Schmollers Jahrbuch 133 (2013), 595 – 606
Duncker & Humblot, Berlin

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 send 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.

"Familien in Deutschland" – FiD*

By Mathis Schröder, Rainer Siegers, and C. Katharina Spieß**

1. Introduction

There are various independent studies evaluating family policy measures in Germany. So far, a systematic evaluation considering the different goals inherent to these measures was missing. The evaluation of family policy measures on behalf of the Federal Ministry for Family Affairs, Senior Citizens, Women and Youth (BMFSFJ) and the Federal Ministry of Finance (BMF) is thus the

* For a more detailed description of FiD, see Schröder/Siegers/Spieß (2013).

** The authors gratefully acknowledge funding from the "Bundesministerium für Familie, Senioren, Frauen und Jugend" (BMFSFJ) and the "Bundesministerium für Finanzen" (BMF). Thanks are especially in order for the BMFSFJ for financing the data collection in 2013 and thus providing the possibilities for a long-term panel.

first systematic overall study. In a feasibility study conducted for this overall evaluation in late 2008, one of the main conclusions was that "*Without additional data only a limited number of policies regarding families and children can be evaluated"* (authors' translation, see Beninger et al., 2008). The available data sets were not sufficient for in-depth analyses, especially regarding specific family types which might be rare in the German population, but still important as targets for the ministries' policies. Such families are especially single parents, large families, low-income families, and families with very young children. The main panel studies to evaluate family policies existing at the time were the Socio-economic Panel (SOEP, see Wagner / Frick / Schupp, 2007), and the Panel Analysis of Intimate Relationships and Family Dynamics (*pairfam*, see Huinink et al., 2011). While the targeted groups are present in the SOEP (general population survey) and in *pairfam* (family survey), the actual case numbers in these studies are far too low to provide sufficient statistical power for an evaluation of family policy measures.

This was the initiation for the data collection effort "Familien in Deutschland" (short FiD, for "Families in Germany"). This project started to collect data in 2010, with the focus on single parents, low income families, large families with three or more children, and families with particularly young children, namely those born between 2007 and 2010. FiD was initially financed by the two federal ministries in charge of the overall evaluation. The funding covered three waves of data collection, spanning the years 2010 – 2012, which were used in various studies for the overall evaluation.[1] A further wave of data collection in 2013 was funded by the Federal Ministry for Family Affairs. As FiD collects longitudinal data very similar in content and structure to the SOEP data, the data collection will subsequently become part of the regular SOEP. However, it is already possible to jointly use the two data sets with sampling weights provided for this particular purpose. With the integration, which will take place in 2014 (i.e. the FiD households will become part of the regular SOEP in the data collection of 2014), FiD will further strengthen the base of family research in the SOEP data.

2. Data Description

2.1 Sampling[2]

The goal for FiD was to provide data on four groups particular important for family policies: single parents, low income families, large families with three

[1] For an overview of studies, see the homepage of the BMFSFJ (accessed June 2013): http://www.bmfsfj.de/BMFSFJ/familie,did=195944.html.

[2] All details on the sampling procedure are available in the documentation by TNS Infratest (Jänsch et al., 2011).

or more children, and families with young children. With respect to the last group, sampling was relatively easy, because even though Germany does not have a central registry, local registries exist and provide sample draws based on certain characteristics, among them year of birth. Hence a sample of individuals born between 2007 and March 2010 was drawn in 160 sample points, which were stratified by state ("Bundesland"), administrative region ("Regierungsbezirk") and a region's population size ("BIK-Regionen").

Due to a low expected response rate of households with a migration background, a decision to include more households of this group than representative for the total population was made early on. Migration households in the register sample were identified in two ways: on the one hand, the registries provide information on the nationality of a person in the register, which was taken to classify non-German households as migrant households. On the other hand, using only nationality would omit all cases who obtained the German nationality later or are immigrants in the second-generation with a German nationality. Thus a second way to identify households with a migration background was employing an onomastic procedure, which basically assigns a linguistic and regional origin to each address based on the person's surname. Among the drawn cases, an oversampling of immigrant households was conducted, such that the percentage of migrants (identified via the register and onomastic procedure) in the sample was doubled for each sampling point. As the sample is characterized by children born in the cohorts from 2007 to 2010, this sample is also referred to as the "Cohort Sample".

