

<p><b>1. Neues, nutzerfreundlich integriertes Datenformat</b></p> <p>Mit der neuen Welle der SOEP-Haupterhebung (SOEP-Core) führen wir die beiden bisher getrennt übermittelten Formate (wide und long) zusammen. Dadurch wollen wir Unklarheiten vermeiden, was in welchem Format vorliegt und die Nutzung der Daten insgesamt erleichtern. Nachdem wir bereits einige Jahre lang SOEPlong als ein nutzerfreundliches ergänzendes Angebot – sowohl für EinsteigerInnen als auch für langjährige NutzerInnen – getestet haben, soll nun die Weitergabe der SOEP-Daten standardmäßig auch alle Datensätze im long-Format enthalten. Das bedeutet, dass das ausgelieferte Datenfile folgende unterschiedlichen Formate der SOEP-Daten enthält, die zum Teil in unterschiedlichen Verzeichnissen gespeichert sind.</p> <p><b>Bitte achten Sie daher darauf, dass Sie beim Entpacken auch die Verzeichnisstruktur mit entpacken.</b></p>	<p><b>1. New, user-friendly integrated data format</b></p> <p>The new wave of the SOEP-Core study incorporates our “wide” and “long” data formats, which used to be provided to users separately. Our aim is to eliminate any confusion about what is available in which format and to make data use easier overall. After several years of testing SOEPlong as an additional service designed to facilitate analysis for both experienced and new users, we will now be providing all datasets in the “long” format as a standard part of our SOEP data release. This means that you will find the different SOEP data formats listed below in your data file, some of which will be contained in separate subdirectories.</p> <p><b>Please make sure that you unpack the entire directory structure when unpacking your data.</b></p>
<p><b>1.1. SOEP im “long”-Format auf der obersten Ebene</b></p> <p>Im obersten Verzeichnis (oder dem Root-Verzeichnis) liegen alle bisher mit SOEPlong ausgelieferten Datensätze (also PL, PPFADL etc.) sowie die bisher nur im wide-Format ausgelieferten Datensätze (z.B. die Biografie- oder spellbezogenen Datensätze, wie BIOPAREN, ARTKALEN usw.). Damit ist der komplette Datenumfang der SOEP-Haupterhebung mit den <u>im obersten Verzeichnis liegenden Datensätzen</u> abgedeckt.</p> <p>Nachdem wir bereits seit einigen Jahren das long-Format als ein nutzerfreundliches ergänzendes Angebot bereitstellen, sind wir überzeugt, dass dieses Format insbesondere für EinsteigerInnen eine sehr viel komfortablere Nutzung der Daten ermöglicht. Deswegen haben wir entschieden, jetzt und in Zukunft dieses Format bevorzugt weiterzugeben.</p>	<p><b>1.1. SOEP in “long” format on the top level</b></p> <p>In the top-level (or root) directory, you will find all of the datasets provided up to now with SOEPlong (PL, PPFADL, etc.) as well as all of the additional datasets formerly provided only in our classic “wide” format (biographical or spell data such as BIOPAREN, ARTKALEN, etc.). All of the data in the main SOEP-Core study are therefore contained in the <u>datasets in the top-level directory</u>.</p> <p>Feedback from experienced and beginning users over the past several years shows that the “long” data offer significant advantages in ease of use, particularly for beginners. We have therefore decided to use this as our primary data format in future data releases.</p>

