy status at the time of the interview. If pregnancy was reported (or was unknown) and a child was born, the year in which the interview took place is contained in BCPREGY. Hence this information is available only for those women in the sample for at least two years with a completed individual interview in the first year and a completed "Mother & Child" questionnaire in the second year. Please note that some mothers are not aware that they are pregnant in the early stages of pregnancy. The time of observation starts in survey year 2003 with the 2002 birth cohort.

BCPREGMO (*bioage long*: pregmo)

Variable label	**"Mother: pregnancy month at interview"**
Variable format	2-digit integer
Bioage File	01
Comment	This variable is based on the exact month of birth (BCPREGMO), the duration of childbearing in weeks (BCSSW) and the interview month of the previous year's personal interview. Hence this information is available only for those women in the sample for at least two years. As the exact day of birth is unknown, this variable remains a close approximation.

BCPREBEG (*bioage long*: prebeg)

Variable label	**"Spell Begin Pregnancy (Month, 01.83=1)"**
Variable format	3-digit integer
Bioage File	01
Comment	The variable BCPREBEG contains information on the beginning of pregnancy (i.e., the month of conception). Information is given in the regular SOEP spell format: values start with 1 for January 1983 (e.g., the earliest spell in *bioage01*, survey year 2010, is 304, which equals April 2008). The variable is based on the exact month of birth (BCPREGMO) and the duration of pregnancy in weeks (BCSSW). Accordingly, information is available only for women who completed the "Mother & Child Questionnaire" and for whom the duration of the pregnancy is known. Note that the month of conception may vary by one month as the exact date of birth remains unknown.

BCPREEND (*bioage long*: pregend)

Variable label	**"Spell End Pregnancy, Birth (Month, 01.83=1)"**
Variable format	3-digit integer
Bioage File	01
Comment	The variable BCPREEND contains information on the end of pregnancy (i.e. the month of birth). Information is given in the regular SOEP spell format: values start with 1 for January 1983 (e.g., earliest spell in *bioage01*, survey year 2010, is 304, which equals April 2008). This variable is based on the exact month of birth (BCPREGMO) and the duration of pregnancy in weeks (BCSSW).

SEX, SEX_C (*bioage long*: sex)

Variable label	**"Gender of child"**
Variable format	1-digit integer
Comment	The sex of the child is not asked for in the "Mother & Child" and "Parent" questionnaires. Information on this variable stems from the *ppfad*.

SEX_E (*bioage long*: sexresp)

Variable label	**"Sex of respondent (parent)"**
Variable format	1-digit integer
Bioage File	08a, 08b
Comment	This variable tells whether the mother or father answered the respective questionnaire. The information for this variable comes from the *ppfad* file.

11 BIOAGE17: The Youth Questionnaire[19]

by Marco Giesselmann *and* Mila Staneva

(Replaces earlier versions by Jürgen Schupp *and* Michael Frühling / Thorsten Schneider *and* Bettina Isengard*)*

A special group of first time respondents are young persons living in a panel household, who reach the surveying age of 17 years. From this specific group of panel entrants, we are able to obtain some more detailed information on youth and socialisation than from other new sample members. At the same time, certain life-course dimensions (as the partnership- or employment biography) have not yet developed in 17 year-olds. With regard to these specifics, the standard biography questionnaire is not appropriate to this group. Thus, we use an independent questionnaire for this special group of first time respondents: the Youth Questionnaire. This instrument is used since the year 2000 and can be understood as an alternative version of the Biography Questionnaire, collecting more comprehensive information on relationships with parents, leisure-time activities, and past achievements in school, as well as on personality characteristics. In addition, there are numerous prospective questions about educational plans and plans for further training, as well as questions about expectations for future career and family.

A number of statements regarding specific circumstances—including the expectations for the future mentioned above—are directly related to the time at which the questionnaire was completed. However, they provide a multifaceted background for long-term analyses since these young people will continue to be interviewed in subsequent years like other SOEP respondents. The Youth Questionnaire also contains retrospective questions, for example, at what age the teenager started his or her first job or first music lessons, what recommendations he or she received regarding choice of secondary school level, and which grades he or she repeated.

11.1 Genesis and Target Population of the Youth Questionnaire

The Youth Questionnaire is aimed at youths who have reached the surveying age of 17 years[20] and are therefore being interviewed for the first time. This questionnaire takes the place of the supplementary Biography Questionnaire, since the latter does not apply to the young people's family or career situations. As a rule, information on social origin can be obtained from the parents' Individual Questionnaire, in case the youth lives together with the respective parent. If the teenager does not live with either parent, the Youth Questionnaire

[19] In earlier SOEP-data releases BIOAGE17 was called BIOYOUTH.
[20] More precisely, this refers to youths who live in an already existing panel household and are or will turn 17 years old in the year of the survey. They are therefore 16 or 17 years old at the time of the interview.

collects information on the missing parent(s). Young people who immigrated to Germany are also given the standard questions on immigration from the supplementary Biography Questionnaire. This guarantees that all important information collected in the Biography Questionnaire is also available on these young people.

A preliminary version of the Youth Questionnaire was tested in 2000 in samples A-E on individuals born in 1983. An expanded and revised questionnaire entered the field one year later, in 2001, for all samples (A-F). In samples A-E, young people born in 1984 were surveyed, and in sample F, those born in the years 1982 to 1984. With the expansion of the number of birth cohorts, entries for the birth year 1983 are also collected for sample F (data previously existed only for samples A-E), which also creates a clear increase in the number of entries. In the following years, also the youths from additional samples have been interviewed. For an overview of the target population in each survey year, see Table 1. In total, we have gathered interview data from 3,946 analysable observations up to the present.

Table 1: Target Population for the Youth Questionnaire by year, sample and age

survey year / sample	A-E	F	G	H	I	J	frequency	percent
• 2000	17 years						232	5.88
• 2001	17 years	17-19 years					618	15.66
• 2002	17 years	17 years					352	8.92
• 2003	17 years	17 years	17 years				365	9.95
• 2004	17 years	17 years	17 years				373	9.45
• 2005	17 years	17 years	17 years				368	9.33
• 2006	17 years	17 years	17 years				307	7.78
• 2007	17 years	17 years	17 years	17 years			346	8.77
• 2008	17 years	17 years	17 years	17 years			261	6.61
• 2009	17 years	17 years	17 years	17 years			243	6.16
• 2010	17 years	17 years	17 years	17 years	17 years		214	5.42
• 2011	17 years	17 years	17 years	17 years		17 years	267	6.77

Status: up to wave BB (2011)

In 2006, a new questionnaire on cognitive potential was introduced. Like the Youth Questionnaire, this instrument is aimed at youths who have reached the surveying age of 17 years. In order to keep the Interview at an acceptable length, the standard Individual Questionnaire is now left out for respondents of the Youth Questionnaire. The 2006 data on cognitive potentials was provided for secondary analysis in 2009 (dataset COGDJ).

11.2 Contents and Structure of the Data Set BIOAGE17

From a technical perspective, four different types of questions are asked in the Youth Questionnaire:

A) Questions used to complete certain biographical files (BIOIMMIG, BIOPAREN). These questions are identical to questions in the standard Biography Interview. This applies to the topic blocks 'Origin' (questions 60 to 71) and 'Childhood and Parents' House' (questions 72-85). The corresponding variables are *not* included in BIOAGE17, but combined with biographical information from non-youth new entrants in the files BIOIMMIG and BIOPAREN.

B) Questions that are similar to items in the standard Biography Interview, but go further into detail. This applies to the topic blocks 'Relationships' (questions 12-14), 'Free time and Sport' (questions 15-25) and 'Education and Career plans' (questions 26-55). These variables are stored in BIOAGE17. Corresponding Variables obtained from other new sample members (with a standard Biography Interview) are included in the dataset BIOSOC. Depending on the complexity and scope of the analysis, the user might want to combine corresponding data from BIOAGE17 and BIOSOC in order to access all panel members.

C) Questions that specifically relate to young persons and therefore have no equivalent in the standard Biography Interview. This applies to the topic blocks 'Residence' (questions 1-3), 'Jobs and Money' (questions 4-11), 'Future' (question 59) and 'Attitudes and Opinions' (questions 86-87). These Variables are stored in BIOAGE17 and have no equivalent for other panel entrants in BIOSOC.

D) Since 2006, selected time-variant questions from the unanswered regular individual questionnaire (which is not handed out to first time panel respondents from existing panel households, see 12.1) are added to the Youth Questionnaire. This refers to the questions 56 to 58, 89, 90 and the topic block 'personality' (questions 91 to 99). This data is *not* included in BIOAGE17[21], but stored in an additional dataset $PAGE17.

The design of the dataset BIOAGE17 is patterned after the 2001 Youth Questionnaire, which is the standard version for subsequent years. As in the biographical data survey, every youth answers the Youth Questionnaire only once. The data is therefore presented in column form, just as it would be in a cross-sectional record. The variable ERHEBJ makes it possible to quickly identify the year of the survey. The entries to the questions that were only asked in 2000 and not in 2001 are not included in BIOAGE17. The complete dataset from 2000 is provided free of charge upon request. However, all entries from 2000 that are also included in 2001 are contained in BIOAGE17![22] With the integration of questions from the Individual Questionnaire in 2006, some changes have occurred in the Youth Questionnaire, especially in the numbering of the questions.

Table 2 (at the end of this chapter) lists all of the variables for the dataset BIOAGE17. The first column contains the name of each variable, the second a brief specification of its content,

[21] The first ten items in question 90 are still stored in BIOAGE17, for details see 12.3.
[22] In the event that a question was asked in 2001 but not in 2000, the variable will have the value -3 for the persons who were surveyed in 2000.

and the third the number of the question as it appears in the Youth Questionnaire distributed in 2000, wave Q. The fourth column lists the corresponding questions in the Youth Questionnaires 2001 to 2005. Since the Youth Questionnaire was not altered further in the following years, the questionnaire numbers reported are identical[23] and therefore function as a reference for the variables in the dataset BIOAGE17. In the last column, the question number from 2006 to 2011 youth questionnaires is noted. The variables containing the identification of the person surveyed and the interview situation have no corresponding number because they do not originate from the regular section of the Youth Questionnaire.

11.3 Special Features of Some Questions and Variables

The question regarding the support received by these young people from their parents (question 14) is based on the Supportive Parenting Scale of Simons et al. (1992)[24], which was transformed for Germany by Schwarz and Walper (1997)[25]. The instrument used to compile career orientation (question 54) was taken from Kracke (1996)[26].

Before 2006, problems arose with the question concerning school attendance (question 25 from 2001 to 2005) because of discrepancies between the information from the Youth Questionnaire and the information on the variable "type of general school attended" from the Individual Questionnaire. Since 2006, 17-year-olds no longer receive the Individual Questionnaire, so the question about school type has been integrated into the question on school attendance. For the previous years, the variable was generated using information from the Individual Questionnaire and questions 25 and 45[27] from the Youth Questionnaire.

If the question on school attendance in the Youth Questionnaire is answered with 'yes' when at the same time information from the regular Individual Questionnaire indicates that the youth does not attend the general school system, or vice versa, a recoding is undertaken. In this case the variable BYSCHBES is changed to the value -3 (-3: Entry deleted after intensive examination). Another problem arises if a person states in the Youth Questionnaire that she attends school but does not specify school type in the Individual Questionnaire. In this case the variable BYSCHBES is given the value -1 (-1: no answer).

In question 51, young people are asked whether they know what career they would like to start. If they give a positive answer ('yes, with some certainty', 'yes, with a lot of certainty'),

[23] In 2004 and 2005, only one question was added. It concerns the number of brothers and sisters. However, the consecutive numbering of the relevant questions remains unchanged.
[24] Simons, R.L., F.O. Lorenz, R.D. Conger and C.-I. Wu (1992): Support from spouse as mediator and moderator of the disruptive influence of economic strain on parenting. in: Child Development 63: 1282-1301.
[25] Schwarz, B. and S. Walper (1997): Erziehung aus Sicht von Eltern und Kindern. Erste Erfahrungen mit den Instrumenten der 1. Erhebung. Berichte aus der Arbeitsgruppe "Familienentwicklung nach der Trennung" #19/97. Ludwig-Maximilians-Universität München.
[26] Kracke, B. (1996): Fragebogen zur Berufsorientierung bei Realschülern. University of Mannheim, unpublished manuscript.
[27] For 2000 questions 24 and 45.

then they are asked to specify the occupation in plain text. This plain-text entry is coded according to the classification of occupations of the Federal Statistical Office, Germany, (Statistisches Bundesamt), version 1992, and according to the ISCO 1988. In addition, the values for Ganzeboom's International Socio-Economic Index of Occupational Status (ISEI), for Treiman's Standard International Occupational Prestige Scale (SIOPS) for Erikson's and Goldthorpe's Class Category (EGP)[28] as well as Wegener's Magnitude Prestige Scale (MPS)[29] are also given.

Since 2005 some respondents have a value of –3 in variables BYMUSART, BYMUSMW and BYSPRTMW. This means that they gave more than one answer to the question although only one answer was possible. Because of this, it was not possible to assign a single valid answer.

By extending the questions about personality, we meanwhile ask the questions regarding attitudes about life and the future on a seven-point scale instead of the four-point scale we started with in the earlier version of this battery. From 2006, the variables BYESVERL to BYESENGA are stored with the values 1 (no acceptance) to 7 (total acceptance) and with the values 11 (total acceptance) to 14 (no acceptance) for respondents of previous years. Thus, the normative decision on how to integrate these two scales is up to the user.

[28] For ISCO 88, SIOPS, ISEI and EGP see Ganzeboom, H.B.G. and D.J. Treiman (1996): Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations. in: Social Science Research 25, 201-239.
[29] Frietsch, R. and H. Wirth (2001): Die Übertragung der Magnitude-Prestigeskala von Wegener auf die Klassifizierung der Berufe. in: ZUMA-Nachrichten, 48, 139-163.

Table 2: Description of the data set BIOAGE17

Variable Name	Content of the Variable	Number of Question in Youth Quest-ionnaire 2000[30]	Number of Question in Youth Quest-ionnaire 2001-2005	Number of Question in Youth Quest-ionnaire 2006 - 2011
	Entries for surveyed person			
HHNR	Original household identifier (invariant)			
HHNRAKT	Actual household identifier			
PERSNR	Personal identifier			
BEFRPER	Respondent identifier			
ERHEBJ	Survey year			
BYGEBJAH	Year of birth			
BYMNR	identifier of mother (taken form BIOPAREN; social, not necessarily biological relationship)			
BYVNR	identifier of father (taken from BIOPAREN; social, not necessarily biological relationship)			
	Residence			
BYWOELT	Residing in parents` household (HH)	01	01	01
BYWOZIM	Own room	02	02	02
BYWOWEI	Additional apartment outside of parents' HH	04	03	03
	Jobs and Money			
BYVDEIG	Own income	09	04	04
BYVDART	Type of income	10	05	05
BYJBFRUE	Worked before (on holiday or while in school)	13	06	06
BYJBALT	Age by first job (on holiday or while in school)	14	07	07
BYJBGRUN	Reason for working	-	08	08
BYTGELD	Allowance	15	09	09
BYTGELDW	Amount of allowance per week	16	10	10
BYTGELDM	Amount of allowance per month	16	10	10
BYSPAR	Saving money	17	11	11
BYSPARM	Amount saved every month	17	11	11
BYSPARUN	Sporadic saving	17	11	11
	Relationships			
	Importance of various persons:			
BYWIVA	Father	-	12	12
BYWIMU	Mother	-	12	12
BYWIBS	Brother, Sister	-	12	12
BYWIVW	Other related persons	-	12	12
BYWIFFR	Serious boy/girlfriend	-	12	12
BYWIBFR	Best friend	-	12	12
BYWILEHR	Teacher	-	12	12
BYWICLQ	Clique	-	12	12
BYWISON	Other person	-	12	12
	Frequency of fights with:			
BYSTRVA	Father	-	13	13
BYSTRMU	Mother	-	13	13
BYSTRBS	Brother, Sister	-	13	13
BYSTRFFR	Serious boy/girlfriend	-	13	13
BYSTRBFR	Best friend	-	13	13

[30] If no corresponding question/variable exists, it is assigned a minus sign; numbers/names in parentheses mean that there is no identical question/variable but a corresponding one.

Variable Name	Content of the Variable	Number of Question in Youth Quest-ionnaire 2000[30]	Number of Question in Youth Quest-ionnaire 2001-2005	Number of Question in Youth Quest-ionnaire 2006 - 2011
BYBZ01MU	Talk with mother about personal experiences	-	14	14
BYBZ01VA	Talk with father about personal experiences	-	14	14
BYBZ02MU	Mother addresses problems	-	14	14
BYBZ02VA	Father addresses problems	-	14	14
BYBZ03MU	Mother asks opinion before a decision is made	-	14	14
BYBZ03VA	Father asks opinion before a decision is made	-	14	14
BYBZ04MU	Mother shows approval	-	14	14
BYBZ04VA	Father shows approval	-	14	14
BYBZ05MU	Solve problems together with mother	-	14	14
BYBZ05VA	Solve problems together with father	-	14	14
BYBZ06MU	Mother shows trust	-	14	14
BYBZ06VA	Father shows trust	-	14	14
BYBZ07MU	Mother asks opinion on family issues	-	14	14
BYBZ07VA	Father asks opinion on family issues	-	14	14
BYBZ08MU	Mother justifies decision	-	14	14
BYBZ08VA	Father justifies decision	-	14	14
BYBZ09MU	Mother shows love	-	14	14
BYBZ09VA	Father shows love	-	14	14
	Free time and Sport			
	Frequency of free time activities:			
BYFZFERN	TV, Video	-	15	15
BYFZPC	Computer games	-	15	15
BYFZMUSH	Listen to music	-	15	15
BYFZMUSS	Play music	-	15	15
BYFZSPRT	Do sports	-	15	15
BYFZTANZ	Dance, Theatre	-	15	15
BYFZTECH	Technical work, Programming	-	15	15
BYFZLESE	Read	-	15	15
BYFZEHRE	Volunteer activities	-	15	15
BYFZABH	Do nothing, hang around, day dream	-	15	15
BYFZMFFR	Spend time with boy/girlfriend	-	15	15
BYFZMBFR	Spend time with best friend	-	15	15
BYFZMCLQ	Spend time with clique	-	15	15
BYFZINT	Internet/chatting	-	-	15
BYFZJUGZ	visiting youth center	-	-	15
BYFZRELI	go to church/religious activities	-	-	15
BYMUSSP	Actively make music	-	16	16
BYMUSART	Style of music made	-	17	17
BYMUSMW	Play music with whom	-	17a	18
BYMUSALT	Age starting playing music	-	18	19
BYMUSUNT	Paid music lessons	-	19	20
BYSPRTTR	Participate in sports	20	20	21
BYSPRTAR	Favourite sport	21	21	22
BYSPRTAL	Age started favourite sport	23	22	23
BYSPRTMW	Where and with whom favourite sport	23	23	24
BYSPRTWE	Participation in competitions	23	24	25

Variable Name	Content of the Variable	Number of Question in Youth Quest-ionnaire 2000[31]	Number of Question in Youth Quest-ionnaire2001-2005	Number of Question in Youth Quest-ionnaire 2006 – 2011
	Education and Career Plans			
BYSCHBES[32]	School attendance	24 / PF: 09	25 / PF[33]	26
BYSCHEND	Last year of school	25	26a	27
BYSCHABS	Type of school certificate	26	26b	28
BYSCHZUK	Strive for further school certificate	27	27	29
BYSCHZAR	Type of further school certificate	28	28	30
BYFMD1[34]	1. foreign language	32	29	31
BYFMD2	2. foreign language	32	29	31
BYSCHAUS	School attendance in foreign country	29	30	32
BYSCHPRI	Attendance in a private school	-	31	33
	Activities in school:			
BYENKSPR	Class representative	34	32	34
BYENSSPR	School representative	34	32	34
BYENSZTG	School newspaper	34	32	34
BYENTHEA	Theatre, Dance group	34	32	34
BYENCHOR	Choir, Music	34	32	34
BYENSPRT	Sport group	34	32	34
BYENSONS	Other groups	34	32	34
BYENNEIN	No activities	34	32	34
BYZFINSG	Satisfaction with effort at school (overall)	31	33	35
BYZFDEUT	Satisfaction with effort in German	31	33	35
BYZFMATH	Satisfaction with effort in math	31	33	35
BYZFFMD1	Satisfaction with effort in 1. foreign language	31	33	35
BYEMPFEH	Recommendation after elementary school	-	34	36
BYNTDEUT	Last grade[35] in German	33	35	37
BYNTMATH	Last grade in math	33	35	37
BYNTFMD1	Last grade in 1. foreign language	33	35	37
BYPTDEUT	Total points[36] in German	33	35	37
BYPTMATH	Total points in math	33	35	37
BYPTFMD1	Total points in 1. foreign language	33	35	37
BYGSDEUT	Level of German at comprehensive school[37]	33	35	37

[31] If no corresponding question/variable exists, it is assigned a minus sign; numbers/names in parentheses mean that there is no identical question/variable but a corresponding one.
[32] As mentioned in 13.3, for the years 2000 to 2005 the variable BYSCHBES is generated in consideration of information stemming from the personal questionnaire.
[33] The relevant question from the personal questionnaire differs from year to year: for 2001 question 11, for 2002 question 14, for 2003 question 33, for 2004 question 08, and for 2005 question 09.
[34] Additional category since 2006: value 7 “Spanish”.
[35] Students normally receive grades ranging from 1 to 6, whereby 1 is the best and 6 the worst. This system of assigning grades is used up to the 11th or 12th grade (level II of upper secondary or comprehensive school) depending on the federal state. After that, a new grading system is used. To make the data set more user-friendly, the information given for school grades and the information on points transformed into grades is stored in this variable. Note: No corrections have been made when a person has reported both grades and point scores and when the two types of information do not correctly correspond.
[36] From the 11th or 12th grade on, pupils are awarded points in upper secondary or comprehensive school ranging from 0 to 15, whereby 15 points are the best, 0 points the worst. The link between points and grades is as follows: 0 points: 6; grade of 1 to 3 points: grade of 5; 4 to 6 points: grade of 4; 7 to 9 points: grade of 3; 10 to 12 points: grade of 2; 13 to 15 points: grade of 1.
[37] The subjects German, math and the first foreign language are split up into different levels during the secondary school level I in comprehensive schools. Level A is the highest. The number of levels differ between the federal states.

