

Das Sozio-oekonomische Panel (SOEP): Multidisziplinäres Haushaltspanel und Kohortenstudie für Deutschland – Eine Einführung (für neue Datennutzer) mit einem Ausblick (für erfahrene Anwender)

Gert G. Wagner · Jan Göbel · Peter Krause · Rainer Pischner · Ingo Sieber

Eingegangen: 3. November 2007 / Angenommen: 7. November 2008 / Online: 2. Dezember 2008
© Die Autoren 2008. Dieser Artikel wurde mit Open Access auf Springerlink.com veröffentlicht.

Zusammenfassung Die Längsschnittstudie *Sozio-oekonomisches Panel (SOEP)* stellt seit einem Viertel-Jahrhundert als Teil der (inter-)nationalen Forschungsinfrastruktur jährlich Mikrodaten zur Messung biographischer Verläufe für die Grundlagenforschung und Politikberatung im Bereich der Sozial-, Wirtschafts- und Verhaltenswissenschaften bereit. Das SOEP ist zugleich Haushaltspanel und Kohortenstudie. Der vorliegende Beitrag beschreibt den Hintergrund des SOEP, seine inhaltlichen Schwerpunkte, die datentechnische Konzeption, sowie Hochrechnung und Anwendungsmöglichkeiten. Der Ausblick fokussiert zukünftige Entwicklungen – die Weiterentwicklung der Governance und des Services – im Hinblick auf eine detailreiche Erfassung des Lebensverlaufs mit disziplinübergreifenden Forschungskonzepten unter Einsatz innovativer Erhebungs- und Meßmethoden.

Stichwörter SOEP · Biographische Erhebung · Haushaltspanel · Paneldesign · Kohortenstudie · Lebensverlauf · Stichprobenziehung · Hochrechnung · Datenstruktur · Service

Abstract The longitudinal study *German Socio-Economic Panel (SOEP)* has been working for a quarter of a century as part of international research infrastructure. It offers annual microdata for the measurement of biographical developments for basic research and policy advice in the social, economic, and behavioral sciences. The SOEP is, at the same time, a household panel as well as a cohort study. The present

G. G. Wagner (✉) · J. Göbel · P. Krause · R. Pischner · I. Sieber
Deutsches Institut für Wirtschaftsforschung (DIW Berlin), Mohrenstraße 58, 10117 Berlin,
Deutschland
E-Mail: officewagner@diw.de

G. G. Wagner
MPI for Human Development, Berlin, Deutschland

paper describes the background of the SOEP, its core themes and data structure, as well as the process of weighting and potential applications. The prospect focuses on future developments — the further improvement of governance and service — aiming to gather detailed life course data with interdisciplinary research concepts by the use of innovative survey and measurement methods.

Keywords SOEP · Biographical survey · Household panel · Panel design · Cohort study · Life course · Life span · Sampling procedure · Weighting · Data structure · Service

JEL classification C81, C93, D31, I, J

1 Einleitung

Das Sozio-oekonomische Panel (SOEP) ist eine von der Wissenschaft getragene in Deutschland durchgeführte Längsschnittserhebung, bei der seit 1984 repräsentativ ausgewählte Haushalte und alle darin lebenden Personen einmal pro Jahr statistisch erfasst und alle Personen im Alter ab 17 Jahren persönlich befragt werden („Haushaltspanel“). Diese Erhebung wird von theoriegeleiteten wissenschaftlichen Fragen und Problemen der Politikberatung bestimmt und nicht von amtlichen und politischen Fragestellungen oder Regularien. Das SOEP ist Teil einer weltweiten Forschungs-Infrastruktur, die Längsschnittsdaten für die empirischen Analysen von Personen und Haushalten zur Verfügung stellt (vgl. Butz u. Torrey 2006; Frick et al. 2008). Die Datenbasis SOEP bietet damit hervorragende Analysemöglichkeiten für die Sozial-, Verhaltens- und Wirtschaftswissenschaften und ist darüber hinaus auch für Spezialfragen der Geographie, Umwelt- und Energiewissenschaften bedeutsam. Bislang liegen aus verschiedenen Disziplinen über 5000 Publikationen vor, die auf dem SOEP beruhen.

Aufgrund der wiederholten Messung bei denselben Erhebungseinheiten wird eine Panelerhebung umso informationsreicher, je länger sie andauert. Das SOEP wurde im Jahr 1984 in Westdeutschland begonnen; mit dem Erhebungsjahr 2008 sind davon etwa 2500 Personen bereits 25 mal befragt worden. Unter den Personen, die 1990 in Ostdeutschland hinzugekommenen sind, werden im Jahr 2009 etwa 1500 Befragten bereits zum 20. Mal befragt. Das SOEP bietet so ständig wachsende Möglichkeiten für exzellente Forschung und für die persönliche Weiterqualifizierung von Nachwuchswissenschaftlerinnen und -wissenschaftlern.

Neben ähnlichen Erhebungen in westlichen Industrieländern (so z. B. PSID in den USA, BHPS in Großbritannien, HILDA in Australien und SHP in der Schweiz sowie der in etlichen EU-Staaten angelegten Altersstudie SHARE¹), sind wissenschaftstragene Haushaltspanels auch in Russland (RMLS)² und der Ukraine (UMLS)³ zu finden. Für China ist eine entsprechende Studie in Planung.

¹ Share ist allerdings beschränkt auf die Alterspopulation der 50-jährigen und Älteren.

² <http://www.cpc.unc.edu/projects/rlms/>

³ <http://www.iza.org> im Forschungsbereich „Arbeitsmärkte in Transformations- und Schwellenländern“.

Daher wird seit langem Wert darauf gelegt, dass die SOEP-Mikrodaten auch mit anderen Paneldaten vergleichend auswertbar sind, um so auf Basis der Mikrodaten internationale Vergleiche zu ermöglichen. Um derartige Analysen, die sehr komplex sind, technisch wesentlich zu erleichtern, wurde an der Cornell University in den USA das „Cross National Equivalent File“ (CNEF) geschaffen, das gegenwärtig mehr als zwei Millionen über die Zeit und zwischen den Ländern vergleichend aufbereitete Beobachtungseinheiten für die USA (1980–2005), Kanada (1993–2005), Australien (2001–2006), Großbritannien (1991–2005), die Schweiz (1999–2005) und Deutschland (1984–2006) enthält (vgl. Frick et al. 2008).

Haushaltspanels wurden zunächst vorwiegend zur Analyse der Dynamik von Einkommensverteilung und -armut begonnen.⁴ Entsprechend liegen auch sehr viele Veröffentlichungen vor, die auf Basis des SOEP Verteilungsänderungen in der Gesamtbevölkerung (Querschnitte) repräsentativ beschreiben. Die Spannweite reicht von international renommierten Fachzeitschriften (vgl. Beaudry u. Green 2003) bis hin zum Armuts- und Reichtumsbericht der deutschen Bundesregierung. Mit dem Vorliegen langjähriger Panel-Daten, die entsprechende Ausschnitte von individuellen Lebensläufen im Familien- und Haushaltskontext beobachten, hat auch ein ganz anderer Typus von Fragestellungen und Veröffentlichungen zugenommen: Die Analyse von Lebensverläufen aus Sicht einzelner Gruppen und Kohorten.

Studien aus verschiedenen Bereichen von Sozial- und Wirtschaftswissenschaften und der Psychologie, die auf Basis von SOEP-Daten der Frage der Lebensverläufe der Befragten und ihres Zusammenspiels mit anderen Lebensverläufen nachgehen und untersuchen, wie einzelne Befragte mit den Folgen von Lebensverläufen umgehen, behandeln etwa die Themen Familien-Dynamik (Tamm 2005, 2008; Fertig u. Tamm 2008), Intergenerationale Mobilität (Casey u. Dustmann 2007; Schnitzlein 2008), Verlassen der elterlichen Wohnung (Scherger 2007, 2008; Lois 2008), Heirat (Diener et al. 2006; Lucas u. Clark 2006; Zimmermann u. Easterlin 2006; Stutzer u. Frey 2003), Arbeitsteilung im Privathaushalt (Cooke 2007), Eintritt in die und Austritt aus der Arbeitslosigkeit (Lucas et al. 2004; Lucas 2005; Jürges 2007; Romeu Gordo 2006), Übergang in den Ruhestand (Börsch-Supan u. Jürges 2006), Scheidung (Andreß et al. 2003; Andreß u. Bröckel 2007), Dynamik ehrenamtlicher Tätigkeiten (Erlinghagen 2007), Anpassung von Zufriedenheiten (Clark et al. 2008, Clark, Frijters u. Shields 2008; Headey 2008); Migration (Jürges 2006; Glowsky 2007; Drever u. Hoffmeister 2008), Verwitwung (Burkhauser et al. 2005) und Tod (Gerstorff et al. 2008). Diese Veröffentlichungen zeigen, dass das SOEP nicht nur ein Haushaltspanel sondern zugleich auch eine Kohortenstudie ist, bei der auch die Verteilungen in früheren Jahren repräsentativ (unter Einbezug bereits ausgefallener Kohortenmitglieder) berücksichtigt werden.⁵

Um Nutzern der Mikrodaten und an Ergebnissen Interessierten einen konzentrierten Überblick über neueste Forschungsarbeiten zu geben, wurde im Jahr 2007 eine Pre-Print-Reihe (Diskussionspapiere) eingerichtet, die unter www.diw.de/soeepapers

⁴ Die erste Studie Panel Study of Income Dynamics – PSID – führt diese Fragestellung sogar im Namen.

