

Heidemann, Valeriia; Richter, David

**Research Report**

## SOEP-Core v36 - BIOAGEL & BIOPUPIL: Generated variables from the "mother & child", "parent", "pre-teen", and "early youth" questionnaires

SOEP Survey Papers, No. 1079

**Provided in Cooperation with:**

German Institute for Economic Research (DIW Berlin)

*Suggested Citation:* Heidemann, Valeriia; Richter, David (2021) : SOEP-Core v36 - BIOAGEL & BIOPUPIL: Generated variables from the "mother & child", "parent", "pre-teen", and "early youth" questionnaires, SOEP Survey Papers, No. 1079, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/248494>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by-sa/4.0/>

# 1079<sup>2021</sup>

**SOEP** Survey Papers  
Series D - Variable Descriptions and Coding

## SOEP-Core v36 – BIOAGEL & BIOPUPIL: Generated Variables from the "Mother & Child", "Parent", "Pre-Teen", and "Early Youth" Questionnaires

Valeriia Heidemann, David Richter, and SOEP Group

Running since 1984, the German Socio-Economic Panel (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 Freie Universität Berlin

Prof. Dr. David Richter, DIW Berlin and Freie Universität Berlin

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

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

Dr. Sabine Zinn, DIW Berlin and Humboldt Universität zu Berlin

Please cite this paper as follows:

Valerii Heidemann, David Richter, and SOEP Group. 2021. SOEP-Core v36 – BIOAGEL & BIOPUPIL: Generated Variables from the “Mother & Child”, “Parent”, “Pre-Teen”, and “Early Youth” Questionnaires. SOEP Survey Papers 1079: Series D. Berlin: DIW/SOEP



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

© 2021 by SOEP

ISSN: 2193-5580 (online)

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

[soeppapers@diw.de](mailto:soeppapers@diw.de)

# **SOEP-Core v36 – BIOAGEL & BIOPUPIL: Generated Variables from the “Mother & Child”, “Parent”, “Pre-Teen”, and “Early Youth” Questionnaires**

**Valeriia Heidemann, David Richter, and SOEP Group**

# BIOAGEL & BIOPUPIL: Generated variables from the “Mother & Child”, “Parent”, “Pre-teen”, and “Early Youth” questionnaires

by Valeriia Heidemann & David Richter<sup>1</sup>

## 1 Introduction

The bioage data files are generated using information collected in the “Mother & Child” and “Parent” questionnaires. In 2014, for the first time, the children themselves answered the “Pre-teen” questionnaire. The “Early Youth” questionnaire was introduced in 2016. In addition, the data from the Families in Germany (FiD) study was integrated into the SOEP in 2014 as well. There are six different “Mother & Child” and “Parent” questionnaires and two questionnaires that are answered by the children themselves: the “Pre-teen” and the “Early Youth” questionnaires (see Table 1). Two additional questionnaires (biage02 and bioage10b) were only surveyed in FiD but are integrated into the *bioage* data set, too.

**Table 1: Overview of bioage data, corresponding age group, and respondents**

Bioage	Classification by Age	Respondents	First Wave	Data Set
<i>Bioage01</i>	0-1 years old	Mothers only	2003	Bioage
<i>Bioage02 (FiD only)</i>	1-2years old	Mothers only	2010	Bioage
<i>Bioage03</i>	2-3 years old	Mothers only	2005	Bioage
<i>Bioage06</i>	5-6 years old	Mothers only	2008	Bioage
<i>Bioage08a</i>	7-8 years old	Parent 1	2010	Bioage
<i>Bioage08b</i>	7-8 years old	Parent 2	2010	Bioage
<i>Bioage10</i>	9-10 years old	Parent 1 (SOEP & FiD)	2012 / 2010 (FiD)	Bioage
<i>Bioage10b (FiD only)</i>	9-10 years old	Parent 2 (FiD only)	2010	Bioage
<i>Bioage12</i>	11-12 years old	Children	2014	Biopupil
<i>Bioage14</i>	13-14 years old	Children	2016	Biopupil

The “Mother & Child”, “Parent“, “Pre-teen”, and “Early Youth” questionnaires aim to follow and observe future generations of the SOEP and collect all information in age-specific files, even though the data come from different survey years. As we try to make this process as

---

<sup>1</sup> This documentation is based on earlier versions of documentation materials on bioage data sets and has benefited from previous work by Sebastian Frischholz, Anne Fromm, Stefanie Lenuweit, Katharina Mahne, Christian Schmitt, and Jürgen Schupp.

comprehensive and gapfree as possible, we begin following and documenting the development of children in SOEP households from birth onwards. Since many questions overlap, all bioage data files are covered in this chapter, rather than dedicating a separate chapter of this document to each bioage data file. By doing so, we are able to provide you with an overview of all variables covered in the bioage data files.