For single parents, low income families and families with three or more children there is no sampling frame in Germany. Even though single parents and large families could theoretically be identified by the local registries, data protection rules prohibit such a combination of individual records into household-specific information. Hence sampling had to be conducted using a screening process, for which the starting sample was provided by TNS Infratest Sozialforschung from omnibus studies conducted every month in a representatively drawn sample. Households participating in such an omnibus study are always asked whether they would respond to a future survey. These households pose the gross sample for the screening process for FiD, where households were asked in telephone interviews about their household composition (to identify single parents and large families) and their household income (to identify low income households). Households of the target population were identified and categorized according to the following criteria:[3]

[3] To keep the screening process simple, the definition of "children" and "adults" was strictly based on age. Children are thus all those individuals in the household who – at the beginning of the survey year – were at most 17 years old. Adults were all others, i.e. those who at the beginning of the survey year were at least 18 years old. The actual family relationships were not relevant here, however, most of these households also are

- *Low income* if the household had a monthly income of less than
  - 2500 Euro, when composed of at least two adults and at least two children
  - 2000 Euro, when composed of at least two adults and one child
  - 1500 Euro, when composed of one adult and at least one child.
- *Single parent* if the household is composed of one adult and at least one child.
- *Large family* if the household includes three or more children.

Households meeting any of the three characteristics were asked if they were willing to participate in the FiD study. A positive answer to the participation question was then followed by an invitation to participate in the study and a visit by an interviewer (for details on the selection process and case numbers, see Schröder / Siegers / Spieß, 2013). Due to the selection process, these cases are also referred to as the "Screening Sample 2010".

In the initial phase of the project, it was uncertain whether enough single parent and large family households could be acquired through the screening process. Hence it was planned from the beginning to repeat the screening process in 2011 for an additional sample of single parents and large families. These cases are also referred to as the "Screening Sample 2011".[4] For the years of 2010 to 2012, Table 1 shows the sample sizes (households, individuals, children) captured by FiD.

families with parent-child relationships. When determining the eligibility of households later, the same definitions were kept.

[4] Even though the sample names are slightly misleading in their terminology ("Screening" describes a sampling procedure, "Cohort" a sample characteristic), these definitions are kept to be consistent with the documentation and the field reports by the survey agency.

Table 1

Sample Sizes and Conducted Interviews by Questionnaire, 2010 – 2012

	2010	2011	2012
Interviews			
Household Questionnaire	4,574	4,529	4,186
Person Questionnaire (17+ year-olds)	7,807	7,648	7,165
Youth Questionnaire (16 – 17 year-olds)	190	262	293
Parent Questionnaire 1 (0 – 1 year-olds)	1,321	207	212
Parent Questionnaire 2 (1 – 2 year-olds)	787	644	568
Parent Questionnaire 3 (2 – 3 year-olds)	871	740	555
Parent Questionnaire 4 (5 – 6 year-olds)	473	486	424
Parent Questionnaire 5[a] (7 – 8 year-olds)	425	527	501
Parent Questionnaire 6[a] (9 – 10 year-olds)	404	510	475
Gap (Luecke) Questionnaire[b]			227
Totals			
Persons	17,002	17,129	15,850
Adults 17+	8,301	8,052	7,630
Youth (16 – 17)	190	262	293
Children (0 – 16)	8,511	8,815	7,927

Source: FiDv3.0

Note that these are wave specific net samples. The numbers in this table are restricted to those households with a completed household interview. There are a few households where only a person interview was conducted (these data are available in the regular distribution).

[a] Number of cases with at least one interview. Parent Questionnaires 5 and 6 are answered by both mother and father if applicable, such that two observations exist for many children in these age groups. The actual number of interviews is thus larger than the sample sizes given here.

[b] The Gap Questionnaire is listed here in "2012", as it was filled out in this year. However, information was gathered for the previous year (2011).