<p>Die einzelnen jahresspezifisch vorliegenden Datensätze werden dabei gepoolt über alle Jahre in einem Datensatz zusammengefasst. (z.B. werden alle \$P-Datensätze in einen PL-Datensatz integriert. Hierbei müssen Variablen gegebenenfalls über die Zeit harmonisiert werden. Die Harmonisierung erfolgt, um Variablen über die Zeit konsistent in einer Variable abbilden zu können: So werden beispielsweise auch die Einkommensangaben, die vor 2001 erfolgten, durchgehend in Euro und nicht in DM ausgewiesen, oder bei veränderten Fragebogenfassungen werden gegebenenfalls Kategorien über die Zeit angepasst. Alle Anpassungen werden jeweils für die NutzerInnen erkennbar und nachvollziehbar dargestellt, zudem werden alle modifizierten Variablen in ihrer jeweils ursprünglichen Form bereitgestellt. (siehe unten: <a href="#">_v*-variables</a>). SOEPlong reduziert damit sowohl die Anzahl an Datensätzen als auch die Zahl der Variablen erheblich.</p> <p>Eine detailliertere Beschreibung dieses neuen Formats der SOEP-Core-Datenweitergabe finden Sie in unserem neuen <a href="#">SOEPcompanion</a>.</p>	<p>All available individual year-specific datasets are pooled into a single dataset (e.g., all \$P datasets are integrated into the PL dataset). In some cases, this means that we have to harmonize variables in order to be able to define them consistently over time. For instance, income information is given in euros up to 2001 and not in deutschmarks, and in cases where questionnaires have changed, the categories are modified over time. All changes are presented to users in a clear and understandable way, and if harmonization is necessary, all input variables are provided in their original form (see below <a href="#">_v*-variables</a>). SOEPlong thus significantly reduces the number of datasets and the number of variables.</p> <p>A more detailed description of the format of our SOEP-Core data releases can be found in our new <a href="#">SOEPcompanion</a>.</p>
<p><b>1.1.1 Die wichtigsten Änderungen zu v33 im long-Format</b></p> <ul style="list-style-type: none"> <li>• Folgende neuen Datensätze wurden hinzugefügt:  <b>HBRUTT:</b> HBRUTT\$\$-Datensätze im long-Format  <b>PLUECKEL:</b> \$PLUECKE-Datensätze im long-Format  <b>VPL:</b> \$VP-Datensätze im long-Format</li> <li>• PL und PL2 werden wieder in einem gemeinsamen File (PL) ausgeliefert</li> <li>• Das Variablenschema mit c-Variablen (cross-sectional) und l-Variablen (longitudinal) wurde folgendermaßen abgewandelt: Wenn sich für eine Variable im long-Format die zu Grunde liegenden Variablen im Querschnitt änderten, dann wird für jede Version eine entsprechende <a href="#">_v*-Variable</a> erzeugt. Gleichzeitig wird eine harmonisierte <a href="#">_h-Variable</a> bereitgestellt. Hierzu gibt es weitere Informationen im</li> </ul>	<p><b>1.1.1 Most important changes to v33 in the long format</b></p> <ul style="list-style-type: none"> <li>• The following new files have been added:  <b>HBRUTT: long file of the HBRUTT\$\$ files</b>  <b>PLUECKEL: long file of the \$PLUECKE files</b>  <b>VPL: long files der \$VP files</b></li> <li>• PL and PL2 are being provided again in a combined file (PL).</li> <li>• The variable scheme with c-variables (cross-sectional) and l-variables (longitudinal) has been modified as follows: If the variables on which a variable in the long format is based changed in the cross-section, then corresponding <a href="#">_v*-variables</a> will be created for each version. A harmonized <a href="#">_h-variable</a> is provided as well. Further information can be found in the SOEPcompanion (<a href="#">general description</a>, <a href="#">examples</a>)</li> </ul>

<p>SOEPcompanion (<a href="#">Allgemeine Erklärung, Beispiele</a>)</p> <ul style="list-style-type: none"> <li>• In allen aus verschiedenen querschnittlichen Datensätzen generierten long-Datensätzen gibt es die neue Variable INPUTDATASET.</li> <li>• Das neue gemeinsame Datenweitergabeformat führt dazu, dass einige Datensätze mit long-spezifischen Namen nicht mehr in der Datenlieferung enthalten sind: CDESIGN, CSAMP, CSAMP<sub>FID</sub>, KIDL, PBR<sub>EXIT</sub></li> <li>• Die folgenden Datensätze wurden umbenannt, damit es nicht zu einem Namenskonflikt mit den Daten im raw -Verzeichnis kommt:</li> </ul> <p><b>PPATH ersetzt PPFAD</b></p> <p><b>PPATHL ersetzt PPFADL</b></p> <p><b>HPATH ersetzt HPFAD</b></p> <p><b>HPATHL ersetzt HPFADL</b></p>	<ul style="list-style-type: none"> <li>• All of the long datasets generated from the various cross-sectional datasets contain the new variable: INPUTDATASET.</li> <li>• Due to adjustments to the new joint data release format, some files with “long”-specific names are no longer included in the data release: CDESIGN, CSAMP, CSAMP<sub>FID</sub>, KIDL, PBR<sub>EXIT</sub></li> <li>• The following datasets have been renamed to avoid conflicts with the data names in the raw directory:</li> </ul> <p><b>PPATH replaces PPFAD</b></p> <p><b>PPATHL replaces PPFADL</b></p> <p><b>HPATH replaces HPFAD</b></p> <p><b>HPATHL replaces HPFADL</b></p>
<p><b>1.2 Klassisches Format im Unterverzeichnis raw</b></p> <p>Weil wir uns bewusst sind, dass viele Skripte auf dem Original-Datenformat beruhen und um den Prozess der Generierung der long-Daten nachvollziehbar zu machen, finden Sie alle Datensätze im originalen SOEP-Format im Verzeichnis raw.</p> <p><b>NutzerInnen, die weiterhin das alte Format nutzen wollen, brauchen nur ins Unterverzeichnis raw zu wechseln und die Datensätze zu nutzen, die dort liegen.</b></p> <p>Der einzige Unterschied zu vorher ist, dass es jetzt in allen Datensätzen im raw-Verzeichnis zusätzliche Identifikatoren mit den Namen im long-Format gibt (PID <b>and</b> PERSNR oder HID <b>and</b> \$HHRNAKT) sowie die Variable für das Befragungsjahr (SYEAR), so dass die Nutzenden ganz einfach Variablen aus beiden Datenformaten mergen können.</p>	<p><b>1.2 Classic format in the subdirectory raw</b></p> <p>Since we know that many users have existing scripts that are based on the original data format, and to enable users to understand the process of generating the “long” data, we provide all of the datasets in their original SOEP format in the directory raw.</p> <p><b>Users who want to continue using the old format simply need to switch into subdirectory raw and use the datasets there.</b></p> <p>The only change is that there are now additional identifiers in all of the datasets in the raw directory with the name in the long format ( PID <b>and</b> PERSNR or HID <b>and</b> \$HHRNAKT) and a survey year variable (SYEAR) so that users can easily merge variables from the two data formats.</p>