Variable Name	Content of the Variable	Number of Question in Youth Quest-ionnaire 2000[31]	Number of Question in Youth Quest-ionnaire2001-2005	Number of Question in Youth Quest-ionnaire 2006 – 2011
BYGSMATH	Level of math at comprehensive school	33	35	37
BYLKDEUT	Complementary / main subject[38] in German	33	35	37
BYLKMATH	Complementary / main subject in math	33	35	37
BYLKFMD1	Complementary / main subject in 1. foreign language	33	35	37
BYKLWDJA	Class repeated	35	36	38
BYKLWD1	Class level 1. repeated	36	37	39
BYKLWD2	Class level 2. repeated	36	37	39
BYNACHHI	Paid tutor lessons	37	38	40
BYELKUEM	Parents care about efforts at school	39	39	41
BYELHAUS	Parents help with homework	40	40	42
BYELDIFF	Problems with parents because of effort at school	41	41	43
BYELABEN	Parents attend parents' evening	42	42	44
BYELSPRE	Parents go to parents' day	42	42	44
BYELLEHR	Parents go to see a teacher	42	42	44
BYELVERT	Active as parent representative	42	42	44
BYELNIDA	Parents do not participate in any of these activities	42	42	44
BYKLAUSL	Number of foreign classmates	(43)	43	45
BYBAABGE	Vocational education, Internship, training	44	44	46
BYBABGJ	Vocational introductory year ("Berufsgrundschul- / Berufsvorbereitungsjahr")	45	45	47
BYBABEGL	Vocational integration training ("Berufl. Eingliederungslehrgaenge")	45	45	47
BYBALEH	Vocational education, apprenticeship ("Berufsausbildung, Lehre")	45	45	47
BYBABFS	Full-time vocational school/ School for public health ("Berufsfachschule / Schule des Gesundheitswesen")	45	45	47
BYBAPRAK	Internship ("Praktikum, Voluntariat")	45	45	47
BYZAJA	Vocational / university degree is aspired	46	46	48
	Type of aspired vocational / university degree:			
BYZALEH	Apprenticeship ("Lehre")	47	47	49
BYZABFS	Full-time vocational school/ School for public health ("Berufsfachsule / Schule des Gesundheitswesen")	47	47	49
BYZAFSC	Technical school, school for master of a trade ("Fachschule, Meister-, Technikerschule")	47	47	49
BYZABEA	Training for civil servants (officer) ("Beamtenausbildung")	47	47	49
BYZABAK	Approved vocational academy ("anerkannte Berufsakademie")	47	47	49
BYZAFH	Advanced technical college ("Fachhochschule")	47	47	49
BYZAUNI	University	47	47	49
BYSLBALT	Desired age for financial independence	48	48	50

[38] From the 11th or 12th grade on, pupils can choose their main subjects. At this stage, German, math and foreign languages can be downgraded from major to minor subjects.

Variable Name	Content of the Variable	Number of Question in Youth Quest-ionnaire 2000[31]	Number of Question in Youth Quest-ionnaire2001-2005	Number of Question in Youth Quest-ionnaire 2006 – 2011
BYBWUNJA	Occupation is aspired	49	49	51
	Occupation categories, encoded:			
BYKLAS	Classification of career according to the Federal Statistical Office, Germany, (Statistisches Bundesamt), version 1992	50	50	52
BYISCO88	International Standard Classification of Occupation 1988 (ISCO88)	50	50	52
BYEGP	Erikson and Goldthorpe's Class Category (EGP)	50	50	52
BYISEI	International Socio-Economic Index of Occupational Status after Ganzeboom (ISEI)	50	50	52
BYSIOPS	Treiman's Standard International Occupational Prestige Scale (SIOPS)	50	50	52
BYMPS	Magnitude Prestige Scale after Wegener (MPS)	50	50	52
BYZBINF	Information level of planned career	-	51	53
BYZBELT	Influence of the parents on career choice	-	52	54
BYZBLAS	No specific career in mind	-	52	54
BYZBBES	Intensive thoughts about various careers	-	52	54
BYZBRAU	Still looking for a career	-	52	54
	Important aspects for the career choice:			
BYWBSICH	Secure job	51	53	55
BYWBEINK	High income	51	53	55
BYWBAUF	Promotion opportunities	51	53	55
BYWBANE	Established profession	51	53	55
BYWBFREI	Enough free time	51	53	55
BYWBINT	Interesting activities	51	53	55
BYWBSELB	Working independently	51	53	55
BYWBKONT	Contact with persons	51	53	55
BYWBGSL	Relevant to society	51	53	55
BYWBGSND	Healthy conditions at work	51	53	55
BYWBFAM	Flexibility for family	51	53	55
BYWBHELF	Help others	51	53	55
	Future			
	Probability of future career related and private events:			
BYWAAUSP	To be accepted for a desired apprenticeship / place at university	52	54	59
BYWAERFA	To complete training/ university successfully	52	54	59
BYWAARBP	Job in desired career	52	54	59
BYWABERF	Job-related success	52	54	59
BYWAARBL	Longer unemployment	52	54	59
BYWAZURU	From family related reasons held back in career	52	54	59
BYWASELB	Self-employed	52	54	59
BYWAAUSL	Work in foreign country	52	54	59
BYWAHEIR	To marry	52	54	59
BYWAPART	Live together with partner (not married)	52	54	59
BYWAKID1	Have one child	52	54	59

Variable Name	Content of the Variable	Number of Question in Youth Quest-ionnaire 2000[31]	Number of Question in Youth Quest-ionnaire2001-2005	Number of Question in Youth Quest-ionnaire 2006 – 2011
	Attitudes and Opinions			
BYGLKIND	Happiness: with/without children	-	80	87
	Success in FRG from			
BYEFFLEI	Studiousness	86	81	88
BYEFAUSN	Exploitation of others	86	81	88
BYEFINT	Intelligence	86	81	88
BYEFFAM	Family's origin	86	81	88
BYEFFACH	Technical know-how	86	81	88
BYEFGELD	Money	86	81	88
BYEFSABS	School education	86	81	88
BYEFHART	Being inconsiderate and hard	86	81	88
BYEFBEZ	Networking	86	81	88
BYEFPOLI	Political activities	86	81	88
BYEFMANN	Sex/ 'being a man'	86	81	88
BYEFINI	Being dynamic and taking initiative	86	81	88
BYESVERL	What happens in life, depends on me	-	(82)	90
BYESERRE	Did not reach, what I deserve	-	(82)	90
BYESGLUE	What you achieve, is a matter of luck	-	(82)	90
BYESAND	Others decide about my life	-	(82)	90
BYESHART	You have to work hard for success	-	(82)	90
BYESZWEI	By difficulties, doubt about own abilities	-	(82)	90
BYESSOZU	Chances are determined by social circumstances	-	(82)	90
BYESFAEH	Abilities are more important than efforts	-	(82)	90
BYESKNTR	Little control over events in my life	-	(82)	90
BYESENGA	Change of social circumstances through social/political activities	-	(82)	90
	Specification of Interview Situation			
BYINTA	Type of interview			
BYDAUER1	Duration of personal interview			
BYDAUER2	Duration of interview filled out independently			
BYANW	Presence of other persons			
BYTAGIN	Day of the interview			
BYMONIN	Month of the interview			
BYINTNR	Identifier of the interviewer			

12 BIOSOC: Retrospective Data on Youth and Socialization

by Henning Lohmann, Marco Giesselmann, Mila Staneva *and* Sven Witzke

(Replaces earlier versions by Jürgen Schupp *and* Michael Frühling / Bettina Isengard *and* Thorsten Schneider*)*

The standard supplementary Biography Questionnaire was expanded in 2000, and again in 2001 to include some specific questions on youth and early adulthood. Some of these questions are derived from the independent Youth Questionnaire (for detailed information on this questionnaire, see chapter on BIOAGE 17). The expanded questionnaire asks respondents of all ages to describe aspects of their life at the age of 15, including their relationship with parents, grades in school, the federal state where they last attained educational qualifications, detailed information on vocational qualifications, as well as intentions to complete further education or vocational training (the latter questions were relevant mainly to younger respondents). Questions concerning military and alternative services are also included in this data set.

As these questions are a part of the standard Biography Questionnaire, they are only asked once. Some of these questions can, however, be followed up by the regular data collected in the Individual Questionnaire. For example, if someone was too young to have completed his military service when the Biography Questionnaire was conducted, the user can look at the data set ARTKALEN in later years, where labour force participation is recorded on a monthly basis. Here one can find out if somebody was doing military service at the time or not.

The data set BIOSOC contains information on 22,984 persons, of whom 8,819 stem from the year 2001. The reason for this is that the Biography Questionnaire was directed to sample F, as this was its second survey year. Consequently, the majority of the persons in this data set belong to sample F (42%) or the samples which were included in more recent years (sample G, H, I and J).

Table 1: Survey Year in BIOSOC

survey year	frequency	percent
2000	246	1.07
2001	8,819	38.37
2002	552	2.40
2003	2,328	10.13
2004	450	1.96
2005	299	1.30
2006	223	0.97
2007	2,232	9.71
2008	336	1.46
2009	196	0.85
2010	2,038	8.87
2011	5,265	22.91
Total	17,719	100

Status: up to wave BB (2011)

Table 2: Samples in BIOSOC

sample	frequency	percent
A: Germans (west)	768	3.55
B: Foreigners (west)	225	1.03
C: Germans (east)	546	2.47
D: Immigrants 1984-1993	111	0.49
E: Supplement 1998	236	1.10
F : Innovation 2000	9,641	42.16
G : High-Income 2002	2,201	9.63
H: Supplement 2006	2,121	9.30
I: Incentivation 2009	1,870	8.14
J: Supplement 2011	5,087	22.13

Status: up to wave BB (2011)

12.1 Structure of the Data Set BIOSOC

Respondents are given the Biography Questionnaire only once in a lifetime. Some of the information stored in the new data set BIOSOC is invariant (such as the relationship to parents at the age of 15) or is not surveyed to such an extent in the regular questionnaire (such as last school grades). Consequently there is only one record for each person and updates are not intended for this data set. The variable ERHEBJ makes it possible to quickly identify the year of the survey. Using the variable BSGEBJAH, which contains the year of birth, the user

can determine the respondent's age. If the respondent is of a certain age, one can assume that some of the variables are constant. This applies to variables such as last school grades or military service.

In Table 3 (at the end of the chapter) all variables of the data set BIOSOC are listed. The first column contains the name of the variable, the second a brief specification of its content. The third column contains the number of the question as it appears in the Biography Questionnaire of wave Q (2000). Here, a minus sign means that the variable is not available in a given year and a question number in parenthesis indicates limited comparability. The fourth column contains the number of question in waves R to BB (2001-2011). As one can see, all listed questions were asked in the year 2001. In the last column the corresponding variable in the BIOAGE17 dataset is given, if available.

12.2 Special Features of Some Questions and Variables

The interviewees were asked if they did sports in their youth. If they answered in the affirmative, they were asked to include the sport they participated in most. This information was re-coded to a numeric variable and categorised. Some categories could easily be coded, such as soccer, whereas for others this was not possible. For users interested in specific research questions on sports in youth, the original plain text answers can be provided upon request.

Individuals were asked: "When was the last year you attended school?" (Question 38). If they were still attending school, they had the opportunity to report that they were students. Unfortunately in the years 2001 to 2003, individuals who reported being students or who did not provide any answer to this question skipped over numerous questions due to the questionnaire design. Consequently, for these individuals there is no information on the number of foreign classmates[39], on the school degree aspired to, or planned vocational qualifications. Their record also lacks information on past vocational qualifications. However, this should prove less problematic since most of the students were enrolled in regular school programs throughout the entire time.

[39] Most persons with no valid information on foreign classmates are not students but individuals who finished school abroad. This is because the question only targets those in German schools.

Table 3: Description of the data set BIOSOC

Variable Name	Content of the Variable	Number of Question in Biography Question-naire 2000[40]	Number of Question in Biography Question-naire 2001-2011	Comparable Variable in BIOAGE17[2]
	Entries for Surveyed Person			
HHNR	Original household identifier (invariant)			HHNR
HHNRAKT	Actual household identifier			HHNRAKT
PERSNR	Personal identifier			PERSNR
BEFRPER	Respondent identifier			BEFRPER
ERHEBJ	Survey year			ERHEBJ
BSGEBJAH	Year of birth			BYGEBJAH
	School			
BSELKUEM	Parents took care about efforts at school	-	29	BYELKUEM
BSNTDEUT	Last grade in German	-	30	BYNTDEUT
BSNTMATH	Last grade in maths	-	30	BYNTMATH
BSNTFMD1	Last grade in 1. foreign language	-	30	BYNTFMD1
BSPTDEUT	Total points[41] in German (last class)	-	30[42]	BYPTDEUT
BSPTMATH	Total points in maths (last class)	-	30[20]	BYPTMATH
BSPTFMD1	Total points in 1. foreign language (last class)	-	30[20]	BYPTFMD1
BSGSDEUT	Level of German at comprehensive school[43] (last class)	-	30[20]	BYGSDEUT
BSGSMATH	Level of maths at comprehensive school (last class)	-	30[20]	BYGSMATH
BSGSFMD1	Level of 1. foreign language at comprehensive school (last class)	-	30[20]	BYGSFMD1
BSLKDEUT	Complementary / main subject[44] in German (last class)	-	30[20]	BYLKDEUT
BSLKMATH	Complementary / main subject in maths (last class)	-	30[20]	BYLKMATH
BSLKFMD1	Complementary / main subject in 1. foreign language (last class)	-	30[20]	BYLKFMD1
	Relationships to Parents, Sport and Activities during Youth			
	Frequency of fights when respondent was 15. years old with:			
BSSTRVA	Father	-	31	BYSTRVA
BSSTRMU	Mother	-	31	BYSTRMU

[40] If no corresponding question/variable exists, it is assigned a minus sign; numbers/names in parentheses mean that there is no identical question/variable but a similar one.
[41] To make the data set more user-friendly, the information given on points are transformed into grades and stored in the corresponding variable. The link between points and grades is as follows: 0 points: grade of 6; 1 to 3 points: grade of 5; 4 to 6 points: grade of 4; 7 to 9 points: grade of 3; 10 to 12 points: grade of 2; 13 to 15 points: grade of 1.
[42] Only in survey year 2001 (wave R)
[43] The subjects German, math and the first foreign language are split up into different levels during the secondary school level I of comprehensive schools. Level A is the highest one.
[44] From the 11th or 12th grade on students can choose their main subjects. At this stage, they can reduce German, maths and foreign languages from major to minor subjects.

Variable Name	Content of the Variable	Number of Question in Biography Question-naire 2000[45]	Number of Question in Biography Question-naire 2001-2010	Comparable Variable in BIOAGE17[2]
BSSPRTTR	Participated in sports during youth	-	32[46];33[47]	BYSPRTTR
BSSPRTAR	Favourite sport during youth	-	33[23];34[24]	BYSPRTAR
BSSPRTWE	Participated in competitions during youth	-	34[23];35[24]	BYSPRTWE
BSMUSSP	Played music or sang during youth	-	35[23];32[24]	BYMUSSP
	School Attendance			
BSSCHBES	Still at school	(34)	37	BYSCHBES
BSSCHEND	Year left school	-	37	BYSCHEND
BSSCHWO	Country of last school attendance	(34)	38	-
BSSCHLA	Federal State of last school attendance	-	41	-
BSKLAUSL	Number of foreign classmates	(37)	43	BYKLAUSL
BSSCHZUK	Strive for further school certificate	35	44	BYSCHZUK
BSSCHZAR	Type of further school certificate	36	45	BYSCHZAR
	Attained and Planed Vocational Qualification			
BSBADABG	Vocational / university degree acquired in Germany	38	46	(BYBAABGE)
	Type of vocational / university degree attained in Germany:			
BSBADLEH	Apprenticeship (“Lehre”)	39	47	(BYBALEH)
BSBADBFS	Full-time vocational school / School for public health (“Berufsfachsule / Schule des Gesundheitswesens”)	39	47	(BYBABFS)
BSBADFSC	Technical school, school for master of a trade (“Fachschule, Meister-, Technikerschule”)	39	47	-
BSBADBEA	Training for civil servants (officer) (“Beamtenausbildung”)	39	47	-
BSBADFHA	Advanced technical college (“Fachhochschule”) or approved vocational academy (“anerkannte Berufsakademie”)	39	47	-
BSBADUNI	University degree	39	47	-
BSBADSON	Other vocational qualification	39	47	(BYBABGJ, BYBABEGL, BYBAPRAK)
BSBADEND	Year of attaining vocational / university degree in Germany	-	48	
BSBADEND	Year of attaining vocational / university degree in Germany	-	48	
BSBAAABG	Vocational / university degree acquired abroad	40	49	-
	Type of vocational / university degree attained abroad:			

[45] If no corresponding question/variable exists, it is assigned a minus sign; numbers/names in parentheses mean that there is no identical question/variable but a similar one.
[46] Survey years 2001 and 2002 (waves R, S)
[47] Survey years 2003 to 2011 (waves T, U, V, W, X, Y, Z, BA, BB)

Variable Name	Content of the Variable	Number of Question in Biography Question-naire 2000[48]	Number of Question in Biography Question-naire 2001-2010	Comparable Variable in BIOAGE17[2]
BSBAAFAN	Short-term training in a company	41	50	-
BSBAAFBA	Apprenticeship in a company	41	50	-
BSBAASCH	Vocational or professional school	41	50	-
BSBAAUNI	University degree	41	50	-
BSBAASON	Other vocational qualification	41	50	-
BSBAAEND	Year of attaining vocational / university degree abroad	-	51	-
BSBAAZEU	Certificate for abroad attained qualification	42	52	-
BSBAAZEA	Recognition of aboard attained certificate	42	52	-
BSZAJA	Vocational / university degree is aspired	43	53	BYZAJA
BSZALEH	Type of aspired vocational / university degree: Apprenticeship ("Lehre")	44	54	BYZALEH
BSZABFS	Full-time vocational school/ School for public health ("Berufsfachschule / Schule des Gesundheitswesens")	44	54	BYZABFS
BSZAFSC	Technical school, school for master of a trade ("Fachschule, Meister-, Technikerschule")	44	54	BYZAFSC
BSZABEA	Training for civil servants (officer) ("Beamtenausbildung")	44	54	BYZABEA
BSZABAK	Approved vocational academy ("anerkannte Berufsakademie")	44	54	BYZABAK
BSZAFH	Advanced technical college ("Fachhochschule")	44	54	BYZAFH
BSZAUNI	University degree	44	54	BYZAUNI
	Military and Voluntary Service			
BSDIGEL	Military or alternative service done *(only men)*	58	71[49];74[50]	-
BSDIART	Type of service *(only men)*	58	71;74	-
BSDIGRU[51]	Reason for not serving *(only men)*	58	71;74	-
BSFSJ	Voluntary social service ("Freiwilliges Soziales Jahr")	-	72;73	-
	Specification of Interview Situation			
BSINTA	Type of interview			BYINTA
BSDAUER1	Duration of personal interview			BYDAUER1
BSDAUER2	Duration of interview filled out independently			BYDAUER2
BSTAGIN	Day of the interview			BYTAGIN
BSMONIN	Month of the interview			BYMONIN
BSINTNR	Identifier of the interviewer			BYINTNR

[48] If no corresponding question/variable exists, it is assigned a minus sign; numbers/names in parentheses mean that there is no identical question/variable but a similar one.
[49] survey years 2001 (wave R)
[50] survey years 2002 to 2011 (waves S, T, U, V, W, X, Y, Z, BA, BB)
[51] Since 2007 there are two new values. The former value 2 (unfit / replacement reserve) is now splitted into 11 (unfit) and 12 (replacement reserve).

