SOEP The Socio-Economic Panel

769

SOEP Survey Papers

Series D – Variable Descriptions and Coding

SOEP – The German Socio-Economic Panel at DIW Berlin 2019

SOEP-Core v34 – BIOBIRTH: A Data Set on the Birth Biography of Male and Female Respondents

Christian Schmitt and SOEP Group

DIW SOEP

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

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

The SOEP Survey Papers is comprised of the following series:
Series A – Survey Instruments (Erhebungsinstrumente)
Series B – Survey Reports (Methodenberichte)
Series C – Data Documentation (Datendokumentationen)
Series D – Variable Descriptions and Coding
Series E – SOEPmonitors
Series F – SOEP Newsletters
Series G – General Issues and Teaching Materials

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

Please cite this paper as follows:

Christian Schmitt and SOEP Group. 2019. SOEP-Core v34 – BIOBIRTH: A Data Set on the Birth Biography of Male and Female Respondents. SOEP Survey Papers 769: Series D. Berlin: DIW/SOEP

ISSN: 2193-5580 (online)

DIW Berlin
German Socio-Economic Panel (SOEP)
Mohrenstr. 58
10117 Berlin
Germany

soeppapers@diw.de

SOEP-Core v34 – BIOBIRTH: A Data Set on the Birth Biography of Male and Female Respondents

Christian Schmitt and SOEP Group

BIOBIRTH: A Data Set on the Birth Biography of Male and Female Respondents

by Christian Schmitt

1 Population and purpose of the data set BIOBIRTH

The file BIOBIRTH provides information on fertility histories of adult respondents in the SOEP. Until 2014 (version 30, wave BD) the data was stored in two separate files: BIOBIRTH containing female fertility histories, and BIOBRTHM providing male fertility histories. It is important to note that the latter file only records the male fertility histories for respondents who entered the SOEP in 2001 or later (for more details see below). Since 2015 (version 31, wave BE) all fertility histories of new respondents as well as the old and continuously updated data of the fertility histories collected in BIOBIRTH (until 2014) and BIOBRTHM (until 2014) are stored in a single file - BIOBIRTH (this naming is identical to the previous fertility histories of women). The variable SEX (distinguishing male and female respondents) is added to the BIOBIRTH dataset since wave BE. Moreover, the file BIOBIRTH is also supplemented with the fertility histories of the "Familien in Deutschland" (FiD) panel survey, which was integrated into the SOEP data-base in 2015. Records from the FiD subsamples can be distinguished by inspecting the variable BIOVALID. For more details on integration and extension procedures see below.

Fertility histories in BIOBIRTH provide information on every woman (as well as every man with a panel entry since 2001) who has ever provided at least one successful SOEP interview. Note that the data is right censored for respondents who left the panel early in their fertile life-phase and it is left censored for persons who entered the SOEP at higher ages without ever filling in a biographical questionnaire. The variable BIOVALID provides information on whether individual level information is based merely on information derived from household composition and family relations, or on biographical questionnaire data. The variables EINTRITT and AUSTRITT in ppfad (panel entry and exit) provide information on censoring (left-censoring can be ignored if a biographical questionnaire exists for a given person).

For each of the mentioned adult respondents BIOBIRTH documents the fertility history. The annual update focuses on including new information on becoming a biological parent as based on data collected with the individual or the biographical questionnaire, respectively. Furthermore adults who have been interviewed for the first time but who have not yet provided information on their fertility histories are included. The latter case applies to either new adult household members, or teenagers who have reached the required minimum for a participation

in the personal questionnaire (16 years). BIOBIRTH constitutes an accumulative data set, in which the entire birth biography of all SOEP respondents is presented.

2 Structure of the data set

BIOBIRTH covers the following information:

(1) Person identifier (PERSNR), respondent's year of birth, status information on the origin of the included data, number of children derived from fertility histories, total number of children derived from fertility histories and subsequent consideration of changes in household structure up to the last date of interview.

(2) A sequence of 15 variables relating to 1 out of 15 children, including child's person identifier (KIDPNR[nn], provided the child could be identified within the SOEP's household structure), child's sex, child's year of birth, child's month of birth.

BIOBIRTH contains the following variables for all adult women (and men since 2001):

- HHNR Invariable number of the original household
- PERSNR Invariable personal identifier of the respondent
- SEX Respondent's sex
- GEBJAHR Respondent's year of birth
- BIOVALID Status of the birth biography: (Please mind: The codes in the variable BIOVALID have been extended in 2015 (wave BF)).