2.2 Structure and Contents

In large parts, FiD resembles the SOEP, i.e. household, person and youth questionnaires as well as questionnaires about children in specific age groups are given to the participants. Due to the similar structure in questionnaires, the data sets provided in FiD are also very similar to those distributed in the regular SOEP. Most SOEP datasets also exist in FiD, as long as the information needed is available in FiD (for details, see Schröder / Siegers / Spieß, 2013).

The contents of the FiD study are very similar to the SOEP, i.e. basic information on the household and each person is asked, including education, past and current labour market experiences, earnings and income, housing characteristics, health, some preferences and life satisfaction in general and for specific aspects. In addition, there is a stronger focus on children and partnership:

FiD includes a detailed partnership module, which retrospectively asks for marriages and partnerships lasting longer than six months.[5] Compared to the SOEP, men and women are asked about their biological children in slightly more detail, including information about the partner's location and the marital status at the time of birth. Also, some aspects of child care at the work place are covered.

Completely new in FiD are questionnaires for the 1–2 year-olds, and the 9–10 year-olds, which previously did not exist in the SOEP (as of 2012, the SOEP added a questionnaire for the 9–10 year-olds, which is comparable to the FiD-version). Each of the questionnaires includes a module on child care, which, as the panel grows older, allows comparing child care decisions for one child over time. In these sections, parents are asked to specify the reasons for or against using day care, and in the case they use care, they are asked about their satisfaction with it on different dimensions. Also covered are more detailed "outcome measures" to capture the skill development of children. Overall, the additional questions are designed to be comparable across the different parental questionnaires.

2.3 Interviewing

All personal interviews in FiD are conducted in a face-to-face mode using computer assisted personal interviews (CAPI). The only exceptions are the parent questionnaires, which can either be conducted on the interviewer's laptop or with pen and paper by the respondent herself. Future mode switches to pen and paper interviews (PAPI) are not feasible for the later waves of FiD, because the questionnaire routing depends to some extent on the technical possibilities a computer offers. Using only CAPI interviews promises some benefits for data quality and lowers the amount of time necessary to test and verify the data. On the other hand, it limits the possibilities of obtaining data from reluctant respondents. Because the SOEP allows for interviews via mail (and thus PAPI mode) at the end of the fieldwork period if all other attempts have failed to convince a household to participate, the integration of FiD into the SOEP may lead to mode switches in the future.

To thank the respondents for their participation in the study, FiD could implement an incentive scheme that was especially targeted at families. € 5 are paid for a completed household interview and the first personal interview. Each further completed individual interview is rewarded with an additional € 5. If all questionnaires for eligible persons in the household are completed, there is an additional premium of € 5 for each child in the household. In addition, as the

[5] A similar module is now integrated in the new SOEP-Samples J and K in CAPI mode.

sample consists of households with children, special panel care measures are taken: each household receives a gift for Children's Day on November 20th, where, depending on the age of the child, balloons, washcloths, bibs, reflectors, pencils or similar small gifts are included with a thank you letter.

2.4 Representativity and weighting

The Cohort Sample drawn in 2010 is representative of the population of families in Germany with children born between January 2007 and March 2010. The sampling weights for this sample are constructed relatively easy, as for each household the sampling probability is known through the design of the survey. These design weights are first adjusted for the initial non-response due to refusal or inability to participate of eligible households. In a second step, they are then calibrated by a raking approach using the margins of the most important variables known for the German population from the Mikrozensus. In the waves after 2010, the sample can be regarded as being representative of the same population, although there is a small bias: some households may actually lose their "eligibility" of being a household with children born between 2007 and 2010, because the children may no longer be in the household. Similarly, the sample does not capture those households, which after 2010 include a child born in those years (e.g. through a moved-in partner coming from abroad with a child). However, these fluctuations in and out of the sample population are very small: of the 3,100 households participating in all three waves, only ten (0.3%) would have to be removed from the cohorts 2007–2010. This relationship will be stable at least for the initial years of the sample.