### **1.3 Neuer EU-SILC-Klon im Unterverzeichnis *eu-silc-clone***

Vielen NutzerInnen ist sicher bekannt, dass das SOEP länderübergreifende Analysen mit Hilfe des CNEF durch den Datensatz PEQUIV unterstützt. Wir haben nun ein Datenprodukt entwickelt, das es Ihnen ermöglicht, die SOEP-Daten für vergleichenden Analysen zusammen mit den Daten von EU-SILC (European Union Statistics on Income and Living Conditions) zu nutzen. EU-SILC, das von Eurostat auf Anfrage zur Verfügung gestellt wird, bietet Querschnitt- und Längsschnittinformationen für viele europäische Länder. Für für Deutschland lagen bisher allerdings nur Querschnittinformationen vor. Der EU-SILC-Klon bietet nun Längsschnittinformationen über private Haushalte in Deutschland auf der Grundlage der SOEP-Daten. Alle darin enthaltenen Informationen sind direkt mit den EU-SILC-Längsschnittinformationen anderer europäischer Länder vergleichbar.

Der EU-SILC-Klon ist in die Standard-SOEP-Datenweitergabe (im Unterverzeichnis *eu-silc-clone*) integriert.

Die Dokumentation des EU-SILC-Klons 2005-2016 finden Sie [hier](#).

### **1.3 New EU-SILC clone in the subdirectory *eu-silc-clone***

Many users are undoubtedly aware that the SOEP supports cross-national analysis with CNEF through the dataset PEQUIV. We have now produced a data product that allows you to use the SOEP data in comparative analyses with the EU-SILC (European Union Statistics on Income and Living Conditions) data. EU-SILC, which is provided by Eurostat upon request, offers cross-sectional and longitudinal information for many European countries. Up to now, only cross-sectional information has been available for Germany. The EU-SILC clone offers longitudinal information on private households in Germany based on the SOEP data. All of the information contained in it can be directly compared with the EU-SILC longitudinal information on other European countries.

The EU-SILC clone is integrated into the standard SOEP data release (in subdirectory *eu-silc-clone*).

Documentation on the 2005-2016 EU-SILC clone can be found [here](#).

<p><b>2. Neue Samples in der SOEP-Hauptstudie</b></p> <p>Die neue SOEP-Core-Datenweitergabe (v34) ist die erste, die Daten aus der IAB-BAMF-SOEP-Befragung von Geflüchteten in Deutschland als Sample M5, sowie die Fortsetzung der PIAAC-L-Befragung als Sample N enthält.</p>	<p><b>2. New samples in the main SOEP study</b></p> <p>The new SOEP data release (v34) will be the first to contain data from the IAB-BAMF-SOEP Survey of Refugees in Germany as Sample M5, as well as the continuation of the PIAAC-L Survey, as Sample N.</p>
<p><b>2.1 IAB-BAMF-SOEP Befragung von Geflüchteten (M5)</b></p> <p>In Zusammenarbeit mit dem Institut für Arbeitsmarkt- und Berufsforschung (IAB) und dem Bundesamt für Migration und Flüchtlinge (BAMF) ist es dem SOEP gelungen, eine dritte Stichprobe von Haushalten Geflüchteter (M5) in die SOEP-Studie zu integrieren. Diese Stichprobe wurde 2017 erstmals befragt. Die Grundgesamtheit von M5 besteht aus erwachsenen Geflüchteten, die seit dem 1. Januar 2013 in Deutschland Asyl beantragt haben und derzeit in Deutschland leben. SOEP-Core wurde mit M5 so um 1.519 Haushalte von Geflüchteten, die seit 2013 nach Deutschland gekommen sind, erweitert.</p>	<p><b>2.1. IAB-BAMF-SOEP Survey of Refugees (M5)</b></p> <p>The SOEP, in cooperation with the Institute for Employment Research (IAB) and the Federal Office for Migration and Refugees (BAMF), has succeeded in integrating a third sample of refugee households (M5) into the SOEP study. The survey was launched in 2017. The population of M5 covers adult refugees who have applied for asylum in Germany since January 1, 2013, and are currently living in Germany. M5 added another 1,519 households of refugees who have migrated to Germany since 2013 to the SOEP framework.</p>
<p><b>2.2. Integration der Befragungspersonen aus PIAAC-L als Teilstichprobe N</b></p> <p>2017 wurden 2.314 Haushalte ehemaliger TeilnehmerInnen der Studie „Program for the International Assessment of Adult Competencies“ (PIAAC und PIAAC-L) erstmals als SOEP-Haushalte befragt und sind jetzt als Stichprobe N integriert. Dies ist die jüngste Ergänzung zu den SOEP-Core-Samples. Die Feldarbeit für Sample N wurde zwischen Mitte März und Mitte August und damit etwas später als für die Samples A-L1 durchgeführt. Weitere Informationen zum PIAAC-L-Projekt finden Sie auf der <a href="#">Projekthomepage</a>.</p>	<p><b>2.2. Integration of respondents from PIAAC-L as Subsample N</b></p> <p>Sample N integrated 2,314 households of former participants of the “Program for the International Assessment of Adult Competencies” (PIAAC and PIAAC-L) in 2017. This is the most recent addition to the SOEP-Core samples. Fieldwork in sample N was conducted between mid-March and mid-August and thus slightly later than the majority of samples A–L1. More information on the PIAAC-L project can be found on the <a href="#">project homepage</a>.</p>