Starting with the v34 data distribution, all available data is now provided in two data sets: *Bioage1* and *Biopupil*. The separate bioage data files are no longer part of the data distribution.

## 2 Respondents in the ‘Bioage1’ and ‘Biopupil’ Data Sets

The values of the BIOAGE variable in the *bioage1* and *biopupil* data files reflects the age of the children when the respective questionnaire was taken by their parents (bioage01 – bioage 10) or answered by themselves (bioage12 and bioage14). For example, *bioage* = 6 covers children who turned six that survey year, producing a range in ages between 5 years and 1 month and 6 years and 11 months, depending on the birth month and interview month. For information on the questionnaires used with each age group, please see Table 1. Except for bioage08b, bioage12, and bioage14 (and – FiD only – bioage10b), it is usually the mother who completes the questionnaire. In the exceptional case that a mother cannot complete the questionnaire, the father does. The respondents’ personal ID numbers (regardless if fathers, mothers or other care takers) are provided in the variable PERSNRESP in the *bioage1* data set. Note that while the parents are the actual respondents (except for bioage12 & bioage14, covering the “Pre-teen” and “Early Youth” questionnaire), the data are organized under the child’s unchanging personal ID number (PERSNR).

### **Bioage01: “Mother and Child” Questionnaire, Children Aged 0-1 years**

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

### **Bioage02: “Your child between the ages of one and two”, Children Aged 1-2 years (FiD only)**

The questionnaire is given to all mothers whose child turns two in the current survey year. In the exceptional case that a mother cannot complete the questionnaire, the father responds.

### **Bioage03: “Your child between the ages of two and three”, Children Aged 2-3 years**

The questionnaire is given to all mothers whose child turns three in the current survey year. In the exceptional case that a mother cannot complete the questionnaire, the father responds.

### **Bioage06: “Your child between the ages of five and six”, Children Aged 5-6 years**

The questionnaire is given to all mothers whose child turns six in the current survey year. In the exceptional case that a mother cannot complete the questionnaire, the father responds.

### **Bioage08a and Bioage08b: “Parent” Questionnaire, children aged 7-8 years**

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

### **Bioage10: “Your child between the ages of nine and ten”, Children Aged 9-10 years (only FiD: Bioage10b)**

The questionnaire is given to all mothers (FiD: and fathers) whose child turns ten in the current survey year. In the exceptional case that a mother cannot complete the questionnaire, the father responds. Data of parent 1, which is usually the mother, can be found in *bioage10a*. Data of parent 2 (FiD only), which is usually the father, can be found in *bioage10b*.

### **Bioage12: “Pupils between the ages of eleven and twelve”, Children Aged 11-12 years**

The questionnaire is given to all children themselves who turn(ed) twelve in the survey year.

### **Bioage14: “Pupils between the ages of thirteen and fourteen”, Children Aged 13-14 years**

The questionnaire is given to all children themselves who turn(ed) fourteen in the survey year.

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

The data set has grown over the years and now contains data on 20,879 children. At the moment (v36, 2019), there are 242 children for whom information has been provided through all of the “Mother & Child”, “Parent”, “Pre-teen”, and “Early Youth” questionnaires (*bioage01*, *bioage03*, *bioage06*, *bioage08a/b*, *bioage10*, *bioage12*, and *bioage14*). Data are available for 813 children from seven questionnaires, 1,587 children from six questionnaires, 1,465 children from five questionnaires, 1,407 children from four questionnaires, for 2,032 from three questionnaires, and for 5,127 children from two questionnaires. For 8,206 children, data are available from one questionnaire.

**Table 2: Number of respondents to the “Mother & Child”, “Parent”, “Pre-teen”, and “Early Youth” questionnaires**

Bioage	Survey Year																	Total
	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	
\$01	318	247	246	234	205	185	196	1,503	353	378	349	444	302	349	758	666	600	7,723
\$02 (FiD)								787	647	568	187							2,189
\$03			257	222	237	246	186	1,061	914	745	701	454	386	395	560	645	602	8,136
\$06						237	210	687	696	612	858	825	657	686	643	662	630	7,954
\$08a/b								1,055	1,193	1,215	1,086	1,028	1,251	1,263	1,263	1,352	887	12,460
\$10								403	510	699	686	656	616	539	1,001	982	806	7,722
\$10b (FiD)								242	310	291	301							1,144
\$12												606	608	563	643	628	742	4,436
\$14														531	654	628	581	2,894
Total	318	247	503	456	442	668	592	5,738	4,623	4,508	4,168	4,013	3,904	4,314	5,522	5,563	4,848	54,658



Table 3 gives an overview of the number of children per household with information from questionnaires in the specific age range. Note that there might be more children in the household than indicated here: these children are not in the age range of the “Mother & Child”, “Parent”, “Pre-teen”, and “Early Youth” questionnaires and therefore not covered by the bioage and the biopupil data set.

