

## SOEP Survey Papers

Series D – Variable Descriptions and Coding

SOEP – The German Socio-Economic Panel study at DIW Berlin

2018

# SOEP-Core v33.1 – BIOSOC: Retrospective Data on Youth and Socialization

Tabea Naujoks, Marco Giesselmann, and SOEP Group

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

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

**Series A** – Survey Instruments (Erhebungsinstrumente)

**Series B** – Survey Reports (Methodenberichte)

**Series C** – Data Documentation (Datendokumentationen)

**Series D** – Variable Descriptions and Coding

**Series E** – SOEPmonitors

**Series F** – SOEP Newsletters

**Series G** – General Issues and Teaching Materials

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

**Editors:**

Dr. Jan Goebel, DIW Berlin

Prof. Dr. Stefan Liebig, DIW Berlin and Universität Bielefeld

Dr. David Richter, DIW Berlin

Prof. Dr. Carsten Schröder, DIW Berlin and Freie Universität Berlin

Prof. Dr. Jürgen Schupp, DIW Berlin and Freie Universität Berlin

Please cite this paper as follows:

Tabea Naujoks, Marco Giesselmann, and SOEP Group. 2018. SOEP-Core v33.1 – BIOSOC: Retrospective Data on Youth and Socialization. SOEP Survey Papers 586: Series D. Berlin: DIW/SOEP



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

© 2018 by SOEP

ISSN: 2193-5580 (online)

DIW Berlin

German Socio-Economic Panel (SOEP)

Mohrenstr. 58

10117 Berlin

Germany

soeppapers@diw.de

# **SOEP-Core v33.1 – BIOSOC: Retrospective Data on Youth and Socialization**

**Tabea Naujoks, Marco Giesselmann, and SOEP Group**

# BIOSOC: Retrospective Data on Youth and Socialization

by Tabea Naujoks and Marco Giesselmann<sup>1</sup>

The standard supplementary Biography Questionnaire was expanded in 2000, and again in 2001 to include some specific questions on youth and early adulthood. Some of these questions are consistent with content of the independent Youth Questionnaire (for detailed information on the Youth Questionnaire, see chapter 13). The expanded questionnaire asks respondents of all ages to describe aspects of their life at the age of 15, including their relationship with parents, grades in school, the federal state where they last attained educational qualifications, detailed information on vocational qualifications, as well as intentions to complete further education or vocational training. Questions concerning military and alternative services are also included in this data set.

As these questions are a part of the standard Biography Questionnaire, they are only asked once. Some of these questions can, however, be followed up by the regular data collected in the Individual Questionnaire. For example, if someone was too young to have completed his educational career when the Biography Questionnaire was conducted, the user can look at the yearly data in later years or at the BIOEDU dataset to track the complete educational career of a sample member.

survey year	frequency
2000	246
2001	8,819
2002	552
2003	2,328
2004	450
2005	299
2006	223
2007	2,232
2008	336
2009	196
2010	2,038
2011	14,797
2012	2,855
2013	5,370
2014	552
2015	2,129
2016	5,181
<b>Total</b>	<b>48,603</b>

sample	frequency
A: Germans (west)	951
B: Foreigners (west)	264
C: Germans (east)	644
D: Immigrants 1984- 1993	135
E: Supplement 1998	259
F: Innovation 2000	9,877
G: High-Income 2002	2,269
H: Supplement 2006	2,208
I: Incentivation 2009	1,870
J: Supplement 2011	5,457
K: Supplement 2012	2,601
L: Family Samples 11/12	10,035
M: Migrants 13/14	7,217
M: Migrants 2016	4,816

Status: up to wave BG (2016)

<sup>1</sup> Replaces earlier versions by Henning Lohmann and Sven Witzke; Jürgen Schupp and Michael Frühling; Bettina Isengard and Thorsten Schneider and Mila Staneva.

