

DIW Berlin



Deutsches Institut
für Wirtschaftsforschung

SOEP

The German
Socio-Economic
Panel Study

**Biography and Life History Data
in the German Socio Economic Panel
(up to wave V, 2005)**

Editors:

Joachim R. Frick
Jürgen Schupp

Berlin, June 2006

Contents

| | |
|--|-----|
| 1. General Introduction..... | 2 |
| 2. Biographical Information in the Meta File PPFAD (Month of Birth, Year of Death, Immigration Variables, Living in East or West Germany in 1989) | 12 |
| 3. Activity Biography in the File PBIOSPE..... | 24 |
| 4. BIOJOB: Detailed Information on First and Last Job..... | 28 |
| 5. The Biography of Family Status and the Generated Current Family Status (BIOMARSY, BIOMARSM and \$FAMSTD) | 58 |
| 6. BIOBIRTH – A Data Set on the Birth Biography of Female Respondents | 69 |
| 7. BIOBRTHM – The Birth Biography of Male Respondents in the SOEP..... | 80 |
| 8. BIOTWIN – Information on Twins in the SOEP | 84 |
| 9. BIOIMMIG: Generated and Status Variables from SOEP for Foreigners and Migrants | 86 |
| 10. BIOPAREN: Biography Information for the Parents of SOEP-Respondents..... | 131 |
| 11. BIOSOC: Retrospective Data on Youth and Socialization | 144 |
| 12. BIORESID: Variables On Occupancy and Second Residence | 149 |
| 13. BIOAGE17: The Youth Questionnaire | 152 |
| 14. BIOAGE01: Generated Variables from the ‘Mother and Child Questionnaire’ | 161 |
| 15. BIOAGE03: Generated Variables from the supplementary questionnaire “Your child between the ages of two and three”..... | 164 |

1. General Introduction

Based on a set of questions on the individual life course that are compiled 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, on the first job, on social background and on immigration history. The function of this data is, on the one hand, to make important background information available for analyses (i.e. information on fertility as an explanatory variable when analyzing labor market supply of women), and on the other hand, to support self-contained analyses (i.e. on occupational careers or intergenerational transmission of education).

In general, each respondent of the SOEP questionnaire (surveying age starts at 16 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. Here it should be noted that the main objective of surveying the biographical information at the very first interview is, due to the costs involved and other reasons, not consequently continued; for example, in sample C (East Germany, field work started in 1990) the biographical questionnaire was first collected in 1992. Consequently the surveyed persons who left SOEP before 1992 or who refused to complete the biography questionnaire in 1992 have no biographical information included in the SOEP data. In addition, due to survey-related reasons biographical information is not asked of persons who were 16 or 17 years old at the time of the first survey. For this group of persons the biographical information (i.e. marital status and family information, occupation history since the age of 16 and social background) can generally be reconstructed from data taken from the yearly survey (variables from the Individual Questionnaire).

Since 2000, a separate youth questionnaire has been in circulation in which questions are directed specifically towards youth issues. In addition, biographical information is collected on social background, more specifically on parents, in the event that only one (or non) parent lives in the same household and for those who immigrated to Germany.

A completely new instrument for collecting biographical data was implemented in 2003. The target of the questionnaire is to collect information about newborn children and is aimed at their mothers. 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 completely new data with the next release. In 2005, a follow-up questionnaire targeted at children aged 2 to 3 years was implemented. Again, information was collected from the mothers 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 relevant decisions for their individual development (e.g. around ages 6 and 12)¹.

A chronological listing of the various changes related to the survey of biographically relevant information for the time period 1984 to 2005 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' is in 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, housework husband/wife, retired and other; question 62 in the standard Individual Questionnaire in Wave 1).
- 1985 The focus for samples A and B was on retrospectively collecting marital and family status information (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.
- 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 professionally-related 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

¹ **Help for 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 4 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).

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 who were included in this retrospective questioning.
- 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 were collected later on.
- Since 1991 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 the four samples A, B, C, and D1/D2.² The formats of some of the questions were slightly changed, and new questions were added, although

² (A) 'West Germany', (B) 'Guest Workers / Foreigners', (C) 'East Germany', (D1/D2) 'Immigrants since 1984', persons from D2 were first surveyed in 1995.

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, the immigration relevant questions were also given to persons in samples A to C in an identical form.
- 1998 Introduction of the supplementary sample E.
- 1999 The 1996 official form of the Biography Questionnaire 'Lebenslauf' ('life history') was given to all participants 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 children of your own is collected for men too, 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 invented 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 (\approx 3,850 €). This sample was also asked retrospective information on inheritances, which was collected in 2001 for samples A through F.
- 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 'living region before reunification (East or West Germany)'. 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).

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 up to SKMUTTI from the file AKIND to SKIND. '\$\$' stands for the general entry of a year (2 digits) and is used as a suffix: for example, NATION\$\$ stands for NATION84 to NATION02 from the files APGEN to SPGEN.
- The file BIOLELA is mentioned quite often 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 quite complex due to the differences in the surveying procedures mentioned above and is therefore the central input for nearly all of the following variables; however BIOLELA doesn't contain information which is 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 suitable for only 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 to interested users of the SOEP data at request.

The following table displays in a general overview all of the biographical information as surveyed in the Biography Questionnaire 'Lebenslauf' ('life history') in 2004 and the current version (June 2006) of the user-friendly editing of this information. The designated numbers in the Biography Questionnaire 'Lebenslauf' ('life history') refer to the 2004 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.

Table 1: Biographical data in SOEP

| Biography Sub-area | Number of Question in the 'Lebenslauf' Questionnaire (2004) | <i>Comparable Questions in the Youth Questionnaire (2004)</i> | SOEP Target Population | Files in the SOEP Database | Analysis Unit | Update Requirements (Source File for Update) | Status: Available / Not Available (up to Wave V) |
|---|---|---|---------------------------------|----------------------------|---------------|--|--|
| Place of birth | 2, 3 | 55 | All persons surveyed | PPFAD | Individual | No | Available |
| Year of immigration | 4 | 58 | For persons not born in Germany | PPFAD | Individual | No | Available |
| Immigration biography | 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 15a | 57, 59, 60, 61, 62, 63, 64, 65, 65, 66a, 66b | 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 | 67a, 67b | All persons surveyed | BIOPAREN | Individual | No | Available |
| Number of brothers and sisters | 18 | 68 | All persons surveyed | | Individual | Yes | Not available |
| Parents living region, year of birth, year of death, nationality | 21, 22, 23 | 71, 72, 73 | All persons surveyed | BIOPAREN | Individual | Partly (year of death from PPFAD) | Available |

| | | | | | | | |
|---|----------------------------|----------------|----------------------|----------|------------|---|-----------|
| Parents' school + occupational degree, their job + occupation as respondent was 15 years old | 24, 25, 26, 27 | 74, 75, 76, 77 | All persons surveyed | BIOPAREN | Individual | No | Available |
| Religious affiliation of parents | 28 | 78 | All persons surveyed | BIOPAREN | Individual | No | Available |
| Parents took care about efforts at school | 29 | 39 | All persons surveyed | BIOSOC | Individual | No | Available |
| Respondent's last school marks | 30 | 35 | 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, 20, 21, 24 | All persons surveyed | BIOSOC | Individual | No | Available |
| Occupational biography | 36 | - | All persons surveyed | PBIOSPE | Spell | Yes (\$P) | Available |
| Year and place of acquiring a school degree | 37, 38, 41 | 26a | All persons surveyed | BIOSOC | Individual | No (although possible using \$P) | Available |
| Level of school degree | 39, 40, 42 | 26b | All persons surveyed | \$PGEN | Individual | Yes (\$P) | Available |
| Number of foreign classmates in last attended school class | 43 | 43 | All persons surveyed | BIOSOC | Individual | No | Available |
| Target school degree | 44, 45 | 27, 28 | All persons surveyed | BIOSOC | Individual | No | Available |
| Attained vocational degree, year and place of attaining, certificate of degrees attained abroad | 46, 47, 48, 49, 50, 51, 52 | 44, 45 | All persons surveyed | BIOSOC | Individual | No (although partly possible using \$P) | Available |

| | | | | | | | |
|--|---|--------|---|----------------------|------------|--|---|
| Target vocational degree | 53, 54 | 46, 47 | All persons surveyed | BIOSOC | Individual | No | Available |
| First job (age, occupational position, public sector, branch,) | 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 | Only available as given in interview; not updated yet |
| 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 |
| Military or alternative community service (only men) and voluntary service | 73, 74 | - | All persons surveyed | BIOSOC | Individual | No (although partly possible using \$P) | Available |
| 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 |

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, Lena Jacobi and Christian Schmitt

The file PPFAD also includes, among other more survey related variables like responding status, most important demographical information for each person who 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 variables on the year of death (TODJAHR and TODINFO), on the year of 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, the construction of these generated variables will be explained briefly.

2.1 Month of birth

Due to the more recent incorporation of adequate questions on the month of birth in the various SOEP survey instruments, the data set PPFAD contains – in addition to the year of birth (GEBJAHR) – the month of birth (GEBMONAT); target population for this information is made up by all individuals independent of their respective survey status as respondent or child. GEBMONAT 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
 3 derived from data set \$KIND
 4 derived from data set SP
 5 derived from data set \$LELA

The month of birth was asked from SOEP respondents for the first time in wave S individual questionnaire (SP). Furthermore, starting in wave T, the month of birth is being asked in the biography questionnaire (\$LELA). For children up to age 16, the month of birth has been recorded in the household questionnaire of several waves, yielding corresponding data for the file \$KIND: An initial record of the month of birth exists in the file DKIND in 1987. This record was updated in 2003 within TKIND.

While all these sources provide the relevant information for most of the more recent panel members, the information remains missing for some persons, including temporary dropouts or

people who exited in earlier waves. For some of them the month of birth could be reconstructed. This reconstruction remains an approximation and might differ from the true month of birth in individual cases.

Construction of variable

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

- Generated (basis: \$P, \$PBRUTTO \$KIND)
- \$KIND
- SP
- \$LELA

whereas the latter information overrides the former.

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

The generated month of birth could only be reconstructed 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, TLELA or TKIND – the data generation is almost always congruent with the collected data and therefore has proven to be reliable.

2.2 Year of death

Variable TODJAHR Year of death -4 digits-

The Variable TODJAHR contains the four digit year entered as the year of death.

Remarks

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

- (a) from the drop-out file YPBRUTTO, that is, the outcome of the yearly field work
- (b) within the scope of the Infratest-Verbleibstudie (Study conducted by Infratest to follow-up on drop-outs) from 1992
- (c) within the scope of the Infratest-Verbleibstudie (Study conducted by Infratest to follow-up on drop-outs) from 2001

Missing codes

- (-2) Persons currently living or who no longer exist 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 recorded. Furthermore, within the framework of, up to now, two 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 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 “Verbleibstudie” and the information from YPBRUTTO, in principle the information from the “Verbleibstudie” was used.

Variable TODINFO Year of death – source of information

Remarks

- 1 'from continued surveying (YPBRUTTO)'
- 2 'Infratest-Verbleibstudie (Follow-up Study) 1992'
- 3 'Infratest-Verbleibstudie (Follow-up Study) 2001'

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

2.3 Immigration information

Variables IMMIYEAR, GERMBORN and CORIGIN

The objective of these variables, based on all the persons who have been a part of SOEP (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 persons which were not followed up in 2003.

Overview 1: Distribution of the Immigration Information according to SOEP-Samples A to G (Version: Wave V, 2005)

| GERMBORN | Sample | | | | | | | Total |
|--|----------------|---------------|---------------|--------------|---------------|----------------|---------------|----------------|
| | A | B | C | D | E | F | G | |
| No Answer (Codes -1 and -3) | 177 1.3 | 103 2.0 | 14 0.2 | 18 1.2 | 25 1.1 | 30 0.2 | 721 24.8 | 1,088 2.4 |
| Born in Germany or immigrated before 1949 (Code „1“) | 13,015 95.7 | 1,254 24.0 | 6,389 98.5 | 475 32.4 | 2,031 90.4 | 10,930 88.5 | 2,077 71.4 | 36,171 81.7 |
| Immigrated since 1949 (Code „2“) | 414 3.0 | 3,858 74.0 | 86 1.3 | 974 66.4 | 191 8.5 | 1,397 11.3 | 110 3.8 | 7,030 15.9 |
| Total | 13,606 100 | 5,215 100 | 6,489 100 | 1,467 100 | 2,247 100 | 12,357 100 | 2,908 100 | 44,289 100 |

Source: All survey participants with at least one SOEP interview from 1984 to 2005 (n=44,289).

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 realized households; 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 personal data questionnaire or from the course of the SOEP installed variations of the “Biography” and “Lebenslauf” (curriculum vitae) questionnaires:

- \$P (wave specific survey data from personal data 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)

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 effects 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 portion of missing entries for immigration relevant variables, for persons who *have* corresponding information 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 | Lives in Germany since...(Filter) | A B |
| GP10802 | GP | Lives in Germany since...(Year) | A B |
| GP10803 | GP | Region of origin | A B |
| GP106A01 | GPAUSL | Immigration to Germany (Year) | B |
| HP108B01 | HP | Lived in Germany since...(Filter) | A B |
| HP108B02 | HP | Lived in Germany since...(Year) | A B |
| HP108B03 | HP | Geographical area of origin | A B |
| HP106A01 | HPAUSL | Year moved to Germany | B |
| IP10801 | IP | Year moved to East/West Germany | A B C |
| IP10802 | IP | Since when.. | A B C |
| IP10803 | IP | Where lived before | A B C |
| IP106A01 | IPAUSL | Year immigrated to Germany | B |
| JP108B01 | JP | Since when in East/West Germany | A B C |
| JP108B02 | JP | Since which year in East/West Germany | A B C |
| JP108B03 | JP | Where lived before | A B C |
| JP106A01 | JPAUSL | Year immigrated to Germany | B |
| LPGRUPPE | LPBRUTTO | Immigration group | D |
| LPHERKFT | LPBRUTTO | Country of origin | D |
| MP90A01 | MP | German | A B C D |
| MP90A02 | MP | Immigrant of German descent from Eastern Europe | A B C D |
| MP90A03 | MP | Have already lived in Germany since 1984 | A B C D |
| NP101 | NP | Born in Germany | A B C D |
| NP102 | NP | Have already lived in Germany since 1984 | A B C D |
| OP103 | NP | Born in Germany | A B C D E |
| OP104 | NP | Have already lived in Germany since 1984 | A B C D E |
| PP117 | NP | Born in Germany | A B C D E |
| PP118 | NP | Have already lived in Germany since 1984 | A B C D E |
| QP12101 | QP | German citizenship: Born in Germany? | A B C D E F |
| QP12102 | QP | Country of birth? | A B C D E F |
| QP122 | QP | Have already lived in Germany since 1984 | A B C D E F |
| QP124 | QP | Foreign citizenship: Born in Germany? | A B C D E F |
| RP11701 | RP | German citizenship: Born in Germany? | A B C D E F |
| RP11702 | RP | Country of birth? | A B C D E F |
| RP118 | RP | Have already lived in Germany since 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 | Where lived prior to German unification 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 |
| P031Z | BIOLELA | Year of immigration to Germany | D |
| P051Z | BIOLELA | Year of immigration to Germany | D |
| P060Z | BIOLELA | Since immigration, always lived in Germany? | D |
| P04Z | BIOLELA | Country of birth | D |
| P021Z | BIOLELA | Lived where in 1984? | D |
| B34 | BIOLELA | Where lived before | A B C |
| B36 | BIOLELA | Country of birth | A B C |
| B37 | BIOLELA | Year of immigration to Federal Republic of Germany | A B C |
| \$B03Z | \$LELA | Country of birth | A B C D E F G |
| \$B02 | \$LELA | Born in Germany | A B C D E F G |
| \$B041Z | \$LELA | Before 1984 – Immigration year to FRG | A B C D E F G |
| \$B042Z | \$LELA | After 1984 – Immigration year to FRG | A B C D E F G |
| 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 |
| \$J56 | \$JUGEND | Born in Germany? | A B C D E F G |
| \$J57 | \$JUGEND | Country of birth | A B C D E F G |
| \$J58 | \$JUGEND | Year of immigration to Federal Republic of Germany | A B C D E F G |
| 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 only used when no other valid information is available. | A B C D E F G |

In the following, 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?

Remarks

- 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-responders
- 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 GERMBORN relevant variables (The process follows the sequence mentioned 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 in the address protocol in the first survey wave a foreign country is entered as the country of origin (variable LPHERKFT). The variable CORIGIN is also filled with the respective value.
- 1.2 Persons for whom a valid entry in the variable IMMIYEAR exists and who over the course of their survey in SOEP have indicated at least once a non-German nationality. This nationality is alternatively used to generate the variable CORIGIN.
- 1.3 Persons whose mother has immigrated after having given birth to this individual, as long as the mother's immigration occurred no more 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 were basically always living 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 household (as given by HHNR) immigrants reside CORIGIN is given the corresponding value of the nationality; IMMIYEAR remains as a rule as missing ("-1").

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

- 2.1 Persons from sample D for whom in the address protocol in the first survey wave Germany is entered as the country of origin.
- 2.2 Persons whose mother was born in Germany or had immigrated prior to giving birth.
- 2.3 Minor Children who live in a household with no immigrants.
- 2.4 Persons from samples A, C, E, F and G 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 such 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 be made based only on plausibility and by 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 which are 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)

Remarks

1949 ...2005 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 is derivable from all of the original variables

Persons, who have been identified as immigrants and from whose original variables no immigration year could be determined, are assigned as a resource the value

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 has immigrated up until the year the person turned 18 years of age.

Variable CORIGIN Country of origin

Remarks

| | |
|----------|---|
| 1 | Germany |
| 2 to 151 | Turkey, (Ex-)Yugoslavia, Greece, Italy, Spain, Yemen |
| 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 from the original variables (code “2” in GP10803 to JP108B03) it is evident that the person in question immigrated from Eastern Europe, respectively from the former Eastern European German territories. Here this also includes, under certain circumstances, a small number of persons from sample D if they were identified in P070Z as person 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 up until the year the person turned 18 years of age.

2.4 Living in East or West Germany 1989

The variable LOC1989 in the meta-file PPFAD provides information about the geographic area a person lived in *prior* to German unification, 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 = 1 or 2).

Variable LOC1989 “Where did you live in 1989 ?”

Remarks:

- 1 East Germany (German Democratic Republic [DDR] einschl. Berlin-Ost)
- 2 West Germany (Federal Republic of Germany
 [Bundesrepublik Deutschland] einschl. Berlin-West)
- 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 2005, 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 have been assigned parental information, if at all available.

