SOEP The Socio-Economic Panel

873

SOEP Survey Papers

Series D – Variable Descriptions and Coding

SOEP – The German Socio-Economic Panel at DIW Berlin 2020

SOEP-Core v35 – BIOEDU: Data on educational participation and transitions

Paul Schmelzer, Antonia Briel, and SOEP Group

DIW SOEP

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**

Please cite this paper as follows:

Paul Schmelzer, Antonia Briel, and SOEP Group. 2020. SOEP-Core v35 – BIOEDU: Data on educational participation and transitions. SOEP Survey Papers 873: 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 v35 – BIOEDU: Data on educational participation and transitions

Paul Schmelzer, Antonia Briel, and SOEP Group

BIOEDU: Data on educational participation and transitions

Paul Schmelzer and Antonia Briel

The Socio-Economic Panel Study (SOEP) contains a broad range of variables which cover early child education and care, educational participation, educational degrees and other related topics. However, the respective questions are included in different questionnaires (e.g., personal questionnaire, household questionnaire, youth questionnaire) and the variables are not always in a format which is suited for longitudinal analyses. For instance, transitions such as school enrolment or entry into tertiary education are not documented in a single variable but can only be reconstructed by comparing the status in a wave t with the status in a wave t+1 (e.g., a transition into tertiary education took place if a person was not in university in wave t but is in university in wave t+1). Generating such variables is time-consuming and prone to errors. It is the aim of the BIOEDU dataset to provide ready-made variables on educational transitions and related topics in order to support analyses in a longitudinal perspective.

The BIOEDU dataset is primarily based on prospectively collected information. Therefore, it contains most information for those persons who have been part of the survey population at the time when they have attended school or other educational institutions. In total the dataset contains information on 106,471 persons in the wave 35 (2018). This is the part of the SOEP sample for which we have observed an educational transition and/or an educational degree. For the larger part of this group we have observed an educational degree only (n=65,016). These are persons who have not been a part of the sample at the time when they participated in education or experienced educational transitions. The smaller part of the sample is more interesting for longitudinal analyse of educational participation. These are persons who lived in a survey household at the time of educational participation.[1] Depending on the age of the individual the dataset contains variables on:

- early child education and care (ECEC)
- entry into primary school
- transition to secondary school
- first exit from secondary school
- secondary school attendance after first exit from school
- first entry into and exit from vocational training
- vocational training participation after first
- first entry into and exit from tertiary education

[1] Accordingly the first group is much older than the second group. At the time of the first observation in the sample the first group is on average 45 years old while the second group has an average age below 9 years.

- tertiary education participation after first exit
- highest ever obtained educational degrees and last observed educational participation

The SOEP as a general household panel study is not specifically directed at the analysis of educational life courses. Nevertheless, right from the beginning of the panel in 1984 the survey instruments contained questions on the educational attainment of the respondents (aged 17 and older) and children younger than 17 years living in survey households. After more than 30 years of survey duration these data provide a precious source for the reconstruction of educational life courses. In the following we describe how we use these data to reconstruct educational transitions starting before school enrolment and up to post-secondary education.

The reconstruction of transitions is primarily based on yearly information on educational participation (i.e. entry and exit reconstructed from changes in participation). For later transitions there is some more information as explicit questions on the end of general school, vocational training and tertiary education are part of the questionnaires (changes during the year prior to the survey, only for persons aged 17+ years, exception: already obtained degrees before age 17 in youth questionnaire).

One remark on the variable naming conventions: The variable names always begin with "be" which stands for "biography education" (in analogy to other biography datasets). The third and fourth letter denote the type of transition or similar. For instance, t0 stands for variables on the first and t1 for the last year in child care. Variables on starting school contain a t2 and so on (up to t8= exit from tertiary education). Variables containing an x as the third letter contain information on the last observed year in education or on the highest educational degrees ever obtained (x4, x6, x8).

Using this dataset you should keep in mind that most of the information covered by the dataset is not directly asked in the SOEP questionnaires but has been derived from the combination of several variables. In the process of reconstruction assumptions have been made which we try to describe as detailed as possible in our exhaustive documentation of the dataset (see below). The more these assumptions are based on additional knowledge, e.g. provided by strict institutional regulations, the better for the reconstruction of the transitions.

The dataset covers transitions starting in early childhood up to tertiary education. For a part of the sample only one of these transitions or episodes is observed, for others the whole sequence from elementary education until the exit from tertiary education. The variable *beinfo* provides an overview on the frequencies of these different patterns. In total the dataset contains information on more than 90,000 persons. This is the part of the SOEP sample for which we have observed an educational transition and/or an educational degree. For 597 cases we have full information (pattern 811111111).

We have provided a number of variables where we documented the process of data generation and the sources where the data stem from (betXinfo, variables with suffixes _s or _g). You could use these variables as indicators of the degree of uncertainty in the process of the reconstruction of educational transitions. The less the variables could be reconstructed just using the basic algorithm (e.g., bet2info<>"0000|0|0000"), the higher is the degree of uncertainty. The same applies to long durations between an observed exit and the observation of a matching educational degree (e.g., a high value in bet6cert_g). It is certainly advisable to check if certain deviations in the process of data generation "explain" substantial results. E.g., if children living in households where interviewed in August (this information is provided in betXinfo) have a much higher propensity of starting school late (bet2agemo), this might just be a data artefact because it is difficult to decide if the information the household provided referred to the school year which just started in August or to the school year which just ended at the time of the interview. In general, you should expect that there are no such systematic measurement errors in the reconstructed variables. But if you want to have a closer look on potential biases you could use the respective variables which document the data generation process. This documentation describes a version of the dataset (v32_0.1). If you have comments or encounter while using the dataset, please let us know.

This is just a brief introduction to the dataset. Way more detailed information (especially concerning the algorithms used to reconstruct information) is provided in the following publication which can be easily found on the DIW website. It is highly recommended for people interested in working with BIOEDU to have a look at it.

Lohmann, Henning / Witzke, Sven (2011): BIOEDU: Biographical data on educational participation and transitions in the German Socio-Economic Panel Study (SOEP), DIW Data Documentation 58, Berlin.