⁵ Zunehmend wird das SOEP auch für so unterschiedlichen Anwendungsbereiche wie den biologischen Grundlagen menschlichen Verhaltens (vgl. Cawley und Spieß 2008) oder der Verbindung von Umwelt und Ökonomie bedeutsam (vgl. z. B. Rehdanz u. Maddison 2008).

abrufbar ist. Außerdem sind Publikationen, die auf dem SOEP basieren und in Zeitschriften und Büchern erschienen sind, in einer eigenen Literaturdatenbank SOEPLIT dokumentiert.

Dieser Aufsatz ist in erster Linie als Überblick über Inhalte und Aufbau der SOEP-Daten für diejenigen geschrieben, die bislang (noch) nicht mit den SOEP-Daten gearbeitet haben. Insbesondere das Schlusskapitel, das zentrale Weiterentwicklungen des SOEP aufzeigt, ist aber auch für erfahrene SOEP-Nutzerinnen und -nutzer von spezifischem Interesse.

Wegen der intensiven internationalen Nutzung des SOEP sind alle jüngeren Dokumentationen der Erhebung in englischer Sprache verfasst (vgl. zuletzt Wagner et al. 2007). Eine gleichermaßen breite, aber knapp gefasste deutschsprachige Darstellung der Grundstrukturen des SOEP, die bei der Auswertung der Daten hilfreich ist, fehlt seit Jahren. Der vorliegende Aufsatz schließt diese Lücke. Er beginnt in Abschn. 2 mit einer Darstellung der Erhebungsinhalte und des Ziehungsdesigns. In Abschn. 3 werden Probleme der Hochrechnung behandelt. In Abschn. 4 wird die Datenstruktur beschrieben mit Hinweisen auf bestehende und für die Auswertung hilfreiche Dokumentationen. Abschnitt 5 enthält schließlich einen Ausblick über weitere Entwicklungen im SOEP, der sich auch an erfahrene Anwender richtet.

2 Konzeption, Erhebungsinhalte und Stichprobendesign des SOEP

Das SOEP wurde geschaffen, um sozial- und wirtschaftswissenschaftliche Fragestellungen im Rahmen von Haushalten, Lebensläufen und lebenslaufbezogenem Verhalten analysieren zu können (vgl. Krupp 2008)⁶. Lebensverläufe werden dabei nicht nur aus ökonomischer, sondern vor allem auch aus soziologischer Sicht betrachtet. Deshalb sind die erhobenen Daten (z. B. zum Verlauf der Lebenszufriedenheit) auch für verhaltenswissenschaftliche oder psychologische Forschungsfragen interessant. Seit 2002 werden die Erhebungsinstrumente des SOEP systematisch erweitert und um verhaltenswissenschaftliche Konzepte ergänzt (vgl. Schupp et al. 2008; Trommsdorff 2008; Schimmack et al. 2008).

Inhalte der Erhebung

Die zentrale Aufgabe einer sozial- und wirtschaftswissenschaftlichen Panelstudie ist es, repräsentative Mikrodaten über Personen, Haushalte und Familien zu erheben, die es erlauben die Stabilität und den Wandel der Lebensbedingungen zu analysieren. Das SOEP folgt dabei einem Ansatz, bei dem mikro-ökonomische Daten mit Indikatoren aus der soziologischen, psychologischen und politikwissenschaftlichen Forschung zusammengeführt werden.

Das Erhebungsspektrum des SOEP ist aufgeteilt in einen Kernbereich an Variablen, der laufend – in der Regel jährlich – nahezu unverändert abgefragt wird und einem von Erhebung zu Erhebung wechselnden Befragungsschwerpunkt, der nur alle drei bis sechs Jahre wiederholt wird.

⁶ Eine breite Darstellung, mit vielen Details der „SOEP-Geschichte“, findet man in Heft 3/2008 der „Vierteljahrshefte zur Wirtschaftsforschung“.

Zum Kern des Erhebungsprogramms über Kinder und für Erwachsene gehören Fragen aus den folgenden Bereichen:

- Demographie und Wohnsituation
- Persönlichkeitsmerkmale und Grundorientierungen (Präferenzen, Werte, usw.)
- Vorschul- und Schulbildung, Berufliche Bildung und Weiterbildung; Qualifikation
- Arbeitsmarkt- und Berufsmobilität
- Einkommen, Vermögen und soziale Sicherheit
- Gesundheit
- Sorgen und Zufriedenheit (allgemeine Lebenszufriedenheit und Bereichszufriedenheiten)

Die darüber hinaus in jährlich wechselnden Modulen detaillierter erhobenen Befragungsschwerpunkte für Erwachsene umfassen die Bereiche biografische Hintergrundinformationen, soziale Herkunft, Nachbarschaft, soziale Sicherung, Vermögen, Weiterbildung, Zeitverwendung, Familie, Arbeitsbedingungen, Zukunftserwartungen und Umwelt. Tabelle 1 gibt einen Überblick über die Erhebung und Replikation dieser Module zu den jeweiligen Erhebungsjahren und zeigt an, welche Stichproben des SOEP (vgl. Übersicht 1) jeweils dazu befragt wurden. Biografische Hintergrundinformationen (BIO) zum Erwerbsverlauf, zu Familienstandsänderungen, zur sozialen Herkunft und zum Berufseinstieg wurden in den ersten Jahren in den Stichproben A und B noch im Rahmen der Standard-Personenfragebögen und bei neuen Befragungspersonen als Zusatzfragen retrospektiv (R) erhoben. Seit 1988 werden Biografieinformationen in einem eigenen Erhebungsinstrument erfasst, das vor allem infolge der Integration der Ost- und Zuwanderer Stichproben (C und D) bis 2001 immer wieder aktualisiert wurde. Seitdem werden Biografieinformationen standardmäßig im „Lebenslagenfragebogen“ bei allen neuen Haushaltsmitgliedern im Befragungsalter im laufenden Jahr erhoben. Bei neuen Stichproben werden Lebenslauffragebögen jedoch erst in der dritten Befragungswelle eingesetzt, wenn die Befragten mit dem Erhebungsinstrument besser vertraut sind und sich aus der zusätzlichen Befragungsbelastung zur Erfassung des biografischen Hintergrundes keine negativen Auswirkungen auf die weitere Teilnahmebereitschaft mehr ergeben. Neue Stichprobenmitglieder erhalten deshalb einen geringfügig modifizierten Personenfragebogen, bei dem beispielsweise auch Bildungsinformationen⁷ abgefragt werden; zudem wurden in einigen Jahren die Modulfragen zu den Erhebungsschwerpunkten nur an Mitglieder der Altstichproben gerichtet.

Um die Stabilität und den Wandel der Lebensbedingungen adäquat zu beschreiben, beziehen sich die Fragen an die Teilnehmer der Studie auf unterschiedliche Zeiträume:

- Zeitpunktbezogene Fragen (Gegenwart, z. B. derzeitiges Einkommen oder derzeitige Zufriedenheit)
- Einfache retrospektive Frage zu bestimmten Ereignissen in der Vergangenheit (z. B. Wie oft die Arbeitsstelle in den letzten 10 Jahren gewechselt wurde).

⁷ Bildungsabschlüsse werden im SOEP standardmäßig nur bei neuen Personen im Rahmen ihrer Biografie erfasst. Im Personenfragebogen werden dann jedes Jahr Veränderungen der Bildungsinformationen erfragt.

Tabelle 1 Erhebungs-Schwerpunkte im SOEP

aktuelle Stichproben	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Befragungsschwerpunkte (Module)	A+B	A+B	A+B	A+B	A+B	A+B	A+C	A-C	A-C	A-C	A-C	A-D	A-D	A-D	A-E	A-E	A-F	A-F	A-G	A-G	A-G	A-G	A-H	A-H
Erwerbsbiographie	R	R	R	R	R	R	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO
Familienstandsbiographie	A+B	R	R	R	R	R	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO
Soziale Herkunft/Berufseinstieg	A+B	R	R	R	R	R	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO
Nachbarschaft	A+B									A-C														
Soziale Sicherheit und ihre Wahrnehmung	A+B							A-C					A-D											
Vermögen																								
Weiterbildung und Qualifikation																								
Zeitverwendung und Präferenzen																								
Familie und Sozialleistungen																								
Arbeitsbedingungen	A+B	A+B	A+B																					
Zukunftserwartungen																								
Umweltbewußtsein																								

Legende:

Die Buchstaben A bis H kennzeichnen die jeweiligen SOEP-Stichproben (vgl. Übersicht 1); Lesbeispiel: A-H bedeutet alle Stichproben von A bis H. BIO = Biographie Fragebogen, nur für Erstbefragte der Altsichproben sowie für Neustichproben (seit Stichprobe F) in der dritten Erhebungswelle zu beantworten. R = Retrospektiv erfragte biografische Informationen aus dem Personenfragebogen (nur Stichproben A+B).

- Retrospektiver Lebensverlauf seit dem 15. Lebensjahr (z. B. Ausbildungs- und Erwerbsverlauf)
- Monatlicher Kalender zu Einkommen und Erwerbstätigkeit im letzten Jahr
- Fragen zu einer bestimmten Zeitperiode (z. B. Änderung des Familienstandes oder Wohnungssituation seit der letzten Befragung)
- Fragen bezüglich zukünftiger Erwartungen (z. B. Erwartungen zum Beruf oder zur Entwicklung der Zufriedenheit in einem Jahr oder fünf Jahren).

