

To the SOEP users

(English version of German letter) May 30, 2011| Page 1/6

The revised SOEP variables that include annual household income, which were provided to our users in October 2009 (in particular, the variable I11102\$\$), recently made headlines in Germany: The revision of the variables had led to a substantial reduction in the poverty rate for households with children. In the following, we summarize the media discussion and the revision of the SOEP data.

It is very important that all SOEP be assured of the following:

This revision *only* affects analyses the using generated variables for annual household income found in the file \$PEQUIV. It *does not affect* any other income data—e.g., the variables on monthly net household income (the “income screener” in the \$HGE files). It also *does not affect* analyses of specific income components (e.g., gross and net wages in the \$P files).

What started the discussion?

The *Financial Times Deutschland* published a cover story on May 8, 2011, that attracted a great deal of media attention and sparked a public debate on the quality of the SOEP income data, or more specifically, on the quality of the indicators of child poverty in Germany published by the OECD. Here, it is important to distinguish between two issues:



Sozio-oekonomisches Panel
Prof. Dr. Jürgen Schupp
Leitung

T +49 30 897 89 -238
F +49 30 897 89 -9238
jschupp@diw.de

**DIW Berlin – Deutsches Institut
für Wirtschaftsforschung e.V.**
Mohrenstraße 58, 10117 Berlin
Postanschrift:
DIW Berlin, 10108 Berlin
T +49 30 897 89 -0
F +49 30 897 89 -200
www.diw.de

Vorstand
Dr. Cornelius Richter
Prof. Dr. Gert G. Wagner (*Vorsitzender*)
Prof. Georg Weizsäcker, Ph.D.

Kuratorium
Prof. Dr. Dr. h. c. Bert Rürup (*Vorsitzender*)

Rechtsform
Eingetragener Verein
Sitz in Berlin
Amtsgericht Charlottenburg
95 VR 136 NZ

USt-IdNr. DE 136622485
Steuer-Nr. 27 640 50519

Bankverbindungen
Berliner Volksbank eG
Bankleitzahl 100 900 00
Kontonummer 8 848 124 002
BIC (SWIFT-Code) BEVODEBBXXX
IBAN DE17 1009 0000 8848 1240 02

Berliner Bank AG
Bankleitzahl 100 708 48
Kontonummer 512 366 600
BIC (SWIFT-Code) DEUTDEDB110
IBAN DE28 1007 0848 0512 3666 00

- The first issue is the measurement of “annual household income” in the SOEP and the improved variable included in the revised SOEP data released in 2009 (see also *SOEPNewsletter* 86, October 2009, p. 8). As stated above, the only users affected by this revision, or improvement, of the data are those using the household income variables generated by the SOEP group from the \$PEQUIV files.
- The second issue is the publication of international comparative indicators by the OECD. We will start by addressing this point, which triggered the public discussion.

The 2011 OECD publication using poverty risk indicators from 2009

For many years now, the SOEP group has been providing the OECD with time series for selected indicators as a free service. These include poverty risk indicators differentiated by socio-demographic characteristics. The poverty rates that were the subject of the recent public discussion were affected particularly strongly by a revision of the SOEP data; the revision was released to SOEP users in November 2009. The OECD only recently took this revision (and the improved weighting factors, which were also released in November 2009) into account in its internationally comparative statistics. The revised OECD statistics were first published in April 2011. As is usually the case, the OECD did not specifically point out revisions in the data, which occur on a regular basis in the data for numerous countries. As a result, careful readers of the OECD report *Doing better for families* were perplexed by the fact that the revision of the SOEP data resulted in a substantial downward revision of the poverty rate for families with children (according to the definitions by OECD: from 16.3 % for 2004 [*pre-revision*] to 10.0 % [*post-revision*] and 8.3% for the year 2008 [*post-revision*]).

In general, the OECD publishes the data provided to them at a substantial delay, since—understandably—they often have to wait for data producers who deliver late. It is also common and understandable that the OECD does not inform the data providers about the precise date of forthcoming publications.

Since the OECD's publication procedures can easily lead to misunderstandings (as the recent case has shown), we have begun working with the Berlin office of the OECD to develop ideas for improving communication between DIW Berlin and the OECD.

On the necessity of the 2009 SOEP revision

Why was this revision needed? Since 2000, the percentage of non-respondents within surveyed SOEP households has risen to 5% of all adult respondents. This is causing an increasing problem for the collection of data on annual income, since these individuals are also income-earners. At this stage, this group of non-respondents is too large to ignore in the calculation of annual household income (especially the variable I11102\$\$). The procedure that was previously used in the SOEP was long the norm in other household panel studies as well (e.g., BHPS). One reason for the higher percentage of partially-interviewed households is that the percentage of non-respondents in the new SOEP sample F, which started in 2000, was higher than in all preceding samples.

