

DIW Berlin, 10108 Berlin

An die Nutzerinnen und Nutzer des SOEP

Berlin, den 30. Mai 2011 | Seite 1/7

Die ab Oktober 2009 an die Nutzerinnen und Nutzer ausgelieferte Revision der Variablen mit den jährlichen Haushalts-Einkommen (insbesondere Variable I11102\$\$) hat kürzlich Schlagzeilen gemacht, da sich durch die Revision der Variablen die Armutsquote für Haushalte mit Kindern deutlich nach unten verändert. Wir erläutern hier die Zeitungs-Berichterstattung und die Datenrevision.

Wichtig für alle Nutzerinnen und Nutzer ist:

betroffen sind *ausschließlich* Analysen, die die generierten Variablen für das jährliche Haushalts-Einkommen verwenden, die im File \$PEQUIV zu finden sind. *Nicht betroffen* sind alle anderen Einkommens-Angaben; z. B. die Variablen für das monatliche Netto-Haushaltseinkommen (die „Income Screener“ in den \$HGEN-Files). Auch Analysen einzelner Einkommensbestandteile (z. B. die des Brutto- und Netto-Arbeitseinkommens (in den \$P -Files) sind *nicht betroffen*.

Ausgangspunkt Eine Titelgeschichte der *Financial Times Deutschland* vom 8.Mai 2011 sorgte in den Medien für beachtliche Aufmerksamkeit und bescherte der SOEP-Gruppe eine Debatte um die Qualität ihrer Einkommensdaten bzw. insbesondere um die von der OECD

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) BEVODE33XXX
IBAN DE17 1009 0000 8848 1240 02

Berliner Bank AG

Bankleitzahl 100 708 48
Kontonummer 512 366 600
BIC (SWIFT-Code) DEUTDE33XXX
IBAN DE28 1007 0848 0512 3666 00

Seite 2/7

veröffentlichten Indikatoren zur Kinderarmut in Deutschland. Dabei sind zwei Ebenen zu unterscheiden:

- Erstens geht es um die Messung von „Haushalts-Jahres-einkommen“ und die in einer Revision mit der Datenweitergabe 2009 (siehe auch den SOEP-Newsletter No. 86/October 2009, S. 8) hierzu von der SOEP-Gruppe zur Verfügung gestellte verbesserte Variable. Von der Revision bzw. der Verbesserung der Daten sind – wie gesagt – nur diejenigen NutzerInnen betroffen, die die von der SOEP-Gruppe generierten Haushaltseinkommen aus den \$PEQUIV-Files nutzen.
- Zweitens ist die Bedeutung der Veröffentlichung von international vergleichenden Indikatoren durch die OECD zu bedenken. Auf diesen Punkt, der die öffentliche Diskussion auslöste, soll zuerst eingegangen werden.

Aktuelle OECD-Veröffentlichung 2011 mit Indikatoren zu Armutsrisiken aus 2009

Die SOEP-Gruppe stellt seit vielen Jahren im Rahmen eines kostenlosen Services der OECD Zeitreihen für ausgewählte Indikatoren zur Verfügung. Hierzu zählen auch Indikatoren zu Armutsrisiken differenziert nach sozio-demographischen Merkmalen. Die Armutsquoten, die jetzt im Mittelpunkt der öffentlichen Diskussion standen, waren von einer Revision der SOEP-Daten besonders betroffen; diese Revision wurde freilich bereits ab November 2009 an die NutzerInnen des SOEP ausgeliefert. Die OECD hat diese Revision - sowie die verbesserte Gewichtung, die auch bereits ab November 2009 ausgeliefert wurde - bei ihren internationalen Vergleichsmessungen aber erst jetzt berücksichtigt und im April 2011 erstmals revidierte Zahlen veröffentlicht. Wie bei der OECD üblich wurden Revisionen in den Daten, die für etliche Länder ständig vorkommen, nicht besonders kenntlich gemacht. Dadurch waren gründliche Leserinnen und Leser des OECD-Berichtes „*Doing better for families*“ irritiert, da es durch die Revision der SOEP-Berechnungen zu einer in der Tat sehr deutlichen Revision der Armutsquoten für Familien mit Kindern kam (in der

Seite 3/7

Abgrenzung der OECD: von 16,3 % für das Jahr 2004 [*vor Revision*] auf 10,0 % [*nach Revision*] und 8,3% für das Jahr 2008 [*nach Revision*]).

Wie bei der OECD üblich erfolgt die Veröffentlichung von gelieferten Daten oft mit großer Zeitverzögerung, da verständlicherweise auf NachzüglerInnen bei der Datenanlieferung gewartet werden muss. Ebenso ist es üblich und verständlich, dass die OECD die Daten-Zulieferer nicht über den kommenden Zeitpunkt der Veröffentlichung informiert.