The data set BIOSOC contains information on all sample members that entered the sample via refreshments or household formation since 2000. These are 48,603 persons. Biographical data collected before 2000 are stored in the dataset BIOL. The dataset BIOL contains additional 26,259 persons from 1984 to 1999.

**In 2014, Data from the SOEP-related FID-study from the years 2010 to 2014 has been integrated in the SOEP (Sample L). Therefore, the number of cases in BIOSOC has increased with SOEP Version 31 retrospectively for the years 2011 to 2013, compared with previous versions. You might want to use the `psample`-variable as filter, if you want to exclude these cases and reproduce your old sample.**

## 1 Structure of the Data Set BIOSOC

Respondents are given the Biography Questionnaire only once in a lifetime. Some of the information stored in the new data set BIOSOC is invariant (such as the relationship to parents at the age of 15) or is not surveyed to such an extent in the regular questionnaire (such as last school grades). Consequently there is only one record for each person and updates are not intended for this data set. The variable SYEAR makes it possible to quickly identify the year of the survey. Using the variable BSGEBJAH, which contains the year of birth, the user can determine the respondent's age. If the respondent is of a certain age, one can assume that some of the variables are constant. This applies to variables such as last school grades or military service.

In **Table 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 of wave Q (2000). The columns four to eight contain the question numbers in waves R to BG (2001-2016). The next four columns list question numbers of the Migration Questionnaire. The penultimate column contains the number of question for the refugee samples. In the last column the corresponding variable in the BIOAGE17 data set is given, if available. A minus sign means that the variable is not available in a given years and a question number in parenthesis indicates limited comparability.

## 2 Special Features of Some Questions and Variables

- The interviewees were asked if they did sports in their youth. If a respondents answer was positive, (s)he were asked to include the sport they participated in most frequently. This information was re-coded to a numeric variable and categorised. Some categories could easily be coded, such as soccer, whereas for others this was not possible. For users

interested in specific research questions on sports in youth, the original plain text answers can be provided upon request.

- Respondents were asked: “When was the last year you attended school?” (Question 46). If they were still attending school, they had the opportunity to report that they were students. Unfortunately in the years 2001 to 2003, individuals who reported being students or who did not provide any answer to this question skipped over numerous questions due to the questionnaire design. Consequently, for these individuals there is no information on the number of foreign classmates<sup>2</sup>, on the school degree aspired to, or planned vocational qualifications. Their record also lacks information on past vocational qualifications. However, this should prove less problematic since most of the students were enrolled in regular school programs throughout the entire time.
- In 2011, the Biography Questionnaire was integrated in the Individual Questionnaire for the samples J and K. For these samples a new variable was introduced – BSDAUER3 – which shows the interview duration with the integrated questionnaire version. All other samples are coded with “-6” on this variable.
- In 2012, sample M containing only respondents with migration background was added in SOEP (*SOEP Survey Paper 216*). This sample was surveyed with a modified version of the Biography Questionnaire which focuses on the immigration history and therefore misses some of the standard questions from the original Biography Questionnaire (*SOEP Survey Paper 218*). The last column in **Table 1** gives an overview of the variables in BIOSOC available for the migration sample. A minus sign means that the variable is not available for sample M (in the Data, we used the missing code “-5” in such cases). A question number in parenthesis indicates that the question in the Migration Questionnaire is comparable, but not identical to the question in the original Biography Questionnaire.
- In 2016 a new refugee sample was integrated into BIOSOC. The questions about vocational training and university degree acquired abroad and their recognition in Germany were asked in loops: Questions about the application and result of the recognition process were asked for every acquired degree. Therefore, the variables BSBAEND, SBAAZEU, SBAAARA, SBAAAGA, SBAAAMB and SBAAAJB are the condensed variables. The underlying selection procedure is described by the example SBAAARA (result recognition). If a person reports one acquired degree, “SBAAARA” contains the code related to this degree. If a person has more than one degree, we determined the code of SBAAARA according to a fixed set of rules: If a person has only one certificate and several more trainings without certificate, the code of the training with a certificate was taken. It also appears that a

