

## European Data Watch

This section offers descriptions as well as discussions of data sources that are of interest to social scientists engaged in empirical research or teaching courses that include empirical investigations performed by students. The purpose is to describe the information in the data source, to give examples of questions tackled with the data and to tell how to access the data for research and teaching. We focus on data from German speaking countries that allow international comparative research. While most of the data are at the micro level (individuals, households, or firms), more aggregate data and meta data (for regions, industries, or nations) are included as well. Suggestions for data sources to be described in future columns (or comments on past columns) should be sent to: Joachim Wagner, Leuphana University of Lueneburg, Institute of Economics, Campus 4.210, 21332 Lueneburg, Germany, or e-mailed to (wagner@leuphana.de). Past “European Data Watch” articles can be downloaded free of charge from the homepage of the German Council for Social and Economic Data (RatSWD) at: <http://www.ratswd.de>.

### The German Socio-Economic Panel (SOEP) as Reference Data Set

By Thomas Siedler, Jürgen Schupp, C. Katharina Spiess,  
and Gert G. Wagner\*

#### 1. Introduction

This Data Watch note discusses how household panels in general – and the German Socio-Economic Panel (SOEP) in particular – can serve as reference data for researchers collecting datasets that do not represent the full universe of the population of interest (e.g., through clinical trials, intervention studies, laboratory and behavioural experiments, and cohort studies).<sup>1</sup>

---

\* For helpful comments, we thank Silke Anger, Joachim Frick, Jan Goebel, Markus Grabka, Henning Lohmann, Martin Kroh and Quirin Schimeta.

<sup>1</sup> A longer version of this note with recommended SOEP questions is published as a discussion paper, see Siedler et al. (2008).

Researchers often collect their own data for use in planning and designing a research study. An important issue that arises for many of these datasets is whether the sample represents the general population, or conversely, whether it is selective (for example, by design or through choice-based sampling). Household panels might offer a useful reference point for such studies, thanks to their longitudinal character and the sampling of all household members – for example, the British Household Panel (BHPS), the new UK household panel Understanding Society, and the German Socio-Economic Panel (SOEP).

This Data Watch note is addressed to researchers who collect their own datasets (mainly in Germany) and are interested in comparing their sample with a representative sample of the German population. It aims to provide them and their international co-authors and partners with recommendations on which questions to incorporate into their questionnaire in order to facilitate comparison with the German Socio-Economic Panel (SOEP), a large, ongoing representative longitudinal survey of the German population, including their families, and households.<sup>2</sup>

This approach offers researchers several benefits:

- First, by asking participants similar questions to those in the SOEP, researchers can compare their sample with either a sub-sample drawn from the SOEP or with the whole SOEP as a representative sample of the German population. By comparing their sample with the SOEP on key socioeconomic dimensions, they can then determine how representative their own sample is.
- Second, the SOEP could also serve as a “control sample” for longitudinal intervention studies.
- Third, in contrast to many of the scales and questionnaire instruments developed by psychologists, for instance, the SOEP questions are not copyrighted and can be used by other researchers free of charge. Thus, the SOEP can be a valuable source for the design of new questionnaires.
- Fourth, the SOEP questionnaire can serve as a general guideline for the selection of questions which are relevant for socio-economic research.

In section 2, we discuss potential benefits of using questions similar to those in the SOEP for studies in which researchers are interested in using household panels as reference data. In section 3, we present a list of topics (“SOEP core questions”) that we recommend be considered when researchers are collecting their own data. Of course the selection of a minimum set of questions depends on the research question under consideration. We then offer general advice for the selection of variables to researchers interested in comparing their own data with the SOEP.

---

<sup>2</sup> For further information about the SOEP, see Wagner et al. (2007), and Haisken-DeNew/Frick (2005). For an outlook on the future of SOEP, see Anger et al. (2008, 2009).

## 2. Two Recent Examples

Two recent studies show the potential for using questions from a panel survey when researchers collect their own data. The first example is the study entitled “Chances of employment in a population of women and men after surgery of congenital heart disease: Gender-specific comparisons between patients and the general population” by Geyer et al. (2009). The authors examine whether individuals aged 17–45 with operated congenital heart disease have adverse employment chances compared to people without heart problems. The authors compare their sample of patients (N=314; treatment group) with a sample drawn from the SOEP, which serves as a comparison group. The treatment group consisted of women and men who had a congenital heart disease and were operated on at the University Hospital of Göttingen. The authors conducted a face-to-face interview with patients using several SOEP questions.