LOC1989 Where did you live in 1989?

| LOC1989 | count row pct col pct | at least one interview (1984-2005) | | Total |
|---|-----------------------------|---------------------------------------|---------------------------|----------------------------|
| | | no | yes | |
| -2 | | 8,799 100.00 70.16 | 0 0.00 0.00 | 8,799 100.00 15.48 |
| -1 | | 1,057 19.67 8.43 | 4,317 80.33 9.75 | 5,374 100.00 9.46 |
| 1 East Germany / DDR incl. Berlin-Ost | | 525 5.44 4.19 | 9,126 94.56 20.61 | 9,651 100.00 16.98 |
| 2 West Germany / Bundesrepublik incl. Berlin-West | | 2,076 6.63 16.55 | 29,246 93.37 66.04 | 31,322 100.00 55.12 |
| 3 Abroad / Ausland | | 84 4.99 0.67 | 1,599 95.01 3.61 | 1,683 100.00 2.96 |
| Total | | 12,541 22.07 100.00 | 44,288 77.93 100.00 | 56,829 100.00 100.00 |

Source: Population in PPFAD as of wave V (2005).

3. Activity Biography in the File *PBIOSPE*

by Rainer Pischner

The spell file *PBIOSPE* is based on the information of the professional life which is collected in the Biography-Questionnaire from every interviewed person as a matrix. The observations start with the age of 15 and end with the actual age (up to age 65). This information on occupational status is only valid up to the time the biography is collected (variable *ERHEBJ*). To update the on-going occupational career in *PBIOSPE*, information of the yearly Individual-Questionnaire is also used. In this Questionnaire the occupational status is always asked for every month of the previous year. Therefore, the information which is collected on a monthly basis and stored in the file *ARTKALEN* is aggregated into yearly values.³ For persons who temporarily could not be questioned, it is sometimes possible to fill the gaps in their occupational status. If these persons fill out the additional questionnaire for temporary drop outs later on, then we can use the information collected there (see files *\$PLUECKE*).

The information on the occupational status is transformed into spells, whereby each spell is defined through the duration of a given status (e.g., unemployment). The starting and ending time of a spell are given by the respondent's age as well as by the calendar year. The information listed in the following table is found in detail in the file *PBIOSPE*.

Of special interest is the variable *SPELLTYP* which gives the occupational status of a person at a single age. The corresponding specifications of the professional status are also listed in the table.

| Serial number | Label (Spelltyp) | Label (Spell Type) |
|---------------|-----------------------|---------------------------|
| (1) | Schule, Studium | School/University |
| (2) | Lehre, Ausbildung | Apprenticeship/Training |
| (3) | Wehr-, Zivildienst | Military/Civilian service |
| (4) | Voll berufstaetig | Full-time employed |
| (5) | Teilzeit beschaeftigt | Part-time employed |
| (6) | Arbeitslos | Unemployed |
| (7) | Hausfrau, Hausmann | House-Husband/Wife |
| (8) | Im Ruhestand | Retired |
| (9) | Andere Tätigkeit | Other |

³ 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.

The following table shows by an example the resulting spell system of some SOEP-respondent using the procedure mentioned above.

Print out of the first cases in PBIOSPE

| HH NR | PERS NR | SPELL NR | SPELL TYP | BEGIN | END | BEGIN BIO | END BIO | BEGIN KAL | END KAL | BEGIN Y | END Y | BEGIN YB | END YB | BEGIN YK | END YK | ZEN SOR | SPELL INF | ERHEBJ | FEHL CODE |
|-------|---------|----------|-----------|-------|-----|-----------|---------|-----------|---------|---------|-------|----------|--------|----------|--------|---------|-----------|--------|-----------|
| 19 | 101 | 1 | 2 | 15 | 18 | 15 | 18 | -2 | -2 | 1945 | 1948 | 1945 | 1948 | -2 | -2 | 2 | 1 | 84 | 0 |
| 19 | 101 | 2 | 4 | 19 | 58 | 19 | 54 | 53 | 58 | 1949 | 1988 | 1949 | 1984 | 1983 | 1988 | 3 | 3 | 84 | 0 |
| 19 | 102 | 1 | 1 | 15 | 15 | 15 | 15 | -2 | -2 | 1955 | 1955 | 1955 | 1955 | -2 | -2 | 2 | 1 | 84 | 0 |
| 19 | 102 | 2 | 4 | 16 | 22 | 16 | 22 | -2 | -2 | 1956 | 1962 | 1956 | 1962 | -2 | -2 | 1 | 1 | 84 | 0 |
| 19 | 102 | 3 | 7 | 23 | 48 | 23 | 44 | 43 | 48 | 1963 | 1988 | 1963 | 1984 | 1983 | 1988 | 3 | 3 | 84 | 0 |
| 19 | 103 | 1 | 1 | 15 | 16 | 15 | 16 | -2 | -2 | 1978 | 1979 | 1978 | 1979 | -2 | -2 | 2 | 1 | 84 | 0 |
| 19 | 103 | 2 | 2 | 17 | 20 | 17 | 20 | -2 | -2 | 1980 | 1983 | 1980 | 1983 | -2 | -2 | 1 | 1 | 84 | 0 |
| 19 | 103 | 3 | 4 | 20 | 23 | 21 | 21 | 20 | 23 | 1983 | 1986 | 1984 | 1984 | 1983 | 1986 | 3 | 3 | 84 | 0 |
| 27 | 201 | 1 | 1 | 15 | 19 | 15 | 19 | -2 | -2 | 1941 | 1945 | 1941 | 1945 | -2 | -2 | 2 | 1 | 84 | 0 |
| 27 | 201 | 2 | 2 | 20 | 24 | 20 | 24 | -2 | -2 | 1946 | 1950 | 1946 | 1950 | -2 | -2 | 1 | 1 | 84 | 0 |
| 27 | 201 | 3 | 4 | 25 | 29 | 25 | 29 | -2 | -2 | 1951 | 1955 | 1951 | 1955 | -2 | -2 | 1 | 1 | 84 | 0 |
| 27 | 201 | 4 | 7 | 30 | 39 | 30 | 39 | -2 | -2 | 1956 | 1965 | 1956 | 1965 | -2 | -2 | 1 | 1 | 84 | 0 |
| 27 | 201 | 5 | 7 | 57 | 58 | -2 | -2 | 57 | 58 | 1983 | 1984 | -2 | -2 | 1983 | 1984 | 1 | 2 | 84 | -2 |
| 27 | 201 | 6 | 8 | 40 | 65 | 40 | 58 | 57 | 65 | 1966 | 1991 | 1966 | 1984 | 1983 | 1991 | 3 | 3 | 84 | 16 |
| 27 | 202 | 1 | 1 | 15 | 25 | 15 | 25 | -2 | -2 | 1971 | 1981 | 1971 | 1981 | -2 | -2 | 2 | 1 | 87 | -2 |
| 27 | 202 | 2 | 1 | 28 | 28 | -2 | -2 | 28 | 28 | 1984 | 1984 | -2 | -2 | 1984 | 1984 | 1 | 2 | 87 | -2 |
| 27 | 202 | 3 | 4 | 27 | 31 | 27 | 31 | 28 | 30 | 1983 | 1987 | 1983 | 1987 | 1984 | 1986 | 3 | 4 | 87 | -2 |
| 27 | 202 | 4 | 5 | 26 | 26 | 26 | 26 | -2 | -2 | 1982 | 1982 | 1982 | 1982 | -2 | -2 | 1 | 1 | 87 | -2 |
| 27 | 202 | 5 | 5 | 30 | 30 | -2 | -2 | 30 | 30 | 1986 | 1986 | -2 | -2 | 1986 | 1986 | 1 | 2 | 87 | -2 |
| 27 | 202 | 6 | 6 | 29 | 29 | 29 | 29 | -2 | -2 | 1985 | 1985 | 1985 | 1985 | -2 | -2 | 1 | 1 | 87 | -2 |
| 27 | 202 | 7 | 7 | 29 | 29 | -2 | -2 | 29 | 29 | 1985 | 1985 | -2 | -2 | 1985 | 1985 | 1 | 2 | 87 | -2 |

The variables BEGIN and END indicate the beginning and the end of a spell, the variables BEGINBIO and BEGINKAL as well as ENDBIO and ENDKAL indicate the “source” of the utilized original information (one-time BIOgraphy-Questionnaire or (KAL) calendar information from the yearly survey). The variables described here are age entries. Analogously there is the block BEGINY and ENDYK. With this block the corresponding Spell data is not longer only in the form of the age, but also coded with the calendar year. Y stands for Year, B for BIO and K for KAL.

Missing information on the beginning or the end of a spell causes so called censoring problems. If both time points are known, then the spell is not censored (1); otherwise the spell is identified as ‘left censored’ (2) when the beginning time is missing, and as ‘right censored’ (3) when the ending time is missing (see the variable ZENSOR).

For surveyed person number 201 in household 27, information over the total life span collected here is available (BEGIN and END). The first activity (SPELLNR=1) was between the ages of 15 to 19 respectively from 1941 to 1945 (BEGINY and ENDY); the respondent went to school (SPELLTYP=1). This information was taken from the Biography-Questionnaire (BEGINBIO=15 and ENDBIO=19; BEGINYB and ENDYB respectively), which was conducted in the year 1984 (ERHEBJ=84). In the ‘calendar for occupations’ there is no relevant information found (BEGINKAL=-2). The information is “left censored” (ZENSOR=2) since there is no information available before the age of 15.

As the second activity (SPELLNR=2) of the respondent Apprenticeship/Training (SPELLTYP=2) between the ages of 20 and 24 has been entered. This data was also found in the Biography-Questionnaire of the year 1984.

Interesting to point out are the 4th, 5th and 6th activities of respondent number 201. For the ages of 30 to 39 the respondent entered the occupational status “Home-Husband/Wife” (SPELLTYP=7) in the Biography-Questionnaire. This information is not censored (ZENSOR=1), that is, status information before and after this activity is available. The 5th spell-activity from the age 57 to 58 is taken from the ‘calendar of occupation’ (SPELLTYP=7; Home-Husband/Wife); in spite of the overlapping time periods this information was not contained in the Biography-Questionnaire. Here the respondent merely specified (SPELLNR=6) that from age 40 to 58 he/she was retired. This status was extended up to the age of 65 using the entries in the ‘calendar of occupation’ of the following years. Thus this activity is ‘right censored’.

The variable SPELLINF names the source of information for the single spells: all spells which are based on the answers given in the biographical schema are encoded with SPELLINF=1, those from the calendar with 2, and combined spells with information from the biographical schema and the calendar (in the above mentioned example SPELLNR=6) are encoded with SPELLINF=3.

The variable FEHLCODE identifies eventual problems by the generation of spells due to missing information for certain/some ages (see the value labels for FEHLCODE).

4. BIOJOB: Detailed Information on First and Last Job

by Tanja Schmidt

(based on work of 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.

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).

Contents of BIOJOB:

Population: All persons with an entry in any LELA-/YOUTH-file up to 2005, even if information on employment is missing.

number of cases: 41174 *waves:* A(84) - V(05) *samples:* A, B, C, D, E, F,G

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
- (-2) does not apply

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 2004), 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 2004), 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 (05) | - | 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 (84-05) | 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 Questionnaires 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 Questionnaires 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 (05) | - | 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), LJSELF (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 | LJSELF | LJWHITE | LJCIVS |
|----------------------------------|--------|--------|---------|--------|
| Sample A,B,C,D, E,F,G (94-05) | 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

LJSELF/LJSEFSIZ

The LJSELF 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 seldomly). 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 (internal DIW files with biographical information up to wave V), QJUGEND, RJUGEND, SJUGEND, TJUGEND, UJUGEND, VJUGEND (internal DIW youth biography files), QP, RP, SP, TP, UP and VP (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, LJSELF, 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 LJSELF, 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

4.4 Results

Overall, 41174 people have an entry in BIOJOB. The following tables display descriptive statistics of all variables of BIOJOB.

AGEFJOB

| | |
|-----------------------|-------|
| Number of valid cases | 36334 |
| Mean | 19,53 |
| Standard Deviation | 3,89 |

| | |
|-----------|----|
| Quantiles | |
| 100% Max | 62 |
| 75% Q3 | 21 |
| 50% Med | 19 |
| 25% Q1 | 17 |
| 0% Min | 6 |
| Mode | 18 |

| | Frequency | Percent | Cumulative Percent |
|---------------------|-----------|---------|-----------------------|
| (-1) No answer | 1368 | 3,3 | 91,6 |
| (-2) Does not apply | 3472 | 8,4 | 100,0 |

NOJOB

| | Frequency | Percent | Cumulative Percent |
|-------------------------|-----------|---------|-----------------------|
| (1) Never been employed | 3779 | 9,2 | 9,2 |
| (-1) no answer | 1368 | 3,3 | 12,5 |
| (-2) does not apply | 36027 | 87,5 | 100,0 |

AGEINFO

| | Frequency | Percent | Cumulative Percent |
|-------------------------------------|-----------|---------|-----------------------|
| (1) LELA-files | 27759 | 67,4 | 67,4 |
| (2) PBIOSPE agefjob<15 begin>15 | 814 | 2,0 | 69,4 |
| (3) PB not worked+later begin | 2449 | 5,9 | 75,8 |
| (4) PB n.w.+earlier begin | 885 | 2,1 | 77,5 |
| (5) PB agefjob>27+earlier begin | 392 | 1,0 | 78,4 |
| (6) Inconsistent info | 54 | 0,1 | 78,6 |
| (7) PB n.w.,agefjob not answ. | 2337 | 5,7 | 84,3 |
| (8) PB n.w.,agefjob,spl n.a. | 1266 | 3,1 | 87,3 |
| (9) Completely missing | 2906 | 7,1 | 94,4 |
| (10) SP no info in LELA+PBIOSPE | 4 | 0,0 | 94,4 |
| (11) Youth files | 6 | 0,0 | 94,4 |
| (12) PB Youth info+earlier begin | 6 | 0,0 | 94,4 |
| (13) No reg.emp.(youth)+later begin | 91 | 0,2 | 94,4 |
| (14) \$P inval.info+later empl. (y) | 11 | 0,0 | 94,7 |
| (15) \$P empl. n.a.+ spl (youth) | 0 | 0,0 | 94,7 |
| (16) missing (youth) | 1923 | 4,7 | 99,3 |
| (17) Inconsistent info (youth) | 264 | 0,6 | 100,0 |
| (18) SP no info YOUTH+PBIOSPE | 7 | 0,0 | 100,0 |

OCCFJOB

| | Frequency | Percent | Cumulative Percent |
|-------------------------|-----------|---------|-----------------------|
| (1) blue collar worker | 14863 | 36,1 | 36,1 |
| (2) self-employed | 765 | 1,9 | 38,0 |
| (3) white collar worker | 11264 | 27,4 | 65,4 |
| (4) civil servant | 1047 | 2,5 | 67,9 |
| (-1) no answer | 9456 | 23,0 | 90,9 |
| (-2) does not apply | 3779 | 9,1 | 100,0 |

FULLTIME

| | Frequency | Percent | Cumulative Percent |
|------------------------------------|-----------|---------|-----------------------|
| (0) part-time work / marginal emp. | 4205 | 10,2 | 10,2 |
| (1) fulltime work | 32350 | 78,6 | 88,8 |
| (-1) no answer | 1224 | 3,0 | 91,8 |
| (-2) does not apply | 3395 | 8,2 | 100,0 |

FJBLUE

| | Frequency | Percent | Cumulative Percent |
|----------------------------------|-----------|---------|-----------------------|
| (10) un-/semiskilled:farmer(C/D) | 301 | 0,7 | 0,7 |
| (11) unskilled worker | 2935 | 7,1 | 7,8 |
| (12) semiskilled worker | 2781 | 6,8 | 14,6 |
| (20) skilled worker | 8751 | 21,3 | 35,9 |
| (30) forem./master:farmer(C/D) | 14 | 0,0 | 35,9 |
| (31) foreman (A/B) | 21 | 0,0 | 35,9 |
| (32) foreman (C/D) | 11 | 0,0 | 35,9 |
| (40) mastercraftsman | 25 | 0,1 | 36,0 |
| (41) farmers(C/D)mid./high.man | 24 | 0,1 | 36,1 |
| (-1) no answer | 9456 | 23,0 | 59,1 |
| (-2) does not apply | 16855 | 40,9 | 100,0 |

FJSELFE

| | Frequency | Percent | Cumulative Percent |
|-----------------------------|-----------|---------|-----------------------|
| (10) independent farmer | 167 | 0,4 | 0,4 |
| (20) free lance | 130 | 0,3 | 0,7 |
| (30) other self employed | 176 | 0,4 | 1,1 |
| (40) within family business | 292 | 0,7 | 1,8 |
| (-1) no answer | 9456 | 23,0 | 24,8 |
| (-2) does not apply | 30953 | 75,2 | 100,0 |

FJSEFSIZ

| | Frequency | Percent | Cumulative Percent |
|--------------------------------------|-----------|---------|-----------------------|
| (10) fsize <= 9 | 69 | 0,2 | 0,2 |
| (11) fsize = 0 | 89 | 0,2 | 0,4 |
| (12) fsize 1-9 | 63 | 0,2 | 0,6 |
| (20) fsize > 9 | 16 | 0,0 | 0,6 |
| (30) fsize <= 10 (special cases C/D) | 18 | 0,0 | 0,6 |
| (40) fsize > 10 (special cases C/D) | 1 | 0,0 | 0,6 |
| (-1) no answer | 9456 | 23,0 | 23,6 |
| (-2) does not apply | 31462 | 76,4 | 100,0 |

FJWHITE

| | Frequency | Percent | Cumulative Percent |
|------------------------------|-----------|---------|-----------------------|
| (10) industrial foreman | 623 | 1,5 | 1,5 |
| (20) unskilled labour | 4241 | 10,3 | 11,8 |
| (21) as (20), without degree | 73 | 0,2 | 12,0 |
| (22) as (20), with degree | 148 | 0,4 | 12,4 |
| (30) skilled labour | 4663 | 11,3 | 23,7 |
| (40) professional labour | 1494 | 3,6 | 30,3 |
| (50) managerial labour | 22 | 0,1 | 30,4 |
| (-1) no answer | 9456 | 23,0 | 53,4 |
| (-2) does not apply | 20454 | 49,6 | 100,0 |