---

<sup>2</sup> Most persons with no valid information on foreign classmates are not students but individuals who finished school abroad. This is because the question only targets those in German schools.

person has multiple degrees and more than one certificate. In this case, we took the code of the most recent degree with certificate. If a person has multiple degrees and multiple certificates in the same year, we took the value of the most prestigious certificate.<sup>3</sup> The majority of the respondents have only one degree (n=1049). If you are interested in a detailed analyses of the degrees acquired abroad and their recognition in Germany, you should use the current data set \$P with all variables and not the simplified version in BIOSOC.

- Besides the simplification of educational information for the sample M, we harmonized also different versions of questions. Especially the whole questions about occupational training and university acquired abroad and their recognition in Germany have different answer categories in the questionnaire for samples A – L, M1/M2 and M3/M4.
- Concretely, the variable BSBAAZEU was harmonized. The core samples were asked if they received a certificate for their degree with the option to answer yes or no. The samples M1 and M2 were only asked if they completed their occupational training / university degree successfully. M3 and M4 had three answer categories (*did not complete the training, ended without qualification and ended with qualification*). To unify all the answers we used the core questionnaire as a starting point. The category “yes” of the variable BSBAAZEU are the category “completed successfully” for M1/M2 and the category “ended with qualification” for M3/M4. For the answer “no” of BSBAAZEU we had to recode the categories “ended without qualification” and “did not complete the training” for M3/M4 into one category.
- The reason for no recognition of abroad attained certificate was also harmonized. Most of the categories are the same or pretty similar. We unified the category “not important for legal reasons” and “not important, no further benefits” to the category “not important”. To do so, we recoded the category 8 and 9 for M1/M2 to category 1.
- The variables BSBAABA, BSBAARA, BSBAAMB, BSBAAMJ and BSBAAGA were repeated in the Individual Questionnaire after they were asked the first time in the Biography Questionnaire. BIOSOC contains only the information of the Biography Questionnaire. If you are interested in the current status of the recognition, please use the current data set \$P.

---

<sup>3</sup> The order of the occupational trainings and university degrees is the following: Other vocational training < short-term training in a company < apprenticeship in a company < vocational or professional school < university degree.

**Table 1: Description of the data set BIOSOC**

Variable Name	Content of the Variable	Number of Question in Biography Questionnaire...						Comparable Question in Migration Questionnaire...				Comparable Question in Refugee Questionnaire ...	Comparable Variable in BIOAGE17
		2000	2001 - 2010	2011 - 2012	2013	2014	2015 - 2016	2013	2014	2015	2016	2016	
HHNR	<b>Entries for Surveyed Person</b> Original household identifier (invariant)												HHNR
HHNRAKT		Actual household identifier											HHNRAKT
PERSNR		Personal identifier											PERSNR
SYEAR		Survey Year											
PSAMPLE		Subsurvey											
SEX		Sex											
BEFRPER		Respondent identifier											BEFRPER
ERHEBJ		Survey year											ERHEBJ
BSGEBJAH	Year of birth											BYGEBJAH	
	<b>School</b>												
BSELKUEM	Parents took care about efforts at school	-	29	30	32	36	38	-	52	128	132		BYELKUEM
BSNTDEUT	Last grade in German	-	30	31	33	37	39	-	53	142	-		BYNTDEUT
BSNTMATH	Last grade in maths	-	30	31	33	37	39	-	53	142	143		BYNTMATH
BSNTFMD1	Last grade in 1. foreign language	-	30	31	33	37	39	-	53	142	143		BYNTFMD1
BSPTDEUT	Total points <sup>1</sup> in German (last class)	-	30 <sup>2</sup>										BYPTDEUT

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

<sup>2</sup> Only in survey year 2001 (wave R)