  10: no birth biographical entries

  20: youth biography questionnaire, no fertility histories on children

  30: birth biography questionnaire, no children in fertility history

  31: birth biography questionnaire, one or more children in fertility history

  40: FiD: no birth biography, no data on children

  41: FiD: no birth biography, information on existing children in FiD parent/couple questionnaire

  50: FiD: birth biography questionnaire, no children in fertility history

  51: FiD: birth biography questionnaire, one or more children in fertility history.

- BIOYEAR — Survey year when birth biography questionnaire was completed (1985ff.). Code "-2" is assigned if the respondent never completed a birth biography questionnaire.
- BIOAGE — Age of the woman at the time of the birth biography survey. Code "-2" is assigned if the respondent never completed a birth biography questionnaire.
- 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+subsequent births during panel participation).
- BIOKIDS — Total number of children identified in the birth biography. Code "-2" is assigned if the respondent never completed a birth biography questionnaire.
- KIDGEB[nn] — Year of birth of the child [nn] (for the first child up to the fifteenth child).
- KIDSEX[nn] — Sex of the child [nn] (for the first child up to the 15th child).
- KIDPNR[nn] — Personal number of the child [nn] (for the first child up to the 15th child), given it is identifiable in the SOEP.
- KIDMON[nn] — Month of birth of a child [nn] (for the first child up to the 15th child). With respect to the variables KIDGEB[nn], KIDSEX[nn], KIDPNR[nn], and KIDMON[nn] identical missing codes apply: The code "-2" is assigned if there's no [nn]th child identified for a mother/father. The code "-1" applies if an [nn]th child can be identified but the information on the birth year and/or sex and/or personal identifier is unavailable, or if it could not be identified.

For every respondent a maximum of up to 15 children are considered. The sequence of children within BIOBIRTH is recorded with regards to the birth order in terms of age of the children. The order ranks from the oldest child specified under KIDPNR01 to the youngest child. If the age is missing it is listed in the first record (KIDPNR01), and in subsequent records following KIDPNR01 if more than one child's personal identifier remains missing.

3 Information basis of the birth biography

The main basis of the individual fertility history considered in BIOBIRTH is the information collected with the biography questionnaire[1], in which the number, birth year and sex of the biological children for every adult respondent are collected. For adults with information on children stemming from the biography questionnaire the BIOVALID code "31" is assigned. Women who completed this questionnaire, but did not report any biological children receive the code "30". In correspondence to this, the samples of the FiD panel integrated into the SOEP since 2015 (wave BE) contribute the codes 50 and 51 for respondents with biography questionnaires with, or without children, respectively. Codes 10 and 40 correspond accordingly for SOEP and FiD data. 41 is a qualitatively new code for parents in the FiD who did not fill in the birth biography but who have provided reliable information on their status as a parent in other sources of FiD (this relates to specific questions in the FiD parent and couple questionnaire).

A minority of respondents did not provide any information on fertility histories by filling in the biographical questionnaire for several reasons[2]. For these cases the variable BIOVALID is assigned the code 10 (original SOEP samples), and 40 (new FiD subsamples), respectively. This group is subject to a risk of underestimating the total number of births, particularly since births prior to the entry into the SOEP cannot be identified unless current household structure provides substantive evidence, usually reflected by parent-child co-residence. Respondents without a valid fertility history (codes 10 and 40) can be distinguished in three major groups:

- Respondents who were 16 years of age at the time of the first interview. In most cases these respondents participate in the biography survey at a later date. Thus, the parent-child relationship recorded earlier in BIOBIRTH (as based on household structure) can be verified and supplemented with data on their fertility histories at a later date.
- Respondents who were at about 30 years of age or younger at the time of first interview. In this sub-population, children are not yet adults and still reside in the parental home in most cases. Since information from the biographical questionnaire is missing, a final distinction in social and biological children is difficult particular for records considered in earlier SOEP waves when the intra-household relationships ($STELL) were less refined, compared to the recent setup. Hence these older records have a higher likelihood of misspecifying a social as a biological parent-child relationship.
- Respondents who were well over 30 years of age at the time of the first interview. In these cases some of the children are likely to already have left the parental home, and therefore are

[1] The information collected over the course of the biography survey for every adult contains the number of children, the year of birth, the sex, residence status of the child, and, if applicable the year of death of the biological child. The biography data is stored in wave specific files ($LELA), which are not provided with the SOEP distribution.

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

no longer part of the survey population. For that reason, the number of biological children might be underestimated in this group of respondents to a larger extent as compared to younger women.