Das SOEP bietet darüber hinaus eine Vielzahl an Möglichkeiten, regionalbezogene Informationen bis hin zu „Nachbarschaftsdaten“ bei der Analyse zu berücksichtigen. Mit Hilfe der regionalen Zuordnung der Haushalte ist es möglich, dem SOEP auf der Ebene der Bundesländer (NUTS⁸-1), der Raumordnungsregionen (oder NUTS-2), der Kreiskennziffern (oder NUTS-3), und seit einigen Jahren auch auf den Ebenen, der Gemeindekennziffern, der Postleitzahlen und der „Straßenabschnitte“ regionale Indikatoren „zuzuspielen“. Die SOEP-Daten können so über entsprechende Verknüpfungsschlüssel (Code für Bundesländer, Kreiskennziffern, etc.) für entsprechende regionale Einheiten mit Regionalmerkmalen der amtlichen Statistik (z. B. Bevölkerungszahl und Mobilität, Arbeitslosigkeit, Krankenhausdichte, Umweltbelastung) wesentlich angereichert werden.⁹ Im Standardverfahren werden allerdings nur Bundeslandinformationen an Nutzer weitergegeben. Mit tiefer gehenden geografischen Informationen erhöhen sich auch die datenschutzrechtlichen Anforderungen. Mit zunehmender Sensibilität der Daten müssen so je nach Ebene der Regionalinformationen weitergehende Sicherheitsvorkehrungen eingehalten werden (für Details vgl. Goebel et al. 2008).

Durch die Zuordnung von (kommerziellen) Nachbarschafts-Daten sind auch Analysen von Interaktionseffekte mit „Meso-Variablen“ möglich; aus Datenschutzgründen sind diese kleinräumigen Informationen aber nur innerhalb des DIW Berlin – für jeden registrierten Nutzer – auswertbar. Dies gilt auch für die sozialstrukturelle Analyse der Vornamen der Befragten (vgl. Gerhards u. Hans 2006).

Konzeption der Erhebung

Es gibt im SOEP eine Vielzahl von Instrumenten, um Personen im Kontext ihrer Familie und ihres Haushaltes im zeitlichen Verlauf erheben zu können. Das SOEP erfasst durch eine einmal pro Kalenderjahr wiederholt stattfindende Begehung von Haushalten und der Befragung der Personen möglichst lange Ausschnitte aus Lebensläufen. In zufällig und repräsentativ ausgewählten Haushalten werden alle Erwachsenen (17-jährige und Ältere) mit Hilfe von Personenfragebögen direkt befragt. Zusätzlich, werden von der „Hauptauskunftsperson“ im Haushaltsfragebogen Merkmale über den gesamten Haushalt (z. B. die Wohnung) erfragt. Hinzu kommen spezielle Fragebögen zum Lebenslauf bei neu erfassten Personen oder Fragebögen, in denen Mütter Angaben über ihre (kleinen) Kinder machen.

⁸ NUTS – Nomenclature des unités territoriales statistiques (internationale Systematik für Gebietseinheiten).

⁹ Zum Beispiel gibt das Statistische Bundesamt eine DVD mit dem Titel „Statistik regional“ heraus und das Bundesamt für Bauwesen und Raumordnung die „INKAR CD“ (Indikatoren und Karten).

Während in Psychologie und Medizin Kohortenstudien einzelne Lebensläufe üblicherweise von der Geburt an verfolgen, liefert das SOEP aufgrund der Erhebung im Haushaltskontext zudem bereits Informationen über das Leben der Mutter für die Zeit vor der Geburt eines Kindes (und auch des Vaters, wenn er – was meist der Fall ist – mit der Mutter zusammenlebt). Für ein Kind, das in einen „SOEP-Haushalt“ hineingeboren wird, wird über den Haushaltsfragebogen und spezielle Fragebögen, die an Mütter gerichtet sind, dessen Weg durch die Kindheit und Jugend beobachtet. Ab dem 17. Lebensjahr wird ein Jugendlicher zum persönlichen Befragten. Bei der erstmaligen Befragung wird – seit 2001 – ein spezieller Jugendfragebogen eingesetzt und seit 2006 wird bei Jugendlichen auch die kognitive Leistungsfähigkeit mit einem halbstündigen Instrument getestet.

Während des Erwachsenenlebens werden zu einer Vielzahl von Lebensbereichen objektive Indikatoren (z. B. Einkommen, Aus- und Weiterbildung, Kinderbetreuung, Erwerbstätigkeit, Arbeitslosigkeit, ausgewählte Freizeitaktivitäten) und subjektive Bewertungen durch die Befragten (z. B. Lebenszufriedenheit, Sorgen) erhoben. Schließlich werden auch die letzten Lebensjahre eingehend erfasst. Selbst im Falle eines Umzugs in ein Alten- oder Pflegeheim wird – soweit möglich – eine eigenständige Befragung durchgeführt, zumindest aber die Tatsache dieses Umzugs erhoben. Aufgrund der Erhebung im Haushaltskontext ist es schließlich möglich, auch noch nach dem Tode eines Befragten die materiellen und mentalen Konsequenzen für die Hinterbliebenen zu erfragen (z. B. Hinterbliebenenrenten, Erbschaften und Auswirkungen auf die Lebenszufriedenheit).¹⁰

Ziehungsdesign der Erhebung

Realisiert wird dieses Konzept der Erhebung von Lebensläufen durch die Ziehung einer Haushaltsstichprobe, bei der alle Personen in diesen Haushalten (genauer: Privathaushalten) selbst zu Erhebungs-Einheiten werden. Auch ungeborene Kinder gehören virtuell zu dieser Stichprobe. Ziehen Personen aus einem Befragungshaushalt aus, so werden diese weiter verfolgt. Ebenso werden Personen, die in einen Befragungshaushalt einziehen, auf Dauer in das SOEP einbezogen. Durch den Einbezug von Nicht-Original-Stichproben-Mitgliedern (vgl. dazu Tabelle 3.1) entsteht im Prinzip eine Schneeballstichprobe. Dies ist beabsichtigt, da so die Lebensläufe der Original-Stichprobenmitglieder besser im Kontext analysierbar sind. Dies gilt auch für Geschiedene, deren Lebenswege sich (scheinbar) völlig trennen (vgl. Spiess et al. 2008). Der Schneeball-Effekt wird durch eine entsprechende Gewichtung der Daten berücksichtigt. Faktisch wird durch dieses „Weiterverfolgungskonzept“ die Stichprobe in der Regel aber ohnehin nicht größer, da den Schneeball-Befragten jene gegenüberstehen, die nicht mehr bereit sind am SOEP weiter teilzunehmen („Panel-Ausfälle“). Auch diese Ausfälle werden im Laufe der Panel-Laufzeit durch Gewichtungen berücksichtigt, um unverzerrte Rückschlüsse auf die Grundgesamtheit ziehen zu können (vgl. Galler 1987).

¹⁰ Mit Hilfe eines speziellen Fragebogens für Hinterbliebene werden ab 2009 die Informationen nach einem Sterbefall noch verbessert werden (über die letzte Lebensphase des verstorbenen SOEP-Teilnehmers und über die Trauerarbeit).

Das Weiterverfolgungskonzept bildet die endogene Bevölkerungsdynamik vollständig ab.¹¹ SOEP-Teilnehmer, die sterben, stellen keine unerwünschten oder verzerrenden Ausfälle dar, sondern gehören zur natürlichen Bevölkerungsdynamik in der Stichprobe. Während durch dieses Konzept Zuwanderung im Zuge von Familiennachzug in einen bereits erfassten Haushalt automatisch einbezogen wird, wird Zuwanderung nicht erfasst, wenn sie in einen noch nicht erfassten Haushalt erfolgt. Zur Sicherstellung der Querschnittsrepräsentativität müssten deswegen – z.B. im Hinblick auf die Erfassung der Einkommensverteilung – idealerweise jährlich Zusatzstichproben für Zuwanderer gezogen werden, die mit bestehenden Haushalten nicht oder nur mit geringer Wahrscheinlichkeit in Kontakt kommen. Faktisch geschah dies bisher nur einmal gezielt (1994/95) (vgl. Schupp u. Wagner 1995). In den Jahren 1998, 2000 und 2006 wurde diesem Problem durch repräsentative „Auffrischungstichproben“ für die gesamte Bevölkerung Rechnung getragen.

Die Erhebung von Lebensläufen im Haushaltskontext erlaubt es, dass die jahresbezogenen Erhebungsdaten für die Erhebungsjahre zwischen 1984 und 2007 und künftiger Jahre auf die Grundgesamtheit aller Personen und (Privat-)Haushalte in Deutschland hochgerechnet werden können (vergleiche dazu auch den Abschnitt zur Hochrechnung des SOEP). Erst dadurch wird zum Beispiel die Analyse der Dynamik von relativer Einkommensarmut, die sich am Medianeinkommen der Bevölkerung orientiert, möglich.

Die Stichproben und Befragungsmodi des SOEP

Mit der 2006 gestarteten Ergänzungsstichprobe H umfasst das SOEP nunmehr acht Teil-Stichproben und – für die Stichproben A und B – 24 auswertbare Wellen (1984–2007).¹² Ein Ende der Erhebungen ist nicht geplant. Übersicht 1 zeigt, wann welche Teilstichprobe für welche Grundgesamtheit begonnen wurde (für eine ausführliche Darstellung vgl. Wagner 2008).

In jeder Stichprobe werden die Teilnehmer am SOEP über den Haushalt ausgewählt. Der Modus der Kontaktaufnahme ist bei allen Teilstichproben gleich: ein Schreiben der Feldarbeits-Organisation (TNS Infratest Sozialforschung, München), mit dem das SOEP und ein Interviewer angekündigt werden. Die Adressen wurden unterschiedlich ermittelt. Meist per Random-Walk (Stichproben A, E, F und H), einmal per Register (Stichprobe B), einmal per Register und Interviewer (Stichprobe C) und zweimal durch Screenen von Haushalten nach spezifischen Bevölkerungsgruppen (Stichprobe D: Zuwanderer per Standard-Random-Walk; Stichprobe G: Hocheinkommenshaushalte per Standard-Telefon-Interview). Alle Adressen wurden nicht direkt befragt, sondern die ermittelten Adressen wurden an die Infratest-Zentrale gemeldet und den Interviewern, die die eigentliche Befragung durchführten, als Adressen vorgegeben. Auf dieser entscheidenden Stufe arbeitet das SOEP also mit derselben Qualität wie eine echte Register-Stichprobe.