Variable Name	Content of the Variable	Number of Question in Biography Questionnaire...						Comparable Question in Migration Questionnaire...				Comparable Question in Refugee Questionnaire ...	Comparable Variable in BIOAGE17	
		2000	2001 - 2010	2011 - 2012	2013	2014	2015 - 2016	2013	2014	2015	2016	2016		
BSPTMATH	Total points in maths (last class)	-	30 <sup>2</sup>											BYPTMATH
BSPTFMD1	Total points in 1. foreign language (last class)	-	30 <sup>2</sup>											BYPTFMD1
BSGSDEUT	Level of German at comprehensive school <sup>3</sup> (last class)	-	30 <sup>2</sup>											BYGSDEUT
BSGSMATH	Level of maths at comprehensive school (last class)	-	30 <sup>2</sup>											BYGSMATH
BSGSFMD1	Level of 1. foreign language at comprehensive school (last class)	-	30 <sup>2</sup>											BYGSFMD1
BSLKDEUT	Complementary / main subject <sup>4</sup> in German (last class)	-	30 <sup>2</sup>											BYLKDEUT
BSLKMATH	Complementary / main subject in maths (last class)	-	30 <sup>2</sup>											BYLKMATH
BSLKFMD1	Complementary / main subject in 1. Foreign language (last class)	-	30 <sup>2</sup>											BYLKFMD1
	<b>Relationships to Parents, Sport and</b>													

<sup>3</sup> The subjects German, math and the first foreign language are split up into different levels during the secondary school level I of comprehensive schools. Level A is the highest one.

<sup>4</sup> From the 11<sup>th</sup> or 12<sup>th</sup> grade on students can choose their main subjects. At this stage, they can reduce German, maths and foreign languages from major to minor subjects.

Variable Name	Content of the Variable	Number of Question in Biography Questionnaire...						Comparable Question in Migration Questionnaire...				Comparable Question in Refugee Questionnaire ...	Comparable Variable in BIOAGE17
		2000	2001 - 2010	2011 - 2012	2013	2014	2015 - 2016	2013	2014	2015	2016	2016	
	<b>Activities during Youth</b>												
	<i>Frequency of fights when respondent was 15. years old with:</i>												
BSSTRVA	Father	-	31	32	34	38	40	-	54	129	133	-	BYSTRVA
BSSTRMU	Mother	-	31	32	34	38	40	-	54	129	133	-	BYSTRMU
BSSPRTRR	Participated in sports during youth	-	32/33	34	36	40	42	57	56	131	135	-	BYSPRTRR
BSSPRTAR	Favourite sport during youth	-	33/34	35	37	41	43	-	57	132	136	-	BYSPRTAR
BSSPRTWE	Participated in competitions during youth	-	34/35	36	38	42	44	-	58	133	137	-	BYSPRTWE
BSMUSSP	Played music or sang during youth	-	35/32	33	35	39	41	56	55	130	134	-	BYMUSSP
	<b>School Attendance</b>												
BSSCHBES	Still at school	(34)	37	38	40	44	46	69	60	135	139	221	BYSCHBES
BSSCHEND	Year left school	-	37	38	40	44	46	69	60	135	139	221	BYSCHEND
BSSCHWO	Country of last school attendance	(34)	38	39	41	45	47	70	61	136	140	222	-
BSSCHLA	Federal State of last school attendance	-	41	42	44	48	50	73	64	140	144	-	-
BSKLAUSL	Number of foreign classmates	(37)	43	44	46	50	52	75	66	143	147	-	BYKLAUSL
BSSCHZUK	Strive for further school certificate	35	44	45	47	51	53	76	67	144	148	229	BYSCHZUK
BSSCHZAR	Type of further school certificate	36	45	46	48	52	54	77	68	145	149	230	BYSCHZAR