Because of this surveying problem (which does not affect the monthly "income screener" for household income and personal income), we recalculated seven income variables in the \$PEQUIV files and imputed the income for specific respondents in survey households who had refused to participate. We then added up all of the personal incomes in each household and, as we have done for many years, used a microsimulation model to determine the amount of taxes and social contributions paid by each household.

Missing income components for individuals who do not respond to the survey are imputed from the SOEP's usual longitudinal perspective, a procedure that is discussed extensively at international conferences. A description of the incidence of partial unit non-response and the methods used to correct for this can be found in: Joachim R. Frick, Markus M. Grabka, and Olaf Groh-Samberg, "Dealing with Incomplete Household Panel Data in Inequality Research," forthcoming in *Sociological Methods and Research* (SMR) (see also *SOEPpaper* No. 290, Berlin 2010).

SOEP data users regularly informed about revisions

These updated variables—as well as the improved weighting factors—were already provided by SOEP one year ago with the October 2009 data release. We reported on the new and revised variables and the improved weighting procedures and documentation in *SOEP Newsletter* No. 86, p. 8, and on our website:

http://www.diw.de/de/diw_02.c.238122.de/aenderungen_am_datensatz.html.

It is important to note that the improved data generation process did not cause the temporal evolution of income inequality to change. However, it did result in a reduction in income inequality and in all poverty rates, which was discussed at length in various publications (see, e.g., *Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung: Jahresgutachten 2009/10*, *DIW Wochenbericht* No. 51-52/2009, *DIW Wochenbericht* 7/2010). What we underestimated was the general problem posed by the country rankings included in the OECD reports. The rankings are based on sample data that are provided by the individual countries and revised as a matter of course. Here, the revision of child poverty caused Germany to move up significantly in the rankings. It is clear that better information could have been provided to the public by those involved—both the DIW Berlin and the OECD (see below “A few lessons for the future”).

For users of the SOEP microdata, it is important that only those using the annual income variables generated by the SOEP group from the \$PEQUIV files are affected by the revision of the data. Those using the directly surveyed current household net income (“income screener” I1HINC\$\$, HINC\$\$, AHINC\$\$ from the \$HGEN files) are not affected. (We would like to point out, however, that since the 2008 data release, one variable, I_HINC\$\$, has been generated for the “income screener” that does not contain missing values because the income screener has been imputed since that release.)

Not all distribution analyses are affected by the revision

When the income screener is used, by definition, the challenges of dealing appropriately with partially interviewed households do not arise. All analyses relating to personal income are also not affected.

Also, the microsimulation models that are widely used for income and distribution analyses are not affected by the revision described above, since only completely interviewed households are used for the usual microsimulation models.

Data revisions are not “snafus”

In DIW press releases, we have emphatically rejected press accusations that the DIW released erroneous data and thus provided the foundation for “misguided policy measures.”¹ To the contrary: our revision of the data utilized new scientific findings on the imputation of missing values and more efficient estimation procedures to achieve an improvement in the data set. The revision of old data is a positive development and not a “snafu.”

A few lessons for the future

We, as the SOEP group, have drawn a number of lessons from this debate. In the future, when it comes to larger revisions of our imputation procedures, but also of our weighting and estimation procedures, we plan to provide more comprehensive information to both SOEP users and the general public, using key socio-demographic indicators to show the concrete effects of the revisions.

Here, we will pay special attention to reaching users who do not work directly with the SOEP microdata but who use the tables based on SOEP data that are provided by DIW Berlin or other organizations and included in other studies or statistics (such as those of the OECD).

¹ See: http://www.diw.de/sixcms/detail.php?id=diw_01.c.372305.de (in German) and

http://www.diw.de/en/diw_01.c.100319.en/press/press_releases/press_releases.html?id=diw_01.c.372865.en

These activities are integral to the approach to revisions generally used by official statistical agencies, and the SOEP plans to make use of their experiences in the future.



We also plan not only to announce our annual data release to the scientific community but also to make this information available to the media. We also intend to highlight the results published in the SOEP Monitor with each new wave. These include, along with various weighted analyses based on income indicators, a range of key welfare indicators (labor market, health, etc.). These also make it possible to clearly identify impacts of improvements and revisions of the data.

Upcoming issues of the SOEP Monitor will not only present the new estimates of the marginal distributions, but also any differences between the new data and those from the previous year.

The use of a reference code for the specific version of the SOEP data (we are currently working on SOEP v26), as introduced with the last data release, will dovetail well with this increased transparency. Future versions of the SOEP data will also be clearly identifiable with a DOI number.

The 2011 Census – A reason for SOEP to revise weighting and estimation procedures

Please note: the population census currently underway in Germany will result in a fundamental revision of the SOEP weighting factors. The revision will probably affect the data release in two or three years. The results of the 2011 census will almost certainly lead to substantial revisions of the regional and socio-demographic distribution of the SOEP sample, as is the case with all other individual and household samples.

A handwritten signature in black ink, appearing to read 'Jürgen Schupp', is written in a cursive style.

Jürgen Schupp
Head of the SOEP Research Infrastructure at DIW Berlin