13 BIOPAREN: Biography Information for the Parents of SOEP-Respondents

by Anne Fromm, Sebastian Frischholz and Daniel D. Schnitzlein

(This documentation is based on earlier versions of the BIOPAREN documentation and has benefited from previous work by Charlotte Büchne, Stefanie Lenuweit, Katharina Mahne, Matthias Pollmann-Schult, Jürgen Schupp and Verena Tobsch)

13.1 Short summary

The aim of the data file BIOPAREN is to make the biography entries on the parents and on the social origin of the respondent available.

13.2 How biography information has been collected in the SOEP

In the third wave (1986) intergenerational aspects of the persons surveyed were included for the first time by means of a special group of questions in the Individual Questionnaire. This deals with statements made about the education or professional training of the parents, the parents' residency, and their year of birth and death. For Sample B, only the education, residency, year of birth and death of the parents were asked. In 1988 the complete collection of biography questions (history of labor force participation, marriage and family biography, career start, and social origin) were included in the Individual Questionnaire for individuals surveyed for the first time. At the same time, a follow-up survey was given to those participants who had not yet received all or part of this collection of questions. This survey was continued in this form each of the following years until 1991, when the separate Biography Questionnaire was introduced. Since 1994, the biography was collected using the Personal History Questionnaire ('Lebenslauf-Fragebogen'), a slightly modified version of the Biography Questionnaire.

The Biography was included in Sample C in the third survey wave, that is, in 1992. The biographies of the persons in Samples D1 and D2 were collected during the first survey in 1994 and 1995. In 1999 the biography was collected for Sample E. In 2001 the follow-up survey was completed for Sample F and was followed by Sample G (High-Income) in 2003. The retrospective data of the sample H was collected in 2007. In 2009 the new subsample I with valid interview data on about 2.500 adults was introduced in the SOEP. These new respondents did not fill in the biography questionnaire in order to reduce response burden in the first wave. Their data have been integrated in their second wave (2010). Starting with sample J which is introduced in 2011 the Biography Questionnaire has been reintegrated in the Personnel Questionnaire and is answered again in the first year.

In addition to the Biography Questionnaire, there has been an independent questionnaire (Youth Questionnaire) in SOEP since 2000 for the group of survey participants who are 16-17 years old and are being interviewed for the first time; this questionnaire is mostly identical to the Biography Questionnaire (see chapter BIOYOUTH). 52

Table 1 gives an overview on the development of the number of respondents that enter BIOPAREN in each year.

13.3 How is BIOPAREN generated?

The information available in BIOPAREN is obtained in two different ways. On the one hand, BIOPAREN includes the children's proxy entries on the parents from the Biography Questionnaire and the Youth Questionnaire. On the other hand, it contains the direct entries from the parents in the case the respondent lives in the same household as his parents. Every respondent is asked for information on the regional mobility of the children, as well as on the religious affiliation of the parents. However, information on the year of birth, as well as the education and occupational training of the parents, additional to the professional position and occupation of the father are not collected due to the filter command in the questionnaire when the parents (or the father) live in the same household as the child at the time of the survey. In this case, the direct entries of the parents are used.

The identification of the parents occurs first of all through the variable $$STELL (relationship to head of household). The possible values of the variable $$STELL are listed in Table 2. The combinations of these characteristics of the $$STELL-variable and their assigned interpretation for the generation of parent identifiers are describe in Table 3.

The second source of information is the population of the file $$KIND, which includes all children under the age of 16. The file contains the personal number of the mother , as well as the personal number of the partner of the mother. Through both variables the latest mother, as

[52] A more precise representation of the development of the instruments used to collect the Personal History, including the social origin, can be found in the introduction to this documentation.

Table 1: Number of observations in BIOPAREN

Year of data collection	N	Samples									
		A	B	C	D	E	F	G	H	I	J
1984	1,682	1,079	603	-	-	-	-	-	-	-	-
1985	11,087	8,369	2,718	-	-	-	-	-	-	-	-
1986	501	355	146	-	-	-	-	-	-	-	-
1987	464	310	154	-	-	-	-	-	-	-	-
1988	380	245	135	-	-	-	-	-	-	-	-
1989	384	246	138	-	-	-	-	-	-	-	-
1990	4,811	221	139	4,451	-	-	-	-	-	-	-
1991	506	219	118	169	-	-	-	-	-	-	-
1992	497	202	142	153	-	-	-	-	-	-	-
1993	471	201	134	136	-	-	-	-	-	-	-
1994	930	213	97	148	472	-	-	-	-	-	-
1995	1,067	217	76	135	639	-	-	-	-	-	-
1996	483	214	88	144	37	-	-	-	-	-	-
1997	487	194	98	144	51	-	-	-	-	-	-
1998	419	192	65	127	35	-	-	-	-	-	-
1999	2,047	206	62	130	35	1,614	-	-	-	-	-
2000	455	168	60	146	35	46	-	-	-	-	-
2001	9,433	156	46	116	33	46	9,036	-	-	-	-
2002	836	150	38	117	42	50	439	-	-	-	-
2003	2,677	155	53	125	33	56	258	1,997	-	-	-
2004	807	145	50	134	26	26	281	145	-	-	-
2005	663	136	50	94	24	36	248	75	-	-	-
2006	529	115	30	92	16	36	170	70	-	-	-
2007	2,577	148	43	89	29	41	202	47	1,978	-	-
2008	594	122	35	48	11	24	162	44	148	-	-
2009	439	103	28	47	19	22	134	32	54	-	-
2010	2252	97	21	41	8	17	108	35	34	1,819	-
2011	5,531	99	30	34	8	30	109	30	30	-	5,161
Total	53,009	14,277	5,397	6,820	1,553	2,044	11,147	2475	2,244	1,891	5,161

Table 2: Characteristics of the variable $STELL "relationship of the person to the head of the household"

Code	Description
0	HH
1	Marital partner of the HH
2	partner of the HH
3	Daughter/son (also adopted/stepchild) of the HH
4	Foster child of the HH
5	Daughter/son–in-law of the HH
6	Father/mother of the HH
7	Father/mother-in-law of the HH
8	Brother/sister, brother/sister-in-law of the HH
9	Grandchild of the HH
10	Other relationship to the HH
11	Not related to the HH
12	Daughter/son of the partner of the HH
13	Marital partner of the HH (same sex)

Table 3: Possible Parent-Child Relationships based on $$STELL

Relationship of the child to the HH	Relationship of the parent to the HH	Person is ...
3	0	Child of HH
3	1 or 2	Child of marital/ partner of HH
4	0	Foster child of HH
4	1 or 2	Foster child of marital/ partner of HH
12	2 or 3	Child of partner of HH
9	3 or 4	Child of child/foster child of HH
0	6	Child is HH, lives with parents in same household
1 or 2	7	Marital partner/partner of HH (child of in laws of HH)
9	5	Grandchild of HH (child of son/daughter-in-law of HH)

well as the latest partner of the mother are identified, ideally, at the time when the child is 16 years old and thus one year before the first survey of the child. In the case the parents could not be identified by the $$STELL variable, this information is used.

In a further step the biological mother is identified through the mother-child relationship in the file BIOBIRTH. In the event that still no personal number for the mother exists, the number from BIOBIRTH is used. Since 2001, an extra BIOBIRTH data-set exists for fathers (BIOBRTHM), which provides a new way of identifying fathers of SOEP-respondents.

13.4 What's new in the current wave?

In the current wave (BB) seven new variables are added to BIOPAREN.

VAORT11, MAORT11 give the recent place of residence of father and mother. This information is collected all five years in the survey focus family and integrated in BIOPAREN.

GESCHW, GESCHWUP, NUMS, NUMB and TWIN are information on siblings. This information was collected in wave T for all participants and from then on the questions were integrated in the Biography questionnaire and the youth questionnaire. The variable GESCHW gives the information if the respondent ever had any siblings at the time of the interview. GESCHWUP gives the information about the year the sibling information was collected. NUMB and NUMS gives the information how many brothers or sisters the respondent reports and TWIN gives you the information if any of these are TWIN siblings (and of which type) to the respondent.

13.5 Complete list of variables in BIOPAREN

HHNR	Number of the original household
PERSNR	Personal number of the respondent (all persons)
VNR	Personal number of the father of the respondent
MNR	Personal number of the mother of the respondent
VGEBJ	Year of birth of the father
MGEBJ	Year of birth of the mother
VTODJ	Year of death of the father
MTODJ	Year of death of the mother
VAORT91	Residency of the father 1991 (Survey focus: family)
MAORT91	Residency of the mother 1991 (Survey focus: family)
VAORT96	Residency of the father 1996 (Survey focus: family)
MAORT96	Residency of the mother 1996 (Survey focus: family)

VAORT01	Residency of the father 2001 (Survey focus: family)
MAORT01	Residency of the mother 2001 (Survey focus: family)
VAORT06	Residency of the father 2006 (Survey focus: family)
MAORT06	Residency of the mother 2006 (Survey focus: family)
VAORT11	Residency of the father 2011 (Survey focus: family)
MAORT11	Residency of the mother 2011 (Survey focus: family)
VSBIL	Education of the father
MSBIL	Education of the mother
VBBIL	Vocational training of the father
MBBIL	Vocational training of the mother
VSINFO	Origin of the information on father's education
MSINFO	Origin of the information on mother's education
VBINFO	Origin of the information on father's vocational training
MBINFO	Origin of the information on mother's vocational training
VRELI	Religious affiliation of the father
MRELI	Religious affiliation of the mother
VNAT	Nationality of the father
MNAT	Nationality of the mother
VBSTELL	Professional position of the father (when the respondent was 15 years old)
VBSINFO	Origin of the information on the professional position of the father
MBSTELL	Professional position of the mother (when the respondent was 15 years old)
MBSINFO	Origin of the information on the professional position of the mother
VISCO88	Professional occupation of the father (when the respondent was 15 years old)
MISCO88	Professional occupation of the mother (when the respondent was 15 years old)
VISEI	Prestige score of father – concept of Ganzeboom
MISEI	Prestige score of mother – concept of Ganzeboom
VMPS	Prestige score of father – Magnitude scale – Wegener
MMPS	Prestige score of mother – Magnitude scale – Wegener
VSIOPS	Prestige score of father – Treiman standard score
MSIOPS	Prestige score of mother – Treiman standard score

VEGP	Prestige score of father – Erikson – Goldthorpe class category
MEPG	Prestige score of mother – Erikson – Goldthorpe class category
VBKLAS	Occupational coding scheme father according German statistical office
MBKLAS	Occupational coding scheme mother according German statistical office
ORTKINDH	Place of childhood
ORTKIND1	Still lives in place of childhood?
ORTKIND2	Year moved out of parents' household *(since 2000 no longer collected)*
ORTKIND3	Still lives in parents' household *(since 2000 no longer collected)*
LIVING1	No. of years living with both parents
LIVING2	No. of years living alone with mother
LIVING3	No. of years living with mother and new partner of mother
LIVING4	No. of years living alone with father
LIVING5	No. of years living alone with father and new partner of father
LIVING6	No. of years living with other relatives
LIVING7	No. of years living with foster parents
LIVING8	No. of years living in youth center
VSTREIT	Conflict with father
MSTREIT	Conflict with mother
VAORTAKT	Father‘s place of residence
MAORTAKT	Mother‘s place of residence
BIOYEAR	Year of the Biography Survey
BIO	Origin of the information (Biography or Youth Questionnaire)
ALTER	Age of the respondents
VALTER	Age of the respondent's father
MALTER	Age of the respondent's mother
MORIGIN	Country of origin of the respondent's mother
VORIGIN	Country of origin of the respondent's father
GESCHW	Sibling yes/no
GESCHWUP	Year of update of GESCHW
NUMS	Number of sisters
NUMB	Number of brothers
TWIN	Twin sister/brother

Variable **VNR / MNR**

Label: Personal number of the father of the respondent / Personal number of the mother of the respondent

Values: (-1) PERSNR father / mother unkown

(-2) Does not apply

(-3) Answer improbable

Description: The personal ID of the parents (VNR and MNR) is generated in three steps.

1. The parents of the respondent are identified by the relationship to the head of the household ($STELL in $BRUTTO). Ideally, the children's parents are identified at the time of the first survey of the child, i.e., when the child is 17 years old. Furthermore, the social parents and not necessarily the biological parents are identified.

2. The parents of the respondent are identified via the mother's ID as well as the mother's partner ID in $KIND. By using these variables the "oldest" parents are identified. Ideally, these are the parents at the time the child is 16 years old (one year before the first survey).

3. The mother-ID of the respondent can be identified in BIOBIRTH and the father-ID in BIOBRTHM.

As BIOPAREN aims at identifying the parents that live in the household when the child is 17 years old, the steps above are carried out in the hierarchy 1-3 with step 1 having the highest priority. If one is interested in only biological parents, please have a look at the information in BIOBIRTH and BIOBRTHM.

Variable **VGEBJ / MGEBJ**

Label: Year of birth of the father / Year of birth of the mother

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

Description: In a first step the information of the year of birth comes from the Biography Questionnaire. Due to a filter command, the children's proxy entries are only available for these variables when the parents or one parent and the child do not live in the same household at the time of the survey.

After the parents' personal numbers have been identified the information can be compared with the entries in PPFAD. If there are differences of +/- two years the VNR / MNR will be set as missing.

For the missing entries the information of the parents' year of birth is taken from PPFAD.

Variable **VTODJ / MTODJ**

Label: Year of death of the father / Year of death of the mother

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

Description: The variables are generated as usual using the information from the Youth Questionnaire or the Biography Questionnaire and the parents' direct-entries from PPFAD. As a next step the annual proxy information on a parent's death from the $P-files are used. Furthermore we use information of the month of death of a parent from the year before. That means we have information on the death of a father or a mother for the years 2002 onwards. With this data a wrong marking as "no death in 2002" / "death in 2003" can be corrected if there is data from 2003 indicating that one parent died e.g. in October 2002.

The variables VTODJ and MTODJ will be updated with new survey information. They are updated as long as the father or the mother is part of the SOEP sample. Since 2003 we additionally use the annual proxy information of respondents about reported life events of the last year.

Variable **VAORT91/ MAORT91 to VAORT11/ MAORT11**

Label: Residency of the father/ Residency of the mother in 1991, 1996, 2001, 2006 and 2011

Values:

(-1) No answer	(4) Lives Same Town
(-2) Does not apply	(5) Lives Other Town
(-3) Answer improbable	(6) Lives Elsewhere In Germany
(0) Has Died	(7) Lives Elsewhere
(1) Lives In Same HH	(8) Lives Else E Germany
(2) Lives In Same Housing	(9) Lives Else W Germany
(3) Lives Neighborhood	(10) Lives Foreign Country

Description: The information on the residency of the parents stems from the Youth and Biography Questionnaires as well as from the Person Questionnaire.

The information from the $P-files have a higher priority. For more details see also the information on VAORTAKT / MAORTAKT.

Variable **VAORTAKT/ MAORTAKT**

Label: Father‘s place of residence /
Mother‘s place of residence

Values:

(-1) No answer	(4) Lives Same Town
(-2) Does not apply	(5) Lives Other Town
(-3) Answer improbable	(6) Lives Elsewhere In Germany
(0) Has Died	(7) Lives Elsewhere
(1) Lives In Same HH	(8) Lives Elsewhere E Germany
(2) Lives In Same Housing	(9) Lives Elsewhere W Germany
(3) Lives Neighborhood	(10) Lives Foreign Country

Description: The variables VAORTAKT and MAORTAKT contain the latest available information about the parents’ residence and on whether or not they are deceased, respectively.

For persons without identified parents who answered the biography questionnaire up to the year 2011, the most recent available information from the Person Questionnaire in 1991, 1996, 2001, 2006 or 2011 was assumed.

For those persons whose parents are identified in the SOEP, the information on the year of death in PPFAD was used for updating. If the year of death lies chronologically after the latest available information, VAORTAKT and MAORTAKT were put on “deceased”.[53]

[53] In gathering the information from different data sets, inconsistencies occurred. On the one hand, some parents had been reported as deceased in the early waves, while information about their residence at a later date was available. In this case, the information about the parents’ residence was not accepted.

Variable **VAORTUP / MAORTUP**

Label: Year of update of VAORTAKT/MAORTAKT

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

Description: The variable contains the year, in which the information stored in VAORTAKT and MAORTAKT has been updated.

Variable **VSBIL / MSBIL**

Label: Education of the father / Education of the mother

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

(0) Do Not Know

(1) Secondary School Degree

(2) Intermediate School Degree

(3) Technical School Degree

(4) Upper Secondary School Degree

(5) Other Degree

(6) No School Degree

(7) School Not Attended

Description: The parents' education is generated with information from the Youth Questionnaire, the Biography Questionnaire and direct entries from the $PGEN-files. Due to the filter command, the children's proxy entries are only available for VSBIL / MSBIL when the parents or one parent and the child do not live in the same household at the time of the survey.

Along with other variables[54] already contained in BIOPAREN, there will be an update with new survey information, insofar as no valid values exist in BIOPAREN.

[54] These are: VBBIL, MBBIL, VSINFO, MSINFO, VBINFO, MBINFO.

Variable **VBBIL/ MBBIL**

Label: Vocational training of the father / Vocational training of the mother

Values:

(-1) No answer
(-2) Does not apply
(-3) Answer improbable
(0) Do Not Know
(10) No Vocational Degree
(20) Vocational Degree
(21) Trained in Foreign Company
(22) Trained long Time in Foreign Company
(23) Foreign Vocational School
(24) Trade, Farming Apprentice
(25) Business Apprentice
(26) Health Care School
(27) Special Technical School
(28) Civil Service Training
(30) Tech Engineer School
(31) Foreign Collage
(32) College, University
(40) Other Training
(50) Currently in Vocational Training
(51) Currently in Schooling

Description: The parents' vocational training is generated the same way as the education variables (see VSBIL / MSBIL).