**Table 3: Number of children in the household with information from questionnaires in the specific age range**

Bioage File	Number of Children in the Household							
	1 Child	2 Children	3 Children	4 Children	5 Children	6 Children	7 Children	8 Children
<i>Bioage01</i>	4,215	1,340	204	40	6	3		1
<i>Bioage02</i> ( <i>FiD only</i> )	1,671	256	20					
<i>Bioage03</i>	4,381	1,406	242	39	6	4	1	
<i>Bioage06</i>	3,933	1,482	276	40	9	4		
<i>Bioage08a/b</i>	3,782	1,486	303	43	7	2		
<i>Bioage10</i>	3,700	1,398	326	54	4	2		
<i>Bioage10b</i> ( <i>FiD only</i> )	783	170	7					
<i>Bioage12</i>	2,617	728	110	7	1			
<i>Bioage14</i>	2,054	375	26	3				

### 3 Topics and Variables

The *bioage* data contain information regarding

- Pregnancy and childbirth
- Child health
- Childcare situation
- Changes in living circumstances since birth of child
- Child’s abilities
- Parenting experiences
- Expectations about the child’s success in school
- Educational goals and aspirations of the parents
- Educational behavior of the parents
- Self-conception of the role of parents
- Self-report: independence and social experience
- Self-report: importance of social contacts
- Self-report of savings
- Self-report: interest in politics

The rules used to generate the variables from the questionnaires are consistent over the various *bioage* questionnaires. In the integrated *bioage* and *biopupil* data set, the data is presented in

“long” format, i.e. the *bioage1* dataset contains information from *bioage01*, *bioage03*, *bioage06*, *bioage08a* and *bioage08b*, as well as *bioage10* and the *biopupil* dataset contains information from *bioage12* as well as *bioage14*. In addition, all variables from the FiDStudy are included as well: if there was a corresponding SOEP Variable, the data was combined into the SOEP variable; if there was no corresponding SOEP Variable or if the data was coded differently, the information is provided in separate variables with the suffix *\_fid*.

A few variables are generated by combining the information from two or more variables. The present documentation provides detailed information on variables generated using information from other files. An overview of the specific variables and the rules by which they were generated is given in section 10.4 “Generated Variables”.

## 4 Generated Variables

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

### AGE

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

### PREGY

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

pregnancy. The time of observation starts in survey year 2003 with the 2002 birth cohort.

## **PREGMO**

Variable label  
Variable format  
*Bioage* File  
Comment

### **“Mother: pregnancy month at interview”**

2-digit integer

01

This variable is based on the exact month of birth (BCPREGMO), the duration of childbearing in weeks (BCSSW) and the interview month of the previous year’s personal interview. Hence this information is available only for those women in the sample for at least two years. As the exact day of birth is unknown, this variable remains a close approximation.

## **PREBEG**

Variable label  
Variable format  
*Bioage* File  
Comment

### **“Spell Begin Pregnancy (Month, 01.83=1)”**

3-digit integer

01

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

## **PREEND**

Variable label  
Variable format  
*Bioage* File  
Comment

### **“Spell End Pregnancy, Birth (Month, 01.83=1)”**

3-digit integer

01

The variable BCPREEND contains information on the end of pregnancy (i.e. the month of birth). Information is given in the regular SOEP spell format: values start with 1 for January 1983 (e.g., earliest spell in *bioage01*, survey year 2010, is 304, which equals April 2008). This variable is based on the exact month of birth (BCPREGMO) and the duration of pregnancy in weeks (BCSSW).

## **SEX**

Variable label  
Variable format  
Comment

### **“Sex of child”**

1-digit integer

The sex of the child is not asked for in the “Mother & Child” and “Parent” questionnaires. Information on this variable stems from the *ppfad*.

**SEXRESP**

Variable label

**“Sex of respondent (parent)”**

Variable format

1-digit integer

Comment

This variable tells whether the mother or father answered the respective questionnaire. The information for this variable comes from the *ppfad* file.