FJCIVS

| | Frequency | Percent | Cumulative Percent |
|---------------------------------|-----------|---------|-----------------------|
| (10) Low level civil servant | 155 | 0,4 | 0,4 |
| (20) middle level civil servant | 335 | 0,8 | 1,2 |
| (30) high level civil servant | 347 | 0,8 | 2,0 |
| (40) executive civil servant | 210 | 0,5 | 2,5 |
| (-1) no answer | 9456 | 23,0 | 25,5 |
| (-2) does not apply | 30671 | 74,5 | 100,0 |

REQEDUC

| | Frequency | Percent | Cumulative Percent |
|---------------------------------------|-----------|---------|-----------------------|
| (10) No training | 2064 | 5,0 | 5,0 |
| (20) completed vocational training | 6625 | 16,1 | 21,1 |
| (30) voc. college / university degree | 43 | 0,1 | 21,2 |
| (31) vocational college degree | 457 | 1,0 | 22,2 |
| (32) university degree | 850 | 2,1 | 24,3 |
| (-1) no answer | 5085 | 12,4 | 36,7 |
| (-2) does not apply | 26050 | 63,3 | 100,0 |

CIVILSFJ

| | Frequency | Percent | Cumulative Percent |
|---------------------|-----------|---------|-----------------------|
| (1) Yes | 2109 | 5,1 | 5,1 |
| (2) no | 8415 | 20,4 | 25,5 |
| (-1) no answer | 4600 | 11,2 | 36,7 |
| (-2) does not apply | 26050 | 63,3 | 100,0 |

OCCMOVE

| | Frequency | Percent | Cumulative Percent |
|------------------------------|-----------|---------|-----------------------|
| (1) never changed occupation | 13422 | 32,6 | 32,6 |
| (2) Once | 7705 | 18,7 | 51,3 |
| (3) more than once | 5267 | 12,8 | 64,0 |
| (-1) no answer | 5598 | 13,6 | 77,6 |
| (-2) does not apply | 9182 | 22,4 | 100,0 |

AGEATMV

| | |
|-----------------------|-------|
| Number of valid cases | 12583 |
| Mean | 30,70 |
| Standard Deviation | 9,14 |

| | |
|-----------|----|
| Quantiles | |
| 100% Max | 65 |
| 75% Q3 | 36 |
| 50% Med | 29 |
| 25% Q1 | 23 |
| 0% Min | 11 |
| Mode | 25 |

| | Frequency | Percent | Cumulative Percent |
|---------------------|-----------|---------|-----------------------|
| (-1) no answer | 5987 | 14,5 | 45,1 |
| (-2) does not apply | 22604 | 54,9 | 100,0 |

INTEDEDUC1

| | Frequency | Percent | Cumulative Percent |
|--------------------------------|-----------|---------|-----------------------|
| (1) Apprenticeship | 825 | 2,0 | 2,0 |
| (2) Vocational school | 216 | 0,5 | 2,5 |
| (3) Technical school | 362 | 0,9 | 3,4 |
| (4) Education as civil servant | 60 | 0,1 | 3,5 |
| (5) Corporate education | 152 | 0,4 | 3,9 |
| (6) Technical/prof. college | 385 | 0,9 | 4,8 |
| (7) University | 876 | 2,1 | 6,9 |
| (-1) no answer | 462 | 1,2 | 8,1 |
| (-2) does not apply | 37836 | 91,9 | 100,0 |

INTEDEDUC2

| | Frequency | Percent | Cumulative Percent |
|--------------------------------|-----------|---------|-----------------------|
| (1) Apprenticeship | 195 | 0,5 | 0,5 |
| (2) Vocational school | 29 | 0,1 | 0,6 |
| (3) Technical school | 36 | 0,1 | 0,7 |
| (4) Education as civil servant | 19 | 0,0 | 0,7 |
| (5) Corporate education | 26 | 0,1 | 0,8 |
| (6) Technical/prof. college | 81 | 0,2 | 1,0 |
| (-1) no answer | 0 | 0,0 | 1,0 |
| (-2) does not apply | 40788 | 99,1 | 100,0 |

INTEDEDUC3

| | Frequency | Percent | Cumulative Percent |
|--------------------------------|-----------|---------|-----------------------|
| (1) Apprenticeship | 36 | 0,1 | 0,1 |
| (2) Vocational school | 7 | 0,0 | 0,1 |
| (3) Technical school | 9 | 0,0 | 0,1 |
| (4) Education as civil servant | 3 | 0,0 | 0,1 |
| (5) Corporate education | 8 | 0,0 | 0,1 |
| (-1) no answer | 0 | 0,0 | 0,1 |
| (-2) does not apply | 41111 | 99,9 | 100,0 |

INTEEDUC4

| | Frequency | Percent | Cumulative Percent |
|--------------------------------|-----------|---------|-----------------------|
| (1) Apprenticeship | 13 | 0,0 | 0,0 |
| (2) Vocational school | 1 | 0,0 | 0,0 |
| (3) Technical school | 0 | 0,0 | 0,0 |
| (4) Education as civil servant | 0 | 0,0 | 0,0 |
| (-1) no answer | 0 | 0,0 | 0,0 |
| (-2) does not apply | 40462 | 100,0 | 100,0 |

CURREMPL

| | Frequency | Percent | Cumulative Percent |
|---------------------|-----------|---------|-----------------------|
| (1) Yes | 7701 | 18,7 | 18,7 |
| (2) No | 5512 | 13,4 | 32,1 |
| (-1) no answer | 138 | 0,3 | 32,4 |
| (-2) does not apply | 27823 | 67,6 | 100,0 |

YEARLAST

| | |
|-----------------------|---------|
| Number of valid cases | 5391 |
| Mean | 1988,78 |
| Standard Deviation | 12,80 |

| | |
|-----------|------|
| Quantiles | |
| 100% Max | 2005 |
| 75% Q3 | 1997 |
| 50% Med | 1993 |
| 25% Q1 | 1986 |
| 0% Min | 1923 |
| Mode | 2000 |

| | Frequency | Percent | Cumulative Percent |
|---------------------|-----------|---------|-----------------------|
| (-1) no answer | 259 | 0,6 | 13,7 |
| (-2) does not apply | 35524 | 86,3 | 100,0 |

SCOPELJ

| | Frequency | Percent | Cumulative Percent |
|----------------------------------|-----------|---------|-----------------------|
| (1) full-time employed | 4259 | 10,3 | 10,3 |
| (2) part-time employment | 786 | 1,9 | 12,2 |
| (3) marg. / irregular employment | 286 | 0,7 | 12,9 |
| (-1) No answer | 319 | 0,8 | 13,7 |
| (-2) does not apply | 35524 | 86,3 | 100,0 |

CIVILSLJ

| | Frequency | Percent | Cumulative Percent |
|---------------------|-----------|---------|-----------------------|
| (1) Yes | 1383 | 3,3 | 3,3 |
| (2) No | 3987 | 9,7 | 13,0 |
| (-1) No answer | 281 | 0,7 | 13,7 |
| (-2) does not apply | 35523 | 86,3 | 100,0 |

OCCLJOB

| | Frequency | Percent | Cumulative Percent |
|-------------------------|-----------|---------|-----------------------|
| (1) blue collar worker | 2190 | 5,3 | 5,3 |
| (2) self-employed | 274 | 0,7 | 6,0 |
| (3) white collar worker | 2540 | 6,1 | 12,1 |
| (4) civil servant | 314 | 0,8 | 12,9 |
| (-1) no answer | 332 | 0,8 | 13,7 |
| (-2) does not apply | 35524 | 86,3 | 100,0 |

LJBLUE

| | Frequency | Percent | Cumulative Percent |
|----------------------------------|-----------|---------|-----------------------|
| (10) un-/semiskilled:farmer(C/D) | 58 | 0,1 | 0,1 |
| (11) unskilled worker | 413 | 1,0 | 1,1 |
| (12) semiskilled worker | 772 | 1,0 | 3,0 |
| (20) skilled worker | 811 | 2,0 | 5,0 |
| (30) forem./master:farmer(C/D) | 1 | 0,0 | 5,0 |
| (31) foreman (A/B) | 73 | 0,2 | 5,2 |
| (32) foreman (C/D) | 0 | 0,0 | 5,2 |
| (40) mastercraftsman | 60 | 0,1 | 5,3 |
| (41) farmers(C/D)mid./high.man | 2 | 0,0 | 5,3 |
| (-1) no answer | 340 | 0,8 | 6,1 |
| (-2) does not apply | 38635 | 93,9 | 100,0 |

LJSELFE

| | Frequency | Percent | Cumulative Percent |
|-----------------------------|-----------|---------|-----------------------|
| (10) independent farmer | 42 | 0,1 | 0,1 |
| (20) Free lance | 38 | 0,1 | 0,2 |
| (30) other self employed | 145 | 0,4 | 0,6 |
| (40) within family business | 32 | 0,1 | 0,7 |
| (-1) no answer | 349 | 0,8 | 1,5 |
| (-2) does not apply | 40568 | 98,5 | 100,0 |

LJSEFSIZ

| | Frequency | Percent | Cumulative Percent |
|--------------------------------------|-----------|---------|-----------------------|
| (10) fsize <= 9 | 19 | 0,0 | 0,0 |
| (11) fsize = 0 | 81 | 0,2 | 0,2 |
| (12) fsize 1-9 | 92 | 0,2 | 0,4 |
| (20) fsize > 9 | 21 | 0,1 | 0,5 |
| (30) fsize <= 10 (special cases C/D) | 7 | 0,0 | 0,5 |
| (40) fsize > 10 (special cases C/D) | 1 | 0,0 | 0,5 |
| (-1) no answer | 332 | 0,8 | 1,3 |
| (-2) does not apply | 40621 | 98,7 | 100,0 |

LJWHITE

| | Frequency | Percent | Cumulative Percent |
|------------------------------|-----------|---------|-----------------------|
| (10) industrial foreman | 23 | 0,1 | 0,1 |
| (20) unskilled labour | 447 | 1,1 | 1,2 |
| (21) as (20), without degree | 567 | 1,4 | 2,6 |
| (22) as (20), with degree | 931 | 2,3 | 4,9 |
| (30) skilled labour | 451 | 1,1 | 6,0 |
| (40) professional labour | 118 | 0,3 | 6,3 |
| (50) managerial labour | 3 | 0,0 | 6,3 |
| (-1) no answer | 349 | 0,8 | 7,1 |
| (-2) does not apply | 38285 | 92,9 | 100,0 |

LJCIVS

| | Frequency | Percent | Cumulative Percent |
|---------------------------------|-----------|---------|-----------------------|
| (10) low level civil servant | 27 | 0,1 | 0,1 |
| (20) middle level civil servant | 86 | 0,2 | 0,3 |
| (30) high level civil servant | 102 | 0,2 | 0,5 |
| (40) executive civil servant | 99 | 0,2 | 0,7 |
| (-1) no answer | 349 | 0,8 | 1,5 |
| (-2) does not apply | 40511 | 98,5 | 100,0 |

5. The Biography of Family Status and the Generated Current Family Status (BIOMARSY, BIOMARSM and \$FAMSTD)

by Rainer Pischner

One of the most important pieces of demographical information of every population sample is family status. As in the occupational biography (in the file PBIOSPE, see chapter 3), the family status biography is saved in a spell format. Every spell is defined by the beginning and the end of a family status. In SOEP the family status episodes are on a monthly and on a yearly basis (files BIOMARSM and BIOMARSY). So to speak, as a side effect we get the generated yearly family statuses (variable \$FAMSTD in the file \$PGEN), which can differ from the original, annual survey information.

5.1 Collection of Family Status

Entries for the family status are, as is explained in the following, collected in the SOEP in various sections.

A retrospective recordation of the family status

Marital biographical information was first collected in the Individual Questionnaire in the second Wave in 1985, and later on in the supplementary biographical questionnaire “Lebenslauf” (life history) as follows:

Question I

Are you currently or were you previously married?

(0) Yes, I am or was married. (0) No. I am single ==> continue with question ...

Since when have you been, respectively from when to when, were you married?

(In the event that you are in your second or further marriage, please answer the question for all of your previous marriages.)

| | | | |
|------------------------------------|------|------|------|
| The marriage was completed in | 19__ | 19__ | 19__ |
| I am still married | (o) | (o) | (o) |
| The marriage ended in | 19__ | 19__ | 19__ |
| - through a divorce | (o) | (o) | (o) |
| - through the death of the partner | (o) | (o) | (o) |
| - no answer to question I | | | |

Changes in family situation

After collecting this retrospective information once, changes in the family situation were asked in subsequent years. The question was:

Question II

Has something changed in your family situation since the beginning of 19\$\$?

Please indicate whether one of the following items applies, and when yes, indicate the date.

| | 19\$\$ in month | 19\$\$+1 in month |
|------------------------------|--------------------|----------------------|
| Got married | _____ | _____ |
| Moved in with my partner | _____ | _____ |
| Got divorced | _____ | _____ |
| Separated from my partner | _____ | _____ |
| My partner has died | _____ | _____ |
| Son / daughter has moved out | _____ | _____ |
| Child was born | _____ | _____ |
| Other (please enter)..... | _____ | |
| No. None of the above. | (o) | |

Note: In the event of a temporary period of non-response this question will be additionally collected later on (see the files \$PLUECKE).

Family status at the date of the interview

Except in 1985 it is standard that the current family status is surveyed every year.

Question III

What is your family status?

- Married, live with husband/wife (o)
- Married, living separated (o)
- Single (o)
- Divorced (o)
- Widowed (o)

Note: Up to and including Wave 13 foreigners were additionally asked whether their husband/wife also lives in Germany. In the event that the husband/wife lives in the home country the status variable \$FAMSTD in the file \$PGEN was coded with the value (6) "Husband/wife in home country".

Family status accounted for

For generating the spells of family status it is only taken into account the family status “single”, “married”, “divorced”, and “widowed”. A distinction is not made between married couples living together or separated.

The family status is for that reason defined and coded in the following manner:

| Code | Family Status |
|------|---|
| 1 | single |
| 2 | married |
| 3 | widowed |
| 4 | divorced |
| 5 | no longer married (in the event that 3 and 4 cannot be distinguished from each other) |
| 9 | Non-recoverable gap due to temporary non response or missing information |

Definition of the spells of family status

A spell system is a rectangular schema, which contains information for the surveyed object (here: persons) on statuses (here: family status) which were adopted during the defined time period.

Each single spell within the spell system is defined by the surveyed individual, the current status of the individual (family status) as well as by the beginning and the end of the status. Thus, a spell consists of a minimum of four variables. In addition to the identifier of the survey unit, there is a serial number for explicit identification and often further variables including such information as error codes. A spell system describes all statuses of the surveyed object within the defined time period and is made up of at least one spell.

5.2 Generation of the spells of family status

With the determination of the chronological demarcation, tax related as well as family related aspects are taken into account. A chronological demarcation based only on the tax regulations would treat a person who was married for only one day out of the year as if he/she had been married for an entire year, neglecting the family related aspects. For this reason it was ensured that every observable family status, both in the monthly spells as well as in the yearly spells, was documented. If, for example, a person got divorced in August and got married again in the same month, you would be able to find the status of being divorced, even when it is short, in the spell system.

Basic Procedure

The following principle underlies the compilation of monthly spells, annual spells and generated family statuses:

1. Create a raw, monthly time series from birth up to the date of the last interview for every person who was surveyed at least once.
2. Fill in recoverable gaps in the time series.
3. Eliminate all inconsistencies in the time series.
4. Generate the monthly spells.
5. Generate the annual spells.
6. Generate the annual family statuses.

Note to step 1.) Create a raw, monthly time series from birth up to the date of the last interview for every person who was surveyed at least once

It is assumed that every person is single, who is under 17 years of age; this means that the first 200 months of a person's life are set to single. If a family situation or change is surveyed, this information is added by prolonging the surveyed family status from the reported date up to the month of the last survey. In doing so, possible inconsistencies are not considered. This is based on the analyses of questions I thru III, mentioned above, as well as the benchmark data from PPFAD.

Note to step 2.) Fill in recoverable gaps in the time series

The following types of gaps have been successfully filled:

- For missing information of various lengths between two months for which the family status "single" was indicated, the gap is filled with the status "single".
- If the family status is missing for one year, then the gap can be filled if the family status of the previous year is identical to the family status of the immediately following year and if there was no indication of a change in the family status in the supplementary questions answered by temporary drop outs.

Note to step 3.) Eliminate all inconsistencies in the time series

Between two marriages or between the last marriage and the end of the survey time period a person can only be divorced or widowed. In some cases this situation is not correctly reported by the respondents. Especially by older persons, inconsistencies in the answers arise. Here the family status for the time between two marriages is determined according to the frequency

of the given answers. The first answer receives a slightly higher weighting than the subsequent answers.

The following combination would not be corrected:

„single“ => “married“ => “single“ => “married“

since this is theoretically possible (annulment of a marriage). In contrast, for example, the following combination:

„single“ => “married“ => “single“ => „divorced“ => “married“

would be changed to:

„single“ => “married“ => “divorced“ => „divorced“ => “married“

Note to step 4.) Generate the monthly spells

From the corrected time series, the monthly spells BIOMARSM are generated for the time period, which spans from the month of the first survey to the month of the last survey. Since the months in BIOMARSM should correspond to the months of the spell systems ARTKALEN and EINKALEN, the first spell of a family status starts with the value 1 at earliest. This value corresponds to the month of January 1983.

The monthly spells overlap themselves in each case by one month, since a person can have two different family statuses during a month if a family status change occurred.

The structure of a family status spell on a monthly basis:

- | | | |
|----|----------|--|
| 1. | HHNR | CASE-ID |
| 2. | PERSNR | Person number |
| 3. | SPELLNR | Serial number of the spell |
| 4. | SPELLTYP | Family status |
| | | 1 = single |
| | | 2 = married |
| | | 3 = divorced |
| | | 4 = widowed |
| | | 5 = no longer married |
| | | 9 = code for a gap |
| 5. | BEGIN | Beginning of the family status in ... months |
| 6. | END | End of the family status in ... months |
| 7. | REMARK | Error code (see below explanations) |

Note to step 5.) Generate the annual spells

From the corrected time series, the annual spells BIOMARSY are generated for the time period, which spans from the year of birth to the year of the last survey.

The episodes are stored in two time formats: “Age in years” and calendar year. There is a simple relationship between this two time formats. Age is constructed as the difference between calendar year and year of birth, independent of the month of birth. The annual spells also overlap themselves, since a person can have two (or more) different family statuses during a year if a family status change occurred.