### **3. Übersetzungsfehler in einigen Fragebogensprachen**

In der IAB-BAMF-SOEP-Befragung Geflüchteter (M3-M5) gab es in einigen Fragen zu Einkommenskomponenten in übersetzten Versionen des Haushaltsfragebogens Übersetzungsfehler. Die Antworten für diese Variablen sind daher nicht mit anderen Antworten vergleichbar. Die entsprechenden Variablen wurden auf -3 gesetzt.

### **3. Translation errors in some questionnaire languages**

In the IAB-BAMF-SOEP Survey of Refugees (M3-M5), there were translation errors in some of the questions on income components in translated versions of the household questionnaire. Answers for these variables are therefore not comparable with other answers. The corresponding variables were set to -3.

#### **4. Löschung von Interviews, die nicht nach den Standards der IAB-BAMF-SOEP-Gruppe bei der IAB-BAMF-SOEP-Befragung von Geflüchteten (M3/M4) durchgeführt wurden.**

Im Rahmen der Datenaufbereitung wurden drei InterviewerInnen identifiziert, die die Erhebung nicht nach den Standards der IAB-BAMF-SOEP-Gruppe durchgeführt haben (weitere Informationen [hier](#)). Die identifizierten InterviewerInnen waren 2016 für 88 Haushalte und 2017 für 112 Haushalte verantwortlich. Die in der ersten Welle der Umfrage (2016) betroffenen Haushalte wurden vollständig aus dem Datensatz entfernt. Die 2017 betroffenen Haushalte, die zum zweiten Mal befragt werden sollten, wurden für 2017 gelöscht, aber im Datensatz für 2016 belassen. Es gibt keine Hinweise darauf, dass die ersten Interviews (die von einer anderen Person durchgeführt worden waren) nicht nach den Standards des IAB-BAMF-SOEP durchgeführt wurden. Die aus der Datenweitergabe gelöschten Interviews und Fälle können auf Anfrage an einem Gastarbeitsplatz im SOEP-FDZ für surveymethodologische Analysen zur Verfügung gestellt werden. Nach den vorgenommenen Löschungen aus allen Datensätzen wurden folgende Anpassungen vorgenommen:

- Die Löschung der Haushalts- und Einzelinterviews erforderte eine Aktualisierung der Gewichte (Datensatz HHRF und PHRF), die nun die leicht reduzierten Fallzahlen in den Erhebungsjahren 2016 und 2017 berücksichtigen.
- Aktualisierung / Aufnahme der neuen Gewichte in die Datensätze BGPEQUIV und BHPEQUIV.

#### **4. Deletion of interviews not conducted in line with the standards of the IAB-BAMF-SOEP group in the IAB-BAMF-SOEP Survey of Refugees (M3/M4)**

In the process of data preparation, three interviewers were identified who had not conducted interviews in line with the standards of the IAB-BAMF-SOEP group (more information [here](#)). The interviewers in question were responsible for 88 households in 2016 and 112 households in 2017. The households affected in the first wave of the survey (2016) were completely removed from the dataset. The households affected in 2017, who were supposed to be interviewed for the second time, were deleted for 2017 but left in the dataset for 2016. There are no indications that the first interviews (by a different interviewer) were not conducted in line with IAB-BAMF-SOEP standards. The interviews and cases deleted from the data release may be accessed upon request from a guest work station at the SOEP-RDC for survey methodological analysis.

After these lines were deleted from all datasets, the following adjustments were made:

- The deletion of the household and individual interviews required an update of the weights (dataset HHRF and PHRF), which now take account of the slightly reduced case numbers in survey years 2016 and 2017.
- Update / inclusion of the new weights in the datasets BGPEQUIV and BHPEQUIV.

## 5. Erweiterte Namenskonvention für Variablen

Die erweiterte Namenskonvention für Variablen wird nur auf Datensätze ab Welle BH angewendet und gilt nur für die Datensätze \$P, \$H, \$H, \$KIND. Wir haben jeweils Unterstriche zwischen den Identifikatoren für die Analyseeinheit, die Frage und das Item hinzugefügt, um Analyseeinheit, Frage und Item visuell klar zu trennen. Zusätzlich wurde eine Fragebogenkennung eingeführt, die ebenfalls durch einen Unterstrich vom Item getrennt ist. Diese neue Version der Namensvariablen wird nur verwendet, wenn sich der Fragebogen vom "ursprünglichen" SOEP-Core-Fragebogen unterscheidet.

Wegen der unterschiedlichen Stichproben im SOEP gibt es einige Befragte, die stichprobenspezifische Fragen erhalten, wie z.B. in der Stichprobe von Geflüchteten, die 2016 begann. Für diese spezielle Gruppe gibt es einen erweiterten Personenfragebogen mit einigen spezifischen Fragen, die über die Standardfragen des SOEP, die jedes Jahr gestellt werden, hinausgehen. Für die spezifischen Fragen können Sie die Instrumentvariable (INSTRUMENT) verwenden, um die Quelle der Variablen anzuzeigen.