4 A new source of biographical information – the youth questionnaire

From wave T onwards the data within BIOBIRTH includes information of a further biographical instrument: the youth biography. The youth-questionnaire has been in circulation 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 dataset 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, H, I, J, K, M, and subsequent samples				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 these few can be identified in the household context (as long as they remain within the SOEP), this does not pose any problem for compiling the birth-biography of the respondents. Nevertheless, a few changes to the BIOBIRTH dataset have to be outlined:

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

5 The fertility histories of male respondents in BIOBIRTH

Since 2001 the fertility histories of male respondents are collected with the SOEP survey and have been integrated in a special data set BIOBRTHM since 2003 (wave T). Since 2015 (wave BE) this data is merged with the female fertility history and collected in the file BIOBIRTH, while the variable SEX distinguishes between the fertility histories of men and women. Contents and updating procedures for male fertility histories are identical to those of their female pendants with two important exceptions:

- First: only information about men with at least one completed questionnaire in 2001 or later is considered in the BIOBIRTH file.
- Second: information from the birth-biography will only be added for new panel members who joined since 2000, 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 on male fertility histories was introduced in 2001. Therefore, most men in subsample "F" (which started in 2000) have completed the birth-biography, since biographical questionnaires are usually completed one wave after the starting wave with only a marginal rate of non-response for this life-course related questionnaire. Note that for men within BIOBIRTH who did not complete a biography questionnaire (BIOVALID code 10), the information about fatherhood is underestimated as only the context of the household is available to determine the respondents biological children. Underestimation of this type is generally more severe for men than for women, since children have a distinctively higher likelihood to co-reside with their mothers after separation, in which case a father will appear to be childless, thus leading to a misspecification as being childless.

6 Integration of "Familien in Deutschland" – FiD

In 2015 (wave BE) data on respondents from the "Familien in Deutschland" (FiD) panel survey was integrated into the SOEP data-base. While the design of this survey was oriented on a close congruence to the SOEP questionnaires and data structure, FiD introduces some additional information. The focus of the FiD survey was put on couples, families and lone parents. As a consequence, there is even more information on parent-child relationships compared to the original SOEP samples. Hence the derivation of parent-child relationships is quite reliable for the FiD subsamples (see variable PSAMPLE in the file PPFAD for a distinction). Almost all adult respondents completed a specific biographical questionnaire including data on their fertility histories. These respondents were assigned BIOVALID codes 50 (without children), and 51 (with children). Additionally, particular questions in a couple, as well as a parent-child-

questionnaire provide data on biological children. Hence, the original BIOVALID code 10, which corresponds to the code 40 for the FiD population, shows only few case numbers, while a new code 41 was introduced for parents, who lack detailed information on their fertility histories, but which provide data on children in FiD's couple or parent questionnaire. For the update of the BIOBIRTH population and data in subsequent waves, the FiD population will be treated identical to any other of the SOEP subsamples (see 1.10 in particular).

7 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). Until wave 29, the variable $STELL had 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 parent-child relationship between a female adult and another person can be assumed.

Table 1: Potential parent-child relationships as a combination of the variable $STELL

$STELL of the		Potential parent-child relationship
woman	another person	In this case the person is the…
0	3	Child of reference person (reference person = head of the 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 6	0 8	Child is reference person, lives with his mother/father in the same household Child is the sister / brother of reference person, the siblings live with their mother/father in the same household
7 7	1 8	Child is spouse of reference person and lives together with spouse and mother/father in the same household Child is daughter / son of the mother/father 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, parent 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/father and child have another relation to reference person
11	11	Child and mother/father are in no way related to reference person

With SOEP wave 29 a more complex representation of the variable $STELL in the file $PBRUTTO has been implemented, describing the following types of relationships to the head of household:

Code	Label
0	Head Of Household
11	Spouse Of HH Head
12	Same-Sex Spouse
13	Life Partner
21	Son, Daughter
22	Stepchild (Child of the Partner)
23	Adoptive Child
24	Foster Child
25	Grandchild
26	Great-Grandchild
27	Son, Daughter-In-Law
31	Father, Mother
35	Parent-In-Law
36	Grandparents
41	Brother, Sister
42	Half-Brother, Half-sister
43	Stepbrother, Stepsister
51	Brother, Sister –in Law (Spouse/ Life Partner of Siblings)
52	Brother, Sister -in Law (Siblings of Spouse/ Life Partner)
61	Aunt, Uncle
62	Niece/ Nephew
63	Cousin/Cousine
64	Other Relative
71	Others
99	Relationship to Head of HH Unknown

Based on these new representations of relationship patterns considered since wave BC, the algorithm, scanning for potential parent-child dyads has been adjusted and implemented in generating the birth-biographical information that relies solely on the household composition.

It should be noted that the majority of parent-child relations provided with the BIOBIRTH biography file are still derived from personal information given in the biographical core-questionnaires (see below).