Variable Name	Content of the Variable	Number of Question in Biography Questionnaire...						Comparable Question in Migration Questionnaire...				Comparable Question in Refugee Questionnaire ...	Comparable Variable in BIOAGE17
		2000	2001 - 2010	2011 - 2012	2013	2014	2015 - 2016	2013	2014	2015	2016	2016	
BSBADABG	<b>Attained and Planed Vocational Qualification</b> Vocational / university degree acquired in Germany <i>Type of vocational / university degree attained in Germany:</i>	38	46	47	49	53	55	80	69	146	150	-	(BYBAABGE)
BSBADLEH	Apprenticeship ("Lehre")	39	47	48	50	53	56	81	70	147	151	-	(BYBABFS)
BSBADBFS	Full-time vocational school / School for public health ("Berufsfachschule / Schule des Gesundheitswesens")	39	47	48	50	54	56	81	70	147	151	-	(BYBABFS)
BSBADFSC	Technical school, school for master of a trade ("Fachschule, Meister-, Technikerschule")	39	47	48	50	54	56	81	70	147	151	-	-
BSBADBEA	Training for civil servants (officer) ("Beamtenausbildung")	39	47	48	50	54	56	81	70	147	151	-	-
BSBADFHA	Advanced technical college ("Fachhochschule") or approved vocational academy ("anerkannte Berufsakademie")	39	47	48	50	54	56	81	70	147	151	-	-
BSBADUNI	University degree	39	47	48	50	54	56	81	70	147	151	-	-

Variable Name	Content of the Variable	Number of Question in Biography Questionnaire...						Comparable Question in Migration Questionnaire...				Comparable Question in Refugee Questionnaire ...	Comparable Variable in BIOAGE17
		2000	2001 - 2010	2011 - 2012	2013	2014	2015 - 2016	2013	2014	2015	2016	2016	
BSBADPRO	Doctorate ("Promotion")	-		48 <sup>5</sup>	50	54	56	81	70	147	151	-	-
BSBADSON	Other vocational qualification	39	47	48	50	54	56	81	70	147	151	-	(BYBABGJ, BYBABEGL)
BSBADEND	Year of attaining vocational / university degree in Germany	-	48	48	50	54	56	81	70	147	151	-	-
BSBAAABG	Vocational / university degree acquired abroad <i>Type of vocational / university degree attained abroad:</i>	40	49	49	51	55	57	82	71	148	152	231	-
BSBAAFAN	Short-term training in a company	41	50		52	56	58	(83)	71	149	152	232	-
BSBAAFBA	Apprenticeship in a company	41	50		52	56	58	(83)	71	149	152	232	-
BSBAASCH	Vocational or professional school	41	50		52	56	58	(83)	71	149	152	232	-
BSBAAUNI	University degree	41	50		52	56	58	(83)	71	149	152	232	-
BSBAASON	Other vocational qualification	41	50		52	56	58	(83)	71	149	152	232	-
BSBAAEND	Year of attaining vocational / university degree abroad	-	51		53	57	59	(83)		149	154	-	-
BSBAAZEU	Certificate for abroad attained qualification	42	52		54	58	60	(84)	73	150	154	233	-
BSBAAABA	Applied for recognition					59	61		74	151	155	236 <sup>6</sup>	

<sup>5</sup> This item was integrated in the Biography Questionnaire in 2011.

<sup>6</sup> The variables bsbaaaba, bsbaaara bsbaaamb and bsbaaajb were asked for every single attained qualification. The question number of these variables refers to the attained qualification "Short-term training in a company".