¹¹ Im SOEP werden alle Haushaltsmitglieder einschließlich später zugezogener Personen, sogenannte „Nicht-Original-Stichproben-Teilnehmer“, weiterverfolgt; im Unterschied dazu verfolgen andere Panelstudien neben Original-Stichprobenmitgliedern nur geschiedene Elternteile/Ex-Partner weiter, wenn gemeinsame Kinder geboren wurden.

¹² Eine aktuelle Beschreibung des SOEP findet sich bei Wagner et al. (2007).

Die SOEP-Befragung wird in der ersten Erhebungswelle einer jeden Stichprobe noch ausschließlich Face-to-Face durchgeführt (mit Paper und Pencil [PAPI]) oder auch – seit 1998 – als Computer Assisted Personal Interviewing [CAPI]). Ab der zweiten Welle ist auch ein Selbstausfüllen mit telefonischer Betreuung möglich. Dies wird aber faktisch erst seit Welle 7 (1990) tatsächlich nennenswert genutzt. Der Anteil der Selbstausfüller liegt im Sample A nach 22 Wellen bei 20%, in Sample F nach 5 Wellen erst bei ca. 6%.

Übersicht 1

Stichprobe A:	(Deutsche) Haushalte ^a in der Bundesrepublik Deutschland (Hauptstichprobe, Start 1984; $N = 4528$ realisierte Haushalte)
Stichprobe B:	Ausländische Haushalte ^b in der Bundesrepublik Deutschland (Start 1984; $N = 1393$ realisierte Haushalte)
Stichprobe C:	Privathaushalte in der DDR (Start im Juni 1990; $N = 2179$ realisierte Haushalte)
Stichprobe D:	Zuwanderer-Privathaushalte in Deutschland (Start 1994/95; $N = 374$ realisierte Haushalte)
Stichprobe E:	Haushalte ^c in Deutschland (Ergänzungsstichprobe, Start 1998; $N = 1056$ realisierte Haushalte)
Stichprobe F:	Haushalte ^c in Deutschland (Ergänzungsstichprobe, Start 2000; $N = 6052$ realisierte Haushalte)
Stichprobe G:	Hocheinkommens-Privathaushalte in Deutschland (Hocheinkommensstichprobe, Start 2002; $N = 1224$ realisierte Haushalte)
Stichprobe H:	Haushalte ^c in Deutschland, Ergänzungsstichprobe (Start 2006; $N = 1506$ realisierte Haushalte)

^a Genauer: Haushalte, deren Haushaltsvorstand nicht türkischer, italienischer, jugoslawischer, griechischer oder spanischer Nationalität war. Dies waren ganz überwiegend (99 %) deutsche Haushaltsvorstände.

^b Genauer: Haushalte, deren Haushaltsvorstand türkischer, italienischer, jugoslawischer, griechischer oder spanischer Nationalität war.

^c Anstaltshaushalte sind nicht repräsentativ eingeschlossen; sie werden zwar auch nicht ausgeschlossen, wenn sie beim random walk gelistet werden und sind insofern in dem Bruttobestand enthalten, werden aber bei der Durchführung der Befragung in der Regel bei neuen Samples nicht berücksichtigt.

Die verschiedenen Befragungs-Modi sind für jeden einzelnen Befragten und Datensatz kodiert und damit für jeden Nutzer auswertbar. Allein aufgrund der verschiedenen Befragungs-Modi ist das SOEP für surveymethodische Fragen eine wahre Fundgrube (vgl. Schräpler 2007). Hierzu kommt die Identifikation eines jeden einzelnen Interviewers (vgl. Schräpler u. Wagner 2000), so dass potenzielle Interviewereffekte leicht analysierbar sind (vgl. Schräpler 2004).¹³

In Spezial-Datensätzen sind auch Angaben zu Haushalten vorhanden, die nicht bereit waren an der Erhebung teilzunehmen, oder im Laufe der Zeit verweigerten¹⁴. Für Re-Analysen stehen in einem speziellen Datensatz auch Daten zur Verfügung, die als vom Interviewer gefälscht identifiziert wurden (vgl. Schräpler und Wagner 2005). Solche Fälschungen, die im Nachhinein wegen der Längsschnittstruktur der Daten aufgedeckt werden konnten, machen allerdings weniger als ein halbes Prozent der Befragten aus.

¹³ Da ein nicht-experimentelles Design vorliegt, sind Befragungsartefakte zwar kontrollierbar, aber eine kausale Zuschreibung auf Einzeleffekte bedarf der Setzung von Annahmen.

¹⁴ „Brutto“-Daten; diese enthalten beispielsweise Regionalinformationen oder Interviewereinschätzungen zum Wohnumfeld.

Datennutzung im Längsschnitt

Durch die Vielzahl sozial- und verhaltenswissenschaftlicher Fragestellungen, die mit den SOEP-Daten wissenschaftlich bearbeitet werden, ist nicht nur die „lebenslaufbezogene“ Konzeption des SOEP immer mehr in den Vordergrund gerückt. Auch die Mikrodaten selbst werden im Hinblick auf „Event-Analysen“ immer nutzerfreundlicher aufbereitet. Im Einzelnen bedeutet dies Folgendes.

Das SOEP wird nicht an einem einzigen Stichtag (bzw. Berichtstag) erhoben. Um die Bereitschaft zu wiederholter Teilnahme hoch zu halten, zieht sich die Feldarbeit vielmehr über mehrere Monate hin. Deshalb kann das SOEP anhand von Tagesangaben (z. B. Zufriedenheit mit dem Leben) bzw. Wochenangaben (z. B. Arbeitszeit) auch für unterjährige Analysen genutzt werden. Aufgrund des Schwerpunkts der Feldarbeit im ersten Quartal stehen nur für diese Periode pro Monat oder Woche (teilweise sogar pro Tag) genügend Stichtags- bzw. Stichwochenfälle zur Verfügung. Normalerweise sind im Laufe des März bereits aber schon rund 50% aller Interviews abgeschlossen.

Freilich werden zu einem Befragungs-Zeitpunkt nicht nur Daten zum Stichtag bzw. einer normalen Woche erhoben, sondern auch retrospektiv über das vergangene Kalenderjahr. Auf Basis derartiger Vorjahresangaben werden auch differenzierte Jahreseinkommen konstruiert, die eine Unterscheidung nach Einkommenskomponenten in Form von Markteinkommen, Transfers, direkte Steuern und Sozialabgaben erlauben (vgl. Grabka 2007; Frick u. Grabka 2003). Außerdem können aufgrund von „Kalender-Angaben“ zu Aktivitäten, die pro Monat ausgeübt werden (z. B. Erwerbstätigkeit, Ausbildung, Wehr- oder Zivildienst), Daten über „Spells“, d. h. zeitliche Lage und Länge dieser Aktivitäten erzeugt werden.¹⁵ Damit stehen seit Beginn des Erhebungszeitraums (Januar 1983 – erhoben in der ersten Welle im Frühjahr 1984) bis Dezember 2006 für maximal 276 Kalendermonate Aktivitätsangaben (Spelldaten) zur Verfügung. Für 2704 Personen liegen diese tatsächlich ohne Missings für alle 276 Monate vor. Für fast 11 Tausend Personen liegen für mindestens 120 Monate (10 Jahre) Spelldaten vor und für mehr als 25 Tausend Personen gibt es Kalendarien die wenigsten 60 Monate umfassen. Tabelle 2 zeigt überdies, dass über 43.000 Beobachtungen vorliegen, für die über mindestens 12 Monate lückenlose Monatsangaben erfasst sind.

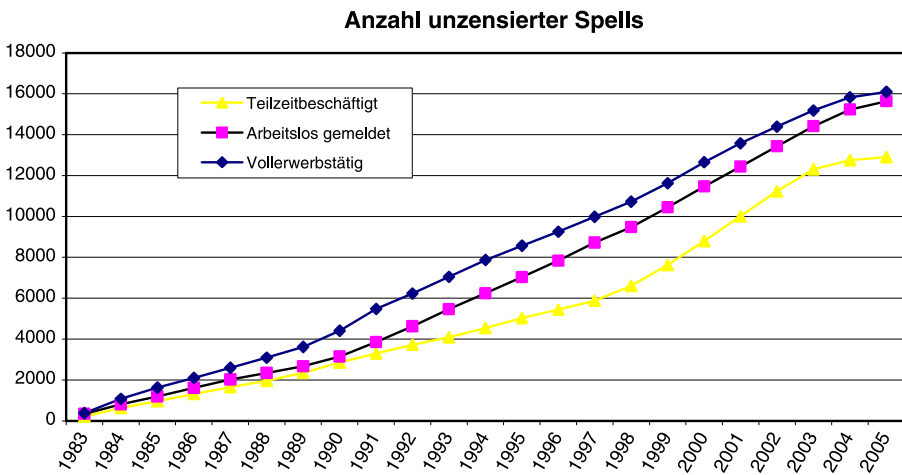
Eine für (potentielle) Nutzer der SOEP-Daten aussagekräftige Darstellung der Reichhaltigkeit und Aussagekraft der Spell-Daten ergibt sich, wenn man die Zahl nicht-zensierter Spells darstellt, d. h. die Zahl von Aktivitäts-Episoden, deren Beginn und deren Ende beobachtet wurden, die also weder links- noch rechtszensiert sind. Wenn z. B. ein Befragter in den Erhebungsjahren 1990 und 1991 für die Monate 7/1990 bis 3/1991 Arbeitslosigkeit als Aktivität angibt, und er davor erwerbstätig und danach im Ruhestand war, dann ist das ein Spell von 9 Monaten für den Aktivitäts-Typus Arbeitslosigkeit. Abbildung 1 zeigt beispielhaft für drei Aktivitäten die kontinuierliche Zunahme der unzensierter Spells auf Monatsbasis.