Da es durch die Veröffentlichungs-Routinen der OECD jedoch im Falle von Revisionen leicht zu Missverständnissen kommen kann (wie jetzt ganz offenkundig wurde) haben wir mit dem Berliner Büro der OECD begonnen über eine verbesserte Kommunikation zwischen DIW Berlin und OECD nachzudenken.

Zur Notwendigkeit der SOEP-Revision im Jahr 2009

Was war der Grund für diese Revision? Bei der Erhebung der Jahreseinkommen zeigt sich seit dem Jahr 2000 zunehmend das Problem, dass der Anteil von Personen, die in realisierten SOEP-Haushalten nicht mehr mitmachen, auf ca. 5 % aller Erwachsenen gestiegen ist. Da es sich dabei auch um Einkommensbezieher handelt, kann diese Gruppe bei der Berechnung von Jahreseinkommen auf der Haushaltsebene (insbesondere Variable I11102\$\$) nicht mehr vernachlässigt werden. Das frühere Vorgehen war auch in anderen Haushaltspanelstudien lange Zeit üblich (so auch in der BHPS). Einer der Gründe für den höheren Anteil an lediglich teilrealisierten Haushalten liegt auch darin, dass in der im Jahr 2000 begonnenen neuen SOEP-Stichprobe F der Anteil der „Non-RespondentInnen“ höher lag als in den Altstichproben.

Aufgrund dieses Erhebungsproblems (das den monatsbezogenen „Income Screener“ für den Haushalt und individuelle Einkommen nicht berührt) wurden – rückwirkend von der ersten Welle an – in den \$PEQUIV-Files sieben individuelle Einkommensinformationen neu berechnet und einzelnen Verweigerern in Befragungshaushalten Einkommen zugewiesen („imputiert“). Alle individuellen Einkommen in einem Haushalt werden anschließend aufsummiert und wie seit langem

Seite 4/7

üblich mit Hilfe eines Mikrosimulationsmodells wird die Belastung des Haushalts mit Steuern- und Sozialabgaben bestimmt.

Die Imputation von fehlenden Einkommensbestandteilen bei nicht befragungsbereiten Personen wird in der SOEP-üblichen Längsschnittperspektive durchgeführt und wurde auch auf internationalen Konferenzen intensiv diskutiert. Eine entsprechende Beschreibung der Inzidenz von so genanntem Partial-Unit-Nonresponse und der entsprechende Korrekturmethode findet sich in: Frick, Joachim R.; Markus M. Grabka and Olaf Groh-Samberg: Dealing with Incomplete Household Panel Data in Inequality Research, erscheint in Kürze in Sociological Methods and Research (SMR) (siehe auch: SOEPPaper No. 290, Berlin 2010).

SOEP-DatennutzerInnen schon seit langem über Revisionen informiert

Mit der Datenauslieferung ab Oktober 2009 wurden diese aktualisierten Variablen – wie auch verbesserte Hochrechnungsfaktoren – bereits vor über einem Jahr allen Datennutzern zur Verfügung gestellt. Im SOEP-Newsletter, No 86, S. 8 haben wir auf die neuen und revidierten Variablen und die verbesserte Hochrechnung sowie die Dokumentation hingewiesen und auch auf unserer Internetseite darüber berichtet (http://www.diw.de/de/diw_02.c.238122.de/aenderungen_am_datensatz.html)

Wichtig ist, dass die verbesserte Datengenerierung nicht zu anderen zeitlichen Verläufen der Einkommensungleichheit geführt hat. Allerdings werden die Einkommensungleichheit und alle Armutsquoten niedriger ausgewiesen, was bereits in verschiedenen Publikationen kommuniziert wurde (siehe z.B. Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung: Jahresgutachten 2009/10, DIW Wochenbericht Nr. 51-52/2009 und DIW Wochenbericht 7/2010). Was wir unterschätzt haben: im OECD-Vergleich werden auf Basis von Stichproben-Ergebnissen Rankings von Ländern erstellt. Und im Länder-Ranking hat durch die Revision der Kinder-Armutsquoten Deutschland tatsächlich einen großen Sprung nach oben gemacht. Dies hätte von allen

Seite 5/7

Beteiligten – DIW Berlin und OECD – offenkundig besser kommuniziert werden können (siehe auch unten „*Einige Lehren für die Zukunft*“).

Für NutzerInnen der SOEP-Mikrodaten ist wichtig: von der Revision bzw. der Verbesserung der Daten sind nur diejenigen NutzerInnen betroffen, die die von der SOEP-Gruppe generierten Jahreseinkommen aus den \$PEQUIV-Files benutzen. Wer das direkt erfragte aktuelle monatliche Haushaltsnettoeinkommen („Income-Screener“, I1HINC\$\$, HINC\$\$, AHINC\$\$ aus den \$HGEN-Files) benutzt, ist nicht betroffen (wobei wir aber hier nochmals ausdrücklich darauf hinweisen wollen, dass seit der Datenauslieferung 2008 auch eine Variable (I_HINC\$\$) für den „Income-Screener“ generiert wird, die keine fehlenden Werte (missing values) enthält, da auch seit der Datenauslieferung 2008 beim Income Screener imputiert wird).