Variable Name	Content of the Variable	Number of Question in Biography Questionnaire...						Comparable Question in Migration Questionnaire...				Comparable Question in Refugee Questionnaire ...	Comparable Variable in BIOAGE17	
		2000	2001 - 2010	2011 - 2012	2013	2014	2015 - 2016	2013	2014	2015	2016	2016		
BSBAAARA	Result recognition					59	61		78	161	165 <sup>7</sup>	239	-	
BSBAAAGA	Reason no recognition					59	61		83	163	167 <sup>8</sup>	-		
BSBAAAMB	Month of assessment					59	61		76	153	160	240		
BSBAAAJB	Year of assessment					59	61		76	153	160	240		
BSBAAZEA	Recognition of aboard attained certificate	42	52		54	-	-	(85 to 89)	-	(161)	-			
BSZAJA	Vocational / university degree is aspired <i>Type of aspired vocational / university degree:</i>	43	53		55	60	62	95	84	164	168	290		BYZAJA
BSZALEH	Apprenticeship (“Lehre”)	44	54		56	60	62	96	85	165	16	291		BYZALEH
BSZABFS	Full-time vocational school/ School for public health (“Berufsfachschule / Schule des Gesundheitswesens”)	44	54		56	60	62	96	85	165	169	291		BYZABFS
BSZAFSC	Technical school, school for master of a trade (“Fachschule, Meister-, Technikerschule”)	44	54		56	60	62	96	85	165	169	291		BYZAFSC
BSZABEA	Training for civil servants (officer) (“Beamtenausbildung”)	44	54		56	60	62	96	85	165	169	291		BYZABEA

<sup>7</sup> We used question 165 as well as question 159 to generate bsbaaara.

<sup>8</sup> The third category for M1/M2 is “I don’t know how to finance the recognition”. The third category for the core samples is “costs for recognition are too high”. It is questionable that these categories aim for the same construct. If a person thinks that the costs are too high does not automatically mean that a person does not know how to finance it. As nobody in M1/M2 used the category three, you can use these category without constraints.

Variable Name	Content of the Variable	Number of Question in Biography Questionnaire...						Comparable Question in Migration Questionnaire...				Comparable Question in Refugee Questionnaire ...	Comparable Variable in BIOAGE17
		2000	2001 - 2010	2011 - 2012	2013	2014	2015 - 2016	2013	2014	2015	2016	2016	
BSZABAK	Approved vocational academy ("anerkannte Berufsakademie")	44	54		56	60	62	96	85	165	169	291	BYZABAK
BSZAFH	Advanced technical college ("Fachhochschule")	44	54		56	60	62	96	85	165	169	291	BYZAFH
BSZAUNI	University degree	44	54		56	60	62	96	85	165	169	291	BYZAUNI
	<b>Military and Voluntary Service</b>												
BSDIGEL	Military or alternative service done (only men)	58	71/74	74	79	78	80	-	-	-	-	-	-
BSDIART	Type of service (only men)	58	71/74	74	79	78	80	-	-	-	-	-	-
BSDIGRU	Reason for not serving (only men)	58	71/74	74	79	78	80	-	-	-	-	-	-
BSFSJ	Voluntary social service ("Freiwilliges Soziales Jahr")	-	72/73	73	78	77	79	-	-	-	-	-	-
BSBFD	Federal voluntary service ("Bundesfreiwilligendienst")	-		3	78	77	79	-	-	-	-	-	-
BSFWD	Voluntary military service ("Freiwilliger Wehrdienst")	-		73	78	77	79	-	-	-	-	-	-
	<b>Specification of Interview Situation</b>												
BSINTA	Type of interview												BYINTA
BSDAUER1	Duration of personal interview												BYDAUER1

Variable Name	Content of the Variable	Number of Question in Biography Questionnaire...						Comparable Question in Migration Questionnaire...				Comparable Question in Refugee Questionnaire ...	Comparable Variable in BIOAGE17
		2000	2001 - 2010	2011 - 2012	2013	2014	2015 - 2016	2013	2014	2015	2016	2016	
BSDAUER2 BSDAUER3 BSTAGIN BSMONIN INTID	Duration of interview filled out independently Interview duration with the integrated questionnaire version Day of the interview Month of the interview Identifier of the interviewer												BYDAUER2  -  BYTAGIN BYMONIN INTID