¹⁵ Diese Kalendarien werden in Form einer Matrix erhoben, bei der für alle Monate des zurückliegenden Jahres der jeweilige Erwerbsstatus festgehalten wird.

Tabelle 2 Zahl der Fälle mit vollständigen monatlichen Erwerbskalendarien nach Stichproben und Zahl der erfassten Monate für die ersten 23 Wellen

Erfasste Dauer	276 Monate	240 Monate und mehr	180 Monate und mehr	120 Monate und mehr	60 Monate und mehr	12 Monate und mehr
Stichproben						
Insgesamt	2.704	3.611	7.216	10.945	25.758	43.593
A	2.278	2.969	4.166	5.797	8.174	12.076
B	426	642	1.106	1.817	2.789	4.440
C			1.944	2.790	4.028	5.869
D				541	856	1.366
E					1.309	2.203
F					7.110	12.184
G					1.492	2.839
H						2.616

Quelle: Sozio-oekonomisches Panel, Wellen Q – W; eigene Berechnungen.

**Abb. 1** Anzahl unzensierter Erwerbstätigkeit-Spells

Anzahl unzensierter Erwerbstätigkeit-Spells

Das besondere Analysepotential eines Haushaltspanels für die Auswertung von Ereignissen und Lebensläufen wird auch an der Geburt von „SOEP-Enkeln“ deutlich. Dies sind Kinder, deren Eltern am SOEP teilnehmen und von denen auch ein Großelternanteil in der SOEP-Stichprobe enthalten ist. Seit 1984 wurden bereits mehr als 1000 Kinder im SOEP erfasst, deren Eltern und Großeltern bereits selbst an der SOEP-Befragung teilgenommen haben; einige sind nach 16 Jahren (2001) bereits selbst ins Befragungsalter „hineingewachsen“. Im Jahr 2005 sind noch mehr als zwei Drittel der erfassten Enkelkinder im Erhebungsbestand.

3 Hochrechnung des SOEP

Die Komplexität des SOEP-Datensatzes bedeutet für deskriptiv angelegte Analysen, dass eine Berücksichtigung der Hochrechnung unbedingt erforderlich ist. Dies hat zwei Gründe, zum einen sind wegen der unterschiedlichen Stichprobendesigns die Ziehungswahrscheinlichkeiten und damit die Designgewichte der Beobachtungen in den einzelnen Teilstichproben unterschiedlich.¹⁶ Hinzu kommt ein unterschiedliches Antwortverhalten nach Welle 1 (d. h. es gibt selektive Ausfälle über die Zeit). Dabei treten Unterschiede zwischen den Teilstichproben auf. So sind z. B. in den Stichproben B und D höhere Emigrationsraten zu beobachten. Übersicht 1 (vgl. Abschn. 2) beschreibt die jeweilige Grundgesamtheit, aus der die bisherigen acht Stichproben des SOEP gezogen wurden.

Die verschiedenen Teilstichproben sind teilweise noch weiter nach dem jeweiligen Ziehungsdesign differenziert (vgl. Tabelle 3): Stichprobe B nach fünf Nationalitäten der Haushaltsvorstände; Zuwandererstichprobe D nach Aus- und Übersiedlern; Stichprobe F nach Deutschen und überproportional erfasste Migranten; die Hocheinkommensstichprobe G differenziert nach dem Einkommensschwellenwert. Diese Differenzen der Ziehungswahrscheinlichkeiten und der daraus sich ergebenden „Hochrechnungsfaktoren“ variieren damit „per Design“. Entsprechend variieren die in Tabelle 3 ausgewiesenen Designgewichte aufgrund der Unterschiede in der Ziehung der Stichproben (vgl. Goebel et al. 2008).

Die Schätzung von Gewichten bzw. Hochrechnungsfaktoren erfolgt – wie international üblich – nach dem Ansatz von Horwitz und Thompson (1952) über den Kehrwert der Auswahlwahrscheinlichkeit der jeweiligen Stichprobeneinheit. Hochrechnungsfaktoren unterscheiden sich von Gewichten lediglich durch einen Skalarmultiplikator. Während die nach Horwitz und Thompson berechnete Summe der Hochrechnungsfaktoren der Zahl der Einheiten der Grundgesamtheit entspricht, stimmt die Summe der Gewichte mit der Fallzahl der Stichprobe überein.

Horwitz und Thompson entwickelten ihren Schätzer indes nur für reine Querschnittsdaten ohne hierarchische Datenebenen wie Haushalte und Personen. Stichprobeneinheiten sind beim SOEP primär die Haushalte, sekundär die Personen, die in diesen Haushalten leben. Der Ansatz von Horwitz und Thompson wurde von Galler (1987) für Paneldaten weiterentwickelt, indem die Schätzung von Auswahlwahrscheinlichkeiten für die Startwelle um die Schätzung der Verbleib- bzw. Antwortwahrscheinlichkeit für die folgenden Wellen erweitert wurde (vgl. Rendtel 1995). Hierzu muss zunächst die Wahrscheinlichkeit einer erneuten Kontaktaufnahme, darauf folgend die Wahrscheinlichkeit einer erneuten Antwortgewährung bestimmt werden. Eng verknüpft mit dieser Aufgabe ist die Analyse des Ausfallverhaltens der Stichprobeneinheiten von Welle zu Welle.¹⁷ Dieses Konzept ermöglicht es, sowohl Querschnittsgewichte ab Welle 2 jeder Stichprobe sowie Längsschnittgewichte theoretisch konsistent und praktisch zufriedenstellend zu ermitteln. Die Details können

¹⁶ D. h., die jeweiligen Beobachtungen in den einzelnen Teilstichproben repräsentieren unterschiedlich viele Personen bzw. Haushalte in der jeweiligen Grundgesamtheit; so sind z. B. in den Stichproben B und D Ausländer überrepräsentiert.

¹⁷ Siehe hierzu Kroh u. Spiess (2008).

Tabelle 3 Designgewichte im sozio-oekonomischen Panel

Stichprobe/ Teilstichprobe	Kurz-Beschreibung der Stichprobe (Startjahr)	Kennziffern für die Designgewichte von Haushalten mit <i>positiven</i> Querschnittsgewichten		
		Mittelwert	Standardabweichung	Fallzahl im Startjahr
A	Deutsch (1984)	3.344	0	4528
B	Ausländisch (1984)	546	305	1393
B1	Türkisch	772	332	397
B2	Italienisch	624	250	194
B3	Griechisch	355	158	196
B4	Jugoslawisch	487	211	306
B5	Spanisch	261	117	200
C	DDR (Juni 1990)	1.900	0	2179
D	Zuwanderer (1994/1995)	3.279	331	252*
D1	Übersiedler	2.946	0	127*
D2	Aussiedler	3.405	0	44*
D3	Sonstige	3.693	0	81*
E	Ergänzung (1998)	19.081	0	1056
F	Ergänzung (2000)	3.406	402	6052
F1	Deutsch	3.519	0	5607
F2	Ausländisch	1.980	0	445
G	Hocheinkommenshaushalte mit Haushaltsnettoein- kommen > 4.500 € (2000)	738	213	998*
G1	West: < 5.113 €	956	0	483*
G2	Ost: < 5.113 €	638	0	55*
G3	West: ≥ 5.113 €	524	0	419*
G4	Ost: ≥ 5.113 €	490	0	41*
H	Ergänzung (2006)	10.432	0	1506

* Folgende Stichproben enthalten weitere Haushalte ohne positive Gewichte für die Querschnittsgewichtung:

122 Haushalte in D, davon 49 in D1, 37 in D2, 36 in D3. Diesen Haushalten konnten kein Designgewicht zugeordnet werden, da ihre Auswahlwahrscheinlichkeit aufgrund der Ziehung mit Hilfe eines „Schneeballsystems“ nicht zu bestimmen ist.

226 Haushalte in G, davon 152 in G1, 23 in G2, 43 in G3 und 8 in G4. Diese Haushalte überschritten in der zweiten Welle nicht die für die Längsschnitterhebung vorgeschriebene Einkommensgrenze von 4.500 €.

Quelle: Sozio-oekonomisches Panel. Wellen A–W; eigene Berechnungen.

hier nicht dargestellt werden, dafür sei auf Rendtel (1995) und Pannenberg et al. (2005) verwiesen.

Tabelle 4 zeigt – exemplarisch für Haushalte¹⁸ – welche Hochrechnungs-Variablen für die verschiedenen Teilstichproben standardmäßig zur Verfügung stehen. Dieser Überblick zeigt einerseits, dass gewichtete Analysen wohlüberlegt durchgeführt werden sollten, da die einzelnen Hochrechnungsfaktoren des SOEP nicht immer alle Teilstichproben umfassen.¹⁹ Andererseits reicht für die meisten Analysen die Benutzung der Standard-Gewichtungsvariable \$xHRF\$ aus. Jede Beobachtung, die für ein bestimmtes Jahr, das der Welle \$ entspricht, analysiert wird, ist mit diesem Gewicht zu multiplizieren. In den Standard-Statistikprogrammen wie SPSS und

¹⁸ Die Personengewichte folgen derselben Logik.

¹⁹ Für Details vergleiche den „Desktop Companion“ (Haisken-DeNew u. Frick 2005).