Beispiele und detailliertere Beschreibungen finden [Sie im Kapitel zu diesem Thema im SOEPcompanion](#).

## 5. Extended variable naming convention

The extended variable naming convention is applied only to data sets from wave BH onwards and only applicable for the datasets \$P, \$H, \$KIND. We added underscores between unit of analysis, question identifier, and item identifier to clearly separate the analysis unit, question, and item visually. In addition, a questionnaire identifier was introduced, which is also separated by an underscore from the item. This new version of naming variables is only used if the survey instrument differs from the "original" SOEP-Core instrument.

Due to our different samples in the SOEP, there are some respondents that receive sample-specific questions, such as the refugee sample that started in 2016. For that specific group, we created an extended individual questionnaire with some specific questions along with the standard SOEP questions that are asked every year. For the specific questions, you can use the instrument variable to see the source of the variables.

Examples and more detailed descriptions can be found in the [chapter on this subject in the SOEPcompanion](#).



<p><b>6. Änderungen bestimmter Variablen</b></p> <p>Neue Variablen für das Interviewjahr: HIYEAR in HGEN, HPATHL und PIYEAR in PGEN, PPATHL. Diese neuen Variablen geben für alle Erhebungsjahre die Haushalts- und Einzelinterviews an, die nach (oder vor) dem Erhebungsjahr abgeschlossen wurden (Variable SYEAR), welches das Bezugsjahr für die Fragebögen und die Datenerhebung ist.</p>	<p><b>6. Changes in specific variables</b></p> <p>New variables for interview year: HIYEAR in HGEN, HPATHL, and PIYEAR in PGEN, PPATHL. These new variables indicate, for all survey years, the household and individual interviews that were finalized after (or before) the survey year (variable SYEAR), which is the reference year for the questionnaires and for data collection.</p>
<p><b>6.1. Datensatz PPATH/PPATHL (in raw: PPFAD)</b></p>	<p><b>6.1. Dataset PPATH/PPATHL (in raw: PPFAD)</b></p>
<p><b>6.1.1. SEXOR</b></p> <p>Die vorherige Datenlieferung war die erste, die die Variablen SEXOR (sexuelle Orientierung) und SEXORINFO (Informationsquelle zur sexuellen Orientierung) enthielt. Der Wert -1 "unzureichende Informationen" wurde in 2 "unzureichende Informationen" geändert.</p>	<p><b>6.1.1. SEXOR</b></p> <p>The previous data release was the first to include the variables SEXOR (sexual orientation) and SEXORINFO (source of information on sexual orientation). The value -1 "insufficient information" has been changed to 2 "insufficient information".</p>
<p><b>6.1.2. PARINFO</b></p> <p>Der Wert -1 "unklar" wurde in 5 "unklar" geändert.</p>	<p><b>6.1.2. PARINFO</b></p> <p>The value -1 "unclear" has been changed to 5 "unclear".</p>
<p><b>6.1.3. Informationen zu Migration</b></p> <p>Die Codierung von GERMBORN, CORIGIN, IMMIYEAR und MIGBACK wurde auf Grund von inkonsistenten Fällen geändert (mehr Informationen erhalten Sie in der PPATH/PPFAD Dokumentation).</p>	<p><b>6.1.3. Migration information</b></p> <p>The coding of GERMBORN, CORIGIN, IMMIYEAR and MIGBACK was changed for inconsistent cases (for more information, see the PPATH/PPFAD documentation).</p>
<p><b>6.1.4. Asylsuchende und Geflüchtete</b></p> <p>Die Variablen für Asylsuchende und Geflüchtete [AREBACK, AREFINFO] wurden umbenannt (in v33: REFBACK, REFINFO) und überarbeitet. Die Variable AREFINFO erlaubt nun auch die Identifikation von bestimmten Untergruppen (mehr Informationen dazu finden Sie in der Dokumentation)</p>	<p><b>6.1.4. Asylum-Seekers and Refugees</b></p> <p>The variables for asylum-seekers and refugees [AREBACK, AREFINFO] have been renamed (in v33: REFBACK, REFINFO) and revised. The variable AREFINFO now also allows identification of specific subgroups (more information is available in the documentation).</p>