The structure of a family status spell on an annual basis:

- | | | |
|----|----------|--|
| 1. | HHNR | CASE-ID |
| 2. | PERSNR | Person number |
| 3. | SPELLNR | Serial number of the spell |
| 4. | SPELLTYP | Family status |
| | | 1 = single |
| | | 2 = married |
| | | 3 = divorced |
| | | 4 = widowed |
| | | 5 = no longer married |
| | | 9 = code for gaps |
| 5. | BEGIN | Beginning of the family status at the age of ... years |
| 6. | END | End of the family status at the age of ... years |
| 7. | BEGINY | Beginning of the family status (calendar year) |
| 8. | ENDY | End of the family status (calendar year) |
| 9. | REMARK | Error code (see explanation below) |

Note to step 6.) Generate the annual family statuses

An annual family status can also be constructed from the monthly time series, which describes the family status (\$FAMSTD) for the date of the interview. This value is accepted as the generated family status. There is a peculiarity for the family status “married”. Although in the spell system the status “married” is only generally given, the annual, generated variable family status (\$FAMSTD) breaks this status down into “live together”, “live separately” and “partner lives in home country”. These subdivisions are accepted as long as the surveyed, as well as the family status in BIOMARSM are “married”. If gaps are closed and thus the generated variable gets the status “married”, then there will be no further differentiation, even if the person in the previous year and in the subsequent year, for example, has answered, “married, live separately”.

The variable **REMARK**

The variable **REMARK** is assigned to every accounted spell, which indicates particularities and errors in the spell system. The meaning of these variables is identical for monthly and annual spells:

REMARK = -1

Entries on family status are missing within the observed period.

REMARK = 0

If there is not discovered any inconsistencies or errors by generating the spell, then the variable **REMARK** is set to 0.

REMARK = 1,2,3,4

Number of inconsistencies or errors, which were corrected by generating a spell. Now, the spell seems to be correct.

REMARK = 9

Entries on family status are missing within the observed time period. The resulting gap could not be closed.

REMARK = 13,14,15,21,31,41,43

An unreliable change between two family statuses has been discovered (previous family status = first digit, current family status = second digit). An error exists in the spell system. It was not attempted to correct the error.

In detail:

- (13): Transition single => divorced
- (14): Transition single => widowed
- (15): Transition single => divorced or widowed

- (21): Transition married => single

- (31): Transition divorced => single

- (41): Transition widowed => single
- (43): Transition widowed => divorced

5.3 Result Files

The files **BIOMARSM** and **BIOMARSY** contain samples A to G up to the year 2005, Wave V.

The spell system based on years contains 44,291 surveyed persons with a total of 94,064 spells. In the following tables it is given the distribution of the annual spells according to the number of spells per person, the frequency of errors and the distribution of the error codes REMARK.

Distribution of the annual spells by persons

| Spells per person | Persons | Percent |
|-------------------|---------|---------|
| 1 | 11,667 | 26,3 |
| 2 | 20,411 | 46,0 |
| 3 | 8,730 | 19,7 |
| 4 | 2,478 | 5,6 |
| 5 | 677 | 1,5 |
| 6 | 253 | 0,6 |
| 7 | 45 | 0,1 |
| 8 | 22 | 0,1 |
| 9 | 4 | 0,0 |
| 10 | 3 | 0,0 |
| 13 | 1 | 0,0 |
| Total | 44,291 | 100,0 |

79 percent of all annual spells have the remark or error code 0; they are formally all right. Additionally, 11% are formally all right, too, but were corrected. The most frequent error codes are due to gaps (code -1 with 4.9%). These errors occur most if the beginning of the first marriage cannot be given an exact date. Changes in the family status are only in a few cases actually inconsistent. Through the correcting process, the proportion of inconsistencies remains to be only for few cases.

*Distribution of the error code REMARK
(Annual Spells)*

| REMARK | Number | Percent |
|--------|---------|---------|
| -1 | 4,624 | 4.9 |
| 0 | 74,2133 | 78.9 |
| 1 | 8,228 | 8.8 |
| 2 | 1,881 | 2.0 |
| 3 | 95 | 0.1 |
| 4 | 2 | 0.0 |
| 13 | 200 | 0.2 |
| 14 | 128 | 0.1 |
| 19 | 4,619 | 4.9 |
| 29 | 1 | 0.0 |
| 31 | 11 | 0.0 |
| 34 | 40 | 0.0 |
| 41 | 1 | 0.0 |
| 43 | 21 | 0.0 |
| Total | 94,064 | 100.0 |

*Printing of the first cases in the file BIOMARSY
(Version 2005, Wave V)*

| HHNR | PERSNR | SPELL NR | SPELL TYP | BEGIN | END | BEGINY | ENDY | REMARK |
|------|--------|-------------|--------------|-------|-----|--------|------|--------|
| 19 | 101 | 1 | 1 | 0 | 24 | 1930 | 1954 | 0 |
| 19 | 101 | 2 | 2 | 24 | 28 | 1954 | 1958 | 0 |
| 19 | 101 | 3 | 3 | 28 | 32 | 1958 | 1962 | 0 |
| 19 | 101 | 4 | 2 | 32 | 59 | 1962 | 1989 | 0 |
| 19 | 102 | 1 | 1 | 0 | 22 | 1940 | 1962 | 0 |
| 19 | 102 | 2 | 2 | 22 | 49 | 1962 | 1989 | 0 |
| 19 | 103 | 1 | 1 | 0 | 24 | 1963 | 1987 | 0 |
| 27 | 201 | 1 | 1 | 0 | 27 | 1926 | 1953 | 0 |
| 27 | 201 | 2 | 2 | 27 | 39 | 1953 | 1965 | 0 |
| 27 | 201 | 3 | 3 | 39 | 79 | 1965 | 2005 | 2 |
| 27 | 202 | 1 | 1 | 0 | 31 | 1956 | 1987 | 0 |
| 27 | 203 | 1 | 1 | 0 | 45 | 1960 | 2005 | 0 |
| 35 | 301 | 1 | 1 | 0 | 24 | 1960 | 1984 | 0 |
| 35 | 301 | 2 | 2 | 24 | 33 | 1984 | 1993 | 0 |
| 35 | 302 | 1 | 1 | 0 | 23 | 1961 | 1984 | 0 |
| 35 | 302 | 2 | 2 | 23 | 32 | 1984 | 1993 | 0 |

The spell system based on months, BIOMARSM, contains 44,288 surveyed persons with a total of 61,563 spells. In the following tables it is given the distribution of the monthly spells according to the number of spells per person, the frequency of errors and the distribution of the error codes REMARK.

Distribution of the monthly spells by persons

| Spells per person | Persons | Percent |
|-------------------|---------|---------|
| 1 | 30,967 | 69.9 |
| 2 | 10,566 | 23.9 |
| 3 | 1,851 | 4.2 |
| 4 | 703 | 1.6 |
| 5 | 137 | 0.3 |
| 6 | 46 | 0.1 |
| 7 | 11 | 0.0 |
| 8 | 4 | 0.0 |
| 9 | 2 | 0.0 |
| 11 | 1 | 0.0 |
| Total | 44,288 | 100.0 |

Nearly 90 percent of all monthly spells have the error code 0; they are formally all right. Additionally, 10.2 % are formally all right, too, but were corrected. With monthly spells, relatively few cases of gaps occur, since the completeness of the monthly spell is not dependent on the subsequent collection of the marriage history.

*Distribution of the error code REMARK
(Monthly spells)*

| REMARK | Number | Percent |
|--------|---------|---------|
| -1 | 42 | 0.1 |
| 0 | 55,012 | 89.4 |
| 1 | 4,253 | 6.9 |
| 2 | 2,001 | 3.3 |
| 13 | 159 | 0.3 |
| 14 | 33 | 0.1 |
| 31 | 6 | 0.0 |
| 34 | 38 | 0.1 |
| 41 | 2 | 0.0 |
| 43 | 17 | 0.0 |
| Total | 61,5630 | 100.0 |

*Printing of the first cases in the file BIOMARSM
(Version 2005, Wave V)*

| HHNR | PERSNR | SPELLNR | SPELLTYP | BEGIN | END | REMARK |
|------|--------|---------|----------|-------|-----|--------|
| 19 | 101 | 1 | 2 | 1 | 79 | 0 |
| 19 | 102 | 1 | 2 | 1 | 79 | 0 |
| 19 | 103 | 1 | 1 | 1 | 55 | 0 |
| 27 | 201 | 1 | 3 | 1 | 256 | 2 |
| 27 | 202 | 1 | 1 | 1 | 55 | 0 |
| 27 | 203 | 1 | 1 | 1 | 267 | 0 |
| 35 | 301 | 1 | 1 | 1 | 23 | 0 |
| 35 | 301 | 2 | 2 | 23 | 123 | 0 |
| 35 | 302 | 1 | 1 | 1 | 23 | 0 |
| 35 | 302 | 2 | 2 | 23 | 123 | 0 |

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 interview³
- (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)

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

³ 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.

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 stemming 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.

For the variables KIDGEB[n], KIDSEX[n] and KIDPNR[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 questionnaire⁴, in which the number, birth year and sex of the biological children for every woman are collected. For women with information on children stemming 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 reasons⁵. In these 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:

- Women 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.

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 circulation

⁴ 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.

⁵ Beside the reason ‘refusal’, the collection date of the life history biographies differ among SOEP sub-samples.

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 | | | | 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 BIODIDS 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 BIODIDS).

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 child |
| 2 | 12 | 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 of reference person |
| 9 | 10 | 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 2005 / up to Wave V)
(For an up-to-date version of these frequencies also refer to <http://panel.gsoep.de/soepinfo2005>)*

BIOVALID Status of the birth biography

| | | Frequency | Percent | Val. Percent | Cum. Percent |
|-------|---|-----------|---------|--------------|--------------|
| Valid | 10 No Birthbio. - No Kids from Bio. | 1579 | 7,0 | 7,0 | 7,0 |
| | 20 Youthbio - No Kids from Bio. | 1161 | 5,1 | 5,1 | 12,1 |
| | 30 Birthbio - No Kids from Bio. | 7840 | 34,7 | 34,7 | 46,8 |
| | 31 Birthbio - Kids according to Bio. | 12006 | 53,2 | 53,2 | 100,0 |
| | Complete | 22586 | 100,0 | 100,0 | |

BIOYEAR Year of the survey of the birth biography

| | | Frequency | Percent | Val. Percent | Cum. Percent |
|-------|-----------------|-----------|---------|--------------|--------------|
| Valid | -2 No birthbio. | 1579 | 7,0 | 7,0 | 7,0 |
| | 1985 | 6619 | 29,3 | 29,3 | 36,3 |
| | 1986 | 83 | ,4 | ,4 | 36,7 |
| | 1987 | 104 | ,5 | ,5 | 37,1 |
| | 1988 | 220 | 1,0 | 1,0 | 38,1 |
| | 1989 | 211 | ,9 | ,9 | 39,0 |
| | 1990 | 202 | ,9 | ,9 | 39,9 |
| | 1991 | 161 | ,7 | ,7 | 40,6 |
| | 1992 | 2636 | 11,7 | 11,7 | 52,3 |
| | 1993 | 231 | 1,0 | 1,0 | 53,3 |
| | 1994 | 592 | 2,6 | 2,6 | 56,0 |
| | 1995 | 528 | 2,3 | 2,3 | 58,3 |
| | 1996 | 255 | 1,1 | 1,1 | 59,4 |
| | 1997 | 231 | 1,0 | 1,0 | 60,4 |
| | 1998 | 203 | ,9 | ,9 | 61,3 |
| | 1999 | 1033 | 4,6 | 4,6 | 65,9 |
| | 2000 | 243 | 1,1 | 1,1 | 67,0 |
| | 2001 | 4928 | 21,8 | 21,8 | 88,8 |
| | 2002 | 467 | 2,1 | 2,1 | 90,9 |
| | 2003 | 1321 | 5,8 | 5,8 | 96,7 |
| | 2004 | 410 | 1,8 | 1,8 | 98,5 |
| | 2005 | 329 | 1,5 | 1,5 | 100,0 |
| | Complete | 22586 | 100,0 | 100,0 | |

SUMKIDS Sum of kids in BIOBIRTH

| | | Frequency | Percent | Val. Percent | Cum. Percent |
|----------|-------|-----------|---------|--------------|-----------------|
| Valid | ,00 | 8497 | 37,6 | 37,6 | 37,6 |
| | 1,00 | 4310 | 19,1 | 19,1 | 56,7 |
| | 2,00 | 5996 | 26,5 | 26,5 | 83,3 |
| | 3,00 | 2389 | 10,6 | 10,6 | 93,8 |
| | 4,00 | 841 | 3,7 | 3,7 | 97,6 |
| | 5,00 | 304 | 1,3 | 1,3 | 98,9 |
| | 6,00 | 128 | ,6 | ,6 | 99,5 |
| | 7,00 | 57 | ,3 | ,3 | 99,7 |
| | 8,00 | 40 | ,2 | ,2 | 99,9 |
| | 9,00 | 11 | ,0 | ,0 | 99,9 |
| | 10,00 | 9 | ,0 | ,0 | 100,0 |
| | 11,00 | 1 | ,0 | ,0 | 100,0 |
| | 12,00 | 3 | ,0 | ,0 | 100,0 |
| Complete | | 22586 | 100,0 | 100,0 | |

BIOKIDS Sum of kids derived from birth biography

| | | Frequency | Percent | Val. Percent | Cum. Percent |
|----------|----|-----------|---------|--------------|-----------------|
| Valid | -2 | 1579 | 7,0 | 7,0 | 7,0 |
| | 0 | 9001 | 39,9 | 39,9 | 46,8 |
| | 1 | 3792 | 16,8 | 16,8 | 63,6 |
| | 2 | 4996 | 22,1 | 22,1 | 85,8 |
| | 3 | 2002 | 8,9 | 8,9 | 94,6 |
| | 4 | 720 | 3,2 | 3,2 | 97,8 |
| | 5 | 267 | 1,2 | 1,2 | 99,0 |
| | 6 | 115 | ,5 | ,5 | 99,5 |
| | 7 | 52 | ,2 | ,2 | 99,7 |
| | 8 | 40 | ,2 | ,2 | 99,9 |
| | 9 | 12 | ,1 | ,1 | 100,0 |
| | 10 | 7 | ,0 | ,0 | 100,0 |
| | 11 | 1 | ,0 | ,0 | 100,0 |
| | 12 | 2 | ,0 | ,0 | 100,0 |
| Complete | | 22586 | 100,0 | 100,0 | |

NEWKIDS Sum of kids identified through \$PBRUTTO

| | | Frequency | Percent | Val. Percent | Cum. Percent |
|----------|---|-----------|---------|--------------|-----------------|
| Valid | 0 | 19657 | 87,0 | 87,0 | 87,0 |
| | 1 | 1705 | 7,5 | 7,5 | 94,6 |
| | 2 | 953 | 4,2 | 4,2 | 98,8 |
| | 3 | 221 | 1,0 | 1,0 | 99,8 |
| | 4 | 42 | ,2 | ,2 | 100,0 |
| | 5 | 6 | ,0 | ,0 | 100,0 |
| | 6 | 2 | ,0 | ,0 | 100,0 |
| Complete | | 22586 | 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-sample. 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 the 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 stemming 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)).

- BIODKIDS 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.

For the variables KIDGEB[n], KIDSEX[n] and KIDPNR[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 BIOBIRM 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 fatherhood was introduced in 2001. Therefore, the majority of 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 entry 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 for further details). Furthermore, this underestimation is more severe for men who did not complete the biographical questionnaire than it is for women as the mother child dyad is most likely to remain intact after a union dissolution.

7.3 Construction of variables - Particularities

The construction of variables for the data-set BIOBIRM 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 variables in the file BIOBIRTH (Version 2005 / up to Wave V)
(For an up-to-date version of these frequencies also refer to <http://panel.gsoep.de/soepinfo2004>)*

BIOVALID Status of the birth biography

| | | Frequency | Percent | Val. Percent | Cum. Percent |
|-------|---|-----------|---------|--------------|--------------|
| Valid | 10 No Birthbio. - No Kids from Bio. | 6593 | 47,9 | 47,9 | 47,9 |
| | 20 Youthbio - No Kids from Bio. | 1027 | 7,5 | 7,5 | 55,4 |
| | 30 Birthbio - No Kids from Bio. | 2043 | 14,8 | 14,8 | 70,2 |
| | 31 Birthbio - Kids according to Bio. | 4101 | 29,8 | 29,8 | 100,0 |
| | Complete | 13764 | 100,0 | 100,0 | |

BIOYEAR Year of the survey of the birth biography

| | | Frequency | Percent | Val. Percent | Cum. Percent |
|----------|-----------------|-----------|---------|--------------|--------------|
| Valid | -2 No birthbio. | 6593 | 47,9 | 47,9 | 47,9 |
| | 2001 | 118 | ,9 | ,9 | 48,8 |
| | 2002 | 4501 | 32,7 | 32,7 | 81,5 |
| | 2003 | 432 | 3,1 | 3,1 | 84,6 |
| | 2004 | 1370 | 10,0 | 10,0 | 94,6 |
| | 2005 | 412 | 3,0 | 3,0 | 97,5 |
| Complete | 338 | 2,5 | 2,5 | 100,0 | |

SUMKIDS Sum of kids in BIOBRTHM

| | | Frequency | Percent | Val. Percent | Cum. Percent |
|----------|-------|-----------|---------|--------------|--------------|
| Valid | ,00 | 6568 | 47,7 | 47,7 | 47,7 |
| | 1,00 | 2505 | 18,2 | 18,2 | 65,9 |
| | 2,00 | 3131 | 22,7 | 22,7 | 88,7 |
| | 3,00 | 1113 | 8,1 | 8,1 | 96,8 |
| | 4,00 | 313 | 2,3 | 2,3 | 99,0 |
| | 5,00 | 86 | ,6 | ,6 | 99,7 |
| | 6,00 | 28 | ,2 | ,2 | 99,9 |
| | 7,00 | 10 | ,1 | ,1 | 99,9 |
| | 8,00 | 8 | ,1 | ,1 | 100,0 |
| | 9,00 | 1 | ,0 | ,0 | 100,0 |
| | 11,00 | 1 | ,0 | ,0 | 100,0 |
| Complete | | 13764 | 100,0 | 100,0 | |

BIOKIDS Sum of kids derived from birth biography

| | | Frequency | Percent | Val. Percent | Cum. Percent |
|----------|----|-----------|---------|--------------|--------------|
| Valid | -2 | 6593 | 47,9 | 47,9 | 47,9 |
| | 0 | 3070 | 22,3 | 22,3 | 70,2 |
| | 1 | 1204 | 8,7 | 8,7 | 79,0 |
| | 2 | 1846 | 13,4 | 13,4 | 92,4 |
| | 3 | 728 | 5,3 | 5,3 | 97,7 |
| | 4 | 217 | 1,6 | 1,6 | 99,2 |
| | 5 | 65 | ,5 | ,5 | 99,7 |
| | 6 | 25 | ,2 | ,2 | 99,9 |
| | 7 | 7 | ,1 | ,1 | 99,9 |
| | 8 | 9 | ,1 | ,1 | 100,0 |
| Complete | | 13764 | 100,0 | 100,0 | |

NEWKIDS Sum of kids identified through \$PBRUTTO

| | | Frequency | Percent | Val. Percent | Cum. Percent |
|----------|----|-----------|---------|--------------|--------------|
| Valid | 0 | 10514 | 76,4 | 76,4 | 76,4 |
| | 1 | 1536 | 11,2 | 11,2 | 87,5 |
| | 2 | 1254 | 9,1 | 9,1 | 96,7 |
| | 3 | 354 | 2,6 | 2,6 | 99,2 |
| | 4 | 82 | ,6 | ,6 | 99,8 |
| | 5 | 18 | ,1 | ,1 | 100,0 |
| | 6 | 3 | ,0 | ,0 | 100,0 |
| | 7 | 2 | ,0 | ,0 | 100,0 |
| | 11 | 1 | ,0 | ,0 | 100,0 |
| Complete | | 13764 | 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 found 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 potentially monozygotic or if the sex of the siblings differs.