STATA geschieht dies mit einem einfachen Befehl, bzw. mit einer entsprechenden Befehlsoption.²⁰ Tabelle 4 ist wie folgt zu lesen:

Jede Gewichtungs-Variable wird mit \$xHRFy bezeichnet:

Es bedeuten:

\$ = Wellenkennzeichen A, B, ..., W für die Jahre 1984, 1985, ..., 2006.

x = Unterscheidung nach Haushalten ($x = H$) und Personen ($x = P$)

HRF kennzeichnet die Variable im Datensatz als Hochrechnungsfaktor

y = Eine Zusatzkennung, die die Art des Gewichts beschreibt

y = <leer>, also nicht besetzt, bezeichnet Standardhochrechnungsfaktoren. Standardgewichte umfassen sämtliche Samples mit Ausnahme der Hocheinkommensstichprobe G. Diese Gewichte sind für sämtliche Wellen verfügbar.

y ≠ <leer> bezeichnet vom Standardhochrechnungsfaktor abweichende Varianten.

Die mit $y = 1$ gekennzeichneten Variablen sind normalerweise identisch mit den Standardgewichten; allerdings werden die Gewichte von Stichproben in ihrer ersten Welle auf Null gesetzt. Zur erstmals mit der Datenauslieferung 2007 durchgeführten Null-Setzung der Querschnittsgewichte der samplespezifischen ersten Wellen sei angemerkt: Es ist inzwischen bekannt, dass neue Befragte bei der Erhebung von Merkmalen wie z.B. Lebenszufriedenheit und Haushaltseinkommen in den ersten Wellen eines Panels signifikante Lerneffekte zeigen.²¹ Deswegen empfiehlt es sich, erste Wellen bei der Analyse solcher Variablen nicht in deskriptive Analysen einzubeziehen.²² Durch das Nullsetzen der Hochrechnungsfaktoren geschieht dies gewissermaßen automatisch²³ für alle Teilstichproben – nur nicht für Sample C, da in der damaligen DDR das Einkommen sehr einfach strukturiert war und deshalb eine Analyse auch der ersten Welle sinnvoll erscheint.

y = ALL, umfasst sämtliche erhobene Stichproben des SOEP

y = D kennzeichnet die isolierte Zuwanderer-Stichprobe D

y = G kennzeichnet die isolierte Hocheinkommensstichprobe G

In Tabelle 4 findet sich eine zusammenfassende Darstellung aller verfügbaren Hochrechnungsfaktoren auf Haushaltsebene. Die ausgewiesenen Fallzahlen wie die Summe der Gewichte beziehen sich nur auf die Privathaushalte; Anstaltshaushalte, die – wie in den meisten amtlichen Stichproben – nicht repräsentativ im SOEP erfasst sind, bleiben in der Darstellung unberücksichtigt.

Ausgehend von den Haushaltsgewichten werden Personengewichte bestimmt. Da grundsätzlich sämtliche Personen, die 17 Jahre und älter sind, an der Befragung teilnehmen sollen, gilt für repräsentative Stichproben prinzipiell die Gleichheit von Haushalts- und Personengewicht. Somit können die geschätzten Haushaltsgewichte zunächst 1 : 1 auf sämtliche Personen des Haushalts (einschließlich deren Kinder) übertragen werden. Anschließend erfolgen zwei Korrekturen:

Zum einen werden in der Startwelle Personen, die einen zweiten Wohnsitz haben, nur mit der Hälfte ihres eigentlichen Gewichts versehen, da sie eine doppelte

²⁰ In SPSS mit Hilfe des Befehls „weight by“ und in STATA mit der Befehlsoption „[weight]“.

²¹ Vgl. Frick et al. 2006.

²² Erste Welle-Effekte können indes mit multivariaten Verfahren kontrolliert werden.

²³ Dies hat zur Konsequenz, dass mit Hilfe von \$HRF1 keine Analysen für das Jahr 1984 möglich sind.

Tabelle 4 Hochrechnungsfaktoren für Privathaushalte im Sozio-oekonomischen Panel für die Wellen A–W (1984–2006) – in den Gewichten enthaltene Stichproben und Eckdaten –

Welle	Jahr	\$HHRF	\$HHRF1	\$HHRFALL	\$HHRF ^{y*}	Zahl der Privathaushalte in der Standard-Stichprobe**	Zahl sämtlicher Haushalte**	Hochgerechnete Privathaushalte in der Grundgesamtheit in Tsd.
A	1984	AB	=0			5853	5921	26.076
B	1985	AB	AB			5238	5322	26.367
C	1986	AB	AB			4991	5090	26.739
D	1987	AB	AB			4920	5026	27.006
E	1988	AB	AB			4719	4814	27.402
F	1989	AB	AB			4602	4690	27.793
G	1990	ABC	ABC			6722	6819	34.848
H	1991	ABC	ABC			6581	6699	35.256
I	1992	ABC	ABC			6564	6665	35.700
J	1993	ABC	ABC			6537	6637	36.230
K	1994	ABC	ABC			6459	6559	36.695
L	1995	ABCD	ABC			6656	6768	36.938
M	1996	ABCD	ABCD		D	6591	6698	37.281
N	1997	ABCD	ABCD		D	6508	6617	37.456
O	1998	ABCDE	ABCD		D	7359	7486	37.532
P	1999	ABCDE	ABCDE		D	7905	7215	37.794
Q	2000	ABCDEF	ABCDE		D und F	12.905	13.078	38.123
R	2001	ABCDEF	ABCDEF		D	11.667	11.783	38.455
S	2002	ABCDEF	ABCDEF	ABCDEF	D und G	11.202	12.308	38.720
T	2003	ABCDEF	ABCDEF	ABCDEF	D und G	10.987	11.910	38.945
U	2004	ABCDEF	ABCDEF	ABCDEF	D und G	10.641	11.642	39.121
V	2005	ABCDEF	ABCDEF	ABCDEF	D und G	10.321	11.294	39.178
W	2006	ABCDEFH	ABCDEF	ABCDEF	D und G	11.409	12.361	39.178***

* $y = D$: Zuwanderer. $y = F$ Ergänzung 2000. $y = G$: Hocheinkommensstichprobe.

** Nur Haushalte mit positivem Gewicht.

*** vorläufig. Die Daten basieren auf dem Mikrozensus 2005.

Auswahlwahrscheinlichkeit besitzen. Zum anderen werden die Personengewichte – getrennt nach alten und neuen Ländern – an die Alterstruktur der Personen in Privathaushalten am Hauptwohnsitz an den Mikrozensus angepasst.

In Abb. 2 ist die Entwicklung der erhobenen Fallzahlen der Entwicklung der „effektiven Fallzahlen“ gegenübergestellt. Die so genannten effektiven Fallzahlen berücksichtigen die Varianz der Gewichtungsfaktoren. Bei einer Varianz der Gewichtungsfaktoren von Null wäre die Effizienz der Stichprobe 100% und die effektive Fallzahl gleich der erhobenen Fallzahl. Mit einer Erhöhung der Varianz in den Gewichtungsfaktoren sinkt die Effizienz der Stichprobe und damit auch die effektive Fallzahl, da sie sich aus dem Produkt von erhobener Fallzahl und Effizienz der Stichprobe berechnet. Aus Abb. 2 ersieht man, dass in den letzten Jahren in jeder Welle ca. 12.000 Haushalte befragt wurden. Die zweite Zeitreihe darunter zeigt dagegen die effektiven Fallzahlen. Sie liegen mit Werten um 5000 deutlich unter denen der erhobenen Haushalte. Der Quotient der effektiven Fallzahlen zu den erhobenen kann als Maß für die Effizienz einer Stichprobe dienen. Dieser Quotient, der offenbar stets Werte zwischen 0 und 1 annimmt, lag für die Welle 23 bei 0,42.

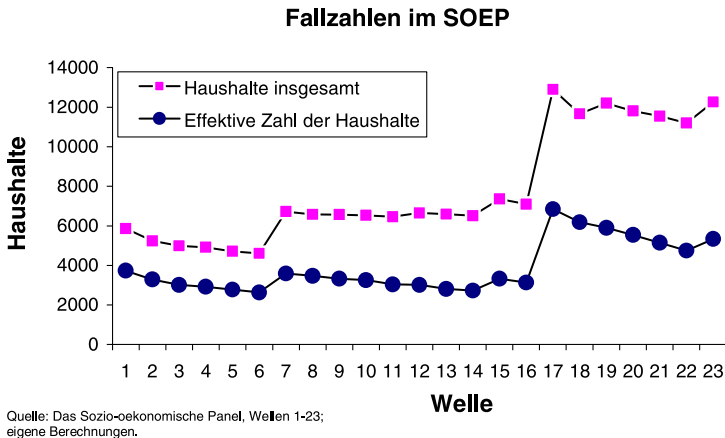


Abb. 2 Entwicklung der Fallzahlen im SOEP nach 23 Wellen

4 Datenstruktur und Dokumentation

Der SOEP-Datensatz besteht in der 2008 für die Nutzer verfügbaren Version aus 297 unterschiedlichen Datensätzen (Files/ Tabellen), die mit 4.751.065 Daten-Zeilen zusammen nahezu 5 Millionen Beobachtungen enthalten, in denen 39.550 Variablen gespeichert sind. Im Grundsatz werden alle direkt erhobenen Personen- und Haushaltsdaten (\$P und \$H)²⁴ sowie die damit korrespondierenden Feldinformationen²⁵ (\$PBRUTTO und \$HBRUTTO) Jahr für Jahr als Querschnittsdatsätze²⁶ abgelegt und als solche auch nahezu unverändert an die Nutzer weitergegeben. Um das Re-Identifikationsrisiko zu minimieren sind von der allgemeinen Weitergabe lediglich Klartextangaben²⁷ und kleinräumige Regional- und Nachbarschaftsinformationen ausgeschlossen. Derartige Informationen können nur unter spezifischen Datenschutzvorkehrungen oder gar nur am DIW Berlin vor Ort ausgewertet werden.

Informationen über Kindergarten- und Schulbesuch zu den noch nicht direkt befragten Kindern werden auf Haushaltsebene erfasst und anhand der für alle Personen

²⁴ Das \$ Zeichen steht hierbei für den wellenspezifischen Präfix, beginnend mit A im Jahr 1984.