Variable	**VSINFO/ MSINFO**
Label:	Origin of the information on father's education / Origin of the information on mother's education
Values:	(-1) No answer
	(-2) Does not apply
	(-3) Answer improbable
	(0) Do Not Know
	(1) Biography-Proxy
	(2) $P-Individual Info

Description: The variable contains the origin of the information on parental education.

Variable **VBINFO/ MBINFO**

Label: Origin of the information on father's vocational training /
Origin of the information on mother's vocational training

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

(0) Do Not Know

(1) Biography-Proxy

(2) $P-Individual Info

Description: The variable contains the origin of the information on parental vocational training.

Variable **VRELI/ MRELI**

Label: Religious affiliation of the father / Religious affiliation of the mother

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

(0) Do Not Know – Proxy

(1) Catholic

(2) Protestant

(3) Other Christian Denomination

(4) Islamic Denomination

(5) Other Denomination

(6) No Denomination

Description: The questions about the religious affiliation of the parents are only asked to youngsters who are not living in the household of their parents. In order to provide as much information as possible we gather data from the Person Questionnaires. The reconstruction of the information on the parents' religion is restricted to the survey years 1997, 2003 and 2007 as they are the only years where information of a respondent's religious affiliation is available. In all survey years the question was formulated differently but the information is made as consistent as possible.

Variable	**VNAT/ MNAT**
Label:	Nationality of the father / Nationality of the mother
Values:	(-1) No answer
	(-2) Does not apply
	(-3) Answer improbable
	(1) German
	(2) Other

Description: The information on the parents' nationality is generated similar to VRELI / MRELI. The question is only asked to youngsters who are not living in the same household as their parents. $PGEN information is used to compute a variable with data from 2006 onwards. In a further step the parents' personal numbers are used to match information on parents' nationality with data from the $PGEN-files in the case if there are missing entries.

Variable **VBSTELL/ MBSTELL**

Label: Professional position of the father (when the respondent was 15 years old) / Professional position of the mother (when the respondent was 15 years old)

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

Description: The children's proxy entries on professional position and occupation of the father (VBSTELL) as well as VISCO88 and all prestige scores are available when the father and the child do not live in the same household at the time of the survey and if the father lived in Germany when the child was 16 years old. Since 2000, the same applies to the entries of the mother.

Besides the proxy entries parents' direct information from the $P-files are used.

Variable **VBSINFO/ MBSINFO**

Label: Origin of the information on the professional position of the father / Origin of the information on the professional position of the mother

Values:

(-1) No answer

(-2) Does not apply

(-3) Answer improbable

(0) do Not Know-Proxy

(1) Biography-Proxy

(2) $P-Individual Info

Description: The variables VBSINFO / MBSINFO are indicator variables. They tell whether the information is from the Biography or Youth or Person Questionnaires. This information is generated at the same steps as it is done with the VBSTELL / MBSTELL variables.

Variable **VISCO88/ MISCO88**

Label: Professional occupation of the father (when the respondent was 15 years old) / Professional occupation of the mother (when the respondent was 15 years old)

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

Description: The variables contain the ISCO88 code for the father and mother.

Variable **VISEI/ MISEI**

Label: Prestige score of father – concept of Ganzeboom /
Prestige score of mother – concept of Ganzeboom

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

Description: The variables contain the ISEI code for the father and mother.

Variable **VMPS/ MMPS**

Label: Prestige score of father – Magnitude scale – Wegener /
Prestige score of mother – Magnitude scale – Wegener

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

Description: The variables contain the prestige scores (magnitude scale - Wegener) for the father and mother.

Variable **VSIOPS/ MSIOPS**

Label: Prestige score of father – Treiman standard score /
Prestige score of mother – Treiman standard score

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

Description: The variables contain the prestige scores (Treiman standard score) for the father and mother.

Variable **VEGP/ MEGP**

Label: Prestige score of father – Erikson – Goldthorpe class category/
Prestige score of mother – Erikson – Goldthorpe class category

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

Description: The variables contain the prestige scores (EGP) for the father and mother.

Variable **VBKLAS/ MBKLAS**

Label: Occupational coding scheme father according German statistical office/
Occupational coding scheme mother according German statistical office

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

Description: The variables contain the occupational code for the father and mother according to the coding scheme of the German statistical office.

Variable **ORTKINDH / ORTKIND1 / ORTKIND2 / ORTKIND3**

Label:		
	ORTKINDH	Place of childhood /
	ORTKIND1	Still lives in place of childhood? /
	ORTKIND2	Year moved out of parents' household (since 2000 no longer collected) /
	ORTKIND3	Still lives in parents' household (since 2000 no longer collected)

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

ORTKINDH:	**ORTKIND1:**
(1) Large City	(1) Yes, Still
(2) Medium City	(2) Yes, Again
(3) Small City	(3) No
(4) Countryside	

Description: The variables provide information on the place of childhood.

Variable **LIVING1 / LIVING2 / LIVING3 / LIVING4 / LIVING5 / LIVING6 / LIVING7 / LIVING8**

Label:		
	LIVING1	No. of years living with both parents
	LIVING2	No. of years living alone with mother
	LIVING3	No. of years living with mother and new partner of mother
	LIVING4	No. of years living alone with father
	LIVING5	No. of years living alone with father and new partner of father
	LIVING6	No. of years living with other relatives
	LIVING7	No. of years living with foster parents
	LIVING8	No. of years living in youth center

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

Description: The variables show the total number of years for different categories of where the child lived during his childhood.

Variable **VSTREIT/ MSTREIT**

Label: Conflict with father /
Conflict with mother

Values:
(-1) No answer

(-2) Does not apply

(-3) Answer improbable

(1) Very Often

(2) Often

(3) Sometimes

(4) Seldom

(5) Never

(6) Person Not Present

Description: The variables provide information on the frequency of conflicts with the parents.

Variable **BIOYEAR**

Label: Year of the Biography Survey

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

Description: The variable BIOYEAR tells in which year the information was surveyed.

Variable **BIO**

Label: Form Of Biography Questionnaire

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

(0) Participation before 2001

(1) Youth

(2) Biolela blue

Description: Since 2008 the variable BIO is generated to indicate the origin of the information in BIOPAREN (Biography or Youth Questionnaire). This information is valid for all cases from 2001 onwards.

Variable **ALTER** / V**ALTER / MALTER**

Label: Age of the respondents / Age of the respondent´s father / Age of the respondent`s mother

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

Description: Since 2008 the variables ALTER, VALTER and MALTER were added to BIOPAREN.

The variable ALTER gives the age of the respondent at the moment of the interview. VALTER gives the age of the respondents' father when the respondent answered the Biography Questionnaire or the Youth Questionnaire. The same was applied for the mothers with the variable MALTER. In order to generate the variables the information for the parents who are identified in the SOEP was gained with data from PPFAD. The proxy entries from BIOPAREN were used when there weren't any information of the respondents parents available. If the year of death lies chronologically before the latest available information, MALTER and VALTER were put on "deceased".

Variable V**ORIGIN/ MORIGIN**

Label: Country of origin of the respondent's father /
Country of origin of the respondent's mother

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

Description: These variables give information about the country of origin of the respondents mother (MORIGIN) and father (VORIGIN). This information is collected in the Youth Questionnaires since 2007. Another source of information can be found in PPFAD by the direct-entries of the parents in the variable CORIGIN. These two kinds of information, proxy- and direct-entries, are used to generate MORIGIN and VORIGIN. In a first step we use the proxy-information for all the parents whose children made an entry in the Youth Questionnaire. For all the parents where there are no proxy-information available, we then use the direct-entries of the parents from the PPFAD-variable CORIGIN.

Variable **GESCHW**

Label: Siblings yes/no

Values:
(-1) No answer

(-2) Does not apply

(-3) Answer improbable

(1) Yes

(2) No

Description: GESCHW contains the information if a respondent has siblings or not. The question is asked since 2003 in the Biography and Youth Questionnaire. In 2003 the question was asked in the Person Questionnaire.

For more Information about siblings see also BIOSIB, BIOTWIN, and the data in the $p-Files from the family focus questions in the Person Questionnaire (1991, 1996, 2001, 2006 and 2011).

Variable **GESCHWUP**

Label: Time of update - siblings

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

Description: GESCHWUP contains the year, in which the sibling information was surveyed. This is either 2003 (if the information comes from the Person Questionnaire) or later (if the information comes from the Biography or Youth Questionnaire).

Variable **NUMS**

Label: Number of sisters

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

Description: Nums contains the number of sisters. The question is asked since 2003 in the Biography and Youth Questionnaire. In 2003 the question was asked in the Person Questionnaire.

Variable **NUMB**

Label: Number of brothers

Values: (-1) No answer

(-2) Does not apply

(-3) Answer improbable

Description: Nums contains the number of brothers. The question is asked since 2003 in the biography and youth questionnaire. In 2003 the question was asked in the Person Questionnaire.

Variable **TWIN**

Label: Twin sister/brother

Values:
(-1) No answer
(-2) Does not apply
(-3) Answer improbable
(1) Yes, monozygotic
(2) Yes, dizygotic
(3) No

Description: Twin contains information whether the respondent has a twin sibling. The question is asked since 2007 in the Biography and Youth Questionnaire.

14 BIOIMMIG: Generated and Status Variables from SOEP for Foreigners and Migrants

by Jan Goebel and Anke Böckenhoff

14.1 Content

The variables contained in BIOIMMIG deal with questions related to foreigners in (and migrants to) Germany. Specifically, questions concerning desire to return to the home country, the presence of relatives in the home country, reasons for coming to Germany, and conditions upon initial arrival in Germany. A complete list of variables is shown in the table with German and English labels.

14.2 Status Variables and Carrying Forth of Information

The data available in this file are longitudinal, that is to say, the same variable name refers to different time periods, differentiated by the variable ERHEBJ. The data is stacked for each person, such that the unit of observation is a person-year. Thus for every person, there are as many observations as interviews given by this person. Much of the information was asked only once, and „carried“ forth in the following years. Frequencies can be found in SOEPINFO.

The sample in the dataset is defined by taking all available information and deleting all those persons who:

are born in Germany *and*
have German nationality *and*
have no valid BIOIMMIG information in any wave that they were observed.

As the data consists of person-year observations, if a person is excluded from the sample, then for all years. However if a person once belonged to the sample, then he is always included (say, even after receiving German citizenship).

List of Variables

Variable	German	*English*
persnr	Personennummer	Person Number
hhnr	Ursprüngliche HH-Nummer	Original HH Number
hhnrakt	Aktuelle HH-Nummer für ERHEBJ	Current HH Number for ERHEBJ
erhebj	Jahr/Erhebungsjahr	Current Year / Year Answered
biimgrp	BI: Status bei Einwanderung in Dt.	BI: Immigration Group
Biresper	BI: Status Aufenthaltserlaubnis	BI: Residence Status
bicamp	BI: Aufnahmelager: J/N	BI: Refugee Residence Y/N
bicampw	BI: Aufnahmelager: Wochen	BI: Refugee Residence: Weeks
bicampm	BI: Aufnahmelager: Monate	BI: Refugee Residence: Months
biwfam	BI: Eingereist als Familienangehöriger	BI: Already had Family in Country
bifamc	BI: Vor Einreise Kontakte mit Pers.	BI: Contacts with Family in Germany
Bifamcl	BI: Zuzug in Wohnort der Bekannten	BI: Moved to Same City/Town as Family
birbetr	BI: Gründe Zuzug D: Besser	BI: Reason Migrate: Better
birmoney	BI: Gründe Zuzug D: Geld	BI: Reason Migrate: Money
birfree	BI: Gründe Zuzug D: Freiheit	BI: Reason Migrate: Freedom
birfam	BI: Gründe Zuzug D: Familie	BI: Reason Migrate: Family
birpoor	BI: Gründe Zuzug D: Armut	BI: Reason Migrate: Poor
birwar	BI: Gründe Zuzug D: Krieg	BI: Reason Migrate: War
birjust	BI: Gründe Zuzug D: Einfach So	BI: Reason Migrate: Just So
birothr	BI: Gründe Zuzug D: Sonstiges	BI: Reason Migrate: Other
biexpr	BI: Vorstellungen von Dt.	BI: Expectations in Germany
Biexprlv	BI: Eigene Wohnung finden	BI: Expectations: Find Apt
biexprac	BI: Von Arbeitskollegen akzeptiert	BI: Expectations: Accepted by Coworker
biexpran	BI: Von Nachbarn akzeptiert	BI: Expectations: Accepted by Neighbor
birelh	BI: In Heimatland Familienmitglieder	BI: Family Abroad
birelhp	BI: In Heimat: Eltern	BI: Family Abroad: Parents
birelhgp	BI: In Heimat: Großeltern	BI: Family Abroad: Grandparents
birelhc	BI: In Heimat: Kinder	BI: Family Abroad: Children
birelhbs	BI: In Heimat: Bruder, Schwester	BI: Family Abroad: Brother/Sister
birelhdr	BI: In Heimat: Entferntere Verwandte	BI: Family Abroad: Distant Relatives
birelhsp	BI: In Heimat: Ehepartner, Verlobte(r)	BI: Family Abroad: Spouse
birelhfr	BI: In Heimat: Bekannte/Freunde	BI: Family Abroad: Friends
birelhmi	BI: Personen gern nach Dt. holen?	BI: Persons abroad bring to Germany
birelhs2	BI: Ehepartner in Deutschland	BI: Spouse in Germany
birelhc2	BI: Kinder unter 18 J. nicht in D	BI: Underage Children not in Germany
Bigoback	BI: Rückkehr Heimat (ab 1994)	BI: Go back home ?
biSTAY	BI: Wunsch in D zu bleiben	BI: Desire to Stay in Germany
biSTAYy	BI: Dauer des geplanten Aufenthalts	BI: Years Desired to Stay in Germany
Biscger	BI: In Dt. Schule besucht?	BI: Attended School in Germany
Biscgrad	BI: In welche Klasse in dt. Schule	BI: Which Grade School
biscgerc	BI: Besuch spezieller Vorbereitung	BI: Attended Special Foreigner Prep Class
biscgc	BI: Auch dt. Schüler in Schulklasse	BI: Also German Pupils in Class
biscgcf	BI: Wie viel Mitschüler Ausländer	BI: How many Pupils foreign
biscgcfn	BI: Eine oder mehrere Nationalität	BI: Mix of Nationalities in Class

14.3 Updating of Time-Dependent Information

The variables found in BIOIMMIG are created first using information from the SOEP biography files, the so-called BIOLELA, $LELA (starting with wave M) files. Additionally, starting in 2000 (wave Q), $JUGEND is collected of 16 and 17 year-olds, containing similar information to $LELA. In any given year, a person can have only information from $JUGEND or $LELA, but not both. If valid information is found in the $LELA or $JUGEND files for the given response year, then it is taken. Yearly valid update information is taken from the foreigner specific files APAUSL through $PAUSL and the foreigner specific questions in MP, NP, OP and onwards. Starting with wave M, the foreigner specific variables are found in the regular $P files, as the questionnaire is identical for natives and foreigners. Sometimes there is competing information in the biography and regular yearly person questionnaires. The most recent valid information is taken to be correct. First the $LELA or $JUGEND info is used and then updated with valid/non-missing information from the person questionnaire.

14.4 Using this File

The BIOIMMIG file can be used in cross-section or in panel. The usual matching variables are included.: PERSNR (Person Number), HHNR (Original HH Number), HHNRAKT (Current HH Number for survey year given in ERHEBJ), ERHEBJ (Year). The data is sorted by HHNR, HHNRAKT, PERSNR, ERHEBJ such that there are typically many person-year observations for every person. In that sense, the data are ready to be used/matched to a longitudinal dataset. However, simply by selecting on the appropriate year in ERHEBJ, the file can be used cross-sectionally as well.

The data structure looks like the following (using fictitious data in this example):

PERSNR	HHNR	HHNRAKT	ERHBEBJ	BIIMGRP	BIRESPER
101	19	19	1995	-2	-2
101	19	19	1996	2	1
101	19	19	1997	2	1
101	19	19	1998	2	1
102	19	19	1995	3	2
102	19	19	1996	3	2
102	19	19	1997	3	2

14.4.1 Using BIOIMMIG as a Cross-Section

An example of how to use BIOIMMIG in a cross-section would be as follows:

(A) Open BIOIMMIG, keeping only those observations in BIOIMMIG for a particular year.

in Stata:
```
use bioimmig if erhebj==1984
```

(B) Rename all the desired variables with wave-specific information.

in Stata:
```
rename bicamp camp1984
rename bicampw campw1984
```

(C) Save the ID's and the renamed variables in a temporary file

in Stata:
```
sort hhnr persnr
save /tmp/bioim1984, replace
```

(D) Merge the temporary file to your main dataset

in Stata:
```
merge hhnr persnr using /tmp/bioim1984, nokeep
drop _merge
```

(E) Repeat starting at step (A) for all years of interest, i.e. `erhebj==1985`

14.5 Documentation of the Variables

Below, each variable is listed and its variable and value labels are displayed in both English and German. A list of the main source variables used in the generation is provided for reference purposes. Further, there is also information as to what question the variables correspond to in the Wave 13 -M-1996 Biography Questionnaire.

Problems:

If you encounter problems using this file, first-aid is available from the original STATA source code used to create this file, delivered with the regular SOEP data distribution.

BIIMGRP BI: Status bei Einwanderung in Dt.
BI: Immigration Group

BIO Question: Q5

Comment: The possible groups change in 2000, such that "[1] East German" and "[5] Non EU "are no longer identified starting 2000. However, as information can be carried forth from previous years, there may be valid [1] and [3] values starting 2000, but only if the information was collected before 2000.

German: ***Zu welcher der folgenden Zuwanderergruppen gehörten Sie, als Sie nach Deutschland kamen ?***
"[1] Ostdeutsche (LT 2000) "
"[2] Aussiedler "
"[3] Deutscher, Ausland lebt "
"[4] EG-Mitglied "
"[5] Nicht EG (LT 2000)"
"[6] Asylbewerber "
"[7] Sonstige "

English: ***Which immigrant group did you belong to, when you came to Germany ?***
"[1] East German (LT 2000)"
"[2] Ethnic German living in East Europe "
"[3] German living abroad "
"[4] EU Member "
"[5] Non EU (LT 2000)"
"[6] Asylum Seeker "
"[7] Other "

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P070Z**
1995 L	BIOLELA	**P070Z**
1996 M	MLELA	**MB070Z**
1997 N	NLELA	**NB070Z**
1998 O	OLELA	**OB070Z**
1999 P	PLELA	**PB070Z**
2000 Q	QLELA	**QB070Z**
2001 R	RLELA	**RB070Z**
2002 S	SLELA	**SB05**
2003 T	TLELA	**TB05**
2004 U	ULELA	**UB05**
2005 V	VLELA	**VB05**
2006 W	WLELA	**WB05**
2007 X	XLELA	**XB05**
2008 Y	YLELA	**YB05**
2009 Z	ZLELA	**ZB05**
2010 BA	BALELA	**BAB05**
2011 BB	BBLELA	BBB05
2000 Q	QJUGEND	**QJ57**
2001 R	RJUGEND	**RJ59**
2002 S	SJUGEND	**SJ59**
2003 T	TJUGEND	**TJ59**
2004 U	UJUGEND	**UJ59**
2005 V	VJUGEND	**VJ59**
2006 W	WJUGEND	**WJ64**
2007 X	XJUGEND	**XJ64**
2008 Y	YJUGEND	**YJ64**
2009 Z	ZJUGEND	**ZJ64**
2010 BA	BAJUGEND	**BAJ64**
2011 BB	BBJUGEND	BBJ64

BIRESPER BI: Status Aufenthaltserlaubnis
BI: Residence Status

BIO Question: **Q6**

Comment: The possible groups change in 2000 in QLELA and QJUGEND, such that "[3] German Citizen" is included in the original question. German citizens for the purpose of this question have been recoded to -2 (does not apply). German citizenship is however recorded in NATION$$ in $PGEN as usual.

