

SOEP Survey Papers

Series B - Survey Reports (Methodenberichte)

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

2019

SOEP-Core v33.1 – BIOTWIN: TWINS in the SOEP

Christian Schmitt 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:

Christian Schmitt and SOEP Group. 2019. SOEP-Core v33.1 – BIOTWIN: TWINS in the SOEP. SOEP Survey Papers 595: Series D. Berlin: DIW/SOEP



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

© 2019 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 –
BIOTWIN: TWINS in the SOEP**

Christian Schmitt and SOEP Group

BIOTWIN: TWINS in the SOEP

by Christian Schmitt

1 Population and contents of the data set BIOTWIN

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

- have exactly the same age as his or her sibling (year & month of birth),
- have a relationship to the head of the household that indicates that he or her and a second persons are siblings, and
- have the same mother (as far as a pointer to the mother is available). Furthermore, it is not only twins that are recorded in the BIOTWIN data set, but also triplets or quadruple siblings.

The following variables are stored within the BIOTWIN data set:

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

The central variable PERSNR is assigned to the sibling with the lowest personal identifier in the twin group. The PNRTWIN and – in rare cases if available – PNRTRIP or PNRQUAD contain the personal identifier of second, and third or fourth sibling in the group. This means that every case in the data set consists of a group of twins (or triplets or quadruplets). The code “-2” is assigned to PNRTRIP and/or PNRQUAD if a third or fourth twin sibling doesn’t exist. PERSNR and PNRTWIN however should always contain valid codes. The variable PNRMOTH provides the link to the mother of the group and is derived from the data sets \$KIND (reference to this \$KIND was discontinued in with wave Z / 2009) and/or BIOBIRTH.

2 The twin survey of 2006

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

The result of the survey could widely validate the selection of the twin population, contained in the BIOTWIN data set of the SOEP. More than 80% of households with potential twins as of 2006 could be contacted and were interviewed in the twin survey. Among these only 3 groups of twins turned out to be identified erroneously (those false positives were removed from the BIOTWIN data set). Thus the algorithms of identifying twins within the SOEP could prove to be widely reliable. Additional information that was collected with the twin survey contributed to identifying a number of mothers of twins, for whom the mother-child-link was missing previously. Furthermore the twin survey provided additional information on monozygotic respectively dizygotic twins. The variable BIOMONOZ was extended, in order to reflect this additional information (see below for more details).

3 Construction of variables in the data set BIOTWIN

The variable BIOMONOZ¹ indicates if the group is monozygotic. If the information could be validated in the twin-survey in 2006 the code is set to 1 for monozygotic twins and 2 for dizygotic twins. If the information on being mono- or dizygotic twins could not be validated in the twin survey, which was carried out in 2006, the code is set to 0 if the sex of all the siblings is identical, and this group thus might be monozygotic. Please pay attention to the fact that the labels and values of the variable BIOMONOZ from wave W onwards are not consistent with values and labels from previous waves.

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

INFOTWIN can take the following characteristics:

- 1 Generated up to 2006 – basis: household co-residence, identical parent, year & month of birth – not validated by in the twin survey 2006
- 2 Possible Twin or Triplet – Information not revisable in twin survey 2006
- 3 Possible Twin or Triplet – Answer refused in twin survey 2006

¹ This variable existed before wave W but was restructured to reflect the additional information which became available with the 2006 twin questionnaire.

- 4 Twin or Triplet – Information validated by twin survey 2006
- 5 Twin or Triplet – New since 2006 (congruent year & month of birth)
- 6 Twin or Triplet – New since 2006 (congruent year of birth / missing month of birth)

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

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

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

An overview of central information in the file BIOTWIN (Version 2017 / Wave BH)

Table 1: Siblings in BIOTWIN²

Sibling type	n	Valid Mother Pointers
Twins	1.793	1.518
Triple	64	57
Quadruple	4	4

Source: SOEP v33.

² Please note: sibling groups contribute observations for each individual of the twin-pair/triplet, that is – a pair of twins would provide two entries to the data set, one for each twin.

Table 2: BIOMONOZ Gender Combination of Siblings

		Frequency	Percent
Valid	(-1) No Answers	2	0,11
	0 Possibly Identical Twins	862	48,08
	1 Definitely Identical Twins	197	10,99
	2 Definitely Fraternal Twins	732	40,83
	Total	1.793	100

Source: SOEP v33.

Table 3: INFOTWIN Twin Status: Source of Information

		Frequency	Percent
Valid	1 Generated - not in twin survey 2006	45	2,62
	2 Twinsurvey 2006 (answer not verified)	73	4,07
	3 Twinsurvey 2006 (answer refused)	2	0,11
	4 Twinsurvey 2006 (answer validated)	196	10,93
	5 Gen. since 2007 (basis: year of birth & month)	845	47,13
	6 Gen. since 2007 (basis: year of birth with missing month)	630	35,14
	Total	1.793	100

Source: SOEP v33.