²⁵ Feldinformationen liegen für alle Haushalte und Personen vor, die zur Jahres-spezifischen Bruttopopulation gehören, diese setzt sich zusammen aus der Vorjahrs-Population, minus den Gestorbenen, diejenigen die in das Ausland verzogen sind und auch ohne jene Personen die im letzten Jahr endgültig verweigert haben. Hinzu kommen jedoch Neugeborene und sonstige neu im Haushalt lebende Personen.

²⁶ Querschnittsdatsätze sind jahresweise abgelegte Tabellen oder Analysefiles, die durch Identifikatoren (z. B. Personen- und Haushalts-Ids) eindeutig miteinander verknüpft werden können.

²⁷ Die Erhebungsinstrumente enthalten neben standardisierten Fragen auch offene Textfelder, die als Klartext abgelegt werden. Beispiele sind Bildungsabschlüsse, Berufe und Branchenangaben, Bezeichnungen der jeweiligen Kranken- oder Rentenversicherungen sowie offene Angaben über Sorgen in der Bevölkerung. Diese Angaben werden nur in vercodeter Form weitergegeben (Berufe und Branchen werden gemäß der internationalen Standardklassifikationen als ISCO- und NACE-Codes weitergegeben). Die Klartexte selbst können Hinweise enthalten, die das Reidentifizierungsrisiko erhöhen und unterliegen deshalb besonderen datenschutzrechtlichen Schutzmaßnahmen.

im Haushalt bereitgestellten Personeneinträge (\$PBRUTTO) als disaggregierte Kinderinformationen ebenfalls Jahr für Jahr personenbezogen (\$KIND) abgelegt. Diese Daten sind insofern ein Sonderfall, dass sie Informationen von nicht direkt befragten Personen (Kinder unter 16 Jahren) enthalten. Die zugehörigen Fragen (z. B. Art der besuchten Schule) sind zwar vom Haushaltsvorstand beantwortet worden, die Daten werden jedoch auf die Personenebene (der Kinder) transferiert weitergegeben, um Längsschnittanalysen von Kindern und biographische Übergänge von der Kindheit zum Erwachsenenalter zu erleichtern.

Die Variablenbezeichnung verweist im SOEP bei den Befragungsdaten auf die Fragebogennummer und stellt so den direkten Bezug zum Erhebungsinstrument sicher. Darüber hinaus werden aus den jährlich abgelegten Personen- und Haushaltsdaten über die Zeit vergleichbar aufbereitete Daten mit einheitlichen Namen generiert. Diese stehen ebenfalls als jahresbezogene Datensätze zur Verfügung, zu finden sind die personenbezogenen generierten Variablen in den \$PGEN Datensätzen und entsprechend die haushaltsbezogenen generierten Variablen in den \$HGEN Dateien. Darüber hinaus stehen noch die Informationen zum Vorjahr in den \$PKAL Datensätzen zur Verfügung und für international vergleichende Forschung finden sich speziell aufbereitete Variablen in den Dateien \$PEQUIV. Die einfache Struktur der jahresbezogenen Daten in Form einer ($N \times V$)-Datenmatrix (Untersuchungseinheiten mal Variablen) gewährleistet eine einfache und direkte Kommunikation zwischen Datenproduzenten und Datennutzern. Die direkt erhobenen jahresbezogenen Originaldaten bleiben unverändert erhalten, wogegen sich die daraus abgeleiteten „generierten Daten“ infolge veränderter Fragestellungen und/oder Imputationen ändern können (auch rückwirkend für alle Jahre).

Identifikatoren und Populationsabgrenzung

Die Verknüpfung der Querschnittsdaten erfolgt über anonymisierte „Identifikatoren“, die auf Personen- und Haushaltsebene für alle im Laufe der Erhebung erfassten Untersuchungseinheiten in zentralen Datensätzen (PPFAD, HPFAD) abgelegt sind. Diese Daten bilden den Ausgangspunkt für die Verknüpfung der wellenspezifischen Informationen. Personenbezogene Daten enthalten immer auch Identifikatoren der höher aggregierten Einheiten. Auf diese Weise können Haushalts- und Personeninformationen durch Aggregation oder Disaggregation flexibel miteinander verknüpft werden. Im SOEP werden pro Erhebungsjahr drei grundlegende Identifikatoren unterschieden, die sich bei spezifischen Datensätzen noch weiter differenzieren (Übersicht 2): die so genannte Case-Id (Ursprungshaushaltsnummer), die aktuelle Haushaltsnummer, sowie die unveränderliche Personennummer.

Die Case-ID [HHNR] ist die grundlegende Datenbankeinheit des SOEP. Sie beinhaltet die zuerst vergebene Haushaltsnummer für jeden Haushalt eines jeden Samples und ist im ersten Erhebungsjahr identisch mit der aktuellen Haushaltsnummer. Im Gegensatz zur aktuellen Haushaltsnummer bleibt die Ur-Haushaltsnummer auch bei Haushaltsabsplattungen als Ursprungsinformation unverändert erhalten. Die Case-ID ermöglicht so nicht nur, Personen- und Haushaltsabsplattungen auf die ursprüngliche Ziehungseinheit zurückzuverfolgen, auf dieser Ebene werden unter Heranziehen

der Sample-Point-Information²⁸ auch Varianzschätzer zur Bestimmung von Konfidenzbändern (Randomgroup-/Jackknife-/Bootstrap-Verfahren) gebildet.

Die wellenspezifische aktuelle Haushaltsnummer (\$HHNR) ist erforderlich, um Personendaten mit den jeweils aktuellen Haushaltsinformationen zu verknüpfen. Infolge von Aus- und Umzügen variiert dieser Schlüssel jedoch über die Zeit.

Die unveränderliche Personennummer ist der zentrale Identifikator von Personen. Er wird zusammen mit den anderen Identifikatoren als Primärschlüssel für jede Person beim Eintritt in den Datenbestand des SOEP generiert und ist unveränderlich, unabhängig davon, ob diese Person einen Fragebogen ausgefüllt hat oder nicht, also auch für Kinder ab 0 Jahren. Dieser Identifikator ist unerlässlich für personenbezogene Verknüpfungen über die Zeit.

Für spezifische zeitabhängige Datenformate (vgl. Übersicht 2) sind weitere Identifikatoren wie das Erhebungsjahr oder die fortlaufende Nummer des Ereignisses pro Person (Spellnummer) notwendig, um die Daten auf verschiedenen Ebenen eindeutig miteinander verknüpfen zu können.

Übersicht 2 – Identifikatoren im SOEP

Name	Beschreibung	Erläuterung
Case-Id [HHNR]	Unveränderliche HH-Nummer (Zuerst vergebene HH-Nummer je Sample)	bezeichnet den ursprünglich gezogenen Haushalt, aus dem sich alle nach Abspaltung alle weiteren Haushalte ableiten
HID [\$HHNR]	Aktuelle HH-Nummer	gewährleistet die Verknüpfung zwischen Haushalten sowie von Personen und Haushalten im jeweiligen Jahr
PID [PERSNR]	Unveränderliche Personennummer	erlaubt die Verknüpfung von Personeninformationen über die Zeit
[SPELLNR]	weitergehende Differenzierungen der Population nach Ereignissen	erlaubt die Verknüpfung von Personen- und Ereignisdaten
SVYYEAR [ERHEBJ]	Erhebungsjahr	Zusätzlicher Verknüpfungsschlüssel bei zeitabhängigen Datenformaten (long-form)

Die wellenübergreifenden Dateien PPFAD und HPFAD, die im Zuge der Datenaufbereitung generiert werden, beinhalten neben den Identifikatoren auch zeitlich variable Angaben zum Einsatz von Erhebungsinstrumenten und zur Verfügbarkeit von personen- oder haushaltsbezogenen Informationen sowie zeitinvariante Angaben (Stichprobenzugehörigkeit, Geschlecht, Geburtsjahr). Zudem werden für jede Person deren Eintritt und Erstbefragung festgehalten sowie deren Letztbefragung und der Austritt aus der Panelpopulation (durch Verweigerung oder Tod). Die Letztbefragung wird dabei laufend aktualisiert. Die Datensätze PPFAD und HPFAD umfassen kumulativ alle im SOEP realisierten Personen und Haushalte. In gleicher Weise sind auch die Hochrechnungsfaktoren (PHRF, HHRF) abgelegt.

Die Populationsabgrenzung erfolgt mit Hilfe der Variablen \$NETTO und \$POP: Erstere verweist auf die Art der verfügbaren Personeninformationen (Gehört die Person im aktuellen Jahr zur Panelpopulation? Liegt ein Personeninterview vor? Gab es

²⁸ Sample-Points bezeichnen die ursprünglichen Ziehungseinheiten, ausgehend von zufällig repräsentativ ausgewählten Startadressen. Im Durchschnitt werden etwa acht Haushalte je Startadresse zufällig „gezogen“.

weitere Fragebögen, die von der Person zusätzlich ausgefüllt wurden? Liegt nach einem temporären Ausfall ein nacherhobener Lückefragebogen vor? etc.). Die Variable \$POP grenzt zudem die Personen in Privathaushalten von der im Zuge der weiteren Panelbefragung erfassten Anstaltsbevölkerung (z. B. Altenheim) ab und differenziert nach der Nationalität der Bezugsperson im Haushalt (Haushaltsvorstand).