German: ***Haben Sie heute eine unbefristete Aufenthaltserlaubnis bzw. Aufenthaltsberechtigung oder haben Sie eine befristete Aufenthaltserlaubnis?***
"[1] Unbefristet "
"[2] Befristet "

English: ***Do you right now have a permanent or temporary residence permit ?***
"[1] Permanent "
"[2] Limited "

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P080Z**
1995 L	BIOLELA	**P080Z**
1996 M	MLELA	**MB080Z**
1997 N	NLELA	**NB080Z**
1998 O	OLELA	**OB080Z**
1999 P	PLELA	**PB080Z**
2000 Q	QLELA	**QB080Z**
2001 R	RLELA	**RB080Z**
2002 S	SLELA	**SB06**
2003 T	TLELA	**TB06**
2004 U	ULELA	**UB06**
2005 V	VLELA	**VB06**
2006 W	WLELA	**WB06**
2007 X	XLELA	**XB06**
2008 Y	YLELA	**YB06**
2009 Z	ZLELA	**ZB06**
2010 BA	BALELA	**BAB06**
2011 BB	BBLELA	BBB06
2000 Q	QJUGEND	**QJ58**
2001 R	RJUGEND	**RJ60**
2002 S	SJUGEND	**SJ60**
2003 T	TJUGEND	**TJ60**
2004 U	UJUGEND	**UJ60**
2005 V	VJUGEND	**VJ60**
2006 W	WJUGEND	**WJ69**
2007 X	XJUGEND	**XJ69**
2008 Y	YJUGEND	**YJ69**
2009 Z	ZJUGEND	**ZJ69**
2010 BA	BAJUGEND	**BAJ69**
2011 BB	BBJUGEND	BBJ69

BICAMP BI: Aufnahmelager: J/N
BI: Refugee Residence Y/N

BIO Question: **Q7a**

German: ***Haben Sie nach Ihrer Einreise zunächst in einem Aufnahmelager oder Übergangswohnheim gelebt ?***
"[1] Ja "
"[2] Nein "

English: ***After you arrived in Germany, did you live in temporary refugee/immigrant housing or residence ?***
"[1] Yes "
"[2] No "

See also: **BICAMP, BICAMPW, BICAMPM**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P090Z**
1995 L	BIOLELA	**P090Z**
1996 M	MLELA	**MB090Z**
1997 N	NLELA	**NB090Z**
1998 O	OLELA	**OB090Z**
1999 P	PLELA	**PB090Z**
2000 Q	QLELA	**QB090Z**
2001 R	RLELA	**RB090Z**
2002 S	SLELA	**SB0701**
2003 T	TLELA	**TB0701**
2004 U	ULELA	**UB0701**
2005 V	VLELA	**VB0701**
2006 W	WLELA	**WB0701**
2007 X	XLELA	**XB0701**
2008 Y	YLELA	**YB0701**
2009 Z	ZLELA	**ZB0701**
2010 BA	BALELA	**BAB0701**
2011 BB	BBLELA	BBB0701
2000 Q	QJUGEND	**QJ5901**
2001 Q	RJUGEND	**RJ6101**
2002 S	SJUGEND	**SJ6101**
2003 T	TJUGEND	**TJ6101**
2004 U	UJUGEND	**UJ6101**
2005 V	VJUGEND	**VJ6101**
2006 W	WJUGEND	**n/a**
2007 X	XJUGEND	**n/a**
2008 Y	YJUGEND	**n/a**
2009 Z	ZJUGEND	**n/a**
2010 BA	BAJUGEND	**n/a**
2011 BB	BBJUGEND	n/a

BICAMPW BI: Aufnahmelager: Wochen
BI: Refugee Residence: Weeks

BIO Question: **Q7b**

German: ***Aufnahmelager: Wenn Ja, für wie lange (Wochen) ?***

English: ***Immigrant Residence: If so, then for how long (weeks)?***

See also: **BICAMP, BICAMPW, BICAMPM**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P091Z**
1995 L	BIOLELA	**P091Z**
1996 M	MLELA	**MB091Z**
1997 N	NLELA	**NB091Z**
1998 O	OLELA	**OB091Z**
1999 P	PLELA	**PB091Z**
2000 Q	QLELA	**QB091Z**
2001 R	RLELA	**RB091Z**
2002 S	SLELA	**SB0702**
2003 T	TLELA	**TB0702**
2004 U	ULELA	**UB0702**
2005 V	VLELA	**VB0702**
2006 W	WLELA	**WB0702**
2007 X	XLELA	**XB0702**
2008 Y	YLELA	**YB0702**
2009 Z	ZLELA	**ZB0702**
2010 BA	BALELA	**BAB0702**
2011 BB	BBLELA	BBB0702
2000 Q	QJUGEND	**QJ5902**
2001 R	RJUGEND	**RJ6102**
2002 S	SJUGEND	**SJ6102**
2003 T	TJUGEND	**TJ6102**
2004 U	UJUGEND	**UJ6102**
2005 V	VJUGEND	**VJ6102**
2006 W	WJUGEND	**n/a**
2007 X	XJUGEND	**n/a**
2008 Y	YJUGEND	**n/a**
2009 Z	ZJUGEND	**n/a**
2010 BA	BAJUGEND	**n/a**
2011 BB	BBJUGEND	n/a

BICAMPM BI: Aufnahmelager: Monate
BI: Refugee Residence: Months

BIO Question: **Q7c**

German: ***Aufnahmelager: Wenn Ja, für wie lange (Monate) ?***

English: ***Immigrant Residence: If so, then for how long (months)?***

See also: **BICAMP, BICAMPW, BICAMPM**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P092Z**
1995 L	BIOLELA	**P092Z**
1996 M	MLELA	**MB092Z**
1997 N	NLELA	**NB092Z**
1998 O	OLELA	**OB092Z**
1999 P	PLELA	**PB092Z**
2000 Q	QLELA	**QB092Z**
2001 R	RLELA	**RB092Z**
2002 S	SLELA	**SB0703**
2003 T	TLELA	**TB0703**
2004 U	ULELA	**UB0703**
2005 V	VLELA	**VB0703**
2006 W	WLELA	**WB0703**
2007 X	XLELA	**XB0703**
2008 Y	YLELA	**YB0703**
2009 Z	ZLELA	**ZB0703**
2010 BA	BALELA	**BAB0703**
2011 BB	BBLELA	BBB0703
2000 Q	QJUGEND	QJ5903
2001 R	RJUGEND	RJ6103
2002 S	SJUGEND	SJ6103
2003 T	TJUGEND	TJ6103
2004 U	UJUGEND	UJ6103
2005 V	VJUGEND	VJ6103
2006 W	WJUGEND	n/a
2007 X	XJUGEND	n/a
2008 Y	YJUGEND	n/a
2009 Z	ZJUGEND	n/a
2010 BA	BAJUGEND	n/a
2011 BB	BBJUGEND	n/a

BIWFAM BI: Eingereist als Familienangehoeriger
BI: Already had Family in Country

BIO Question: **Q8**

German: ***Als Sie einreisten, kamen Sie da als Familienangehöriger einer bereits in Deutschland lebenden Familie bzw. Person ?***
"[1] Ja "
"[2] Nein "

English: ***When you immigrated to Germany, was (at least one) a member of your family already living in Germany ?***
"[1] Yes "
"[2] No "

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P100Z**
1995 L	BIOLELA	**P100Z**
1996 M	MLELA	**MB100Z**
1997 N	NLELA	**NB100Z**
1998 O	OLELA	**OB100Z**
1999 P	PLELA	**PB100Z**
2000 Q	QLELA	**QB100Z**
2001 R	RLELA	**RB100Z**
2002 S	SLELA	**SB08**
2003 T	TLELA	**TB08**
2004 U	ULELA	**UB08**
2005 V	VLELA	**VB08**
2006 W	WLELA	**WB08**
2007 X	XLELA	**XB08**
2008 Y	YLELA	**YB08**
2009 Z	ZLELA	**ZB08**
2010 BA	BALELA	**BAB08**
2011 BB	BBLELA	BBB08
2000 Q	QJUGEND	**QJ60**
2001 R	RJUGEND	**RJ62**
2002 S	SJUGEND	**SJ62**
2003 T	TJUGEND	**TJ62**
2004 U	UJUGEND	**UJ62**
2005 V	VJUGEND	**VJ62**
2006 W	WJUGEND	**n/a**
2007 X	XJUGEND	**n/a**
2008 Y	YJUGEND	**n/a**
2009 Z	ZJUGEND	**n/a**
2010 BA	BAJUGEND	**n/a**
2011 BB	BBJUGEND	n/a

BIFAMC BI: Vor Einreise Kontakte mit Pers.
BI: Contacts with Family in Germany

BIO Question: **Q9**

German: ***Hatten Sie vor der Einreise überhaupt Kontakte zu Verwandten oder Bekannte in Deutschland, an die Sie sich wenden konnten ?***
"[1] Ja "
"[2] Nein "

English: ***Before immigrating to Germany, did you have any contact with relatives or friends, who could possibly help you ?***
"[1] Yes "
"[2] No "

See also: **BIFAMC, BIFAMCL**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P110Z**
1995 L	BIOLELA	**P110Z**
1996 M	MLELA	**MB110Z**
1997 N	NLELA	**NB110Z**
1998 O	OLELA	**OB110Z**
1999 P	PLELA	**PB110Z**
2000 Q	QLELA	**QB110Z**
2001 R	RLELA	**RB110Z**
2002 S	SLELA	**SB09**
2003 T	TLELA	**TB09**
2004 U	ULELA	**UB09**
2005 V	VLELA	**VB09**
2006 W	WLELA	**WB09**
2007 X	XLELA	**XB09**
2008 Y	YLELA	**YB09**
2009 Z	ZLELA	**ZB09**
2010 BA	BALELA	**BAB09**
2011 BB	BBLELA	BBB09
2000 Q	QJUGEND	**QJ61**
2001 R	RJUGEND	**RJ63**
2002 S	SJUGEND	**SJ63**
2003 T	TJUGEND	**TJ63**
2004 U	UJUGEND	**UJ63**
2005 V	VJUGEND	**VJ63**
2006 W	WJUGEND	**n/a**
2007 X	XJUGEND	**n/a**
2008 Y	YJUGEND	**n/a**
2009 Z	ZJUGEND	**n/a**
2010 BA	BAJUGEND	**n/a**
2011 BB	BBJUGEND	n/a

BIFAMCL BI: Zuzug in Wohnort der Bekannten
BI: Moved to Same City/Town as Family

BIO Question: **Q10**

German: ***Sind Sie in den Ort in Deutschland gezogen, wo diese Verwandten bzw. Bekannten lebten ?***
"[1] Ja "
"[2] Nein "

English: ***Did you move to the same town/city in Germany where these relatives or friends lived ?***
"[1] Yes "
"[2] No "

See also: **BIFAMC, BIFAMCL**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P120Z**
1995 L	BIOLELA	**P120Z**
1996 M	MLELA	**MB120Z**
1997 N	NLELA	**NB120Z**
1998 O	OLELA	**OB120Z**
1999 P	PLELA	**PB120Z**
2000 Q	QLELA	**QB120Z**
2001 R	RLELA	**RB120Z**
2002 S	SLELA	**SB10**
2003 T	TLELA	**TB10**
2004 U	ULELA	**UB10**
2005 V	VLELA	**VB10**
2006 W	WLELA	**WB10**
2007 X	XLELA	**XB10**
2008 Y	YLELA	**YB10**
2009 Z	ZLELA	**ZB10**
2010 BA	BALELA	**BAB10**
2011 BB	BBLELA	BBB10
2000 Q	QJUGEND	**QJ62**
2001 R	RJUGEND	**RJ64**
2002 S	SJUGEND	**SJ64**
2003 T	TJUGEND	**TJ64**
2004 U	UJUGEND	**UJ64**
2005 V	VJUGEND	**VJ64**
2006 W	WJUGEND	**n/a**
2007 X	XJUGEND	**n/a**
2008 Y	YJUGEND	**n/a**
2009 Z	ZJUGEND	**n/a**
2010 BA	BAJUGEND	**n/a**
2011 BB	BBJUGEND	n/a

BIRBETR BI: Gruende Zuzug D: Besser
BI: Reason Migrate: Better

BIO Question: **Q11a**

Comment: This variable is not defined for youths answering the $JUGEND biography questionnaire.

German: ***Es gibt ja unterschiedliche Gründe, nach Deutschland zu ziehen. Welche der folgenden Gründe spielten bei Ihnen eine Rolle ? -- Ich wollte ein besseres Leben haben: Besser wohnen, mehr kaufen können usw.***
"[1] Besseres Leben "

English: ***There are many reasons to migrate to Germany. Did the following reason playa role ? -- I wanted a better life. Live better, to be able to buy more etc..***
"[1] Better Life "

See also: **BIRBETR, BIRMONEY, BIRFREE, BIRFAM, BIRPOOR, BIRWAR, BIRJUST, BIROTHR**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P151Z**
1995 L	BIOLELA	**P151Z**
1996 M	MLELA	**MB151Z**
1997 N	NLELA	**NB151Z**
1998 O	OLELA	**OB151Z**
1999 P	PLELA	**PB151Z**
2000 Q	QLELA	**QB151Z**
2001 R	RLELA	**RB151Z**
2002 S	SLELA	**SB1401**
2003 T	TLELA	**TB1401**
2004 U	ULELA	**UB1401**
2005 V	VLELA	**VB1401**
2006 W	WLELA	**WB1401**
2007 X	XLELA	**XB1401**
2008 Y	YLELA	**YB1401**
2009 Z	ZLELA	**ZB1401**
2010 BA	BALELA	**BAB1401**
2011 BB	BBLELA	BBB1401
2000 --	$JUGEND	**n/a**

BIRMONEY BI: Gruende Zuzug D: Geld
BI: Reason Migrate: Money

BIO Question: **Q11b**

Comment: This variable is not defined for youths answering the $JUGEND biography questionnaire.

German: ***Es gibt ja unterschiedliche Gründe, nach Deutschland zu ziehen. Welche der folgenden Gründe spielten bei Ihnen eine Rolle ? --Ich wollte arbeiten und Geld verdienen in Deutschland, um meine Familie zu unterstützen und Geld sparen.***
"[1] Geld verdienen "

English: ***There are many reasons to migrate to Germany. Did the following reason play a role ? -- I wanted to work and earn money to support my family and save money.***
"[1] Earn money "

See also: **BIRBETR, BIRMONEY, BIRFREE, BIRFAM, BIRPOOR, BIRWAR, BIRJUST, BIROTHR**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P152Z**
1995 L	BIOLELA	**P152Z**
1996 M	MLELA	**MB152Z**
1997 N	NLELA	**NB152Z**
1998 O	OLELA	**OB152Z**
1999 P	PLELA	**PB152Z**
2000 Q	QLELA	**QB152Z**
2001 R	RLELA	**RB152Z**
2002 S	SLELA	**SB1402**
2003 T	TLELA	**TB1402**
2004 U	ULELA	**UB1402**
2005 V	VLELA	**VB1402**
2006 W	WLELA	**WB1402**
2007 X	XLELA	**XB1402**
2008 Y	YLELA	**YB1402**
2009 Z	ZLELA	**ZB1402**
2010 BA	BALELA	**BAB1402**
2011BB	BBLELA	BBB1402
2000 --	$JUGEND	**n/a**

BIRFREE BI: Gruende Zuzug D: Freiheit
BI: Reason Migrate: Freedom

BIO Question: **Q11c**

Comment: This variable is not defined for youths answering the $JUGEND biography questionnaire.

German: ***Es gibt ja unterschiedliche Gründe, nach Deutschland zu ziehen. Welche der folgenden Gründe spielten bei Ihnen eine Rolle ? -- Ich wollte in der Freiheit leben.***
"[1] In Freiheit leben "

English: ***There are many reasons to migrate to Germany. Did the following reason play a role ? -- I wanted to live in freedom.***
"[1] Live in freedom "

See also: **BIRBETR, BIRMONEY, BIRFREE, BIRFAM, BIRPOOR, BIRWAR, BIRJUST, BIROTHR**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P153Z**
1995 L	BIOLELA	**P153Z**
1996 M	MLELA	**MB153Z**
1997 N	NLELA	**NB153Z**
1998 O	OLELA	**OB153Z**
1999 P	PLELA	**PB153Z**
2000 Q	QLELA	**QB153Z**
2001 R	RLELA	**RB153Z**
2002 S	SLELA	**SB1403**
2003 T	TLELA	**TB1403**
2004 U	ULELA	**UB1403**
2005 V	VLELA	**VB1403**
2006 W	WLELA	**WB1403**
2007 X	XLELA	**XB1403**
2008 Y	YLELA	**YB1403**
2009 Z	ZLELA	**ZB1403**
2010 BA	BALELA	**BAB1403**
2011 BB	BBLELA	BBB1403
2000 --	$JUGEND	**n/a**

BIRFAM BI: Gruende Zuzug D: Familie
BI: Reason Migrate: Family

BIO Question: **Q11d**

Comment: This variable is not defined for youths answering the $JUGEND biography questionnaire.

German: ***Es gibt ja unterschiedliche Gründe, nach Deutschland zu ziehen. Welche der folgenden Gründe spielten bei Ihnen eine Rolle ? -- Ich wollte mit meiner Familie zusammenleben (Ehepartner, Eltern, Kinder).***
"[1] Mit Familie zusammen "

English: ***There are many reasons to migrate to Germany. Did the following reason play a role ? -- I wanted to be together with my family (spouse, parents, children).***
"[1] Live together with family "

See also: **BIRBETR, BIRMONEY, BIRFREE, BIRFAM, BIRPOOR, BIRWAR, BIRJUST, BIROTHR**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P154Z**
1995 L	BIOLELA	**P154Z**
1996 M	MLELA	**MB154Z**
1997 N	NLELA	**NB154Z**
1998 O	OLELA	**OB154Z**
1999 P	PLELA	**PB154Z**
2000 Q	QLELA	**QB154Z**
2001 R	RLELA	**RB154Z**
2002 S	SLELA	**SB1404**
2003 T	TLELA	**TB1404**
2004 U	ULELA	**UB1404**
2005 V	VLELA	**VB1404**
2006 W	WLELA	**WB1404**
2007 X	XLELA	**XB1404**
2008 Y	YLELA	**YB1404**
2009 Z	ZLELA	**ZB1404**
2010 BA	BALELA	**BAB1404**
2011 BB	BBLELA	BBB1404
2000 --	$JUGEND	**n/a**

BIRPOOR BI: Gruende Zuzug D: Armut
BI: Reason Migrate: Poor

BIO Question: **Q11e**

Comment: This variable is not defined for youths answering the $JUGEND biography questionnaire.

German: ***Es gibt ja unterschiedliche Gründe, nach Deutschland zu ziehen. Welche der folgenden Gründe spielten bei Ihnen eine Rolle ? -- In meinem Heimatland herrschte Not und Armut.***
"[1] Not/Armut in Heimat "

English: ***There are many reasons to migrate to Germany. Did the following reason play a role ? -- In my native country there was poverty and hunger.***
"[1] Poverty/Hunger at home "

See also: **BIRBETR, BIRMONEY, BIRFREE, BIRFAM, BIRPOOR, BIRWAR, BIRJUST, BIROTHR**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P155Z**
1995 L	BIOLELA	**P155Z**
1996 M	MLELA	**MB155Z**
1997 N	NLELA	**NB155Z**
1998 O	OLELA	**OB155Z**
1999 P	PLELA	**PB155Z**
2000 Q	QLELA	**QB155Z**
2001 R	RLELA	**RB155Z**
2002 S	SLELA	**SB1405**
2003 T	TLELA	**TB1405**
2004 U	ULELA	**UB1405**
2005 V	VLELA	**VB1405**
2006 W	WLELA	**WB1405**
2007 X	XLELA	**XB1405**
2008 Y	YLELA	**YB1405**
2009 Z	ZLELA	**ZB1405**
2010 BA	BALELA	**BAB1405**
2011 BB	BBLELA	BBB1405
2000 --	$JUGEND	**n/a**

BIRWAR BI: Gruende Zuzug D: Krieg
BI: Reason Migrate: War

BIO Question: **Q11f**

Comment: This variable is not defined for youths answering the $JUGEND biography questionnaire.