## 6.2. Datensatz PGEN

### 6.2.1. Partnerzeiger

- Für die Variable PGPARTZ (PARTZ\$), wurde der Wert -1 (“keine Angabe”) durch den korrekten Wert 5 (“unklar”) ersetzt
- Ab Welle BH wurden neue Prüfprozesse für die Generierung der Partnerzeiger implementiert, die die Qualität der Daten auch für vergangene Wellen rückwirkend verbessern:
- widersprüchliche Partnerschaftsangaben zwischen zwei Partnern werden identifiziert und behandelt
- Partnerschaften mit unterschiedlichen Partnerzeigern (1 „Ehepartner“ oder 2 „Lebenspartner“) innerhalb einer Partnerschaft werden identifiziert und behandelt.
- Fehler beim Vergeben der Werte für PARTZ (1 „Ehepartner“ und 2 „Lebenspartner“) durch unterschiedliche Filterführung der verschiedenen Erhebungsinstrumente wurde korrigiert. Beim Personenbiofragebogen der Samples J+K und beim Personenfragebogen der Samples A-I wurden Ehen unterschiedlich erhoben. Entscheidend für die Korrektur war hierbei das Ausgliedern der Samples J und K, da diese durch eine andere Filterführung Fehler hervorgerufen haben.
- Partnerschaften mit kürzlich verstorbenen Personen wurden identifiziert und aufgelöst.
- Angaben der Befragten zu Scheidung, Trennung oder dem Tod des Lebenspartners/ der Lebenspartnerin innerhalb des letzten Jahres wurden im Generierungsprozess erstmalig berücksichtigt.
- Für die Partnerschaften von Geflüchteten (Samples M3-M5) wurden erstmals zusätzlich Familienstatus, Partnerstatus, Vorname des Partners/der Partnerin (laufende Partnernummer) und Aufenthaltsort des Partners/ der Partnerin berücksichtigt. Im Vorjahr wurden die Partnerschaften nur über die Interviewerangaben (stell Variable) identifiziert.

## 6.2. Dataset PGEN

### 6.2.1. Partner pointer

- For the variable PGPARTZ (PARTZ\$), the value -1 (“no answer”) has been replaced by the correct value 5 (“unclear”).
- Starting with wave BH, the new quality control processes implemented in generating the partner indicator have improved the quality of data from previous waves:
- Contradictory answers between partners regarding their relationship have been identified and corrected.
- Partnerships with differing partner indicators (1 “spouse” or 2 “life partner”) within a relationship have been identified and corrected.
- Errors in the assignment of PARTZ values (1 “spouse” and 2 “life partner”) due to different filter routing in the different survey instruments have been corrected. Marriages were asked differently in the individual biography questionnaire for Sample J+K and in the individual questionnaire for Samples A-I. Separating out samples J and K played a key role in this correction, since these two led to errors due to their different filter routing.
- Partnerships with recently deceased individuals were identified and deleted.
- Respondents’ data on divorce, separation, or the death of a life partner within the past year have been taken into account for the first time in the process of generating the data.
- For the first time, family status, civil status, partner’s first name (permanent partner number) and place of residence of the partner in the case of refugees’ partnerships (Samples M3-M5) have been taken into account in addition to interviewer given relationships between the different household members.

### 6.2.2. Freiwilligendienst und Nebenjob

Die PGEN-(raw: \$\$PGEN-) Datensätze enthalten neun neue Variablen. Im Jahr 2017 hat das SOEP die Befragung von Nebenjobs grundsätzlich überarbeitet. Jetzt können die Befragten zum ersten Mal Antworten auf drei verschiedene Nebenjobs geben. Sie können nun auch die Art der Nebentätigkeit unterscheiden, ob sie ehrenamtlich tätig sind (Variablen HONOR1, HONOR2, HONOR3) und ob sie für einen Arbeitgeber oder freiberuflich tätig sind. (SNDTYP1, SNDTYP2, SNDTYP3). Die Beträge der zusätzlichen Bruttoeinkommen aus Nebenjobs werden als kalkulatorische Informationen bereitgestellt (SNDJOB1, SNDJOB2, SNDJOB3).

**SNDTYP117 : Erste Nebentätigkeit Berufliche Stellung**

**SNDTYP217 : Zweite Nebentätigkeit Berufliche Stellung**

**SNDTYP317 : Dritte Nebentätigkeit Berufliche Stellung**

**SNDJOB117 : Akt. Bruttoverdienst Nebentätigkeit 1 (gen.) in Euro**

**SNDJOB217 : Akt. Bruttoverdienst Nebentätigkeit 2 (gen.) in Euro**

**SNDJOB317 : Akt. Bruttoverdienst Nebentätigkeit 3 (gen.) in Euro**

**HONOR117 : Ehrenamtliche Tätigkeit 1**

**HONOR217 : Ehrenamtliche Tätigkeit 2**

**HONOR317 : Ehrenamtliche Tätigkeit 3**

### 6.2.2. Volunteer work and side jobs

The PGEN (raw: \$\$PGEN) files contain nine new variables. In 2017, the SOEP fundamentally revised how respondents were surveyed about side jobs. Now, for the first time, respondents can provide answers on three different side jobs. They can also now differentiate the type of side job, whether volunteer work (variables HONOR1, HONOR2, HONOR3) and whether they are working for an employer or working freelance (SNDTYP1, SNDTYP2, SNDTYP3). The amount of gross additional income from side jobs is provided as imputed information (SNDJOB1, SNDJOB2, SNDJOB3).