The case is slightly different for the Screening Samples. Here, the sample drawn in 2010 is representative of the population of families in Germany, which are low-income families, large families or single parent families in 2010. While margins for this population in 2010 exist, the initial design weights have to be estimated, as the sampling probabilities are not known for the screening process. Compared to the Cohort Samples, this introduces some uncertainty, although the calibration can then be done in a similar way. The challenges increase in the following years: due to the fluctuations in and out of the three screening groups in every year, this sample does not represent the same three groups in the German population in 2011, or any of the following years. 44% of the households identified as low income move out of this group in at least one of the following years. In terms of their sample characteristics, things look much better for the other two groups: almost 80% of the single parent families remain in this state over the three years. The "large family" characteristic is even more stable: about 11% leave it within the first two years of the panel. Strictly speaking, the Screening Samples should thus be seen as representative of the respective populations in 2010 and 2011, and then can be used to moni-

tor the changes in these populations over time. However, as fluctuations at least in family compositions are not huge, the sampled groups remain close to the targeted groups at least in the first years.

Given these difficulties, constructing cross sectional sampling weights for FiD alone (i.e. the joint Cohort and Screening Samples without the SOEP-Samples) in 2011 is not trivial, but necessary as useful analyses also for future waves are possible only with these sampling weights as a basis. Because our approach to this problem is non-standard, more detail is provided here.

The starting point is the integration of the FiD households into the regular SOEP, which allows for joint analyses of the two datasets. This integration is achieved by treating the FiD cases as any other new sample in the SOEP, which would be integrated by including the old cases with their previous sampling weights (adjusted for attrition) and the new cases with their design weights (adjusted for the initial non-response). After an adjustment according to the number of observations in each group, the calibration follows a raking approach using margins from the general population (information from the "Mikrozensus"). The joint SOEP-FiD-weights from this step form the building block for the cross-sectional weights for the FiD population in 2011, which consists of the following possible types of household characteristics:

– low income household in 2010 (from Screening 2010)

– large family household in 2010 or 2011 (Screening 2010 or 2011)

– single parent household in 2010 or 2011 (Screening 2010 or 2011)

– cohort household 2007 – 2010 in survey year 2010

In principle, the integrated weights for the joint FiD and SOEP samples provide the weights of this population. To calculate weights for the FiD population alone, the SOEP cases have to be removed, and their removal has to be adjusted for in the weights. However, the population described by the four characteristics above is not easily determined for all households available in 2011: For those households not present in 2010, the likelihood of being in one of the groups has to be estimated. This affects new cases from 2011 (Screening 2011 from FiD as well as the new SOEP sample "J") and cases in the SOEP, which did not participate in 2010 but returned in 2011. Of the overall 16,819 cases in the joint FiD-SOEP population, this concerns 4,226 households (Screening 2011: 915 households; Sample J: 3,136 households; temporary dropouts SOEP: 175 cases). For these cases it is assumed that if they have children in the four cohort years in 2011, they also had them in 2010. For the other three categories, the status for 2010 is unknown. Taking the cases which are available in both years, a logit model predicts the likelihood of having the characteristics in 2010. An out of sample prediction then categorizes those cases only observed in 2011.

This procedure assigns all households in the joint FiD-SOEP sample to the four groups above, which also means that with the joint sampling weights the population equivalent of the above groups can be produced. To calculate the weights for the FiD population alone, the SOEP cases in this group need to be removed, and the FiD cases need to be scaled up to still remain at the same population total. While the simplest way would be to multiply each FiD weighting factor with the inverse of the fraction of SOEP cases in the total, this approach would ignore any systematic differences between the SOEP and the FiD cases – which, given the sampling design, are sure to exist. Instead an approach similar to a regular estimation of drop-out probabilities for an attrition analysis is used (see for example Kroh, 2011): By estimating the likelihood in the joint SOEP-FiD population of belonging to the FiD sample only, a household-specific factor (the inverse of this likelihood) is obtained and multiplied with the integrated weights. The sum of these newly achieved factors leads to an estimate of the population similar to the one derived for the joint SOEP-FiD cases.

These cross-sectional weights for 2011 provide the starting point for the future cross-sectional weights. The above method is followed for each year, e.g. the weights of 2011 are adjusted for attrition into 2012, and the integrated SOEP-FiD-data are then again jointly calbrated. These cases provide the estimate of the population equivalent of the FiD-population in 2012.