German: ***Es gibt ja unterschiedliche Gründe, nach Deutschland zu ziehen. Welche der folgenden Gründe spielten bei Ihnen eine Rolle ? -- In meinem Heimatland konnte ich nicht in Sicherheit leben (Verfolgung, Krieg)***
"[1] Krieg in Heimat "

English: ***There are many reasons to migrate to Germany. Did the following reason play a role ? -- In my native country I could not live safely (Oppression, War).***
"[1] War/Oppression at home "

See also: **BIRBETR, BIRMONEY, BIRFREE, BIRFAM, BIRPOOR, BIRWAR, BIRJUST, BIROTHR**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P156Z**
1995 L	BIOLELA	**P156Z**
1996 M	MLELA	**MB156Z**
1997 N	NLELA	**NB156Z**
1998 O	OLELA	**OB156Z**
1999 P	PLELA	**PB156Z**
2000 Q	QLELA	**QB156Z**
2001 R	RLELA	**RB156Z**
2002 S	SLELA	**SB1406**
2003 T	TLELA	**TB1406**
2004 U	ULELA	**UB1406**
2005 V	VLELA	**VB1406**
2006 W	WLELA	**WB1406**
2007 X	XLELA	**XB1406**
2008 Y	YLELA	**YB1406**
2009 Z	ZLELA	**ZB1406**
2010 BA	BALELA	**BAB1406**
2011 BB	BBLELA	BBB1406
2000 --	$JUGEND	**n/a**

BIRJUST BI: Gruende Zuzug D: Einfach So
BI: Reason Migrate: Just So

BIO Question: **Q11g**

Comment: This variable is not defined for youths answering the $JUGEND biography questionnaire.

German: ***Es gibt ja unterschiedliche Gründe, nach Deutschland zu ziehen. Welche der folgenden Gründe spielten bei Ihnen eine Rolle ? -- Ich wollte einfach in Deutschland leben.***
"[1] Einfach in D leben "

English: ***There are many reasons to migrate to Germany. Did the following reason play a role ? -- I just wanted to live in Germany.***
"[1] Just wanted to live in Germany "

See also: **BIRBETR, BIRMONEY, BIRFREE, BIRFAM, BIRPOOR, BIRWAR, BIRJUST, BIROTHR**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P157Z**
1995 L	BIOLELA	**P157Z**
1996 M	MLELA	**MB157Z**
1997 N	NLELA	**NB157Z**
1998 O	OLELA	**OB157Z**
1999 P	PLELA	**PB157Z**
2000 Q	QLELA	**QB157Z**
2001 R	RLELA	**RB157Z**
2002 S	SLELA	**SB1407**
2003 T	TLELA	**TB1407**
2004 U	ULELA	**UB1407**
2005 V	VLELA	**VB1407**
2006 W	WLELA	**WB1407**
2007 X	XLELA	**XB1407**
2008 Y	YLELA	**YB1407**
2009 Z	ZLELA	**ZB1407**
2010 BA	BALELA	**BAB1407**
2011 BB	BBLELA	BBB1407
2000 --	$JUGEND	**n/a**

BIROTHR BI: Gruende Zuzug D: Sonstiges
BI: Reason Migrate: Other

BIO Question: **Q11h**

Comment: This variable is not defined for youths answering the $JUGEND biography questionnaire.

German: ***Es gibt ja unterschiedliche Gründe, nach Deutschland zu ziehen. Welche der folgenden Gründe spielten bei Ihnen eine Rolle ? -- Sonstige Gründe.***
"[1] Sonstige Gruende "

English: ***There are many reasons to migrate to Germany. Did the following reason play a role ? -- Other reasons.***
"[1] Other reasons "

See also: **BIRBETR, BIRMONEY, BIRFREE, BIRFAM, BIRPOOR, BIRWAR, BIRJUST, BIROTHR**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P158Z**
1995 L	BIOLELA	**P158Z**
1996 M	MLELA	**MB158Z**
1997 N	NLELA	**NB158Z**
1998 O	OLELA	**OB158Z**
1999 P	PLELA	**PB158Z**
2000 Q	QLELA	**QB158Z**
2001 R	RLELA	**RB158Z**
2002 S	SLELA	**SB1408**
2003 T	TLELA	**TB1408**
2004 U	ULELA	**UB1408**
2005 V	VLELA	**VB1408**
2006 W	WLELA	**WB1408**
2007 X	XLELA	**XB1408**
2008 Y	YLELA	**YB1408**
2009 Z	ZLELA	**ZB1408**
2010 BA	BALELA	**BAB1408**
2011 BB	BBLELA	BBB1408
2000 --	$JUGEND	**n/a**

BIEXPR BI: Vorstellungen von D realisiert
BI: Expectations in Germany

BIO Question: **Q12**

Comment: This variable is not defined for youths answering the $JUGEND biography questionnaire.

German: ***Haben sich Ihre Vorstellungen, mit denen Sie nach Deutschland gekommen sind, im grossen und ganzen erfuellt ?***
"[1] Ja "
"[2] Nur teilweise "
"[3] Nein, gar nicht "

English: ***Have your original expectations of Germany been fulfilled ?***
"[1] Yes"
"[2] Only partially "
"[3] No, not at all "

See also: **BIEXPR, BIEXPRLV, BIEXPRAC, BIEXPRAN**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P160Z**
1995 L	BIOLELA	**P160Z**
1996 M	MLELA	**MB160Z**
1997 N	NLELA	**NB160Z**
1998 O	OLELA	**OB160Z**
1999 P	PLELA	**PB160Z**
2000 Q	QLELA	**QB160Z**
2001 R	RLELA	**RB160Z**
2002 S	SLELA	**SB15**
2003 T	TLELA	**TB15**
2004 U	ULELA	**UB1501**
2005 V	VLELA	**VB1501**
2006 W	WLELA	**WB1501**
2007 X	XLELA	**XB1501**
2008 Y	YLELA	**YB1501**
2009 Z	ZLELA	**ZB1501**
2010 BA	BALELA	BA**B1501**
2011 BB	BBLELA	BBB1501
2000 --	$JUGEND	**n/a**

BIEXPRLV BI: Eigene Wohnung finden
BI: Expectations: Find Apt

BIO Question: **Q13a**

Comment: This variable is not defined for youths answering the $JUGEND biography questionnaire.

German: ***Auf welchen Gebieten war es leichter oder schwerer, als sie vorher gedacht hatten ? -- Eine eigene Wohnung zu finden.***
"[1] Schwerer "
"[2] Wie erwartet "
"[3] Leichter "
"[4] TNZ "

English: ***In which areas was it harder or easier than you expected ? -- to find your own apartment/housing.***
"[1] Harder"
"[2] Just as expected "
"[3] Easier "
"[4] Not applicable "

See also: **BIEXPR, BIEXPRLV, BIEXPRAC, BIEXPRAN**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P171Z**
1995 L	BIOLELA	**P171Z**
1996 M	MLELA	**MB171Z**
1997 N	NLELA	**NB171Z**
1998 O	OLELA	**OB171Z**
1999 P	PLELA	**PB171Z**
2000 Q	QLELA	**QB171Z**
2001 R	RLELA	**RB171Z**
2002 S	SLELA	**SB1601**
2003 T	TLELA	**TB1601**
2004 U	ULELA	**UB1502**
2005 V	VLELA	**VB1502**
2006 W	WLELA	**WB1502**
2007 X	XLELA	**XB1502**
2008 Y	YLELA	**YB1502**
2009 Z	ZLELA	**ZB1502**
2010 BA	BALELA	**BAB1502**
2011 BB	BBLELA	BBB1502
2000 --	$JUGEND	**n/a**

BIEXPRAC BI: Von Arbeitskollegen akzeptiert
BI: Expectations: Accepted by Coworker

BIO Question: **Q13b**

Comment: This variable is not defined for youths answering the $JUGEND biography questionnaire.

German: ***Auf welchen Gebieten war es leichter oder schwerer, als sie vorher gedacht hatten ? -- Von den Arbeitskollegen akzeptiert zu werden.***
"[1] Schwerer "
"[2] Wie erwartet "
"[3] Leichter "
"[4] TNZ "

English: ***In which areas was it harder or easier than you expected ? -- to be accepted by your collegues at work.***
"[1] Harder"
"[2] Just as expected "
"[3] Easier "
"[4] Not applicable "

See also: **BIEXPR, BIEXPRLV, BIEXPRAC, BIEXPRAN**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P172Z**
1995 L	BIOLELA	**P172Z**
1996 M	MLELA	**MB172Z**
1997 N	NLELA	**NB172Z**
1998 O	OLELA	**OB172Z**
1999 P	PLELA	**PB172Z**
2000 Q	QLELA	**QB172Z**
2001 R	RLELA	**RB172Z**
2002 S	SLELA	**SB1602**
2003 T	TLELA	**TB1602**
2004 U	ULELA	**UB1503**
2005 V	VLELA	**VB1503**
2006 W	WLELA	**WB1503**
2007 X	XLELA	**XB1503**
2008 Y	YLELA	**YB1503**
2009 Z	ZLELA	**ZB1503**
2010 BA	BALELA	**BAB1503**
2011 BB	BBLELA	BBB1503
2000 --	$JUGEND	**n/a**

BIEXPRAN BI: Von Nachbarn akzeptiert
BI: Expectations: Accepted by Neighbor

BIO Question: **Q13c**

Comment: This variable is not defined for youths answering the $JUGEND biography questionnaire.

German: ***Auf welchen Gebieten war es leichter oder schwerer, als sie vorher gedacht hatten ? -- Von den Nachbarn akzeptiert zu werden.***
"[1] Schwerer "
"[2] Wie erwartet "
"[3] Leichter "
"[4] TNZ "

English: ***In which areas was it harder or easier than you expected ? -- To be accepted by your neighbors.***
"[1] Harder"
"[2] Just as expected "
"[3] Easier "
"[4] Not applicable "

See also: **BIEXPR, BIEXPRLV, BIEXPRAC, BIEXPRAN**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P173Z**
1995 L	BIOLELA	**P173Z**
1996 M	MLELA	**MB173Z**
1997 N	NLELA	**NB173Z**
1998 O	OLELA	**OB173Z**
1999 P	PLELA	**PB173Z**
2000 Q	QLELA	**QB173Z**
2001 R	RLELA	**RB173Z**
2002 S	SLELA	**SB1603**
2003 T	TLELA	**TB1603**
2004 U	ULELA	**UB1504**
2005 V	VLELA	**VB1504**
2006 W	WLELA	**WB1504**
2007 X	XLELA	**XB1504**
2008 Y	YLELA	**YB1504**
2009 Z	ZLELA	**ZB1504**
2010 BA	BALELA	**BAB1504**
2011 BB	BBLELA	BBB1504
2000 --	$JUGEND	**n/a**

BIRELH BI: Familienmitglieder im Heimatland oder außerhalb Deutschlands
BI: Family in the home country or abroad

BIO Question: **Q14**

Comment: From 2001 onwards the variable is only identified by the parents in $LELA and missing for $JUGEND. A distinction between abroad and home country is not consistently possible over time.

German: ***Haben Sie in dem Land, aus dem Sie kommen bzw. aus dem Ihre Familie kommt, noch Familienangehörige oder andere Ihnen nahstehende Menschen ?***
"[1] Ja "
"[2] Nein "

English: ***Do you have family members or close friends in the home country you (or your family) come from ?***
"[1] Yes "
"[2] No "

See also: **BIRELH, BIRELHP, BIRELHGP, BIRELHC, BIRELHBS, BIRELHDR, BIRELHSP, BIRELHFR, BIRELHMI**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P18Z**
1995 L	BIOLELA	**P18Z**
1996 M	MLELA	**MB18Z**
1997 N	NLELA	**NB18Z**
1998 O	OLELA	**OB18Z**
1999 P	PLELA	**PB18Z**
2000 Q	QLELA	**QB18Z**
2001 R	RLELA	**RB0703V RB0705M**
2002 S	SLELA	**SB2101 SB2102**
2003 T	TLELA	**TB2101 TB2102**
2004 U	ULELA	**UB2101 UB2102**
2005 V	VLELA	**VB2101 VB2102**
2006 W	WLELA	**WB2101 WB2102**
2007 X	XLELA	**XB2101 XB2102**
2008 Y	YLELA	**YB2101 YB2102**
2009 Z	ZLELA	**ZB2101 ZB2102**
2010 BA	BALELA	**BAB2101 BAB2102**
2011 BB	BBLELA	BBB22V01 BBB22M01
2000 Q	QJUGEND	**QJ66**
2001 --	$JUGEND	**n/a**

BIRELHP BI: Im Ausland: Eltern
BI: Family Abroad: Parents

BIO Question: **Q15a**

Comment: This variable is used to identify any relatives starting 2001 for $LELA and missing for all $JUGEND starting 2001.

German: ***Personen in der Heimat: Was für Personen sind das? Eltern?***
"[1] Eltern "

English: ***Persons in Native Country: Who are they? Parents?***
"[1] Parents "

See also: **BIRELH, BIRELHP, BIRELHGP, BIRELHC, BIRELHBS, BIRELHDR, BIRELHSP, BIRELHFR, BIRELHMI, BIRELHS2, BIRELHC2, BIRELHSP, BIRELHC, BIRELHP**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P191Z**
1995 L	BIOLELA	**P191Z**
1996 M	MLELA	**MB191Z**
1997 N	NLELA	**NB191Z**
1998 O	OLELA	**OB191Z**
1999 P	PLELA	**PB191Z**
2000 Q	QLELA	**QB191Z**
2001 R	RLELA	**RB0703V RB0705M**
2002 S	SLELA	**SB2101 SB2102**
2003 T	TLELA	**TB2101 TB2102**
2004 U	ULELA	**UB2101 UB2102**
2005 V	VLELA	**VB2101 VB2102**
2006 W	WLELA	**WB2101 WB2102**
2007 X	XLELA	**XB2101 XB2102**
2008 Y	YLELA	**YB2101 YB2102**
2009 Z	ZLELA	**ZB2101 ZB2102**
2010 BA	BALELA	**BAB2101 BAB2102**
2011 BB	BBLELA	BBB22V01 BBB22M01
2000 Q	QJUGEND	**QJ6701**
2001 --	$JUGEND	**n/a**

BIRELHGP BI: In Heimat: Grosseltern
BI: Family Abroad: Grandparents

BIO Question: **Q15b**

Comment: This variable is not defined for new entrants starting 2001 (R).

German: ***Personen in der Heimat: Was für Personen sind das ? Grosseltern ?***
"[1] Grosseltern "

English: ***Persons in Native Country: Who are they ? Grandparents ?***
"[1] Grandparents "

See also: **BIRELH, BIRELHP, BIRELHGP, BIRELHC, BIRELHBS, BIRELHDR, BIRELHSP, BIRELHFR, BIRELHMI**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P192Z**
1995 L	BIOLELA	**P192Z**
1996 M	MLELA	**MB192Z**
1997 N	NLELA	**NB192Z**
1998 O	OLELA	**OB192Z**
1999 P	PLELA	**PB192Z**
2000 Q	QLELA	**QB192Z**
2001 --	$LELA	**n/a**
2000 Q	QJUGEND	**QJ6702**
2001 --	$JUGEND	**n/a**

BIRELHC BI: In Heimat: Kinder
BI: Family Abroad: Children

BIO Question: **Q15c**

Comment: This variable is not defined for new entrants starting 2001 (R).

German: ***Personen in der Heimat: Was für Personen sind das ? Kinder ?***
"[1] Kinder "

English: ***Persons in Native Country: Who are they ? Children ?***
"[1] Children "

See also: **BIRELH, BIRELHP, BIRELHGP, BIRELHC, BIRELHBS, BIRELHDR, BIRELHSP, BIRELHFR, BIRELHMI, BIRELHS2, BIRELHC2, BIRELHSP, BIRELHC, BIRELHP**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P193Z**
1995 L	BIOLELA	**P193Z**
1996 M	MLELA	**MB193Z**
1997 N	NLELA	**NB193Z**
1998 O	OLELA	**OB193Z**
1999 P	PLELA	**PB193Z**
2000 Q	QLELA	**QB193Z**
2001 --	$LELA	**n/a**
2000 Q	QJUGEND	**QJ6703**
2001 --	$JUGEND	**n/a**

BIRELHBS BI: In Heimat: Bruder, Schwester
BI: Family Abroad: Brother/Sister

BIO Question: **Q15d**

Comment: This variable is not defined for new entrants starting 2001 (R).

German: ***Personen in der Heimat: Was für Personen sind das ? Bruder/Schwester ?***
"[1] Bruder/Schwester "

English: ***Persons in Native Country: Who are they ? Brother/Sister ?***
"[1] Brother/Sister "

See also: **BIRELH, BIRELHP, BIRELHGP, BIRELHC, BIRELHBS, BIRELHDR, BIRELHSP, BIRELHFR, BIRELHMI**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P194Z**
1995 L	BIOLELA	**P194Z**
1996 M	MLELA	**MB194Z**
1997 N	NLELA	**NB194Z**
1998 O	OLELA	**OB194Z**
1999 P	PLELA	**PB194Z**
2000 Q	QLELA	**QB194Z**
2001 --	$LELA	**n/a**
2000 Q	QJUGEND	**QJ6704**
2001 --	$JUGEND	**n/a**

BIRELHDR BI: In Heimat: Entferntere Verwandte
BI: Family Abroad: Distant Relatives

BIO Question: **Q15e**

Comment: This variable is not defined for new entrants starting 2001 (R).

German: ***Personen in der Heimat: Was für Personen sind das ? Entferntere Verwandte ?***
"[1] Entferntere Verwandte "

English: ***Persons in Native Country: Who are they ? Distant Relatives ?***
"[1] Distant Relatives "

See also: **BIRELH, BIRELHP, BIRELHGP, BIRELHC, BIRELHBS, BIRELHDR, BIRELHSP, BIRELHFR, BIRELHMI**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P195Z**
1995 L	BIOLELA	**P195Z**
1996 M	MLELA	**MB195Z**
1997 N	NLELA	**NB195Z**
1998 O	OLELA	**OB195Z**
1999 P	PLELA	**PB195Z**
2000 Q	QLELA	**QB195Z**
2001 --	$LELA	**n/a**
2000 Q	QJUGEND	**QJ6705**
2001 --	$JUGEND	**n/a**

BIRELHSP BI: In Heimat: Ehepartner, Verlobte(r)
BI: Family Abroad: Spouse

BIO Question: **Q15f**

Comment: This variable is not defined for new entrants starting 2001 (R).

German: ***Personen in der Heimat:***
Was für Personen sind das ? Ehepartner / Verlobte(r)?
"[1] Ehepartner/Verlobte(r) "

English: ***Persons in Native Country: Who are they ? Spouse / Fiance(e)?***
"[1] Spouse/Fiance(e) "

See also: **BIRELH, BIRELHP, BIRELHGP, BIRELHC, BIRELHBS, BIRELHDR, BIRELHSP, BIRELHFR, BIRELHMI, BIRELHS2, BIRELHC2, BIRELHSP, BIRELHC, BIRELHP**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P196Z**
1995 L	BIOLELA	**P196Z**
1996 M	MLELA	**MB196Z**
1997 N	NLELA	**NB196Z**
1998 O	OLELA	**OB196Z**
1999 P	PLELA	**PB196Z**
2000 Q	QLELA	**QB196Z**
2001 --	$LELA	**n/a**
2000 Q	QJUGEND	**QJ6706**
2001 --	$JUGEND	**n/a**

BIRELHFR BI: In Heimat: Persoenliche Bekannte
BI: Family Abroad: Friends

BIO Question: **Q15g**

Comment: This variable is not defined for new entrants starting 2001 (R).