The central variable PERSNR is assigned to the sibling in the group with the lowest personal identifier. PNRTWIN and – in seldom cases – if available PNRTRIP or even PNRQUAD contains the personal identifier of second, and the third and fourth sibling respectively 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 sibling doesn’t exist. PERSNR and PNRTWIN however should always contain valid codes.

The variable PNRMOTH shows the link to the mother of the group and is derived from the data set \$KIND. It is identical to the variable \$KMUTTI. The variable BIOMONOZ indicates if the sex of all the siblings is identical and this group therefore *might* be monozygotic (code “1”). If the sex of the siblings differs, this code is set to “0”.

8.2 Construction of variables in the data set BIOTWIN

The selection of twins within the SOEP for the data set BIOTWIN is based on the month of birth. As a woman might give birth at two different times in a year and social siblings need to be distinguished from biological siblings, the month of birth is of vital importance when determining the twin-groups. Therefore, only people with a) valid month of birth information and b) identical month of birth may be 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 (\$KMUTTI) is checked if available for both siblings to cross check the results of the previous steps. This pointer is then transferred into the variable PNRMOTH. If the variable \$KMUTTI is unavailable or incomplete the variable PNRMOTH is set to -1.

An overview over central information in the file BIOTWIN (Version 2005 / Wave V)

(For an up-to-date version of these info also refer to <http://panel.gsoep.de/soepinfo2005>)

Table 1 Sibling groups in the BIOTWIN

| group size | | valid mother pointers |
|------------|-----|--------------------------|
| twins | 134 | 131 |
| triple | 3 | 3 |
| quadruple | - | - |

BIOMONOZ Possibly monozygotic siblings?

| | | Frequenc y | Percent | Val. Percent | Cum. Percent |
|-------|---|---------------|---------|-----------------|-----------------|
| Valid | 0 different sex | 48 | 35,0 | 35,0 | 35,0 |
| | 1 same sex – possibly monozygotic | 89 | 65,0 | 65,0 | 100,0 |
| | Complete | 137 | 100,0 | 100,0 | |

9. BIOIMMIG: Generated and Status Variables from SOEP for Foreigners and Migrants

by Jan Goebel and Maria Richter

(Replaces earlier versions by John P. Haisken-DeNew)

9.1 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.

9.2 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 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

| <i>Variable</i> | <i>German</i> | <i>English</i> |
|-----------------|--|--|
| 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 |
| BIEXPLV | 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 |

9.3 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 LPAUSL 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.

9.4 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 |

9.4.1 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`

9.5 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 |
| 2000 Q | QJUGEND | QJ57 |
| 2001 R | RJUGEND | RJ59 |
| 2002 S | SJUGEND | SJ59 |
| 2003 T | TJUGEND | TJ59 |
| 2004 U | UJUGEND | UJ59 |
| 2005 V | VJUGEND | VJ59 |

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 |
| 2000 Q | QJUGEND | QJ58 |
| 2001 R | RJUGEND | RJ60 |
| 2002 S | SJUGEND | SJ60 |
| 2003 T | TJUGEND | TJ60 |
| 2004 U | UJUGEND | UJ60 |
| 2005 V | VJUGEND | VJ60 |

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

BIO Question: Q7a

German: *Haben Sie nach Ihrer Einreise zunächst in einem Aufnahmelager oder Übergangwohnheim 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 |
| 2000 Q | QJUGEND | QJ5901 |
| 2001 Q | RJUGEND | RJ6101 |
| 2002 S | SJUGEND | SJ6101 |
| 2003 T | TJUGEND | TJ6101 |
| 2004 U | UJUGEND | UJ6101 |
| 2005 V | VJUGEND | VJ6101 |

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 |
| 2000 Q | QJUGEND | QJ5902 |
| 2001 R | RJUGEND | RJ6102 |
| 2002 S | SJUGEND | SJ6102 |
| 2003 T | TJUGEND | TJ6102 |
| 2004 U | UJUGEND | UJ6102 |
| 2005 V | VJUGEND | VJ6102 |

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 |
| 2000 Q | QJUGEND | QJ5903 |
| 2001 R | RJUGEND | RJ6103 |
| 2002 S | SJUGEND | SJ6103 |
| 2003 T | TJUGEND | TJ6103 |
| 2004 U | UJUGEND | UJ6103 |
| 2005 V | VJUGEND | VJ6103 |

BIWFAM BI: Eingereist als Familienangehöriger
 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 |
| 2000 Q | QJUGEND | QJ60 |
| 2001 R | RJUGEND | RJ62 |
| 2002 S | SJUGEND | SJ62 |
| 2003 T | TJUGEND | TJ62 |
| 2004 U | UJUGEND | UJ62 |
| 2005 V | VJUGEND | VJ62 |

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 |
| 2000 Q | QJUGEND | QJ61 |
| 2001 R | RJUGEND | RJ63 |
| 2002 S | SJUGEND | SJ63 |
| 2003 T | TJUGEND | TJ63 |
| 2004 U | UJUGEND | UJ63 |
| 2005 V | VJUGEND | VJ63 |

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 |
| 2000 Q | QJUGEND | QJ62 |
| 2001 R | RJUGEND | RJ64 |
| 2002 S | SJUGEND | SJ64 |
| 2003 T | TJUGEND | TJ64 |
| 2004 U | UJUGEND | UJ64 |
| 2005 V | VJUGEND | VJ64 |

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 play a 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 |
| 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 |
| 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 |
| 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 |
| 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 |
| 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 |
| 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 |
| 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 |
| 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 erfuehlt ?*
"[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 |
| 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 |
| 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 colleagues 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 |
| 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 |
| 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 |
| 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 |
| 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] Personliche 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 whether 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 whether 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 |
| 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 |
| 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 |
| 2000 Q | QJUGEND | QJ63 |
| 2001 R | RJUGEND | RJ65 |
| 2002 S | SJUGEND | SJ65 |
| 2003 T | TJUGEND | TJ65 |
| 2004 U | UJUGEND | UJ65 |
| 2005 V | VJUGEND | VJ65 |

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, BISC GC, BISC GCF, BISC GCFN**

| Year | File | Variable |
|-------------|---------|----------|
| 1984 A | AP AUSL | n/a |
| 1985 B | BPAUSL | n/a |
| 1986 C | CP AUSL | n/a |
| 1987 D | DPAUSL | n/a |
| 1988 E | EPAUSL | n/a |
| 1989 F | FPAUSL | FP108A |
| 1990 G | GPAUSL | GP108A |
| 1991 H | HP AUSL | HP108A |
| 1992 I | IP AUSL | 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 |
| 2000 Q | QJUGEND | QJ64 |
| 2001 R | RJUGEND | RJ6601 |
| 2002 S | SJUGEND | SJ6601 |
| 2003 T | TJUGEND | TJ6601 |
| 2004 U | UJUGEND | UJ6601 |
| 2005 V | VJUGEND | VJ6601 |

BISCGERC BI: Besuch spezieller Vorbereitung
 BI: Attended Special Foreigner Prep Class

BIO Question: Q20

German: *Haben Sie vorher eine spezielle Vorbereitungs-klasse 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 | AP AUSL | n/a |
| 1985 B | BPAUSL | n/a |
| 1986 C | CP AUSL | n/a |
| 1987 D | DPAUSL | n/a |
| 1988 E | EPAUSL | n/a |
| 1989 F | FPAUSL | FP109A |
| 1990 G | GPAUSL | GP109A |
| 1991 H | HP AUSL | HP109A |
| 1992 I | IP AUSL | 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 |
| 2000 Q | QJUGEND | QJ65 |
| 2001 R | RJUGEND | RJ6602 |
| 2002 S | SJUGEND | SJ6602 |
| 2003 T | TJUGEND | TJ6602 |
| 2004 U | UJUGEND | UJ6602 |
| 2005 V | VJUGEND | VJ6602 |

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 | AP AUSL | 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 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 | 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 |
| 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 |

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 |

10. BIOPAREN: Biography Information for the Parents of SOEP-Respondents

by Katharina Mahne and Jürgen Schupp

(Replaces earlier versions by Verena Tobsch, Matthias Pollmann-Schult and Charlotte Büchner)

10.1 Content of the data file BIOPAREN

The goal of the data file BIOPAREN is to make the biography entries on the parents and on the social origin of the respondent available. In particular, this is information on the year of birth and death, the religious orientation, the education and career training of the parents, as well as the father's job and the regional mobility of the surveyed individual since childhood. This information is collected using the Biography Questionnaire, which every person receives who is accepted into SOEP. Since 2000, there has been an independent questionnaire in SOEP for the group of survey participants who are 16-17 years old and are being interviewed for the first time; with respect to the place of origin, this questionnaire is identical to the Biography Questionnaire (see chapter BIOYOUTH).

The information gained on the parents from the Biography and the Youth Questionnaires are referred to in the following as proxy entries. The proxy entries of the children⁶ for education, year of death and religious affiliations of the parents are indicated by attributes of the parents at the time of the survey. Information on the time of the survey is found in the variable BIOYEAR. Entries on the professional position and occupation of the father should, based on the way the questions in the Biography Questionnaire are formulated, reflect the situation when the respondent was 16 years old. Since 2000, we have also collected data on the mothers of respondents. The religious affiliation of the parents was first included in 1996 within the framework of the survey of biographical data for "new participants" in SOEP. In 2001, information on the nationality of the parents and the professional position as well as occupation of the mother at the age of 15 was added. In view of these extensions, no proxy-information exists for those persons surveyed who had filled out the Biography before the additions were made. Entries on the residency of the parents were asked in 1991, 1996 and 2001 in the Individual Questionnaire. The current version includes new variables on the current place of residence of both the father and the mother (MAORTAKT, VAORTAKT). Furthermore, the variables LIVING1 to LIVING8 indicate where the first 15 years of life were spent. The data file BIOPAREN contains only those persons with at least one realized personal interview with biography information, as well as persons who, up to now, have

⁶ In the following the respondents are marked as children if they have completed the Biography / Youth Questionnaire and in doing so have provided information on their parents.

exclusively filled out the Biography or Youth Questionnaire. Furthermore, the data file is supplemented every year with the newly surveyed persons.

10.2 Inclusion of the Biography

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.

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.

10.3 Generating using BIOPAREN

The information available in BIOPAREN can be obtained in two different ways. On the one hand, BIOPAREN includes the children's proxy entries on the parents, and on the other hand, it contains the direct entries from the 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, in addition 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 parents can be identified by the household number and a certain combination of the \$STELL-Variable (relationship to head of household) between the child and parent (see below).

Variables in BIOPAREN

Population Basis: Surveyed persons taken from PPFAD for whom at least one interview had been completed up to 2004 as well as five additional persons who had completed the Biography Questionnaire but not the Individual Questionnaire.

Event Number: 41.087

Waves: A-V

Subsamples: A, B, C, D1, D2, E, F, G

Variables:

| | |
|------------------|--|
| -> <u>HHNR</u> | Number of the original household |
| -> <u>PERSNR</u> | 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) |
| -> 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 (<i>since 2000 no longer collected</i>) |
| -> ORTKIND3 | Still lives in parents' household (<i>since 2000 no longer collected</i>) |
| -> 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 |
| -> VAORTUP | Year of update of vaortakt |
| -> MAORTUP | Year of update or maortakt |
| -> BIOYEAR | Year of the Biography Survey |

A description of the characteristics of these variables can be viewed under SOEPinfo on the SOEP Website (<http://www.diw.de/soep/soepinfo/>). For all variables on the father's or mother's occupation, their professional position, their prestige score and industry codes, see the detailed documentation on value labels in the documentation of PGEN.

10.3.1 Steps for Generating from BIOPAREN

The information contained in BIOPAREN on the parents comes on the one hand from the children's proxy entries in the Biography Questionnaire or the Youth Questionnaire, and on the other hand, from the direct entries of the parents in the Individual Questionnaire. The children's proxy entries always have a higher priority than the direct entries of the parents. The direct entries are only used when the child, due to the filter command, skips the questions regarding the year of birth, the education and professional training of the parents, as well as the professional position and occupation of the father. According to the filter command, this information is not collected when the parents, or the father, lives in the same household as the child at the time of the survey. The resulting missing entries ("missings") are replaced with the direct entries of the parents. In order to use the parents' individual entries, the personal number (PERSNR) of the parents must be identified.

1. Variables with complete entries in the Biography Questionnaire

The following variables can be generated for all persons from the information collected in the Biography Questionnaire or the Youth Questionnaire: VTODJ, MTODJ, VRELI, MRELI, MNAT, VNAT, ORTKINDH, ORTKIND1, VSTREIT, MSTREIT

2. Variables with incomplete entries in the Biography Questionnaire or Youth Questionnaire

Due to the filter command, the children's proxy entries are only available for the following variables when the parents or one parent and the child do not live in the same household at the time of the survey: VGEBJ, MGEBJ, VSBIL, MSBIL, VBBIL, MBBIL. The children's proxy entries on professional position and occupation of the father (VBSTELL as well as VISCO88 and all prestige scores) are only 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.

3. Generating the parents' personal number (MNR and VNR)

The variables listed under 2 can be generated for those respondents whose parents/father live in the same household only through the direct entries of the parents/father. To do this, the parents must be identified.

The personal number of the parents (VNR and MNR) is generated in three steps. The steps are in order of priority.

1. The parents of the respondent are identified by the relationship to the head of the household (\$STELL in \$BRUTTO). Here the first (oldest) entry has priority. Ideally, the children's parents are identified at the time of the first survey of the child, i.e., when the child is 16 years old. Furthermore, sometimes the social parents and not the biological parents are identified.

2. The parents of the respondent are identified by the mother indicator (\$KMUTTI) as well as the partner indicator (\$KMUP) 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-PERSNR of the respondent can be identified in BIOBIRTH.

Steps 1. and 2. are aimed at identifying the parents that live in the household when the child is 16 years old, whereas through BIOBIRTH the biological mother can be

identified. The personal number found using BIOBIRTH is compared to the identification steps described under (3.1) and (3.2), and any new personal numbers are incorporated.

4. Reasons for missing entries

The current version of the data set includes 34 interviewees who do not have a mother or father indicator, despite the parents having been identified. The reason for this is as follows:

In the generation file biogener2.sps, the information on the parents' years of birth from the Personal History Questionnaire is compared with that information given directly by the parents, provided that they have been identified. This leads to inconsistencies. For the 23 cases, this is solely the case for children who no longer live in the parental home and therefore provide proxy details on their parents. These proxy details always stand out in BIOPAREN and it was therefore decided to leave out the parent indicators for those children for whom the details on their parents' years of birth deviate by more than two years from those contained in PPFAD. However, all other variables for these parents are available.

The following list indicates the individuals affected.

| PERSNR | VNR | MNR |
|---------|---------|---------|
| 22304 | 22306 | |
| 22305 | 22306 | |
| 66703 | 33603 | |
| 101504 | 101505 | |
| 176504 | 918303 | |
| 278304 | 278305 | |
| 300703 | 176505 | |
| 309604 | 1056904 | |
| 309605 | 1082204 | |
| 328303 | 1082204 | |
| 375103 | 1063506 | |
| 531105* | | 531102 |
| 532004* | | 532002 |
| 532005* | | 532002 |
| 712103 | 712104 | |
| 734703 | 328305 | |
| 751303 | 751305 | |
| 826707 | 502304 | |
| 1078103 | | 1008502 |
| 2022702 | 2022703 | |

| | | |
|----------|---------|---------|
| 2028203 | 2306404 | |
| 2124805 | 2124802 | |
| 2180804 | 2180807 | |
| 2203502 | 2203504 | |
| 2510103 | | 2510105 |
| 2527603 | 2527601 | |
| 2756403 | 2756401 | |
| 2800405 | 2800402 | |
| 2858104 | 2858102 | |
| 2991102 | 2991105 | |
| 2992203 | 2992201 | |
| 3042804 | 3042802 | |
| 3083903 | 3083902 | |
| 3828102 | 3828103 | |
| 5002104 | 5002105 | |
| 5030307 | 5030301 | |
| 5040903* | | 5040902 |
| 5044103 | 5044101 | |
| 5198703 | 5198705 | |
| 5409003 | | 5409004 |
| 5603803 | 5210602 | |
| 5604303 | 5053701 | |
| 7006203 | 7006204 | |
| 7234504 | 1126205 | |
| 8008603 | 8008602 | |
| 8008604 | 8008602 | |
| 8039403* | | 8039402 |
| 8041703 | 8041702 | |
| 8114403 | 8114401 | |
| 8127704 | 8127705 | |
| 8137603 | 8137602 | |
| 8158203 | 8158201 | |
| 8159103 | 8159101 | |
| 8159104 | | 8159102 |
| 8214903 | | 8214901 |
| 8215703 | 8215701 | |
| 8226403 | 8226401 | |
| 8233403 | 8233401 | |
| 8233404 | 8233401 | |
| 8245603 | 8245602 | |
| 8245604 | 8245602 | |
| 8248603 | 8248602 | |

The persons marked with ‘*’ are children whose mother is only 15 to 17 years older. In this case, it is not clear whether the details are reliable.

4. Completing missing entries

Missing entries in the variables listed in (2), as well as missing entries on the year of birth and death of the parents, are completed with the corresponding direct entries of the parents in PPFAD as well as in \$PGEN. In generating the variables for education and professional training of the parents, as well as the professional position and occupation of the father, the corresponding, most recent information for the 16-year-olds surveyed for the first time is incorporated.