2.5 Survey quality measures[6]

Some important measures of survey quality were investigated, i.e. the initial response rates, the retention rates, the rate of complete households and the rate of item non-response. Due to the different sampling strategies employed, it is not possible to report a single response rate for FiD. The Cohort Sample allows for a straightforward calculation of the response rate, which was almost 40%. Migrant households (identified as described above) are significantly less likely to participate than German households (a difference of 23 percentage points). For households in the Screening Samples, for which the eligibility could be determined, around 70% participated in the respective first waves of 2010 and 2011. Once the households took part, they were rather likely to stay in the study: the household retention rates are slightly higher in the Screening Sample 2010 than in the Cohort Sample (86 vs. 79% in 2011, 89 vs. 83% in 2012), but both samples show a positive trend as is usual for the beginning of a panel study.

[6] We provide only a brief overview of measures of survey quality here, For details, see Schröder / Siegers / Spieß (2013).

The participation of eligible household members within a participating household is another important measure of completeness. Over the years, there is also a positive trend: the rate of completed household increased from 88% in 2010 to 93% in 2012. Especially successful is the completion rate for the parent questionnaires: in all years, more than 97% of required parent questionnaires were completed. Finally, the rate of item non-response is important for researchers: with higher rates of item non-response, more missing data may reduce the validity of the analysis. FiD has low rates of item non-response: over all years, the median of missing values is slightly above zero, and 90% of all questionnaires contain less than three per cent missing values.

3. Data Access

The data from "Familien in Deutschland" are accessible for the scientific community similar to the SOEP data. Interested researchers can apply for the data usage at the SOEP-group by filling out a two-page form. Following an evaluation of the application, a contract between the researcher and the SOEP needs to be signed before the data are made available to the new user, currently via one-time downloads. For further information on the application process (and all possible changes to it) please consider the information at www.diw.de/fid-soep.

4. Summary and Outlook

The FiD data collection effort has become a success story for the SOEP group at DIW Berlin. An entirely new and relatively large sample was drawn and proved to be of similar longitudinal stability as the regular SOEP samples. With respect to different quality indicators (e.g. item non-response, partial unit non-response) the FiD data adhere to the high standards set by the SOEP. Several new and extended questionnaires related to family specific topics were implemented within a very short period of time, which provide new and improved data on children. As such, FiD allows for more in-depth analyses of families and children.

After FiD was financed by the ministries for three years and a fourth wave was commissioned by the BMFSFJ to continue a full data collection in 2013, the FiD-samples will be integrated into the main SOEP for the data collection of 2014. The data distribution of 2015 will then contain the complete set of SOEP and FiD cases for the first time.

References

Beninger, D. / *Bonin*, H. / *Clauss*, M. / *Hofmann*, H. / *Horstschräer*, J. / *Munz*, S. / *Spieß*, C. K. / *Weding*, M. / *Wrohlich*, K. (2008): Machbarkeitsstudie zur stufenweisen Gesamtevaluation des Gesamttableaus ehe- und familienbezogener Leistungen in Deutschland, Studie im Auftrag der Prognos AG.

Huinink, J. / *Brüderl*, J. / *Nauck*, B. / *Walper*, S. / *Castiglioni*, L. / *Feldhaus*, M. (2011): Panel Analysis of Intimate Relationships and Family Dynamics (*pairfam*): Conceptual framework and design, Zeitschrift für Familienforschung 23 (1), 77 – 100.

Jänsch, A. / *Huber*, S. / *Siegel*, N. A. / *Stimmel*, S. / *Geue*, D. (2011): "Familien in Deutschland" (FiD) 2010 Methodenbericht: Anlage und Ergebnisse der FiD-Stichproben, TNS Infratest, München. (Available at www.diw.de/fid-soep, "FiD-Dokumentation".)

Kroh, M. (2011): Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) (1984 until 2010), DIW Data Documentation 59.

Schröder, M. / *Siegers*, R. / *Spieß*, C. K. (2013): "Familien in Deutschland" – FiD. Enhancing Research on Families in Germany, SOEP-Paper 556 / 2013, DIW Berlin.

Wagner, G. G. / *Frick*, J. R. / *Schupp*, J. (2007): The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements, *Schmollers Jahrbuch* – Journal of Applied Social Science Studies 127 (1), 139 – 169.