**SNDTYP117 : First side job occupational status**

**SNDTYP217 : Second side job occupational status**

**SNDTYP317 : Third side job occupational status**

**SNDJOB117 : Current gross additional income from side job 1 (gen.)  
in euros**

**SNDJOB217 : Current gross additional income from side job 2 (gen.)  
in euros**

**SNDJOB317 : Current gross additional income from side job 3 (gen.)  
in euros**

**HONOR117 : Volunteer work 1**

**HONOR217 : Volunteer work 2**

**HONOR317 : Volunteer work 3**

### 6.2.3. Bildungsabschlüsse

- In v34 basieren CASMIN und ISCED auf zusätzlichen Informationen über im Ausland erworbene Bildungsabschlüsse. Daher weisen einige Personen mit ausländischen Abschlüssen in v33 höhere Ränge auf als in v34.
- Ein Fehler in der Variablen CASMIN in v33 wurde behoben: In v33 wurden Personen mit 2c\_voc (Berufsreifezeugnis) fälschlicherweise als 2c\_gen (allgemeines Reifezeugnis) eingestuft.

### 6.2.3. Educational degrees

- In v34, CASMIN and ISCED are based on additional information on educational degrees obtained abroad. Hence, some individuals with degrees from abroad display higher ranks in v33 than in v34.
- The error in the CASMIN variable in v33 is fixed: In v33, individuals with 2c\_voc (vocational maturity certificate) were mistakenly categorized as 2c\_gen (general maturity certificate).

<p><b>6.2.4. AUTONO</b></p> <ul style="list-style-type: none"> <li>Die Variable AUTONO wurde 2017 nicht mehr generiert, da es schwierig ist, diese Variable mit den üblichen Modellen der Autonomie zu vergleichen. Derzeit wird daran gearbeitet, vergleichbare Definitionen von Autonomie zu erstellen.</li> </ul>	<p><b>6.2.4. AUTONO</b></p> <ul style="list-style-type: none"> <li>The generation of AUTONO was discontinued in 2017 due to the difficulty in comparing this variable with the usual models of autonomy. Work is currently underway to introduce comparable definitions of autonomy.</li> </ul>
<p><b>6.3. Dataset PEQUIV</b></p> <p>Der PEQUIV-(raw: \$\$PEQUIV) Datensatz enthält 6 neue Variablen. Diese sind:</p> <p><b>IAUS117</b> : Altersversorgung aus einem anderem Land  <b>IAUS217</b> : Witwen-/Waisenrente aus einem anderen Land  <b>ASYL17</b> : Leistung nach Asylbewerberleistungsgesetz  <b>FASYL17</b> : Imputation flag: Leistung nach Asylbewerberleistungsgesetz  <b>EDUPAC17</b> : Leistungen aus dem Bildungspaket  <b>FEDUPAC17</b> : Imputation flag: Leistungen aus dem Bildungspaket</p> <p>Weitere Informationen im SOEP Survey Paper <a href="#">Codebook for the \$PEQUIV File 1984-2017</a></p>	<p><b>6.3. Dataset PEQUIV</b></p> <p>The PEQUIV (raw: \$\$PEQUIV) files contain six new variables. These are:</p> <p><b>IAUS117</b> : Pensions from another country  <b>IAUS217</b> : Widows / orphans pension from another country  <b>ASYL17</b> : Asylum-seeker benefit  <b>FASYL17</b> : Imputation flag: Asylum-seeker benefit  <b>EDUPAC17</b> : Benefits from the educational package  <b>FEDUPAC17</b> : Imputation flag: Benefits from the educational package</p> <p>For more details, see the SOEP Survey Paper: <a href="#">Codebook for the \$PEQUIV File 1984-2017</a>.</p>
<p><b>6.4. Datensätze BIOAGEL und BIOPUPIL</b></p> <ul style="list-style-type: none"> <li>Die Variablen aus den Fragebögen für 12- bzw. 14-Jährige werden nun im BIOPUPIL-Datensatz bereitgestellt, um die Unterschiede im Erhebungsmodus widerzuspiegeln (Eltern werden zu ihren Kindern befragt vs. Kinder werden direkt befragt).</li> <li>Variablen aus zusätzlichen Fragen bei den Geflüchteten-Stichproben (M3-M5) wurden in BIOAGEL und BIOPUPIL integriert.</li> </ul>	<p><b>6.4. Dataset BIOAGEL and BIOPUPIL</b></p> <ul style="list-style-type: none"> <li>Variables from questionnaires given to 12-year-olds and 14-year-olds are now provided in BIOPUPIL dataset to reflect the differences in survey mode (parents being asked questions about their children vs. children being surveyed directly).</li> <li>variables from additional questions in refugee samples are integrated in BIOAGEL and BIOPUPIL datasets.</li> </ul>

### **6.5. Datensatz HGEN**

- In den letzten Jahren wurden einige Änderungen bei den Fragen zur Wohnungsmiete vorgenommen. Die erste Änderung fand im Haushaltsfragebogen der Welle BF (2014) statt. Die Frage nach den Mietnebenkosten wurde so detailliert gestellt, dass die Befragten nicht in der Lage waren, korrekte Antworten zu geben. Dies führte zu einer Unterschätzung sowohl der Grundmiete als auch der Nebenkosten. Es zeigte sich, dass dies zu einem leichten Bruch der Zeitreihe führte. Die Miete ist im Laufe der Jahre seit 1984 kontinuierlich gestiegen. In den Jahren 2014 und 2015 sanken die Mietkosten jedoch und sind seit 2016 wieder stark gestiegen. Dieser Bruch lässt sich durch die Änderung des Fragebogens erklären.