5. Updating BIOPAREN

Some variables of the persons already contained in BIOPAREN will be updated with new survey information, insofar as no valid values exist in BIOPAREN. These are: VTODJ, MTODJ, VSBIL, MSBIL, VBBIL, MBBIL, VSINFO, MSINFO, VBINFO, MBINFO.

10.3.2 Identification of the Parents

Identification of the relationship to the head of household

The identification of the parents described above occurs first of all through the variable \$STELL (relationship to head of household). The combinations of the characteristics of the \$STELL-variable listed below describe the possible parent-child relationships for a parent and a child.

Characteristics of the variable \$STELL “relationship of the person to the head of the household”:

Code Label

- 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

Possible Parent-Child Relationships

| Relationship of the child to the HH | Relationship of the parent to the HH | Person is ... |
|--|---|--|
| 3 | 0 | Child of HH |
| 3 | 1 / 2 | Child of marital/ partner of HH |
| 4 | 0 | Foster child of HH |
| 4 | 1 / 2 | Foster child of marital/ partner of HH |
| 12 | 2 | Child of partner of HH |
| 9 | 3 / 4 | Child of child/foster child of HH |
| 0 | 6 | Child is HH, lives with parents in same household |
| 1 / 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) |

Identification through the Mother Indicator / Partner of the Mother Indicator from \$KIND

The population of the file \$KIND includes all children under the age of 16. The file contains the personal number of the mother (\$KMUTTI), as well as the personal number of the partner of the mother (\$KMUP). Through both variables the latest mother, as 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.

Identifying the Mother (the Parents) using BIOBIRTH

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. While it cannot be stated definitively that these pointers link to biological or social fathers, experience shows that the number of cases identified through the parent-child relationship in BIOBIRTH is low. In the current update, all parents were identified using \$STELL and \$KIND.

10.4 Updating

Accepting new survey participants occurs with the help of the collation of the data from the corresponding biographical entries and the individual entries by identifying the parents in SOEP. This updating depends on the process described for generating the “parent-specific” data.

For parents who are already contained in the data file BIOPAREN and are identified with a personal number, an update of the year of death ultimately occurs in 2005 if they have passed away during the term of the SOEP.

10.5 Correction and adaptation

For the current version of BIOPAREN, some modifications of value labels concerning VBSTELL, MBSTELL, VBBIL and MBBIL have been carried out. Furthermore, failures that occurred while generating VBBIL and MBBIL have been eliminated.

The labels of VBSTELL and MBSTELL have been adapted according to the particular questions as follows:

| NEW | VBSTELL/MBSTELL | OLD |
|------------|---|------------|
| 0 | do not know | 0 |
| 1 | has died | 1 |
| | | |
| 10 | not employed | 10 |
| 11 | in training | 11 |
| 12 | unemployed, sick | 12 |
| 13 | retired | 13 |
| 14 | prisoner of war | 14 |
| 15 | military, community work | 15 |
| 100 | apprentice in firm | |
| 110 | apprentices/trainees | |
| 120 | apprentice in industry of technology | |
| 130 | apprentice in trade and commerce | |
| 140 | volunteer, intern, etc. | |
| 150 | aspirants | |
| | | |
| 200 | blue-collar worker | 20 |
| 210 | untrained worker | 21 |
| 220 | Semi-trained worker | 22 |
| 230 | trained worker | 23 |
| 240 | foreman, team leader | 24 |
| 250 | foreman | 25 |
| | | |
| 310 | agricultural worker | 31 |
| 320 | agricultural specialist | 32 |
| 330 | agricultural foreman | 33 |
| 340 | agricultural manager | 34 |
| | | |
| 400 | All types of self-employed | 40 |
| | | |
| 410 | self-employed farmer | 41 |
| 411 | without employees | |
| 412 | 1 to 9 employees | |
| 413 | 10 or more employees | |
| 414 | with employees, no info about number | |
| | | |
| 420 | free-lance professional | 42 |
| 421 | without employees | |
| 422 | 1 to 9 employees | |
| 423 | 10 or more employees | |
| 424 | with employees, no info about number | |
| | | |
| 430 | other self-empl. without or up to 9 empl. | 43 |
| 431 | without employees | |
| 432 | 1 to 9 employees | 46 |
| 433 | 10 or more employees | |

| | | |
|------------|--------------------------------------|-----------|
| 434 | with employees, no info about number | 44 |
| | | |
| 440 | help in family business | 45 |
| | | |
| 500 | white-collar worker | 50 |
| | | |
| 510 | foreman | 51 |
| 520 | white-collar with simple profession | 52 |
| 521 | untrained white-collar worker | |
| 522 | trained white-collar worker | 56 |
| 530 | qualified profession | 53 |
| 540 | high qualified profession | 54 |
| 550 | managerial | 55 |
| | | |
| 600 | civil servant | 60 |
| | | |
| 610 | low-level civil servant | 61 |
| 620 | middle-level civil servant | 62 |
| 630 | high-level civil servant | 63 |
| 640 | executive civil servant | 64 |
| 999 | employed without information | 0 |

The labels of VBBIL and MBBIL have been adapted as follows:

| NE W | VBBIL/MBBIL | OLD |
|-----------------|---|------------|
| 0 | do not know | 0 |
| 10 | no vocational degree | 1 |
| 20 | vocational degree | - |
| 21 | trained in foreign company | 2 |
| 22 | trained long time in foreign company | 3 |
| 23 | foreign vocational school | 4 |
| 24 | trade, farming apprentice | 5 |
| 25 | business apprentice | 6 |
| 26 | health care school | 7 |
| 27 | special technical school | 8 |
| 28 | civil service training | 9 |
| 30 | tech engineer school | 10 |
| 31 | foreign collage | 11 |
| 32 | college university | 12 |
| 40 | other training | 13 |
| 50 | currently in vocational training | 14 |
| 51 | currently in education | 15 |

In the data-set up to wave T, there were failures in the proxy entries. Misleadingly, entries on “vocational degree” had been generated as “trained long time in foreign company” and

“college-university” had been generated as “other training”. Besides, in some cases, “vocational degree” was generated falsely as “other training”. The label “vocational training” is based exclusively on proxy entries, while “other training” is solely generated from the information given directly by the parents in \$PGEN.

10.6 New Variables in BIOPAREN as of data release (2005, up to wave U): VAORTAKT/MAORTAKT and VAORTUP/MAORTUP

The variables VAORTAKT and MAORTAKT contain the latest available information about the parents’ residence and on whether or not they are deceased, respectively. These variables were generated using the information from VAORT91, VAORT96, VAORT01 (accordingly: MAORT91, MAORT96, MAORT01) and data concerning the year of death of those whose parents are identified by a personal identification number.

For persons without identified parents who answered the biography questionnaire up to the year 2001, the most recent available information from the person questionnaire in 1991, 1996 or 2001 was assumed. For persons without identified parents who answered the biography questionnaire from 2002 onwards, the variables contain the information given in the biography questionnaire.

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”.

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 accepted.

On the other hand, 10 persons without identified parents apparently had filled out the biography questionnaire twice. In these 10 cases, the parents obviously had died in the meantime and because of this, the information “deceased” was generated based on the second information from the biography data.

The variables VAORTUP and MAORTUP give the year of updating.

11. BIOSOC: Retrospective Data on Youth and Socialization

by Bettina Isengard and Michael Frühling

The standard supplementary Biography Questionnaire was expanded in 2000, and again 2001 with some specific questions on youth and early adulthood. Part of these questions are derived from the relatively new independent Youth Questionnaire (for detailed information on this questionnaire see chapter 13). The expanded questionnaire asked respondents of all ages to describe aspects of their life including their relationship with parents at the age of 15, school marks, the federal state in which the last educational qualification was attained, detailed information on the vocational qualification, as well as the intention of achieving further educational or vocational qualifications (the last were mostly relevant for younger respondents). Questions concerning military and alternative services are also included in this new 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 for the Individual-Questionnaire, e.g. if someone were 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 the labour force activities are recorded on monthly basis. Thus, one can find out if somebody was conscripted or not.

The data set BIOSOC contains information on 12,694 persons, of which 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, most of the persons in this new data set belong to sample F (73%). Another large proportion comes from sample G (almost 17%) which is comprised of people who filled out the Biography Questionnaire for the ‘first’ time in 2003.

11.1 Structure of the Data Set BIOSOC

The Biography Questionnaire is only asked once in a lifetime. Some of the information stored in the new data set BIOSOC is unvarying (such as the relationship to the parents at the age of 15) or isn’t surveyed to such an extent in the regular questionnaire (such as last school marks). 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 and 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 marks or military service.

In table 1 (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 in 2001, Wave R. In the next two columns it is noted whether the variable exists for the year 2000 or 2002 and later as well. A plus sign means that it did, a plus sign in parenthesis indicates limited comparability and a minus sign means that the variable is not available in a given year. In the last column the comparable number in the Youth Questionnaire is given, if it applies.

11.2 Special Features of Some Questions and Variables

The interviewees were asked if they did sports in their youth. If they answered affirmative, they were asked to write in the sport they participated in the most. This information was recorded in a numeric variable and categorised. Some categories could easily be coded, e.g. as soccer, whereas for others this was not possible. If you are interested in specific research questions regarding sports in youth, it would be possible to give you the original plain text information.

Individuals were asked: ‘When was the last year you attended school?’ (Question number 37, in Wave R). If individuals were still attending school, they had the opportunity to report that they were students. Unfortunately individuals who did this or who didn’t provide any answer to this question skipped over a lot of questions due to the questionnaire design. Consequently, for these individuals who were asked in one of the years 2001 - 2005 (not in 2000), there is no information on the number of foreign classmates⁷, on the aspired school degree and planned vocational qualification. Their record also lacks information on past vocational qualifications. However, this should prove less problematic, as most of the students were on the regular school process the entire time.

⁷ Most persons with no valid information on foreign classmates are not students. These are individuals who finished school abroad. It is in the sense of the questionnaire this is because the question only targets those in German schools.

Table 1: Description of the data set BIOSOC

| Variable Name | Content of the Variable | Variable Name in Biography Questionnaire 2001 ⁸ | Variable Present in the Year 2000 ⁹ | Variable Present in 2002 and Later Years ² | Comparable Question in the Youth Questionnaire 2001 ¹⁰ |
|------------------------------------|--|--|--|---|---|
| Entries for Surveyed Person | | | | | |
| HHNR | Original household identifier (invariant) | | + | + | |
| HHNRAKT | Actual household identifier | | + | + | |
| PERSNR | Personal identifier | | + | + | |
| BEFRPER | Respondent identifier | | + | + | |
| ERHEBJ | Survey year | | + | + | |
| BSGEBJAH | Year of birth | | + | + | |
| School | | | | | |
| BSELKUEM | Parents took care about efforts at school | RB29 | - | + | RY39 |
| BSNTDEUT | Last grade ¹¹ in German | RB30 | - | + | RY35 |
| BSNTMATH | Last grade ⁴ in maths | RB30 | - | + | RY35 |
| BSNTFMD1 | Last grade ⁴ in 1. foreign language | RB30 | - | + | RY35 |
| BSPTDEUT | Total points ⁴ in German (last class) | RB30 | - | - | RY35 |
| BSPTMATH | Total points ⁴ in maths (last class) | RB30 | - | - | RY35 |
| BSPTFMD1 | Total points ⁴ in 1. foreign language (last class) | RB30 | - | - | RY35 |
| BSGSDEUT | Level of German at comprehensive school ¹² (last class) | RB30 | - | - | RY35 |
| BSGSMATH | Level of maths at comprehensive school ⁵ (last class) | RB30 | - | - | RY35 |
| BSGSFMD1 | Level of 1. foreign language at comprehensive school ⁵ (last class) | RB30 | - | - | RY35 |
| BSLKDEUT | Complementary / main subject ¹³ in German (last class) | RB30 | - | - | RY35 |
| BSLKMATH | Complementary / main subject ⁶ in maths (last class) | RB30 | - | - | RY35 |
| BSLKFMD1 | Complementary / main subject ⁶ in 1. foreign language (last class) | RB30 | - | - | RY35 |

⁸ The variable names related to the Youth Questionnaire in Wave R start with “**RB**”. The first letter stands for the survey wave R (2001), the second letter for the surveying instrument, the Biography Questionnaire. The very same naming principle is used for the following waves S, T, etc.

⁹ Variable present: +; Variable not present: -.

¹⁰ Corresponding number in Youth Questionnaire 2001 (**RY**) if a comparable question exists, otherwise: -.

¹¹ To make the data set more user-friendly, the information given on points are transformed into marks and stored in the corresponding variable. The link between marks and points is as follows: 0 points: mark 6; 1 to 3 points: mark 5; 4 to 6 points: mark 4; 7 to 9 points: mark 3; 10 to 12 points: mark 2; 13 to 15 points: mark 1.

¹² 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 one.

¹³ From the 11th or 12th class on pupils can choose main subjects. Thus German, maths and foreign languages can be downgraded to a complementary subject.

... contd.

Relationships to Parents, Sport and Activities during Youth

Frequency of fights as respondent was 15. years old with:

| | | | | | |
|---|--|------|-----|---|--------|
| BSSTRVA | Father | RB31 | - | + | RY13 |
| BSSTRMU | Mother | RB31 | - | + | RY13 |
| BSSPRTRR | Participated in sports during youth | RB32 | - | + | RY20 |
| BSSPRTAR | Favourite sport during youth | RB33 | - | + | RY21 |
| BSSPRTWE | Participated in competitions during youth | RB34 | - | + | RY24 |
| BSMUSSP | Played music or sang during youth | RB35 | - | + | RY16 |
| School Attendance | | | | | |
| BSSCHBES | Still at school | RB37 | - | + | RY25 |
| BSSCHEND | Year left school | RB37 | - | + | RY26a |
| BSSCHWO | Country of last school attendance | RB38 | (+) | + | - |
| BSSCHLA | Federal State of last school attendance | RB41 | - | + | - |
| BSKLAUSL | Number of foreign classmates | RB43 | (+) | + | RY43 |
| BSSCHZUK | Strive for further school certificate | RB44 | + | + | RY27 |
| BSSCHZAR | Type of further school certificate | RB45 | + | + | RY28 |
| Attained and Planed Vocational Qualification | | | | | |
| BSBADABG | Vocational / university degree acquired in Germany | RB46 | + | + | (RY44) |
| Type of vocational / university degree attained in Germany: | | | | | |
| BSBADLEH | Apprenticeship ("Lehre") | RB47 | + | + | (RY45) |
| BSBADBFS | Full-time vocational school / School for public health ("Berufsfachschule / Schule des Gesundheitswesens") | RB47 | + | + | (RY45) |
| BSBADFSC | Technical school, school for master of a trade ("Fachschule, Meister-, Technikerschule") | RB47 | + | + | - |
| BSBADBEA | Training for civil servants (officer) ("Beamtenausbildung") | RB47 | + | + | - |
| BSBADFHA | Advanced technical college ("Fachhochschule") or approved vocational academy ("anerkannte Berufsakademie") | RB47 | + | + | - |
| BSBADUNI | University degree | RB47 | + | + | - |
| BSBADSON | Other vocational qualification | RB47 | + | + | (RY45) |
| BSBADEND | Year of attaining vocational / university degree in Germany | RB48 | - | + | |
| BSBAAABG | Vocational / university degree acquired abroad | RB49 | + | + | - |
| Type of vocational / university degree attained abroad: | | | | | |
| BSBAAFAN | Short-term training in a company | RB50 | + | + | - |
| BSBAAFBA | Apprenticeship in a company | RB50 | + | + | - |
| BSBAASCH | Vocational or professional school | RB50 | + | + | - |
| BSBAAUNI | University degree | RB50 | + | + | - |
| BSBAASON | Other vocational qualification | RB50 | + | + | - |

... contd.

| | | | | | |
|----------|---|------|---|---|------|
| BSBAAEND | Year of attaining vocational / university degree abroad | RB51 | - | + | - |
| BSBAAZEU | Certificate for abroad attained qualification | RB52 | + | + | - |
| BSBAAZEA | Recognition of aboard attained certificate | RB52 | + | + | - |
| BSZAJA | Vocational / university degree is aspired | RB53 | + | + | RY46 |
| | Type of aspired vocational / university degree: | | | | |
| BSZALEH | Apprenticeship ("Lehre") | RB54 | + | + | RY47 |
| BSZABFS | Full-time vocational school/ School for public health ("Berufsfachschule / Schule des Gesundheitswesens") | RB54 | + | + | RY47 |
| BSZAFSC | Technical school, school for master of a trade ("Fachschule, Meister-, Technikerschule") | RB54 | + | + | RY47 |
| BSZABEA | Training for civil servants (officer) ("Beamtenausbildung") | RB54 | + | + | RY47 |
| BSZABAK | Approved vocational academy ("anerkannte Berufsakademie") | RB54 | + | + | RY47 |
| BSZAFH | Advanced technical college ("Fachhochschule") | RB54 | + | + | RY47 |
| BSZAUNI | University degree | RB54 | + | + | RY47 |
| | Military and Voluntary Service | | | | |
| BSDIGEL | Military or alternative service done (<i>only men</i>) | RB71 | + | + | - |
| BSDIART | Type of service (<i>only men</i>) | RB71 | + | + | - |
| BSDIGRU | Reason for not serving (<i>only men</i>) | RB71 | + | + | - |
| BSFSJ | Voluntary social service ("Freiwilliges Soziales Jahr") | RB72 | - | + | - |
| | Specification of Interview Situation | | | | |
| BSINTA | Type of interview | | + | + | |
| BSDAUER1 | Duration of personal interview | | + | + | |
| BSDAUER2 | Duration of interview filled out independently | | + | + | |
| BSTAGIN | Day of the interview | | + | + | |
| BSMONIN | Month of the interview | | + | + | |
| BSINTNR | Identifier of the interviewer | | + | + | |

12. BIORESID: Variables On Occupancy and Second Residence

by Thorsten Schneider and Michael Frühling

In 1994 questions with a focus on occupancy have been introduced in the Biographical Questionnaire asking for the duration of residence in the current dwelling, and on any second residence. Questions on the second residence were also asked before 1994, however, these 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?*¹⁴

(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

Abroad¹⁵

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

¹⁴ In the years 1994 and 1995 the question was “Do you have another home, in Germany, in which you yourself reside in?”

¹⁵ The new category "abroad" was added in 1996.

12.1 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. Due to fieldwork-related reasons, some very few people have answered the Biography Questionnaire twice. For these cases, the first interview is taken as relevant for BIORESID. Further cases are dropped if their information stems from an interview held before the 1994.