The study by Ermisch et al. (2007) demonstrates how a panel survey can help in determining the extent to which a particular sample is representative of the general population. Ermisch et al. (2007) integrate a new experimental trust design into a former sample of the British population and compare their trust sample with a sample from the British Household Panel Survey (e.g., see Ermisch et al., 2007, Section 3.5). By using a questionnaire similar to the BHPS, the authors are able to determine that their trust sample over-represents women, people who are retired, older, divorced, or separated.

The studies by Geyer et al. (2009) and Ermisch et al. (2007) show that household panel studies can serve as useful reference data for researchers collecting their own samples and can help to reveal the representativeness of their own collected data.

## 3. Recommended Topics Covered by “SOEP Core Questions”

This section presents recommendations of selected SOEP topics aimed at measuring the life course of adults and young children. Over the years, the SOEP has asked many more questions, and we should point out that researchers are also free to use SOEP topics not recommended here.

For adults we recommend survey questions in six different areas: (1) demographic and parental characteristics; (2) labour market; (3) health; (4) personality, preferences, and subjective orientations; (5) subjective wellbeing, and (6) political involvement and participation. For young children before they enter school, we recommend a special list of topics and the respective questions.

Our “selection” of topics is defined by the main research purpose of the SOEP: describing the life course and the life span (cf. Anger et al., 2008, 2009). Thus, time use, with a focus on gainful employment, is one of the most

important aspects of SOEP's survey program. In addition, the SOEP measures respondents' personal goals and attitudes (as well as parents' goals for their children) and socio-economic variables such as education, earnings, labour forces status and satisfaction with life. Values and attitudes that have little or nothing to do with individuals' use of time are only covered to a minor extent (for example, political opinions and preferences with regard to the environment and environmental protection<sup>3</sup>). Overview 1 displays these topics. See also Siedler et al. (2008) for recommended SOEP questions of these topics. The detailed list of all survey questions (either in German or English) can be downloaded at <http://www.diw.de/soep-as-reference-data>.

### *Overview 1*

#### **SOEP's core topics on adults life**

Demographic and Parents' Characteristics  
 Labour Market  
 Health  
 Personality, Preferences, and Subjective Orientations  
 Subjective Wellbeing  
 Political Involvement and Participation

Besides the survey instruments used with all adult SOEP respondents, which were introduced with the very first wave of SOEP in 1984, in 2003, the SOEP began collecting mother-child questionnaires. With these new questionnaires, it will be possible to collect information from early childhood up to adulthood. These survey instruments are of particular interest for child-related research and research on intergenerational transmission (see, for example, Cawley/Spiess, 2008 and Coneus/Pfeiffer, 2007). By linking these child-related data with the main SOEP survey, a variety of parental characteristics such as personality and health can be controlled for, and a broad set of family and household context variables can be used. Thus, our child-related topics provide a good source of information helping to explain human behavior from a life course perspective.

Up to now, three age-specific survey instruments have been developed in the SOEP: one questionnaire deals with children in their first year of life, a second one with children aged 2–3, and a third questionnaire, which has been in the field since 2008, deals with children aged 5–6. The SOEP plans to develop and implement further child-related questionnaires for children in primary and secondary school up to the age of early adulthood.<sup>4</sup>

<sup>3</sup> Note that the new household longitudinal survey Understanding Society aims at collecting detailed information on respondents' preferences towards the environment. For further information see: <http://www.iser.essex.ac.uk/>.

The questionnaires for mothers with children five to six years old include an instrument to measure the personality of a child. For this age group, personality measures could be used similarly to the “Big Five Personality Measure” for adults (see Weinert et al., 2007: chapter 4).<sup>5</sup>

Many child-related research questions refer to some kind of *child outcome*. However, outcome measures in early childhood are very age-sensitive. Thus potential outcome measures in the SOEP mother-child-questionnaires vary by age as well. For children two to three years of age, a measure of the child’s adaptive behavior is used, covering four skill dimensions: social skills, language skills, motor skills, and everyday life skills (for a more detailed description of this instrument, see Schmiade et al., 2008).