German: ***Personen in der Heimat: Was für Personen sind das ? Bekannte, Freunde ?***

"[1] Persönliche Bekannte "

English: ***Persons in Native Country: Who are they ? Friends ?***
"[1] Friends "

See also: **BIRELH, BIRELHP, BIRELHGP, BIRELHC, BIRELHBS, BIRELHDR, BIRELHSP, BIRELHFR, BIRELHMI**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P197Z**
1995 L	BIOLELA	**P197Z**
1996 M	MLELA	**MB197Z**
1997 N	NLELA	**NB197Z**
1998 O	OLELA	**OB197Z**
1999 P	PLELA	**PB197Z**
2000 Q	QLELA	**QB197Z**
2001 --	$LELA	**n/a**
2000 Q	QJUGEND	**QJ6707**
2001 --	$JUGEND	**n/a**

BIRELHMI BI: Personen gern nach Dt. holen?
BI: Persons abroad bring to Germany

BIO Question: **Q16**

Comment: This variable is not defined for new entrants starting 2001 (R).

German: ***Personen in der Heimat: Gibt es darunter Personen, die auch nach Deutschland kommen wollen bzw. die Sie gerne nachholen möchten ?***
"[1] Ja "
"[2] Nein "

English: ***Persons in Native Country: Among those mentioned above, do some want to come to Germany, or would you like them to come to Germany ?***
"[1] Yes "
"[2] No "

See also: **BIRELH, BIRELHP, BIRELHGP, BIRELHC, BIRELHBS, BIRELHDR, BIRELHSP, BIRELHFR, BIRELHMI**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P200Z**
1995 L	BIOLELA	**P200Z**
1996 M	MLELA	**MB200Z**
1997 N	NLELA	**NB200Z**
1998 O	OLELA	**OB200Z**
1999 P	PLELA	**PB200Z**
2000 Q	QLELA	**QB200Z**
2001 --	$LELA	**n/a**
2000 Q	QJUGEND	**QJ68**
2001 --	$JUGEND	**n/a**

BIRELHS2 BI: Ehepartner in Deutschland
BI: Spouse in Germany

BIO Question: **Q15f**

Comment: This variable is not defined for new entrants starting 1996 (M) and missing for all $JUGEND.

German: ***Lebt Ihr Ehepartner in Deutschland ?***
"[1] D hier im HH "
"[2] D nicht im HH "
"[3] Nicht in D "

English: ***Does your spouse live in Germany ?***
"[1] Yes, here in the HH "
"[2] Yes, but NOT with me in HH "
"[3] Not in Germany "

See also: **BIRELHS2, BIRELHC2, BIRELHSP, BIRELHC, BIRELHP**

Year	File	Variable
1984 A	APAUSL	**AP58A02**
1985 B	BPAUSL	**n/a**
1986 C	CPAUSL	**CP90A01**
1987 D	DPAUSL	**DP92A01**
1988 E	EPAUSL	**EP85A01**
1989 F	FPAUSL	**FP102A01**
1990 G	GPAUSL	**GP102A01**
1991 H	HPAUSL	**HP102A01**
1992 I	IPAUSL	**IP102A01**
1993 J	JPAUSL	**JP102A01**
1994 K	KPAUSL	**KP102A01**
1995 L	LPAUSL	**LP110A01**
1996 --	$P	**n/a**
1984 --	BIOLELA	**n/a**
2000 --	$JUGEND	**n/a**

BIRELHC2 BI: Kinder unter 18 J. nicht in Deutschland
BI: Underage Children not in Germany

BIO Question: **Q15c**

German: ***Haben Sie Kinder unter 18 Jahren, die nicht in Deutschland leben ?***
"[1] Ja "
"[2] Nein "

English: ***Do you have children under 18, who do not live in Germany ?***
"[1] Yes "
"[2] No "

See also: **BIRELHS2, BIRELHC2, BIRELHSP, BIRELHC, BIRELHP**

Year	File	Variable
1984 A	APAUSL	**AP66A01**
1985 B	BPAUSL	**BP95A01**
1986 C	CPAUSL	**CP86A01**
1987 D	DPAUSL	**DP88A01**
1988 E	EPAUSL	**n/a**
1989 F	FPAUSL	**FP98A01**
1990 G	GPAUSL	**n/a**
1991 H	HPAUSL	**HP98A01**
1992 I	IPAUSL	**n/a**
1993 J	JPAUSL	**JP98A01**
1994 K	KPAUSL	**n/a**
1995 L	LPAUSL	**LP106A01**
1996 M	MP	**MP7406**
1997 N	NP	**NP111A04**
1998 O	OP	**n/a**
1999 P	PP	**PP12904**
2000 --	$PAUSL	**n/a**
1984- A-	BIOLELA	**n/a**
2000 --	$JUGEND	**n/a**

BIGOBACK BI: Rueckkehr Heimat (ab 1994)
BI: Go back home ?

BIO Question: **Q17**

Comment: The question BIGOBACK (using BIOLELA) asks whether one intends to **return** home to the native country whereas BISTAY (using $PAUSL) asks wether one intends to **stay** in Germany. The wording and the answer possibilities are different in both questions. Further, there is no particular reason to believe that the two variables even are consistent. Starting 2001, this is not defined for new entrants.

German: ***Planen Sie selbst, in Ihr Herkunftsland wieder zurückzukehren ?***
"[1] Ja, ganz sicher "
"[2] Ja, wahrscheinlich "
"[3] Eher unwahrscheinlich "
"[4] Nein, sicher nicht "

English: ***Are you planning to go back to live in your native country?***
"[1] Yes, certainly "
"[2] Yes, probably "
"[3] Probably not "
"[4] No, Certainly not "

See also: **BIGOBACK, BISTAY, BISTAYY**

Year	File	Variable
1984-93 A-J	BIOLELA	**n/a**
1994 K	BIOLELA	**P230Z**
1995 L	BIOLELA	**P230Z**
1996 M	MLELA	**MB230Z**
1997 N	NLELA	**NB230Z**
1998 O	OLELA	**OB230Z**
1999 P	PLELA	**PB230Z**
2000 Q	QLELA	**QB230Z**
2001 --	$LELA	**n/a**

2000 Q	QJUGEND	**QJ69**
2001 --	$JUGEND	**n/a**

BISTAY BI: Wunsch in D zu bleiben
BI: Desire to Stay in Germany

BIO Question: **Q17**

Comment: The question BIGOBACK (using BIOLELA) asks whether one intends to **return** home to the native country whereas BISTAY (using $PAUSL) asks wether one intends to **stay** in Germany. The wording and the answer possibilities are different in both questions. Further, there is no particular reason to believe that the two variables even are consistent. This variable is not defined for youths answering the $JUGEND biography questionnaire.

German: ***Wie lange wollen Sie in Deutschland bleiben ?***
"[1] Kehre innerhalb eines Jahres zurück"
"[2] Einige Jahre und zwar ..."
"[3] Für immer in D bleiben "

English: ***How long would you like to stay in Germany ?***
"[1] Go back within 12 months "
"[2] Several years, specifically ..."
"[3] Always stay in Germany "

See also: **BIGOBACK, BISTAY, BISTAYY**

Year	File	Variable
1984 A	APAUSL	**AP67A01**
1985 B	BPAUSL	**BP96A01**
1986 C	CPAUSL	**CP87A01**
1987 D	DPAUSL	**DP89A01**
1988 E	EPAUSL	**EP77A01**
1989 F	FPAUSL	**FP99A01**
1990 G	GPAUSL	**GP96A01**
1991 H	HPAUSL	**HP99A01**
1992 I	IPAUSL	**IP99A01**
1993 J	JPAUSL	**JP99A01**
1994 K	KPAUSL	**KP96A01**
1995 L	LPAUSL	**LP107A01**
1996 M	MP	**MP101A01 MP100A**
1997 N	NP	**NP109A01 NP108A**
1998 O	OP	**OP11401 OP113**
1999 P	PP	**PP12601 PP125**
2000 Q	QP	**QP13401 QP133**
2001 R	RP	**RP12701 RP126**
2002 S	SP	**SP12601 SP125**
2003 T	TP	**TP13301 TP132**
2004 U	UP	**UP13501 UP134**
2005 V	VP	**VP14601 VP145**
2006 W	WP	**WP13601 WP135**
2007 X	XP	**XP14601 XP145**
2008 Y	YP	**YP14501 YP144**
2009 Z	ZP	**ZP14401 ZP143**
2010 BA	BAP	**BAP14601 BAP145**
2011 BB	BBP	BBP14701 BBP146
2000 --	$JUGEND	**n/a**

BISTAYY BI: Dauer des geplanten Aufenthalts
BI: Years Desired to Stay in Germany

BIO Question: **Q17**

Comment: This variable is not defined for youths answering the $JUGEND biography questionnaire.

German: ***Wie lange wollen Sie in Deutschland bleiben ? Einige Jahre und zwar...***

English: ***How long would you like to stay in Germany ? Several years, specifically ...***

See also: **BIGOBACK, BISTAY, BISTAYY**

Year	File	Variable
1984 A	APAUSL	**AP67A02**
1985 B	BPAUSL	**BP96A02**
1986 C	CPAUSL	**CP87A02**
1987 D	DPAUSL	**DP89A02**
1988 E	EPAUSL	**EP77A02**
1989 F	FPAUSL	**FP99A02**
1990 G	GPAUSL	**GP96A02**
1991 H	HPAUSL	**HP99A02**
1992 I	IPAUSL	**IP99A02**
1993 J	JPAUSL	**JP99A02**
1994 K	KPAUSL	**KP96A02**
1995 L	LPAUSL	**LP107A02**
1996 M	MP	**MP101A02**
1997 N	NP	**NP109A02**
1998 O	OP	**OP11402**
1999 P	PP	**PP12602**
2000 Q	QP	**QP13402**
2001 R	RP	**RP12702**
2002 S	SP	**SP12602**
2003 T	TP	**TP13302**
2004 U	UP	**UP13502**
2005 V	VP	**VP14602**
2006 W	WP	**WP13602**
2007 X	XP	**XP14602**
2008 Y	YP	**YP14502**
2009 Z	ZP	**ZP14402**
2010 BA	BAP	**BAP14602**
2011 BB	BBP	BBP14702
2000 --	$JUGEND	**n/a**

BISCGER BI: In Dt. Schule besucht?
BI: Attended School in Germany

BIO Question: **Q18**

Comment: This question asks only if one has **ever** attended a (primary/secondary) school in Germany, but does **not** ask whether one received a certificate/diploma, such as the generated variable $PSBIL in the file **$PGEN**.

German: ***Haben Sie in Deutschland eine Schule besucht?***
"[1] Ja"
"[2] Nein"

English: ***Did you attend school in Germany?***
"[1] Yes"
"[2] No"

See also: **BISCGER, BISCGRAD, BISCGERC, BISCGC, BISCGCF, BISCGCFN**

Year	File	Variable
1984 A	APAUSL	**AP06A01**
1985 B	BPAUSL	**BP100A01**
1986 C	CPAUSL	**CP100B01**
1987 D	DPAUSL	**DP97A01**
1988 E	EPAUSL	**EP90A01**
1989 F	FPAUSL	**FP107A**
1990 G	GPAUSL	**GP107A**
1991 H	HPAUSL	**HP107A**
1992 I	IPAUSL	**IP107A**
1993 J	JPAUSL	**JP107A**
1984-93 A-J	BIOLELA	**B46A**
1994 K	BIOLELA	**P280Z**
1995 L	BIOLELA	**P280Z**
1996 M	MLELA	**MB280Z**
1997 N	NLELA	**NB280Z**
1998 O	OLELA	**OB280Z**
1999 P	PLELA	**PB280Z**
2000 Q	QLELA	**QB280Z**
2001 R	RLELA	**RB280Z**
2002 S	SLELA	**SB11**
2003 T	TLELA	**TB11**
2004 U	ULELA	**UB11**
2005 V	VLELA	**VB11**
2006 W	WLELA	**WB11**
2007 X	XLELA	**XB11**
2008 Y	YLELA	**YB11**
2009 Z	ZLELA	**ZB11**
2010 BA	BALELA	**BAB11**
2011 BB	BBLELA	BBB11
2000 Q	QJUGEND	**QJ63**
2001 R	RJUGEND	**RJ65**
2002 S	SJUGEND	**SJ65**
2003 T	TJUGEND	**TJ65**
2004 U	UJUGEND	**UJ65**
2005 V	VJUGEND	**VJ65**
2006 --	$JUGEND	**n/a**

BISCGRAD BI: In welche Klasse in dt. Schule
BI: Which Grade School

BIO Question: **Q19**

Comment: The question here is **not** on the highest schooling achieved , but rather what was the grade or class when one **first** came to Germany.

German: ***In welche Klasse sind Sie in Deutschland in die Schule gekommen ?***

English: ***Which class/grade did you attend when you came to Germany?***

See also: **BISCGER, BISCGRAD, BISCGERC, BISCGC, BISCGCF, BISCGCFN**

Year	File	Variable
1984 A	APAUSL	**n/a**
1985 B	BPAUSL	**n/a**
1986 C	CPAUSL	**n/a**
1987 D	DPAUSL	**n/a**
1988 E	EPAUSL	**n/a**
1989 F	FPAUSL	**FP108A**
1990 G	GPAUSL	**GP108A**
1991 H	HPAUSL	**HP108A**
1992 I	IPAUSL	**IP108A**
1993 J	JPAUSL	**JP108A**
1984-93 A-J	BIOLELA	**B47A**
1994 K	BIOLELA	**P290Z**
1995 L	BIOLELA	**P290Z**
1996 M	MLELA	**MB290Z**
1997 N	NLELA	**NB290Z**
1998 O	OLELA	**OB290Z**
1999 P	PLELA	**PB290Z**
2000 Q	QLELA	**QB290Z**
2001 R	RLELA	**RB290Z**
2002 S	SLELA	**SB12**
2003 T	TLELA	**TB12**
2004 U	ULELA	**UB12**
2005 V	VLELA	**VB12**
2006 W	WLELA	**WB12**
2007 X	XLELA	**XB12**
2008 Y	YLELA	**YB12**
2009 Z	ZLELA	**ZB12**
2010 BA	BALELA	**BAB12**
2011 BB	BBLELA	BBB12
2000 Q	QJUGEND	**QJ64**
2001 R	RJUGEND	**RJ6601**
2002 S	SJUGEND	**SJ6601**
2003 T	TJUGEND	**TJ6601**
2004 U	UJUGEND	**UJ6601**
2005 V	VJUGEND	**VJ6601**
2006 --	$JUGEND	**n/a**

BISCGERC BI: Besuch spezieller Vorbereitung
BI: Attended Special Foreigner Prep Class

BIO Question: **Q20**

German: ***Haben Sie vorher eine spezielle Vorbereitungsklasse für Ausländer in Deutschland besucht?***
"[1] Ja"
"[2] Nein"

English: ***Did you attend a special preparation class for foreigners in Germany?***
"[1] Yes"
"[2] No"

See also: **BISCGER, BISCGRAD, BISCGERC, BISCGC, BISCGCF, BISCGCFN**

Year	File	Variable
1984 A	APAUSL	**n/a**
1985 B	BPAUSL	**n/a**
1986 C	CPAUSL	**n/a**
1987 D	DPAUSL	**n/a**
1988 E	EPAUSL	**n/a**
1989 F	FPAUSL	**FP109A**
1990 G	GPAUSL	**GP109A**
1991 H	HPAUSL	**HP109A**
1992 I	IPAUSL	**IP109A**
1993 J	JPAUSL	**JP109A**
1984-93 A-J	BIOLELA	**B48A**
1994 K	BIOLELA	**P300Z**
1995 L	BIOLELA	**P300Z**
1996 M	MLELA	**MB48A**
1997 N	NLELA	**NB48A**
1998 O	OLELA	**OB48A**
1999 P	PLELA	**PB48A**
2000 Q	QLELA	**QB48A**
2001 R	RLELA	**RB48A**
2002 S	SLELA	**SB13**
2003 T	TLELA	**TB13**
2004 U	ULELA	**UB13**
2005 V	VLELA	**VB13**
2006 W	WLELA	**WB13**
2007 X	XLELA	**XB13**
2008 Y	YLELA	**YB13**
2009 Z	ZLELA	**ZB13**
2010 BA	BALELA	**BAB13**
2011 BB	BBLELA	BBB13
2000 Q	QJUGEND	**QJ65**
2001 R	RJUGEND	**RJ6602**
2002 S	SJUGEND	**SJ6602**
2003 T	TJUGEND	**TJ6602**
2004 U	UJUGEND	**UJ6602**
2005 V	VJUGEND	**VJ6602**
2006--	$JUGEND	**n/a**

BISCGC BI: Auch dt. Schueler in Schulklasse
BI: Also German Pupils in Class

BIO Question: **Q21a**

Comment: This variable is not defined for new entrants starting 2000 (Q) and $JUGEND.

German: ***Gab es in der Schulklasse, die Sie zuletzt in Deutschland besucht haben, auch deutsche Schüler?***
"[1] Ja"
"[2] Nein"

English: ***Were there also German children present in the class you last attended?***
"[1] Yes"
"[2] No"

See also: **BISCGER, BISCGRAD, BISCGERC, BISCGC, BISCGCF, BISCGCFN**

Year		File	Variable
1984	A	APAUSL	**n/a**
1985	B	BPAUSL	**n/a**
1986	C	CPAUSL	**n/a**
1987	D	DPAUSL	**n/a**
1988	E	EPAUSL	**n/a**
1989	F	FPAUSL	**FP110A01**
1990	G	GPAUSL	**GP110A01**
1991	H	HPAUSL	**HP110A01**
1992	I	IPAUSL	**IP110A01**
1993	J	JPAUSL	**JP110A01**
1984-93	A-J	BIOLELA	**B49A**
1994	K	BIOLELA	**B49A**
1995	L	BIOLELA	**B49A**
1996	M	MLELA	**MB49A**
1997	N	NLELA	**NB49A**
1998	O	OLELA	**OB49A**
1999	P	PLELA	**PB49A**
2000	--	$LELA	**n/a**
2000	--	$JUGEND	**n/a**

BISCGCF BI: Wieviel Mitschueler Auslaender
BI: How many Pupils foreign

BIO Question: **Q21b**

German: ***Wie viele Ihrer Mitschueler waren Auslaender?***
"[1] Die meisten"
"[2] Etwa 1/2"
"[3] Etwa 1/4"
"[4] Weniger als 1/4"
"[5] Ausser mir niemand"

English: ***How many of your fellow students were foreigners ?***
"[1] Most of them"
"[2] Around 1/2"
"[3] Around 1/4"
"[4] Less than 1/4"
"[5] I was only one"

See also: **BISCGER, BISCGRAD, BISCGERC, BISCGC, BISCGCF, BISCGCFN**

Year	File	Variable
1984-88 A-E	$PAUSL	**n/a**
1989 F	FPAUSL	**FP110A02**
1990 G	GPAUSL	**GP110A02**
1991 H	HPAUSL	**HP110A02**
1992 I	IPAUSL	**IP110A02**
1993 J	JPAUSL	**JP110A02**
1984-93 A-J	BIOLELA	**B50A**
1994 K	BIOLELA	**B50A**
1995 L	BIOLELA	**B50A**
1996 M	MLELA	**MB50A**
1997 N	NLELA	**NB50A**
1998 O	OLELA	**OB50A**
1999 P	PLELA	**PB50A**
2000 Q	QLELA	**QB50A**
2001 R	RLELA	**RB50A**
2002 S	SLELA	**SB43**
2003 T	TLELA	**TB43**
2004 U	ULELA	**UB43**
2005 V	VLELA	**VB43**
2006 W	WLELA	**WB43**
2007 X	XLELA	**XB43**
2008 Y	YLELA	**YB43**
2009 Z	ZLELA	**ZB43**
2010 BA	BALELA	**BAB43**
2011 BB	BBLELA	BBB43
2000 Q	QJUGEND	**n/a**
2001 R	RJUGEND	**RJ43**
2002 S	SJUGEND	**SJ43**
2003 T	TJUGEND	**TJ43**
2004 U	UJUGEND	**UJ43**
2005 V	VJUGEND	**VJ43**
2006 W	WJUGEND	**WJ45**
2007 X	XJUGEND	**XJ45**
2008 Y	YJUGEND	**YJ45**
2009 Z	ZJUGEND	**ZJ45**
2010 BA	BAJUGEND	**BAJ45**

BISCGCFN BI: Eine oder mehrere Nationalitaet
BI: Mix of Nationalities in Class

BIO Question: **Q21c**

Comment: This variable is not defined for new entrants starting 2000 (Q) and $JUGEND.