12.2 Population of Interest

The data set BIORESID as of wave 2005 contains information on 18,047 individuals, stemming from samples A-G. The data set is supplemented every year by new respondents filling in the supplementary Biography Questionnaire.

Table 1: Samples in BIORESID

| sample | frequency | percent |
|-------------------------|-----------|---------|
| A: Germans (west) | 1709 | 9,5 |
| B: Foreigners (west) | 637 | 3,5 |
| C: Germans (east) | 1183 | 6,6 |
| D: Immigrants 1984-1993 | 1339 | 7,4 |
| E: Supplement 1998 | 1791 | 9,9 |
| F : Innovation 2000 | 9293 | 51,5 |
| G : High-Income 2002 | 2095 | 11,6 |

Status: up to wave V (2005)

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.

12.3 Variable List of the Data Set BIORESID

Table 2: 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 |

13. BIOAGE17: The Youth Questionnaire¹⁶

by Bettina Isengard and Michael Frühling

Since the year 2000, youth-specific questions have been integrated in the Socio-Economic Panel (SOEP) through an independent instrument, namely the Youth Questionnaire. This questionnaire collects information on the relationship with parents, on free time activities and on past efforts in school as well as information on personality characteristics. In addition, there are numerous prospective questions about educational plans and plans for further training, as well as questions about expectations on a future career and family.

A series of statements on specific circumstances, including the indicated expectations of the future, are indeed directly related to the time at which the questionnaire was administered. However, they provide a multifaceted background for analyses in the long run since the youths will continue to be questioned in subsequent years, as is the case for all SOEP-respondents. The questionnaire also contains retrospective questions, for example, at what age the teenager had his or her first job or first music lesson, which recommendations the teenager may have for the secondary school level I or which class level he or she may have repeated.

13.1 Genesis and Target Population of the Youth Questionnaire

The Youth Questionnaire is aimed at youths who have reached the surveying age of 17 years¹⁷ and who are therefore being questioned for the first time. This survey instrument takes the place of the supplementary Biography Questionnaire, since its questions don't apply to the youths' family and career situation. As a rule, entries for social origin also exist since the parents themselves have been surveyed. In the case where the teenager doesn't live with one or both parents the corresponding information will be collected on the missing parent(s). In addition, the standard questions regarding immigration from the supplementary Biography Questionnaire will be asked of youths who have immigrated. This guarantees that all important information which is collected in the Biography Questionnaire, is also available for the youths.

A preliminary version of the Youth Questionnaire was tested in 2000 in samples A-E on persons who were born in 1983. An expanded and revised questionnaire entered the field one year later in 2001, which included all samples (A-F). In samples A-E, the youths born in 1984 were surveyed, and in sample F those born in the years 1982 to 1984. With the expansion of

¹⁶ In earlier SOEP-data releases BIOAGE17 was called BIOYOUTH.

¹⁷ More precisely this refers to youths who are or will turn 17 years old in the year of the survey. Therefore they can be either 16 or 17 years old at the time of questioning.

the number of birth cohorts, entries for the birth year 1983 are also collected for sample F (data previously only existed for samples A-E), which also creates a clear increase in the number of entries. For an overview of the target population in each survey year see table 1. Therefore we have gathered interview data from 1940 analysable cases up to now.

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

| sample | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 |
|--------------|----------|----------------|----------|----------|----------|----------|
| A-E | 17 years | 17 years | 17 years | 17 years | 17 years | 17 years |
| F | | 17-19 years | 17 years | 17 years | 17 years | 17 years |
| G | | | | 17 years | 17 years | 17 years |
| observations | 232 | 618 | 352 | 365 | 373 | 368 |

13.2 Contents and Structure of the Data Set BIOAGE17

The set-up of the data set BIOAGE17 is orientated on the Youth Questionnaire from 2001 since it is the standard version for subsequent years. As in the biographical data survey, every youth answers the Youth Questionnaire only once. Therefore the data is written in a column as it would be in a cross-sectional record. The variable “erhebj” makes it possible to identify the year of the survey quickly. The entries to the questions which were only asked in 2000 and not in 2001 are not included in BIOAGE17. The complete data set from 2000 will be sent (free of charge) upon request. However, all entries from 2000 which are also included in 2001, are contained in BIOAGE17!¹⁸

Table 2 (at the end of the chapter) lists all of the variables for the data set BIOAGE17. The first column contains the name of the variable, the second a brief specification of its content and the third contains the number of the question as it appears in the Youth Questionnaire from 2001, Wave R, which functions as a reference for the variables. As the Youth Questionnaire wasn't altered in the following year, the reported questionnaire numbers are identical in 2002 to 2005, too (except variable 'byschbes', details below).¹⁹ The variables which contain the identification of the surveyed person and the interview situation have no corresponding number because these entries do not originate from the regular section of the Youth Questionnaire. In the last column it is noted whether the variable existed in the year 2000. A plus sign means that it did, a plus sign in parenthesis indicates a limited comparability, and a minus sign means that the variable was not included in 2000.

¹⁸ 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.

¹⁹ 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.

The topic blocks ‘Origin’ (questions 55 to 66b in Wave R, 2001) and ‘Childhood and Parents’ House’ (questions 67 to 78 in Wave R, 2001) are not in the data set BIOAGE17. The questions asked in these sections are taken from the Biography Questionnaire and the entries are stored in the corresponding biographical files (BIOIMMIG, BIOPAREN).

13.3 Special Features of Some Questions and Variables

The question regarding the support that the youths receive from their parents (question 14 in Wave R, 2001) is based on the Supportive Parenting Scale from Simons et al. (1992)²⁰ which was transformed for Germany by Schwarz and Walper (1997)²¹. The instrument used to compile career orientation (question 52 in Wave R, 2001) was taken from Kracke (1996)²².

One of the questions asked in the Youth Questionnaire is ‘do you still go to school?’ (variable ‘byschbes’, question 25 in Wave R, 2001). Underneath the question the interview directions ‘*Here vocational schools are **Not** meant*’ are given. Yet there are youths who answer this question with a “yes” even though they are no longer in the general school system. This problem can be identified when a teenager answers at the same time that he or she is currently in a vocational school or trainee program (question 45 in Wave R, 2001). In this case the variable ‘byschbes’ is given the value -3 (-3: Entry deleted after intensive examination). Since 2003, this question has an additional category for pupils attending a vocational school. Therefore the former problem doesn’t appear anymore.

A further recoding is undertaken 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. In this case the variable for school attendance will also be changed to -3. In order to make the data set user-friendly, the school type collected by the Individual Questionnaire (variable ‘schulart’) is integrated in the Youth Questionnaire.

In question 49 (Wave R, 2001) the youths are asked whether they know which career they would like to start. If this question is answered positively (‘yes, with some certainty’, ‘yes, with a lot of certainty’), then the aspired occupation is written down 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

²⁰ 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.

²¹ 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.

²² Kracke, B. (1996): *Fragebogen zur Berufsorientierung bei Realschülern.* University of Mannheim, unpublished manuscript.

Erikson's and Goldthorpe's Class Category (EGP)²³ as well as Wegener's Magnitude Prestige Scale (MPS)²⁴ are also given.

In 2005 some persons have the value -3 in variables bymusart (6 persons; 1,63 %), bymusmw (9 persons; 2,45 %) and bysprtmw (40 persons; 10,87 %). It means, that they give more than one answer to the question, although only one answer was possible. Because of referring more categories it was not possible to assign one valid answer.

Table 2: Description of the data set BIOAGE17

| Variable Name | Content of the Variable | Variable Name in Youth Questionnaire 2001 ²⁵ | Variable Present in the Year 2000 ²⁶ |
|------------------------------------|---|---|---|
| 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 from 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) | RY01 | + |
| BYWOZIM | Own room | RY02 | + |
| BYWOWEI | Additional apartment outside of parents' HH | RY03 | + |
| Jobs and Money | | | |
| BYVDEIG | Own income | RY04 | + |
| BYVDART | Type of income | RY05 | + |
| BYJBFRUE | Worked before (on holiday or while in school) | RY06 | + |
| BYJBALT | Age by first job (on holiday or while in school) | RY07 | + |
| BYJBGRUN | Reason for working | RY08 | + |
| BYTGELD | Allowance | RY09 | + |
| BYTGELDW | Amount of allowance per week | RY10 | + |
| BYTGELDM | Amount of allowance per month | RY10 | + |
| BYSPAR | Saving money | RY11 | + |
| BYSPARM | Amount saved every month | RY11 | + |

²³ 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.

²⁴ Frietsch, R. and H. Wirth (2001): Die Übertragung der Magnitude-Prestigeskala von Wegener auf die Klassifizierung der Berufe. in: ZUMA-Nachrichten, 48, 139-163.

²⁵ The variable names related to the Youth Questionnaire in Wave R start with "RY". The first letter stands for the survey wave R (2001), the second letter for the surveying instrument, the Youth Questionnaire. The very same naming principle is used for the following waves S, T, etc.

²⁶ Variable present in 2000: +; Variable not present in 2000: -.

| | | | |
|------------------------------------|---|------|---|
| BYSPARUN | Sporadic saving | RY11 | + |
| Relationships | | | |
| Importance of various persons: | | | |
| BYWIVA | Father | RY12 | - |
| BYWIMU | Mother | RY12 | - |
| BYWIBS | Brother, Sister | RY12 | - |
| BYWIVW | Other related persons | RY12 | - |
| BYWIFFR | Serious boy/girlfriend | RY12 | - |
| BYWIBFR | Best friend | RY12 | - |
| BYWILEHR | Teacher | RY12 | - |
| BYWICLQ | Clique | RY12 | - |
| BYWISON | Other persons | RY12 | - |
| Frequency of fights with: | | | |
| BYSTRVA | Father | RY13 | - |
| BYSTRMU | Mother | RY13 | - |
| BYSTRBS | Brother, Sister | RY13 | - |
| BYSTRFFR | Serious boy/girlfriend | RY13 | - |
| BYSTRBFR | Best friend | RY13 | - |
| BYBZ01MU | Talk with mother about personal experiences | RY14 | - |
| BYBZ01VA | Talk with father about personal experiences | RY14 | - |
| BYBZ02MU | Mother addresses problems | RY14 | - |
| BYBZ02VA | Father addresses problems | RY14 | - |
| BYBZ03MU | Mother asks opinion before a decision is made | RY14 | - |
| BYBZ03VA | Father asks opinion before a decision is made | RY14 | - |
| BYBZ04MU | Mother shows approval | RY14 | - |
| BYBZ04VA | Father shows approval | RY14 | - |
| BYBZ05MU | Solve problems together with mother | RY14 | - |
| BYBZ05VA | Solve problems together with father | RY14 | - |
| BYBZ06MU | Mother shows trust | RY14 | - |
| BYBZ06VA | Father shows trust | RY14 | - |
| BYBZ07MU | Mother asks opinion on family issues | RY14 | - |
| BYBZ07VA | Father asks opinion on family issues | RY14 | - |
| BYBZ08MU | Mother justifies decision | RY14 | - |
| BYBZ08VA | Father justifies decision | RY14 | - |
| BYBZ09MU | Mother shows love | RY14 | - |
| BYBZ09VA | Father shows love | RY14 | - |
| Free time and Sport | | | |
| Frequency of free time activities: | | | |
| BYFZFERN | TV, Video | RY15 | - |
| BYFZPC | Computer games | RY15 | - |
| BYFZMUSH | Listen to music | RY15 | - |
| BYFZMUSS | Play music | RY15 | - |
| BYFZSPRT | Do sports | RY15 | - |
| BYFZTANZ | Dance, Theatre | RY15 | - |
| BYFZTECH | Technical work, Programming | RY15 | - |
| BYFZLESE | Read | RY15 | - |
| BYFZEHRE | Volunteer activities | RY15 | - |
| BYFZABH | Do nothing, hang around, day dream | RY15 | - |
| BYFZMFFR | Spend time with boy/girlfriend | RY15 | - |
| BYFZMBFR | Spend time with best friend | RY15 | - |
| BYFZMCLQ | Spend time with clique | RY15 | - |
| BYMUSSP | Actively make music | RY16 | - |
| BYMUSART | Style of music made | RY17 | - |
| BYMUSMW | Play music with whom | RY17 | - |
| BYMUSALT | Age starting playing music | RY18 | - |

| | | | |
|------------------------|---|--------------------|---|
| BYMUSUNT | Paid music lessons | RY19 | - |
| BYSPTTR | Participate in sports | RY20 | + |
| BYSPTAR | Favourite sport | RY21 | + |
| BYSPTAL | Age started favourite sport | RY22 | + |
| BYSPTMW | Where and with whom favourite sport | RY23 | + |
| BYSPTWE | Participation in competitions | RY24 | + |
| School | | | |
| BYSCHBES ²⁷ | School attendance | RY25 | + |
| BYSCHART | Type of general school attended | RI11 ²⁸ | + |
| BYSCHEND | Last year of school | RY26a | + |
| BYSCHABS | Type of school certificate | RY26b | + |
| BYSCHZUK | Strive for further school certificate | RY27 | + |
| BYSCHZAR | Type of further school certificate | RY28 | + |
| BYFMD1 | 1. foreign language | RY29 | + |
| BYFMD2 | 2. foreign language | RY29 | + |
| BYSCH AUS | School attendance in foreign country | RY30 | + |
| BYSCHPRI | Attendance in a private school | RY31 | - |
| Activities in school: | | | |
| BYENKSPR | Class representative | RY32 | + |
| BYENSSPR | School representative | RY32 | + |
| BYENSZTG | School newspaper | RY32 | + |
| BYENTHEA | Theatre, Dance group | RY32 | + |
| BYENCHOR | Choir, Music | RY32 | + |
| BYENSPRT | Sport group | RY32 | + |
| BYENSONS | Other groups | RY32 | + |
| BYENNEIN | No activities | RY32 | + |
| BYZFINSG | Satisfaction with effort at school (overall) | RY33 | + |
| BYZFDEUT | Satisfaction with effort in German | RY33 | + |
| BYZFMATH | Satisfaction with effort in math | RY33 | + |
| BYZFFMD1 | Satisfaction with effort in 1. foreign language | RY33 | + |
| BYEMPFEH | Recommendation after elementary school | RY34 | - |
| BYNTDEUT | Last grade ²⁹ in German | RY35 | + |
| BYNTMATH | Last grade ¹¹ in math | RY35 | + |
| BYNTFMD1 | Last grade ¹¹ in 1. foreign language | RY35 | + |
| BYPTDEUT | Total points ³⁰ in German | RY35 | + |
| BYPTMATH | Total points ¹² in math | RY35 | + |
| BYPTFMD1 | Total points ¹² in 1. foreign language | RY35 | + |
| BYGSDEUT | Level of German at comprehensive school ³¹ | RY35 | + |
| BYGSMATH | Level of math at comprehensive school ¹³ | RY35 | + |

²⁷ Additional category since 2003: value 3 'yes, vocational school'.

²⁸ The variable „schulart“ stems from the Individual-Questionnaire. The question number that is noted refers to the number in the Individual-Questionnaire from Wave **R**.

²⁹ Pupils normally get marks ranging from 1 to 6, whereby 1 is the best and 6 the worst mark. This system of giving marks is used up until the 11th class or the 12th class (secondary school level II in upper secondary school 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 marks and the information on points transformed into marks is stored in this variable. Note: No corrections have been made when a person has reported marks and points and when both types of information do not correctly correspond.

³⁰ From the 11th or 12th class on, pupils get points in upper secondary school or comprehensive school ranging from 0 to 15, whereby 15 points are the best, 0 points the worst. The link between marks and points is as follows: 0 points: mark 6; 1 to 3 points: mark 5; 4 to 6 points: mark 4; 7 to 9 points: mark 3; 10 to 12 points: mark 2; 13 to 15 points: mark 1.

³¹ 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 one. The number of levels differs between the federal states.

| | | | |
|-----------------------------------|--|------|-----|
| BYGSFMD1 | Level of 1. foreign language at comprehensive school ¹³ | RY35 | + |
| BYLKDEUT | Complementary / main subject ³² in German | RY35 | + |
| BYLKMATH | Complementary / main subject ¹⁴ in math | RY35 | + |
| BYLKFMD1 | Complementary / main subject ¹⁴ in 1. foreign language | RY35 | + |
| BYKLWDJA | Class repeated | RY36 | + |
| BYKLWD1 | Class level 1. repeated | RY37 | + |
| BYKLWD2 | Class level 2. repeated | RY37 | + |
| BYNACHHI | Paid tutor lessons | RY38 | + |
| BYELKUEM | Parents care about efforts at school | RY39 | + |
| BYELHAUS | Parents help with homework | RY40 | + |
| BYELDIFP | Problems with parents because of effort at school | RY41 | + |
| BYELABEN | Parents attend parents' evening | RY42 | + |
| BYELSPRE | Parents go to parents' day | RY42 | + |
| BYELLEHR | Parents go to see a teacher | RY42 | + |
| BYELVERT | Active as parent representative | RY42 | + |
| BYELNIDA | Parents do not participate in any of these activities | RY42 | + |
| BYKLAUSL | Number of foreign classmates | RY43 | (+) |
| Education and Career plans | | | |
| BYBAABGE | Vocational education, Internship, training | RY44 | + |
| BYBABGJ | Vocational introductory year ("Berufsgrundschul- / Berufsvorbereitungsjahr") | RY45 | + |
| BYBABEGL | Vocational integration training ("Berufl. Eingliederungslehrgaenge") | RY45 | + |
| BYBALEH | Vocational education, apprenticeship ("Berufsausbildung, Lehre") | RY45 | + |
| BYBABFS | Full-time vocational school/ School for public health ("Berufsfachschule / Schule des Gesundheitswesens") | RY45 | + |
| BYBAPRAK | Internship ("Praktikum, Voluntary") | RY45 | + |
| BYZAJA | Vocational / university degree is aspired | RY46 | + |
| | Type of aspired vocational / university degree: | | |
| BYZALEH | Apprenticeship ("Lehre") | RY47 | + |
| BYZABFS | Full-time vocational school/ School for public health ("Berufsfachschule / Schule des Gesundheitswesens") | RY47 | + |
| BYZAFSC | Technical school, school for master of a trade ("Fachschule, Meister-, Technikerschule") | RY47 | + |
| BYZABEA | Training for civil servants (officer) ("Beamtenausbildung") | RY47 | + |
| BYZABAK | Approved vocational academy ("anerkannte Berufsakademie") | RY47 | + |
| BYZAFH | Advanced technical college ("Fachhochschule") | RY47 | + |
| BYZAUNI | University | RY47 | + |
| BYSLBALT | Desired age for financial independence | RY48 | + |
| BYSLBHEU | Already financially independent | RY48 | + |
| BYBWUNJA | Occupation is aspired | RY49 | + |
| | Occupation categories, encoded: | | |
| BYKLAS | Classification of career according to the Federal Statistical Office, Germany, (Statistisches Bundesamt), version 1992 | RY50 | + |
| BYISCO88 | International Standard Classification of Occupation 1988 (ISCO88) | RY50 | + |
| BYEGP | Erikson and Goldthorpe's Class Category (EGP) | RY50 | + |
| BYISEI | International Socio-Economic Index of Occupational Status after Ganzeboom (ISEI) | RY50 | + |
| BYSIOPS | Treiman's Standard International Occupational Prestige Scale (SIOPS) | RY50 | + |
| BYMPS | Magnitude Prestige Scale after Wegener (MPS) | RY50 | + |
| BYZBINF | Information level of planned career | RY51 | - |
| BYZBELT | Influence of the parents on career choice | RY52 | - |
| BYZBLAS | No specific career in mind | RY52 | - |