Ab Welle BH werden die Befragten wieder auf die gleiche Weise wie in Welle BG (2016) und in Welle BD (2013) nach der Miete befragt, um die langfristige Vergleichbarkeit zu gewährleisten (siehe Tabelle Miete).

- Darüber hinaus sind mit der Welle BH die neue Migrationsstichprobe M5 und die neue Auffrischungsstichprobe N Teil des SOEP. Da die Stichprobe M5 nicht in vergleichbarer Weise zu den Nebenkosten befragt wurde und viele dieser Befragten wahrscheinlich in Gruppenwohnungen leben oder Zuschüsse zur Deckung der Lebenshaltungskosten erhalten, wurde keine Mietvariable für diese Stichprobe generiert.

### **6.5. Dataset HGEN**

- A number of changes have taken place in recent years in questions on home rental. The first change took place in the household questionnaire of wave BF (2014). The question asked about the costs of utilities in such detail that respondents were not able to provide correct answers. This led to underestimation of both base rent and utilities. It emerged that this led to a slight break in the time series. Rent has increased continuously over the years since 1984. In 2014 and 2015, however, rental costs fell and have been increasing again sharply since 2016. This break can be explained by the change in the questionnaire.

Starting with wave BH, respondents are being asked about rent in the same way as in wave BG (2016) and in wave BD (2013) in order to maintain long-term comparability (see table Rent).

- In addition, with wave BH, the new migration sample M5 and the new refresher sample N are part of the SOEP. Since Sample M5 was not surveyed on utility costs in a comparable way and since many of these respondents probably live in group housing or receive subsidies to cover living costs, no rent variable was generated for them.

### **6.6. Datensatz BIOIMMIG**

- Die Population ist auf Grund einer Kodierung in BIIMGPR gesunken (Nähere Informationen in der Dokumentation BIOIMMIG)

### **6.6. Dataset BIOIMMIG**

- The population of BIOIMMIG shrunk due to a change of coding of BIIMGPR (for more information, see the BIOIMMIG documentation)

## 6.7. Datensatz HHRF/PHRF

- Neue Variablen in HHRF: BHHHRF, BHHBLEIB, BHHHRFAM4, BHHHRFM5, BHHHRFN
- Neue Variablen in PHRF (und ENUMHRF, erhältlich auf Nachfrage): BHPHRF, BHPBLEIB, BHPHRFAM4, BHPHRFM5, BHPHRFN
- Bitte beachten Sie, dass Sie mit unserem neuen integrierten Datenformat alle GewichtungsvARIABLEN jetzt direkt in PPATHL oder HPATHL finden.
- Auf Nachfrage stellen wir gern GewichtungsvARIABLEN für die getrennte Gewichtung der Geflüchtetenstichproben M3, M4 und M5 zur Verfügung (BHPHRFM35, BHHHRFM35).

## 6.7. Dataset HHRF/PHRF

- New variables in HHRF: BHHHRF, BHHBLEIB, BHHHRFAM4, BHHHRFM5, BHHHRFN
- New variables in PHRF (and ENUMHRF, available on request): BHPHRF, BHPBLEIB, BHPHRFAM4, BHPHRFM5, BHPHRFN
- Please note that with our new integrated data format, you'll find all weighting variables now directly in PPATHL or HPATHL.
- On request, we provide stand-alone weighting variables (BHPHRFM35, BHHHRFM35) for the refugee samples M3, M4, and M5.

### 6.7.1. Revisionen und Fehlerbehebungen

- Wegen der Verwechslung der Ländercodes für Iran und Russland im Ziehungsrahmen (Zentralregister der Ausländer, AZR) mussten die Designgewichte für die Samples M3 und M4 sowie deren Querschnittsgewichte für die Welle BG aktualisiert werden.
- In Welle BG hatten wir als Grundgesamtheit der Stichproben M3 und M4 die Geflüchteten angesehen, die zwischen Januar 2013 und Januar 2016 in Deutschland angekommen sind. Tatsächlich sind nur die Geflüchteten, deren Eintragung in das Zentrale Ausländerregister (AZR) bis April 2016 erfolgte, in diese Stichproben aufgenommen worden. In Stichprobe M5 wurden unter anderem diejenigen Geflüchteten befragt, die zwar im gleichen Zeitraum zugewandert waren, aber später registriert wurden. Aus diesem Grund wurde die Summe für die Nachschichtung der zweiten Welle von M3 und M4 um die Anzahl der Geflüchteten mit einem späteren Registrierungsdatum reduziert.

### 6.7.1. Revisions and Bugfixes

- Due to confusion in the country codes for Iran and Russia in the sampling frame (Central Register of Foreign Nationals, AZR), design weights for Samples M3 and M4 as well as their cross-sectional weights for wave BG had to be updated.
- In Wave BG, we interpreted the population of samples M3 and M4 as refugees who immigrated to Germany between January 2013 and January 2016. In fact, only those refugees whose registration at the Central Register of Foreign Nationals (AZR) took place until April 2016 were included in those samples. In Sample M5, among others, those refugees were interviewed who, although they had immigrated in the same period, were registered later. For this reason, the total for the post-stratification of the second wave of M3 and M4 has been reduced by the number of refugees with a later registration date.