8 Identification of the children of parents with completed fertility histories

If a parent mentions either the existence of biological children, the year of birth of a child, the sex of a child or co-residence with this initiates a process of identification. In the first step, the algorithm scans the current household structure and aims to determine the parent's relationship to the reference person and scans potential combinations of parent child relationships within the current household structure. The principle is the same as outlined above (1.6). The goal is to supplement the biographical information on sex and year of birth of a child with a personal identifier, thus enabling more complex analyses of parent-child interaction for the researcher. If a fit is determined between the year of birth and the sex of a potential child in birth biography and household, the individual is considered as the child of the respondent and the child's personal identifier is assigned. Since the majority of the households with children present small nuclear families including one potential mother/father, this kind of identification process is broadly sufficient. In other, rather complex households a careful hand editing is conducted, in order to identify the 'right' child to the 'right' mother/father. The same is done, if the sex or the year of birth of a child mentioned in the biography questionnaire is unspecified, i.e. missing.

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

9 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 parents only, specific relationships among household members are considered. Since the key information from the biography questionnaire is not available, a careful analysis of the composition and the history of the household in which the children live is conducted in order to assign proper parent-child linkages.

10 Updating BIOBIRTH

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 children the parent-child ties are investigated by the same algorithm which has been described in section 1.6. The general logic of updating the BIOBIRTH birth biography follows this pattern:

1) New information from household composition (basis $STELL) updates older information contained in BIOBIRTH (i.e. new-born children, or children moving into the household or otherwise observed in the SOEP for the first time).

2) New or first time information from the biographical questionnaire is displayed as core part of the parent child-ties in BIOBIRTH, or replaces older household-based information contained in BIOBIRTH, in case household structure-based data is inconsistent with the data from the biographical questionnaire.

3) New information, derived from household composition (basis $STELL), may not replace earlier information in BIOBIRTH, in case the previous information on children stems from the biographical questionnaire, and the two sources cover the same period and are inconsistent. In other words: information from the biographical questionnaire always overrides household information.

Overview: Central variables in the file BIOBIRTH (Version 2017 / Up to Wave BH)

Biovalid

		Frequency	Percent
Valid	10 No Birth Biography - No Kids From Bio	11.254	12,76
	20 Youth Biography - No Kids From Bio	7.204	8,11
	30 Birth Biography - No Kids From Bio	23.516	26,47
	31 Birth Biography - Kids From Bio	35.872	40,37
	40 FiD: No Birth Biography	1.063	1,20
	41 FiD: Parent/Partner Questionnaire – Data on Kids	855	0,96
	50 FiD: Birth Biography - No Kids From Bio	642	0,72
	51 FiD: Birth Biography - Kids From Bio	8.450	9,51
	Total	86.194	100

Source: SOEP v34.

BIOYEAR Year of Biography Survey

Year of Biography Survey	Frequency	Percent
-2 (trifft nicht zu)	13.172	14,82
1985	6.618	7,45
1986	83	0,09
1987	104	0,12
1988	220	0,25
1989	211	0,24
1990	202	0,23
1991	161	0,18
1992	2.635	2,97
1993	230	0,26
1994	592	0,67
1995	528	0,59
1996	255	0,29

1997	231	0,26
1998	203	0,23
1999	1.033	1,16
2000	361	0,41
2001	9.429	10,61
2002	899	1,01
2003	2.691	3,03
2004	822	0,93
2005	667	0,75
2006	530	0,60
2007	2.576	2,90
2008	594	0,67
2009	438	0,49
2010	9.597	10,80
2011	6.908	7,77
2012	3.227	3,63
2013	651	0,73
2014	4.606	5,18
2015	2.630	2,96
2016	2.653	2,99
2017	10.422	11,73
2018	2.677	3,01
Total	88.856	100,00

Source: SOEP v33.

SUMKIDS Total Number of Births

Total Number of Children Born	Frequency	Percent
0	32.823	36,94
1	15.908	17,90
2	22.902	25,77
3	10.704	12,05
4	3.870	4,36
5	1.476	1,66
6	559	0,63
7	304	0,34
8	140	0,16
9	75	0,08
10	49	0,06
11	25	0,03
12	13	0,01
13	1	0,00
14	1	0,00
15	1	0,00
16	3	0,00
17	1	0,00
18	0	0,00
19	1	0,00
Total	86.194	100,00

Source: SOEP v33.

BIOKIDS Number of Births from Biography

Births from Birth Biography Questionnaire	Frequency	Percent
-2	13.172	14,82
0	29.594	33,31
1	13.281	14,95
2	18.476	20,79
3	8.826	9,63
4	3.222	3,63
5	1.243	1,40
6	482	0,54
7	274	0,31
8	130	0,15
9	79	0,09
10	39	0,04
11	20	0,02
12	10	0,01
13	2	0,00
14	0	0,00
15	1	0,00
16	3	0,00
17	1	0,00
18	0	0,00
19	1	0,00
Total	86.194	100,00

Source: SOEP v34.