Nicht alle Verteilungsanalysen von Revision überhaupt betroffen

Die Herausforderungen eines sachgerechten Umgangs mit teilrealisierten Haushalten treten bei Verwendung des Income-Screener definitionsgemäß nicht auf. Sämtliche Analysen, die sich auf Einkommen von Personen beziehen sind auch nicht betroffen.

Auch die Mikrosimulationsmodelle, die für Einkommens- und Verteilungsanalysen von vielen NutzerInnen angewandt werden, sind durch die oben geschilderten Revisionen nicht betroffen, da für die gängigen Mikrosimulationsmodelle – unserer Kenntnis nach – nur vollrealisierte Haushalte benutzt werden.

Revisionen sind keine „Panne“

Vorwürfe der Presse, das DIW habe falsche Daten und damit Grundlage für „falsche politische Maßnahmen“ geliefert, haben wir in entsprechenden Pressemitteilungen des DIW Berlin mit Nachdruck zurückgewiesen¹. Die Revision hat vielmehr neue wissenschaftliche

¹ Siehe: http://www.diw.de/sixcms/detail.php?id=diw_01.c.372305.de sowie

Seite 6/7

Erkenntnisse zur Imputation fehlender Werte und effizienteren Hochrechnungsverfahren genutzt, die zu einer Verbesserung des Datensatzes führten. Die Revision alter Daten ist aus wissenschaftlicher Perspektive eine positive Entwicklung und keine ‚Panne‘.

Einige Lehren für die Zukunft

Aus den Debatten ziehen wir als SOEP-Gruppe die Lehre, dass wir bei größeren Revisionen der Imputationsverfahren, aber auch von Gewichtungs- und Hochrechnungsfaktoren nicht nur wie in der Vergangenheit die SOEP-DatennutzerInnen, sondern künftig aktiv noch stärker als bisher auch die breite Öffentlichkeit anhand von sozio-demographischen Kernindikatoren über die Revisionseffekte in Kenntnis setzen müssen.

Besonders zu beachten sind insbesondere solche DatennutzerInnen, die nicht mit den Mikrodaten des SOEP direkt arbeiten, sondern mit vom DIW Berlin (oder anderen) aufbereiteten SOEP-basierten Tabellen, die dann Eingang in Gutachten oder Statistiksammlungen (wie die der OECD) finden.

In der amtlichen Statistik gibt es eine solche Revisions-Kultur; das SOEP will sich diese Erfahrungen in Zukunft zu Eigen machen.

Wir werden also künftig die jährlichen Datenauslieferungen nicht nur „wissenschaftsintern“ sondern auch für die Medien ankündigen. Dazu dient auch ein stärkeres Herausstellen der seit Jahren mit jeder neuen Welle zur Verfügung gestellten Ergebnissen im SOEP-Monitor. Diese umfassen neben einer Vielzahl von gewichteten Analysen auf Basis von Indikatoren zur Einkommens- als auch Wohlfahrtssituation eine Vielzahl von Kernindikatoren (Arbeitsmarkt, Gesundheit, etc.) und erlaubt auch Effekte von verbesserten und revidierten Daten explizit deutlich zu machen.

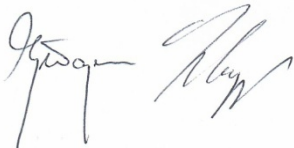
Seite 7/7

Im SOEP-Monitor werden künftig nicht nur die neuen hochgerechneten Randverteilungen dargestellt, sondern auch etwaige Differenzen zur Datenauslieferung aus dem Vorjahr.

In diese Verbesserung der Transparenz fügt sich auch die bereits mit der letzten Datenauslieferung eingeführte explizite Versionierung der Datenauslieferung ein (derzeit wird mit der Version SOEP v26 gearbeitet). Künftig werden die neuen Versionen auch mit einer DOI-Nummer eindeutig identifizierbar sein (mehr dazu wird im nächsten SOEP-Newsletter berichtet werden).

Zensus 2011 – Ein Grund für eine SOEP-Revision bei Gewichtung und Hochrechnung

Bereits jetzt kündigen wir an: Der gerade laufende Zensus (Volkszählung) wird eine gründliche Revision der SOEP-Hochrechnungsfaktoren zur Folge haben. Diese wird sich vermutlich auf die Datenlieferung in zwei oder drei Jahren erstrecken. Die Ergebnisse des Zensus 2011 führen mit an Sicherheit grenzender Wahrscheinlichkeit zu substanziellen Revisionen in der regionalen wie sozio-demographischen Verteilung der SOEP-Stichprobe wie aller anderen Personen- und Haushaltsstichproben.



(Leiter der Infrastruktureinrichtung SOEP im DIW Berlin)