German: ***Gab es in dieser Klasse nur Schüler Ihrer Nationalität oder waren verschieden Nationalitäten gemischt?***
"[1] Nur meine Nationalitaet "
"[2] Gemischt "

English: ***Were there only children of your nationality, or were the nationalites mixed?***
"[1] Only my nationality"
"[2] Mixed "

See also: **BISCGER, BISCGRAD, BISCGERC, BISCGC, BISCGCF, BISCGCFN**

Year	File	Variable
1984 A	APAUSL	**n/a**
1985 B	BPAUSL	**n/a**
1986 C	CPAUSL	**n/a**
1987 D	DPAUSL	**n/a**
1988 E	EPAUSL	**n/a**
1989 F	FPAUSL	**FP110A03**
1990 G	GPAUSL	**GP110A03**
1991 H	HPAUSL	**HP110A03**
1992 I	IPAUSL	**IP110A03**
1993 J	JPAUSL	**JP110A03**
1984-93 A-J	BIOLELA	**B51A**
1994 K	BIOLELA	**B51A**
1995 L	BIOLELA	**B51A**
1996 M	MLELA	**MB51A**
1997 N	NLELA	**NB51A**
1998 O	OLELA	**OB51A**
1999 P	PLELA	**PB51A**
2000 --	$LELA	**n/a**
2000 --	$JUGEND	**n/a**

15 BIORESID: Variables On Occupancy and Second Residence

by Henning Lohmann, Marco Giesselmann *and* Sven Witzke

(Replaces earlier versions by Jürgen Schupp *and* Michael Frühling / Thorsten Schneider*)*

In 1994 questions with a focus on occupancy were introduced to the Biographical Questionnaire asking for the duration of residence in the current dwelling and any second residence. Questions on the second residence were also asked before 1994, but those were collected in the (blue version of the) Individual Questionnaire and therefore the corresponding variables are part of the $P files. The information surveyed in the Biographical Questionnaire is stored in the new file BIORESID.

The variables of BIORESID are based on following questions:

Question I

When did you move into this home?
☐☐☐☐

Year

Question II

Do you have another home in which you yourself reside or spend your vacation? [55]
(0) No (1) Yes => continue with question

Is this second home in western Germany (including West Berlin), in eastern Germany (including East Berlin) or abroad?
Western Germany
Eastern Germany
Abroad[56]

Which home is your main residence?
This one
The other one
I use both about the same

From which residence do you usually go to work?
From this one
From the other one
Not applicable

[55] In the years 1994 and 1995 the question was "*Do you have another home, in Germany, in which you yourself reside in?*"
[56] The new category "abroad" was added in 1996.

15.1 Sources of Variables

The information for the years 1994 and 1995 stem from the file BIOLELA. Information for later years are taken from the wave-specific data sets $LELA.

In principle, SOEP respondents answer the Biography Questionnaire only once, so every person has only one record with wave-specific information in BIORESID. For fieldwork-related reasons, very few people have answered the Biography Questionnaire twice. For these, the first interview is taken as relevant for BIORESID. Further cases are dropped if their information stems from an interview completed before 1994.

15.2 Population of Interest

The BIORESID dataset as of wave 2011 contains information on 28,337 individuals, stemming from samples A-J. The data set is supplemented every year by new respondents filling in the supplementary Biography Questionnaire.

Table 1: Survey Year in BIORESID

Survey year	n	%	% (cum.)
1994	933	3.29	3.29
1995	1,075	3.79	7.09
1996	471	1.66	8.75
1997	480	1.69	10.44
1998	415	1.46	11.91
1999	2,039	7.20	19.10
2000	243	0.86	19.96
2001	8,816	31.11	51.07
2002	508	1.79	52.86
2003	2,319	8.18	61.05
2004	449	1.58	62.63
2005	299	1.06	63.69
2006	223	0.79	64.47
2007	2,232	7.88	72.35
2008	336	1.19	73.54
2009	196	0.69	74.23
2010	2,038	7.19	81.42
2011	5,265	18.58	100.00
Total	23.072	100	

Status: up to wave BB (2011)

The information in BIORESID is treated as time-invariant. Although, in principle, it is possible to update the information on occupancy for some individuals on the basis of more recent information, we abstain from doing so for selectivity reasons.

15.3 Variable List of the Data Set BIORESID

Table 3: Description of the Data Set BIORESID

Variable Name	Content of the Variable
	Entries for Surveyed Person
HHNR	Original household number (invariant)
HHNRAKT	Current wave HH number (wave of biography interview)
PERSNR	Never changing person ID
ERHEBJ	Survey year
	Occupancy
BRMOVEIN	Year person moved in current dwelling
	Second Residence
BRSECHOM	Having a second residence
BRSECREG	Region of second residence
BRSECUSE	Use of second residence
BRSECWOR	Second residence at place of work
	Specification of Interview Situation
BRINTA	Type of interview
BRINTNR	Identifier of the interviewer

16 BIOEDU (beta version): Data on educational participation and transitions

By Henning Lohmann and Sven Witzke

The Socio-Economic Panel Study (SOEP) contains a broad range of variables which cover early child education and care, educational participation, educational degrees and other related topics. However, the respective questions are included in different questionnaires (e.g, personal questionnaire, household questionnaire, youth questionnaire) and the variables are not always in a format which is suited for longitudinal analyses. For instance, transitions such as school enrolment or entry into tertiary education are not documented in a single variable but can only be reconstructed by comparing the status in a wave t with the status in a wave t+1 (e.g., a transition into tertiary education took place if a person was not in university in wave t but is in university in wave t+1). Generating such variables is time-consuming and prone to errors. It is the aim of the BIOEDU dataset to provide ready-made variables on educational transitions and related topics in order to support analyses in a longitudinal perspective.

The BIOEDU dataset is primarily based on prospectively collected information. Therefore, it contains most information for those persons who have been part of the survey population at the time when they have attended school or other educational institutions. In total the dataset contains information on 62,635 persons. This is the part of the SOEP sample for whom we have observed an educational transition and/or an educational degree. For the larger part of this group we have observed an educational degree only (n=46,390). These are persons who have not been a part of the sample at the time when they participated in education or experienced educational transitions. The smaller part of the sample is more interesting for longitudinal analyse of educational participation. These are persons who lived in a survey household at the time of educational participation.[57] Depending on the age of the individual the dataset contains variables on:

- early child education and care (ECEC)
- entry into primary school
- transition to secondary school
- first exit from secondary school
- secondary school attendance after first exit from school
- first entry into and exit from vocational training

[57] Accordingly the first group is much older than the second group. At the time of the first observation in the sample the first group is on average 45 years old while the second group has an average age below 9 years.

- vocational training participation after first
- first entry into and exit from tertiary education
- tertiary education participation after first exit
- highest ever obtained educational degrees and last observed educational participation

The SOEP as a general household panel study is not specifically directed at the analysis of educational life courses. Nevertheless, right from the beginning of the panel in 1984 the survey instruments contained questions on the educational attainment of the respondents (aged 17 and older) and children younger than 17 years living in survey households. After more than 25 years of survey duration these data provide a precious source for the reconstruction of educational life courses. In the following we describe how we use these data to reconstruct educational transitions starting before school enrolment and up to post-secondary education.

The reconstruction of transitions is primarily based on yearly information on educational participation (i.e. entry and exit reconstructed from changes in participation). For later transitions there is some more information as explicit questions on the end of general school, vocational training and tertiary education are part of the questionnaires (changes during the year prior to the survey, only for persons aged 17+ years, exception: already obtained degrees before age 17 in youth questionnaire). In the following we describe in detail how each of the variables is constructed.

One remark on the variable naming conventions: The variable names always begin with “be” which stands for “biography education” (in analogy to other biography datasets). The third and forth letter denote the type of transition or similar. For instance, t0 stands for variables on the first and t1 for the last year in child care. Variables on starting school contain a t2 and so on (up to t8= exit from tertiary education). Variables containing and x as the third letter contain information on the last observed year in education or on the highest educational degrees ever obtained (x4, x6, x8).

Using this dataset you should keep in mind that most of the information covered by the dataset is not directly asked in the SOEP questionnaires but has been derived from the combination of several variables. In the process of reconstruction assumptions have been made which we tried to describe as detailed as possible in the previous sections. The more these assumptions are based on additional knowledge, e.g. provided by strict institutional regulations, the better for the reconstruction of the transitions.

The dataset covers transitions starting in early childhood up to tertiary education. For a part of the sample only one of these transitions or episodes is observed, for others the whole sequence from elementary education until the exit from tertiary education. The variable *beinfo* provides an overview on the frequencies of these different patterns. In total the dataset contains information on more than 60,000 persons. This is the part of the SOEP sample for whom we have observed an educational transition and/or an educational degree. For 434 cases we have full information (pattern 811111111).

We have provided a number of variables where we documented the process of data generation and the sources where the data stem from (betXinfo, variables with suffixes _s or _g). You could use these variables as indicators of the degree of uncertainty in the process of the reconstruction of educational transitions. The less the variables could be reconstructed just using the basic algorithm (e.g., bet2info<>"0000|0|0000"), the higher is the degree of uncertainty. The same applies to long durations between an observed exit and the observation of a matching educational degree (e.g., a high value in bet6cert_g). It is certainly advisable to check if certain deviations in the process of data generation "explain" substantial results. E.g., if children living in households who where interviewed in August (this information is provided in betXinfo) have a much higher propensity of starting school late (bet2agemo), this might just be a data artefact because it is difficult to decide if the information the household provided referred to the school year which just started in August or to the school year which just ended at the time of the interview. In general, you should expect that there are no such systematic measurement errors in the reconstructed variables. But if you want to have a closer look on potential biases you could use the respective variables which document the data generation process. This documentation describes a beta version of the dataset (v28_0.1). If you have comments or encounter while using the dataset, please let us know.

More detailed information is provided in:

Lohmann, Henning / Witzke, Sven (2011): BIOEDU (beta version): Biographical data on educational participation and transitions in the German Socio-Economic Panel Study (SOEP), DIW Data Documentation 58, Berlin.

17 LIFESPELL: Information on the Pre- and Post-Survey History of SOEP-Respondents

by Hannes Neiss and Martin Kroh

Prospective panel surveys typically face the problem that no information is available on units of analysis after respondents have left the survey. The SOEP team therefore regularly conduct drop-out studies to identify the whereabouts of attritors. These studies draw on official register data and allow us to determine whether a person is still alive, is deceased, or has moved abroad since the last SOEP interview. The information is combined in a spell file LIFESPELL. This dataset reports all available information on the pre- and the post-survey history of all persons who have ever been a member of a SOEP household. The LIFESPELL file lends itself particularly to mortality research, migration research, and non-response research. It extends the period under investigation from the last SOEP interview to the last drop-out study and thus reduces the problem of selective observational probabilities of units of analysis. For users less familiar with spell files, we also provide a STATA code for converting a spell file into a long format file (persnr x year) at the end of this chapter.

The file includes spells for every person ever living in a SOEP household. Each spell indicates one of the following states:

- Living abroad
- Living in Germany
- Living in Germany + part of a SOEP HH[58]
- Deceased
- No information about status

[58] This code does not distinguish between active and inactive respondents, such as children and temporary non-respondents. As long as interviewers can contact the household (even though an interview might not take place), we have information on the vital status of respondents. Therefore, all HH members are coded as being in the SOEP until the HH finally drops out of the SOEP and is no longer followed up by the interviewer.

The information comprised by the LIFESPELL file includes the following:

- Year of birth
- Migration background
- Year of immigration*
- Year of entry into SOEP
- Year of exit out of SOEP
- Year of emigration*
- Year of death

The year of birth, migration background, and year of immigration are self-reported (retrospective) information from personal interviews (e.g., p-files). The year of entry and exit are taken from gross information reported by the interviewer (e.g., pbrutto-files). The year of emigration and the year of death can either come from related persons in the household (e.g., deceased-files), the interviewer (e.g., pbrutto-files), or, finally, from drop-out studies.

The following drop-out studies have been conducted in the past:

a) **1992 drop-out study** The study provides information about the dates of death or emigration, or one can infer that a person was still living in Germany until 1992.

b) **2001 drop-out study** The study provides information about the dates of death or emigration, or one can infer that a person was still living in Germany until 2001. Additionally, the study allows us to analyze regional mobility within Germany. It includes HH drop-outs from 1985-1998.

c) **2006 drop-out study** The study is not a register study. Attritors were contacted by mail and responded by mail. It allows us to include the year of death for those whose letters returned by post with the mark "deceased" or whose letters were answered by relatives or friends. Those who answered themselves can also be assumed to still have been living in Germany in 2006. No information is available

* indicates that the fact is only necessary for subgroups of the population

for cases where there was no response. No information about emigration is given. The study includes drop-outs from 2001-2004.

d) **2008 drop-out study** The latest of the studies provides information about the year of death and those still living in Germany in 2008. Emigration is captured insufficiently. It includes drop-outs from 1985-2006.

Year of study	1992	2001	2006	2008
Population at risk				
Attritors between…	1984-1990	1984-1998	2001-2004	1985-2006
Overall	1,089	8,082	4,982	5,838
Identified	1,044	7,106	962	5,591
Information available				
Death	+	+	(+)	+
Emigration	+	+	(+)	(+)
Mobility within Germany		+		

LIFESPELL contains the following variables. Comments are added to clarify certain assumptions made when coding the variables.

SPELLNR

Var Label: SPELLNR **"Number of spell"**

Var format: SPELLNR (I2)

Comment: This variable indicates the spell number.

SPELLTYP

Var Label: SPELLTYP **“Type of spell”**

Value Label: SPELLTYP (1)“Living in Germany”
(2)“SOEP”
(3)“Living abroad”
(4)“Deceased”
(5)“Gap”

Comment: The variable gives information about the current residence and status of the person. “Living in Germany” means that the person lives in Germany, but is not yet or no longer part of a SOEP HH. Persons who are part of a SOEP HH are given the “SOEP” code on this variable. Note that the code refers to the person being part of the gross sample, which also includes children and covers periods of temporary drop-outs as well as the final year when respondents refuse to participate. If a spell is coded “Abroad,” this means a person has not yet immigrated to Germany or has already emigrated. “Deceased” indicates the spell with the year of death and finally “Gap” means that for the time marked by the spell no information about the person’s status can be determined.

BEGIN

Var label: BEGIN **“Year in which spell begins”**

END

Var label: BEGIN **“Year in which spell ends”**

Comment: Two consecutive spells may overlap for one year in individual cases. This happens, for instance, when refusal to respond coincides with emigration in the same year.

CENSORING

Var label: CENSORING **“Censoring: How and why”**

Value Label: CENSORING

(10) “left censored”
(11) “left censored: gap”
(20) “right censored”
(21) “right censored: gap”
(22) “rc: no information after SOEP”
(23) “rc: no information after 1992 study”
(24) “rc: no information after study 2001”
(25) “rc: no information after study 2006”
(26) “rc: no information after study 2008”
(27) “rc: end of survey”
(40) “spell does not exist”
(50) “not censored”
(60) “right+left censored: gap”

Comment: The variable “censoring” gives information for each spell if all necessary information is available or if the spell is (left or right) censored, and why. If some information is missing, it is indicated by “Gap.” There are several instances when one or several facts are missing. For example, not every person reports a valid year of birth or year of immigration. In addition, sometimes we are able to report that a person is deceased but not the year of death. In these cases, a spell called ”Gap” is inserted between the last information which is certain (e.g., exit from SOEP) and the latest point in time by which the change in status must have happened. (For example, death must have occurred before the study was conducted, so the year of the study is the year of death; between the last SOEP participated in and the year of the study is the gap spell for which no information is certain except that the respondent’s death must have happened during this period.)

INFO

Var label: INFO “**Origin of information of spell**”

Value label:

(1) “SOEP survey”
(2) “SOEP fieldwork”
(3) “1992 study”
(4) “2001 study”
(5) “2006 study”
(6) “2008 study”
(7) “no info”

Comment: The variable indicates the information source for a spell. This can be a retrospective self-report (SOEP survey), fieldwork information provided by the interviewer (SOEP fieldwork), or a register-based drop-out study (1992, 2001, 2006, and 2008 studies).

STUDY$$$$

Var label: STUDY$$$$ “**Study $$$$**”
Value label:

(1) “not in study”
(3) “found in study”
(4) “not found in study”

Comment: The variable indicates if a person was part of the drop-out study in $$$$. A further distinction is made whether or not the person could be identified in register data.

FLAG

Var label: FLAG **"Spell information deviates from other SOEP information"**

Comment: In case of conflicting information between the drop-out studies and SOEP-based information, we give higher priority to the information from the drop-out studies. If two or more studies provide different information, that from the latest study is used. In just a few cases (110), the data set includes spells that contain information deviating from other SOEP data ("flag"). This happens, for instance, if register information says persons emigrated one year prior to refusal to take part in the SOEP. In this case, both spells end/begin in the same year with the register information being corrected by one year. A change in the original register information is coded 1 on the flag variable.

The following two tables illustrate the data structure for two individuals in the SOEP. Both individuals were born abroad, moved to Germany, became part of the SOEP sample, refused to participate after several waves, and finally deceased. While in the first case, full information is available (no censoring), in the second case the exact year of immigration and the exact year of death are missing and therefore "gaps" had to be added (left and right censoring).

persnr	spellnr	spelltype	begin	end	censoring	info
7016603	1	Living abroad	1903	1989	not censored	SOEP survey
7016603	2	Living in Germany	1990	1993	not censored	SOEP survey
7016603	3	SOEP	1994	1999	not censored	SOEP fieldwork
7016603	4	Living in Germany	2000	2003	not censored	2008 study
7016603	5	Deceased	2004	2004	not censored	2008 study

persnr	spellnr	spelltype	begin	end	censoring	info
2808901	1	Living abroad	1927	1927	rc: no year of immigration	SOEP survey
2808901	2	Gap	1928	1998	right+left censored (gap)	SOEP survey
2808901	3	Living in Germany	1999	1999	lc: no year of immigration	SOEP survey

2808901	4	SOEP	2000	2001	not censored	SOEP fieldwork
2808901	5	Living in Germany	2002	2002	rc: no exact year of death	2008 study
2808901	6	Gap	2003	2007	right+left censored (gap)	No info
2808901	7	Deceased	2008	2008	lc: no exact year of death	2008 study

Users less familiar with analyzing spell data can easily convert the file into a long-format file (person x year). This is an example of STATA coding:

```
use lifespell.dta, clear
gen spellduration=(end-begin)+1
expand spellduration
bysort persnr spellnr: gen n=_n
gen year = begin+n-1
move year spellnr
keep persnr year spellnr spelltyp  zensor info study1992 study2001 study2006 study2008 flag
compress

save lifelong.dta, replace
```

Index

$

3

A

B

C

E

F

G

H

I

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z