Another child outcome measure for older children, aged five to six, is a modified scale of the Strength and Difficulties Questionnaire – SDQ Scale to measure the socio-emotional skills of a child (see Goodman 1997 for the original scale). A summary of the mother-child survey topics is provided in Overview 2.

Overview 2

**Mother-child topics**

<b>Children in their first year of life / newborns</b>
Sex of the child
Birth year and birth month of the child
Week of pregnancy at birth
Height, weight and head circumference at birth
Health and development status of the child (crude indicators including hospital stays and doctor visits)
Care arrangements
Personality of the child (crude indicators)
Physical and mental status of the mother during pregnancy
Change of circumstances in the life of the mother through birth
Support of partner
<b>Children 2 to 3 years old</b>
Current height and weight
Health of the child (crude indicators including hospital stays and doctor visits)
Care arrangements
Personality of the child
Activities with / of the child, child outcomes

<sup>4</sup> Suggestions for these questionnaires are especially welcome.

<sup>5</sup> At younger ages, the personality is not yet developed enough to measure it with exactly this scale; thus a shorter scale with only four items is used.

In many cases, even the recommended short version of the SOEP questionnaire will be too long and time-consuming, when collected in addition to data from laboratory experiments, for example. Then the question arises: what would be a meaningful short questionnaire? Of course, the selection of a minimum set of questions depends on the research question at hand. Nevertheless, we would recommend collecting the following basic information:

- gender, age, educational level, marital status, household type, number of persons living in the household, employment, occupational and health status. This information allows researchers to compare their own sample with the SOEP with respect to some basic socio-economic characteristics.
- basic demographic and socio-economic characteristics of parents: their age (or age at death), educational level and occupational status (at age 15 of the respondent / subject).
- religion or actual religious activity of parents as a potential indicator for early childhood socialization influences.
- out of a battery of personal trait questions, we recommend asking about the “locus of control” since these questions are significantly related to various economic behaviors (see Schupp et al., 2008). In addition, the Big Five Inventory might be useful for researchers interested in collecting a measure for personality traits.

Alternatively, researchers might have enough time and a large enough budget to conduct a relatively long survey, lasting around 30–40 minutes. We would like to point out that in such circumstances it is possible to use a complete SOEP questionnaire. The full SOEP questionnaires are available online at <http://www.diw.de/soep-questionnaires>. The usage of the questionnaire (either the English or German version) is free of charge for academic research.

#### 4. Conclusions

The use of selected SOEP variables in other studies (e.g., clinical trials, cohort studies, intervention studies, laboratory experiments) is an easy way of making those studies comparable with SOEP data and of using the SOEP – a representative and longitudinal sample of the entire German population – as a reference point (control sample).

While the SOEP questionnaire is an important component of the overall SOEP survey, we would like to point out that other aspects of the survey are also relevant. Thus, in addition to recommendations on which questions to incorporate into a particular survey, the SOEP survey group can also provide advice on other aspects of how to conduct a survey.

For instance, certain survey techniques such as the tracking rule in panel studies make it possible to follow not only cohort members but also their children (and grandchildren), spouses, and other related persons over time.

Moreover, the SOEP survey group in Berlin and TNS Infratest Sozialforschung in Munich can help interested researchers with the layout of a particular questionnaire and the instrument design. The SOEP data is mainly collected in face-to-face interviews using completely standardized survey instruments (“paper and pencil”, PAPI), with the interviewer conducting an oral interview or the respondents filling out questionnaires themselves. In addition, the “Computer Assisted Personal Interview” (CAPI) method has been introduced gradually. The paper version (PAPI) is relatively easy to implement, whereas CAPI is a more complicated instrument design. Whether researchers aim at using some (core) questions of the SOEP and collecting data in a similar manner, or if they have ideas on how to improve the SOEP, we recommend early consultation with the SOEP survey group at DIW Berlin.<sup>6</sup> Using SOEP questions is easy, but selecting the best questions for a given project is often more difficult.