³² From the 11th or 12th class on pupils can choose main subjects. Thus German, math and foreign languages can be downgraded to a complementary subject.

| | | | |
|----------|--|------|---|
| BYZBBES | Intensive thoughts about various careers | RY52 | - |
| BYZBRAU | Still looking for a career | RY52 | - |
| | Important aspects for the career choice: | | |
| BYWBSICH | Secure job | RY53 | + |
| BYWBEINK | High income | RY53 | + |
| BYWBAUF | Promotion opportunities | RY53 | + |
| BYWBANE | Established profession | RY53 | + |
| BYWBFREI | Enough free time | RY53 | + |
| BYWBINT | Interesting activities | RY53 | + |
| BYWBSELB | Working independently | RY53 | + |
| BYWBKONT | Contact with persons | RY53 | + |
| BYWBGSL | Relevant to society | RY53 | + |
| BYWBGSD | Healthy conditions at work | RY53 | + |
| BYWBFAM | Flexibility for family | RY53 | + |
| BYWBHELP | Help others | RY53 | + |
| | Future | | |
| | Probability of future career related and private events: | | |
| BYWAAUSP | To be accepted for a desired apprenticeship / place at university | RY54 | + |
| BYWAERFA | To complete Training/ university successfully | RY54 | + |
| BYWAARBP | Job in desired career | RY54 | + |
| BYWABERF | Job-related success | RY54 | + |
| BYWAARBL | Longer unemployment | RY54 | + |
| BYWAZURU | From family related reasons held back in career | RY54 | + |
| BYWASELB | Self-employed | RY54 | + |
| BYWAAUSL | Work in foreign country | RY54 | + |
| BYWAHEIR | To marry | RY54 | + |
| BYWAPART | Live together with partner (not married) | RY54 | + |
| BYWAKID1 | Have one child | RY54 | + |
| BYWAKIDM | Have more than one child | RY54 | + |
| | Attitudes and Opinions | | |
| BYGLPART | Happiness: live with/without partner | RY79 | + |
| BYGLKIND | Happiness: with/without children | RY80 | - |
| | Success in FRG from | | |
| BYEFFLEI | Studiosness | RY81 | + |
| BYEFAUSN | Exploitation of others | RY81 | + |
| BYEFINT | Intelligence | RY81 | + |
| BYEFFAM | Family's origin | RY81 | + |
| BYEFFACH | Technical know-how | RY81 | + |
| BYEFGELD | Money | RY81 | + |
| BYEFSABS | School education | RY81 | + |
| BYEFHART | Being inconsiderate and hard | RY81 | + |
| BYEFBEZ | Networking | RY81 | + |
| BYEFPOLI | Political activities | RY81 | + |
| BYEFMANN | Sex/ 'being a man' | RY81 | + |
| BYEFINI | Being dynamic and taking initiative | RY81 | + |
| BYESVERL | What happens in life, depends on me | RY82 | - |
| BYESERRE | Did not reach, what I deserve | RY82 | - |
| BYESGLUE | What you achieve, is a matter of luck | RY82 | - |
| BYESAND | Others decide about my life | RY82 | - |
| BYESHART | You have to work hard for success | RY82 | - |
| BYESZWEI | By difficulties, doubt about own abilities | RY82 | - |
| BYESSOZU | Chances are determined by social circumstances | RY82 | - |
| BYESFAEH | Abilities are more important than efforts | RY82 | - |
| BYESKNTR | Little control over events in my life | RY82 | - |
| BYESENGA | Change of social circumstances through social/political activities | RY82 | - |

| 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 | + |

14. BIOAGE01: Generated Variables from the ‘Mother and Child Questionnaire’

by Jürgen Schupp and Katharina Mahne

Since 2003, questions regarding the birth of a child have been integrated in the SOEP with the help of the "mother and child questionnaire". The questionnaire is aimed at all women who, in the current survey year or the year before, gave birth to their own child, as well as at women whose non-biological child was born in the time period mentioned above. This means that at the time of the survey these children are either newborn or they have a maximum age of one and a half years.

The questionnaire is comprised of 19 questions which cover 4 subjects:

- pregnancy
- body measurements and health of the child
- change in living circumstances due to the birth of the child
- circumstances surrounding the care of the child

In the future this additional questionnaire will be used for newborn cohorts on an annual basis. There are plans to follow and record the children's development by surveying them with additional brief questionnaires at certain points in time.

The data set currently consists of 812 children, of which there are 12 pairs of twins and 11 pairs of siblings. This means that information from the mother-child questionnaire has been collected for 789 mothers.

Number of Respondents of Mother & Child Questionnaire

| | <i>Year of Data Collection</i> | | | |
|----------------------|--------------------------------|------|------|--------------|
| <i>Year of birth</i> | 2003 | 2004 | 2005 | Total |
| 2002 | 274 | 2 | | 276 |
| 2003 | 44 | 204 | | 248 |
| 2004 | | 41 | 198 | 239 |
| 2005 | | | 49 | 49 |
| Total | 318 | 247 | 247 | 812 |

14.1 Contents and aims of the BIOAGE01 data file

The aim of the BIOAGE01 data file is to observe the future generation of the SOEP, preferably from birth onwards. The data set is in the order of the never-changing person ID of the child, so that information from the BIOAGE01 data set can be directly linked to the child files (\$KIND). From a mother's point of view, the data set presents a detailed source of information on pregnancy and the changes in life experienced by women who have recently become mothers.

The basis for BIOAGE01 therefore consists of all individuals who are named as children in the "Mother and Child Questionnaire" (PERSNR). With the help of the mother's never-changing person ID, information on the mothers can also be directly linked to the individual information on the mothers.

In the case of multiple children, the mother fills out a respective number of questionnaires, so that siblings can be identified by the identical never-changing person ID of the mother (PERSNRM), month of birth and year of birth.

The variables in BIOAGE01 correspond to their structure in the user-friendly original variables from the mother and child questionnaire. Information is provided on the time and place of birth, as well as the child's height and weight at birth. The variable BCKALTER gives the age of the child at the time of the survey in months. With regards to the children, there is information on any disorders in their development, as well as the health of the child in its first 3 months. BIOAGE01 also includes information on whether or not the pregnancy was planned, if the child is the first and if it is a biological child. For mothers with a non-biological child, the questions are presented in an identical fashion, however questions on pregnancy are left out. Finally, there are questions on the mother's evaluation of the child's behaviour at the time of the interview. The variables BCVERAE1 to 8 refer to the mother's assessment of the new circumstances of life. BIOAGE01 also contains information on the mother's current personal situation, i.e. whether there is a partner present, as well as to what extent the child is cared for by people other than the main care provider.

14.2 Variables in BIOAGE01

Number of cases: 812 Wave: T-V Sub-samples: A, B, C, D1, D2, E, F, G

| VARIABLE NAME | Content of the Variable |
|----------------------|--|
| HHNRAKT | Current wave household number |
| PERSNRM | Never-changing person ID of the mother |
| PERSNR | Never-changing person ID of the child |
| ERHEBJ | Survey year |
| BCKGEBMO | Child's month of birth |
| BCKGEBJA | Child's year of birth |

| | |
|----------|---|
| BCKALTER | Age (in months) of the child at the time of the survey |
| BCKSEX | Child's gender |
| BCENTBIN | Place where birth took place |
| BCSSW | Pregnancy week of birth |
| BCKGEW | Weight of child at birth in grams |
| BCKGROE | Height of child in cm |
| BCKKOPF | Head circumference of the child in cm |
| BCKLETZU | Last medical examination |
| BCKSTOER | Child has confirmed disorders |
| BCKARZT | Medical help - number of times in the first 3 months |
| BCKKRHAU | Length of hospital stay in the first 3 months in days |
| BCKIZAHL | Newborn is 1st, 2nd, 3rd, etc. child |
| BCKLEIBL | Biological child |
| BCSSPLAN | Pregnancy planned/unplanned |
| BCVATER | Father lives in household |
| BCUNTPA | Supported by partner |
| BCHAUPTB | Mother is main provider of care |
| BCBEFIN1 | Physical condition in the final third of the pregnancy |
| BCBEFIN2 | Physical condition in the first 3 months after birth |
| BCBEFIN3 | Mental state in the final third of the pregnancy |
| BCBEFIN4 | Mental state in the first 3 months after birth |
| BCVERAE1 | Circumstances in life have greatly changed |
| BCVERAE2 | Child provides happiness and joy |
| BCVERAE3 | Often close to running out of strength |
| BCVERAE4 | Very satisfied with the role of mother |
| BCVERAE5 | Often unable to cope with tasks/responsibilities |
| BCVERAE6 | Have made new contacts through the child |
| BCVERAE7 | Suffering from being limited to the role of mother |
| BCVERAE8 | Important to provide the child with much affection |
| BCBETRE1 | Cared for by partner (in h/week) |
| BCBETRE2 | Cared for by grandparents (in h/week) |
| BCBETRE3 | Cared for by grandparents by older siblings (in h/week) |
| BCBETRE4 | Cared for by other relatives (in h/week) |
| BCBETRE5 | Cared for by childminder (in h/week) |
| BCBETRE6 | Cared for in crèche/day nursery (in h/week) |
| BCBETRE7 | Cared for by others (in h/week) |
| BCBETRE8 | No use of temporary care |
| BCKGESU1 | Concerns about the child's health |
| BCKGESU2 | Child is generally happy and satisfied |
| BCKGESU3 | Child is easily irritated and often cries |
| BCKGESU4 | Child is hard to console |
| BCKGESU5 | Child is curious and active |

15. BIOAGE03: Generated Variables from the supplementary questionnaire “Your child between the ages of two and three”

By Jürgen Schupp and Katharina Mahne

This questionnaire was used for the first time in 2005 after completion of pretesting in the year 2004. The questionnaire is given to mothers with a child between the ages of two and three at the time of the survey. The year of the child’s birth is the determining factor here: thus, in the 2005 survey, all children born in the year 2002 were included. The questionnaire seeks indicators for the child’s individual course of development and the mother’s specific experiences with her child. It also extends the indicators collected since 2003 through the Mother & Child questionnaire further; for the children in these cohorts—the majority of whom were born in the year 2002—the new supplementary questionnaire acts as a continuation of the information already collected on individual development characteristics. To provide an incentive to participate, after completing the questionnaire, mothers are given beside their individual lottery ticket an additional lottery ticket issued in the child’s name to the *Aktion Mensch* lottery. *Aktion Mensch* provides funding to over 300 social projects targeting developmentally disabled people as well as children and young adults.

The questionnaire consists of 13 questions in the following main thematic areas:

- Body measurements and health
- Recreational activities
- Child’s abilities
- Childcare situation
- Parental experiences with the child

In the future, the questionnaire will be sent only once per year to the cohort of mothers of two to three-year-old children. Ideally, they will already have answered the questionnaire “Mother & Child”.

The data set currently consists of 257 children including 2 pairs of twins. Thus, 255 mothers responded to the questionnaire. For 244 of the children, there is additional information available in BIOAGE03.

Content and goal of the BIOAGE03 dataset

The BIOAGE03 dataset makes up one part of our observation of the young people comprising the new SOEP generation before they reach respondent age themselves, a process we aim to make as comprehensive and gap-free as possible. We sort the data according to the fixed original household number HHNR as well as the unchanged person ID number of the child (PERSNR) so that information from BIOAGE03 can be directly linked to the annual child data (\$KIND) or to PPFAD or PHRF.

The BIOAGE03 dataset is based on all those individuals who are referred to as children in the questionnaire. If a mother has more than one child of this age, she fills out the corresponding number of questionnaires—one per child. It can be determined whether the child has siblings through the identical person ID number of the mother (PERSNRM). The mother's permanent person ID number can also be used to establish links to data from the individual and household files.

The variables in BIOAGE03 correspond to the answers to the questionnaire “Your child between the ages of two and three”. The information provides both the month and year of birth and also the current height and weight of the child. The generated variable BKALTER provides the current age of the child at the point in time of the survey and in that particular year. The data set contains not only information on childhood illnesses, hospital stays, and doctor visits, but also information on the child's behavior and level of development from the mother's point of view. Very detailed data is collected on the child's abilities: linguistic capacities, everyday skills, motor abilities, and social relationships (according to the Vineland Adaptive Behavior Scale). There are also detailed questions regarding the childcare situation. In addition to questions about recreational activities during the last 14 days and the child's TV-watching habits, we ask whether household members usually speak German or another language with the child at home.

Variables in BIOAGE03

Number of observations: 257

Wave: V

Samples: A, B, C, D, E, F, G

| | |
|----------|---|
| HHNRAKT | Current Wave HH Number |
| PERSNR | Permanent Person ID Child |
| PERSNRM | Permanent Person ID Mother |
| ERHEBJ | Survey Year |
| BKGEBMO | Child, Month of Birth |
| BKGEBJA | Child, Year of Birth |
| BKALTER | Age of Child (in Months) at Time of Survey |
| BKGEW | Child, Weight in kg |
| BKGROE | Child, Height in cm |
| BKSTOER | Child Has Confirmed Disorder |
| BKARZT | Child, Medical Care: No. Of Times last 3 months |
| BKKRHAU | Child, Length of Hospital Stays Last 12 Months In Days |
| BKBETRE1 | Cared for by Spouse/Partner: (Hrs/Wk) |
| BKBETRE2 | Cared for by Childs Father (If Not Resident of Same Household) (Hrs/Wk) |
| BKBETRE3 | Cared for by Grandparents: (Hrs/Wk) |
| BKBETRE4 | Cared for by Older Siblings: Hrs Per Wk |
| BKBETRE5 | Cared for by Other Relatives: Hrs Per Wk |
| BKBETRE6 | Cared for in Family Day Care (Hrs/Wk) |
| BKBETRE7 | Cared for by a Nanny or Other In-Home Daycare Provider (Hrs/Wk) |
| BKBETRE8 | Cared for at a Daycare Center (Hrs/Wk) |
| BKBETRE9 | Cared for by Others (Hrs/Wk) |
| BKBETRE0 | Cared for Solely by Respondent |
| BKGESU1 | Child Usually Happy, Content |
| BKGESU2 | Child Irritable / Cries Frequently |
| BKGESU3 | Child Difficult to Console when Crying |
| BKGESU4 | Child Curious, Active |
| BKGESU5 | Child Communicative and Talkative |
| BKGESU6 | Child Shows Empathy when Others Are Sad |
| BKGESU7 | Worried About Childs Health |
| BKAKT1 | Singing Childrens Songs to/with Child – No. of Times during last 14 days |
| BKAKT2 | Taking Walks Outdoors, No. of Times during last 14 days |
| BKAKT3 | Painting or Doing Arts and Crafts - No. of Times during last 14 days |
| BKAKT4 | Reading or Telling Stories - No. of Times during last 14 days |
| BKAKT5 | Looking at Picture Books -No. of Times during last 14 days |
| BKAKT6 | Going to Playground - No. of Times during last 14 days |
| BKAKT7 | Visiting Other Families with Children - No. of Times during last 14 days |
| BKAKT8 | Going Shopping with Child - No. of Times during last 14 days |
| BKAKT9 | Watching Television or Videos with Child - No. of Times during last 14 days |
| BKTV | Child Allowed to Watch Television or Videos Alone |
| BKEIG1 | Child Tends to Be Shy / Outgoing |
| BKEIG2 | Child Tends to Be Focused / Easily Distracted |
| BKEIG3 | Child Tends to Be Obstinate / Obedient |
| BKEIG4 | Child Tends to Be Quick to Learn New Things / to Need More Time |
| BKSPR1 | Child Understands Brief Instructions |
| BKSPR2 | Child Forms Sentences with at Least Two Words |
| BKSPR3 | Child Speaks in Full Sentences (at Least Four Words) |
| BKSPR4 | Child Listens Attentively to a Story for at Least Five Minutes |
| BKSPR5 | Child Can Relate Simple Messages |

| | |
|----------|--|
| BKALLT1 | Child Eats with Spoon without Making a Mess |
| BKALLT2 | Child Blows Nose without Assistance |
| BKALLT3 | Child Uses Toilet to Do Number Two |
| BKALLT4 | Child Puts On Pants and Underpants Frontwards |
| BKALLT5 | Child Brushes Teeth without Assistance |
| BKBEW1 | Child Walks Forwards Down the Stairs |
| BKBEW2 | Child Uses Door Handle to Open Doors |
| BKBEW3 | Child Climbs Jungle Gyms and Other High Playground Equipment |
| BKBEW4 | Child Uses Scissors to Cut Paper |
| BKBEW5 | Child Paints / Draws Recognizable Forms on Paper |
| BKSOZ1 | Child Calls Familiar People by Name |
| BKSOZ2 | Child Plays Games with Other Children |
| BKSOZ3 | Child Participates in Role-Playing Games |
| BKSOZ4 | Child Shows Particular Liking for Certain Playmates or Friends |
| BKSOZ5 | Child Calls His/Her Own Feelings by Name, e.g., Sad, Happy, scared |
| BKDEUT | Language Spoken with Child |
| BKSEX | Childs Gender |
| BKSTOE1 | Asthma |
| BKSTOE2 | Chronic Bronchitis |
| BKSTOE3 | Spastic / Acute Bronchitis |
| BKSTOE4 | Pseudocroup / Croup Syndrome |
| BKSTOE5 | Middle-Ear Inflammation |
| BKSTOE6 | Hayfever |
| BKSTOE7 | Neurodermatitis |
| BKSTOE8 | Vision Impairment |
| BKSTOE9 | Hearing Impairment |
| BKSTOE10 | Nutritional Disorders |
| BKSTOE11 | Motor Impairment |
| BKSTOE12 | Other Impairments or Disorders |