The SOEP survey group is more than happy to provide advice. For those considering the use of SOEP methods of data collection, a larger amount of consultation is necessary, and numerous practical issues must be considered, such as negotiations with the fieldwork organization and data preparation. The latter issue is important because the SOEP survey group at DIW Berlin does not have the capacity to prepare additional data for other studies. Such a task goes far beyond the current manpower of the SOEP group. So: if you want to utilize SOEP for your own study, the SOEP survey group is eager to help and give advice. However, we would like to stress that due to finite staff capacities, consultations must be planned well in advance.

## References

*Anger, S. / Frick, J. R. / Goebel, J. / Grabka, M. M. / Groh-Samberg, O. / Haas, H. / Holst, E. / Krause, P. / Kroh, M. / Lohmann, H. / Pischner, R. / Schupp, J. / Sieber, I. / Siedler, T. / Schmitt, C. / Spiess, C. K. / Spiess, M. / Tucci, I. / Wagner, G. G. (2008): Zur Weiterentwicklung von SOEPsurvey und SOEPservice, Vierteljahrshefte zur Wirtschaftsforschung 77 (3), 157 – 177.*

---

<sup>6</sup> In rare cases, another study could theoretically introduce changes to SOEP, especially in the form of new questions or tests. Up to now, however, this has never occurred. It could happen only if there were an exceptionally strong scientific case made for the change. The SOEP is part of a research infrastructure, and we should stress that changes in SOEP can be made only for the purposes of pure scientific research. Changes aimed at facilitating commercial research are contrary to the constitutional principles of SOEP, and therefore not possible.

- Anger, S.* et al. (2009): Developing the SOEPsurvey and SOEPservices: The (near) future of the German Socio-Economic Panel Study (SOEP), SOEPpapers on Multidisciplinary Panel Data Research No. 155, Berlin: DIW Berlin.
- Cawley, J./Spiess, C. K.* (2008): Obesity and skill attainment in early childhood, *Economics and Human Biology* 6 (3), 388–297.
- Coneus, K./Pfeiffer, F.* (2007): Self-productivity in early childhood. ZEW – Centre for European Economic Research Discussion Paper No. 07–053, Mannheim.
- Ermisch, J./Gambetta, D./Laurie, H./Siedler, T./Uhrig, N.* (2007): Measuring people's trust, ISER Working Paper 2007–32. Colchester: University of Essex. A revised version is forthcoming in the *Journal of the Royal Statistical Society, Series A*.
- Geyer, S./Norozi, K./Buchhorn, R./Wessel, A.* (2009): Chances of Employment in Women and Men after Surgery of Congenital Heart Disease: Comparisons between Patients and the General Population, *Congenital Heart Disease* 4 (1), 25–33.
- Goodman, R.* (1997): The strengths and difficulties questionnaire: A research note, *Journal of Child Psychology and Psychiatry* 38, 581–586.
- Haisken-DeNew, J. P./Frick, J. R.* (2005): Desktop Companion to the German Socio-Economic Panel (SOEP) – Version 8.0, Berlin. Available at: <http://www.diw.de/deutsch/sop/service/dtc/dtc.pdf>.
- Schmiade, N./Spiess, C. K./Tietze, W.* (2008): Zur Erhebung des adaptiven Verhaltens von zwei- und dreijährigen Kindern im Sozio-oekonomischen Panel (SOEP), SOEPpapers on Multidisciplinary Panel Data Research No. 116, Berlin: DIW Berlin.
- Schupp, J./Herrmann, S./Jaensch, P./Lang, F. R.* (2008): Erfassung kognitiver Leistungspotentiale Erwachsener im Sozio-oekonomischen Panel (SOEP), DIW Data Documentation No. 32, Berlin: DIW Berlin.
- Siedler, T./Schupp, J./Spieß, C. K./Wagner, G. G.* (2008): The German Socio-Economic Panel as a Reference Data Set, SOEPpapers on Multidisciplinary Panel Data Research No. 150, Berlin: DIW Berlin.
- Wagner, G. G./Frick, J. R./Schupp, J.* (2007): The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements, *Schmollers Jahrbuch* 127 (1), 161–191.
- Weinert, S./Asendorpf, J. A./Beelmann, A./Doil, H./Frevert, S./Lohaus, A./Hasselhorn, M.* (2007): Expertise zur Erfassung von psychologischen Personmerkmalen bei Kindern im Alter von fünf Jahren im Rahmen des SOEP, DIW Data Documentation No. 20, Berlin: DIW Berlin.