Dokumentation und Metadaten

Ein komplexer Datensatz wie das SOEP ist für einen potentiellen Nutzer nur dann sinnvoll zu nutzen, wenn eine extensive Dokumentation der Daten vorliegt. Für das SOEP ist die komplette Dokumentation über die Internetseite www.diw.de/soep frei zugänglich. Diese Seite beinhaltet unter anderem eine einführende Zusammenfassung (der „SOEP Desktop Companion“), die genaue Dokumentation aller generierten Variablen, die genutzten Fragebögen, Methodenberichte des Feldinstituts und insbesondere die interaktive Webanwendung SOEPinfo. SOEPinfo gibt unter anderem Auskunft über die Häufigkeitsverteilungen aller Variablen und darüber, welche Variablen über die Zeit vergleichbar erhoben wurden (Variablenkorrespondenz); aus den interaktiv zusammengestellten Variablenlisten werden ferner für alle gängigen Auswertungsprogramme (SPSS, SAS, STATA) fertige Syntaxdateien zur Aufbereitung der SOEP Daten bereitgestellt. Außerhalb der SOEP-Survey Gruppe gibt es eine externe von John Haisken-DeNew geschriebene Erweiterung des Statistikpakets STATA mit dem Namen PanelWhiz. Diese Sammlung von ado-files ist explizit für den Umgang mit Paneldaten geschrieben worden und erleichtert den Umgang mit derartigen Datensätzen gerade für Anfänger beträchtlich. PanelWhiz unterstützt neben dem SOEP zudem noch vier weitere Paneldatensätze (siehe: <http://www.panelwhiz.eu/>).

5 Ausblick

Dieser Ausblick²⁹ gibt zum ersten einen knappen Überblick über die Weiterentwicklung des für die Analyse der SOEP-Daten zur Verfügung gestellten Service. Zum zweiten werden Neuerungen bei der Erhebung der Daten skizziert.

Weiterentwicklung des SOEP-Services

Derzeit wird an einer Umstrukturierung der Daten in eine noch stärker nutzungsfreundliche Form gearbeitet. Dazu werden die Jahr für Jahr abgelegten Variablen über die Zeit vergleichbar aufbereitet. Die derzeit überwiegend jährlich bzw. wellenspezifisch abgelegten Querschnittsdatenfiles werden in konsistenter Art und Weise miteinander verknüpft und in ein einheitliches Datenformat überführt („Long-Format“), welches Auswertende effektiv dabei unterstützt, die Panelkomponente der SOEP Daten effizient zu nutzen. Dies wird zu einer wesentlichen Erleichterung bei der Datennutzung und damit zu einer signifikanten Verbesserung des SOEP-Service führen.

²⁹ Vgl. dazu und für das Folgende Anger et al. (2008).

Ab dem Jahr 2009 werden die Mikro-Daten einer 2007 durchgeführten Drop-Out-Erhebung allgemein verfügbar sein, bei der die Wiedergewinnung von Personen nach Teilnahmeausfällen (in der Regel durch Verweigerung) geprüft wurde. Dann werden auch die Ergebnisse einer 2007 durchgeführten Interviewer-Befragung vorliegen, die die Daten über die Interviewer, die aus der Buchhaltung des Umfrageinstituts stammen³⁰, ergänzen. Dies wird deutlich vertiefte Analysen von Interviewereffekten ermöglichen.

In den letzten Jahren wurde der Zugriff von Nutzerinnen und Nutzern auf datenschutzrechtlich sensitive Regionaldaten bzw. geo-kodierte Daten etwa zum Umfeld einer Wohnung immer weiter verbessert. Diese Arbeiten werden auch in den nächsten Jahren fortgeführt. Mit Hilfe von SOEPremote, einem Software-gestützten und automatisierten remote Zugriff per email ist es bereits seit 2006 möglich, über eine sichere Fernrechenverbindung Analysen durchzuführen, ohne direkt am DIW in Berlin vor Ort sein zu müssen. Dadurch sind etwa auch kleinräumige Regionalinformationen auf Ebene der Kreise, die aus datenschutzrechtlichen Gründen nicht Teil der Standarddatenlieferung auf DVD sein können, von auswärts zugänglich. Durch den sicheren Zugriff über diese spezielle Software ist gewährleistet, dass datenschutzrechtlich sensitive Mikrodaten das DIW Berlin nicht verlassen.³¹ Die Berliner SOEP-Surveygruppe arbeitet derzeit daran, relativ lange Rechenzeiten und Kapazitätsengpässe im Umgang mit SOEPremote zu verbessern.

Bei der letzten Nutzerbefragung stellte sich heraus, dass insbesondere junge Wissenschaftler und Doktoranden den Umgang mit den SOEP-Daten als kompliziert und schwer verständlich empfinden. Deshalb kreierte die SOEP-Gruppe am DIW Berlin im Jahr 2007 das neue Service-Element SOEPcampus. Dieses Instrument soll auch dazu beitragen, dass SOEP Daten bereits im Rahmen der Statistik-Ausbildung genutzt werden. Das Angebot richtet sich vorwiegend an Studierende und Promovierende in den Bereichen Ökonomie, Soziologie sowie Psychologie. SOEPcampus, ein durchaus zeitaufwändiges Instrument zur Schulung und Förderung des Umgangs mit der komplexen Datenstruktur des SOEP, wird in den kommenden Jahren ausgebaut werden. In Kooperation mit Wissenschaftlerinnen und Wissenschaftlern anderer Institute und Universitäten in ganz Deutschland bietet die Berliner SOEP-Gruppe zudem praktische Kurse und Vorlesungen zu Nutzung und Analysemöglichkeiten des SOEP-Surveys an.

Es sei ausdrücklich darauf hingewiesen, dass Gäste für das Erlernen des fortgeschrittenen, komplexen Datenhandlings bei der SOEP-Gruppe am DIW Berlin herzlich willkommen sind. Dies gilt insbesondere auch für studentische Praktikantinnen und Praktikanten, denen gute Arbeitsbedingungen geboten werden können.

Weiterentwicklung des SOEP-Surveys

An dieser Stelle kann aus Platzgründen nicht über sämtliche gegenwärtig laufende und bereits in der Implementierungsphase befindliche Weiterentwicklungen des SOEP-Surveys berichtet werden. Inhaltliche Verbesserungen sind ebenso in Arbeit

³⁰ Z. B. Geschlecht und Alter (vgl. Schräpler und Wagner 2000).

³¹ Eine Beschreibung des Vorgehens findet sich in Goebel (2005).

wie eine Vielzahl erhebungsmethodischer Verbesserungen. Einige dieser Verbesserungen sollen – nach grundsätzlichen Überlegungen zur „Governance“ des SOEP – im Folgenden kurz dargestellt werden.

Ein Schwerpunktziel der nächsten Jahre wird es sein, den SOEP-Survey gemeinsam mit der wachsenden Zahl an Forscherinnen und Forschern, gezielt weiter zu entwickeln beziehungsweise spezifisch zu ergänzen. Dazu soll ein „Innovations-Ausschuß“ geschaffen werden. Wobei wie bislang die meisten Innovationsvorschläge ohne aufwändige formalisierte Abstimmungsprozesse umgesetzt und realisiert werden sollen. Der Wissenschaftliche Beirat des SOEP („Survey Beirat“) dient – wie bislang – in erster Linie der Qualitätssicherung.

Die SOEP-Gruppe am DIW Berlin ist interdisziplinär besetzt und umfasst neben den unmittelbar mit der Durchführung des Surveys befassten Mitarbeitern auch Studenten, Doktoranden sowie affilierte Forscher. Die SOEP-Surveygruppe arbeitet seit Jahren zusammen mit externen am SOEP interessierten Wissenschaftlerinnen und Wissenschaftlern an der allgemeinen und speziellen Weiterentwicklung des SOEP. Am Beginn des SOEP stand die Zusammenarbeit mit dem Sonderforschungsbereich (Sfb) 3, der anfänglich die intellektuelle Führung des SOEP inne hatte. Mit dem planmäßigen Auslaufen des Sfb 3 im Jahr 1991 ging dieses Forum jedoch verloren. Der SOEP-Beirat, dem vor allem eine Kontrollfunktion oblag, konnte und wollte diese Lücke nicht schließen, schon allein deswegen, weil er nur einmal pro Jahr tagt. Die SOEP-Gruppe konnte bislang auf Basis ihrer eigenen Forschungsarbeiten auch das inhaltliche Spektrum der SOEP-Erhebung abdecken³². Die SOEP-Surveygruppe in Berlin kann jedoch nicht dauerhaft die voraussichtlich notwendige Kompetenz in sämtlichen Forschungsfeldern abdecken. Und im Hinblick auf die eigene Forschung ist eine gewisse Spezialisierung bzw. Konzentration auf ausgewählte Schwerpunktfelder notwendig.

Nach dem Planungsstand im Herbst 2008 wird angestrebt einen „Innovations-Ausschuss“ einzurichten, der sich – wie der ursprüngliche „Panel-Ausschuss“ des Sfb 3 – aus externen Expertinnen und Experten und Mitgliedern der SOEP-Gruppe zusammensetzen soll. In diesem Ausschuss sollen sich externe wie interne Forscherinnen und Forscher für längere Zeit für die konkrete Weiterentwicklung der SOEP-Erhebung engagieren. Durch diese stärkere Formalisierung von Entscheidungsprozessen wird angestrebt, die Transaktionskosten zu senken und vor allem stärkere Anreize für innovative Ideen von Nutzerinnen und Nutzern zu setzen.

Bei der Erhebung von Lebensläufen, die ja den Fokus der SOEP-Surveys bilden, steht die Zeitverwendung insbesondere in Hinblick auf Erwerbstätigkeit nach wie vor grundsätzlich im Vordergrund. Die Zeitverwendung dominiert nämlich die Analysen, die weltweit mit dem SOEP durchgeführt werden. Künftig soll – angeleitet durch neueste Entwicklungen in Ökonomie, Soziologie und Psychologie – eine noch umfassendere Beschreibung der Zeitverwendung erfolgen. Dazu soll ihre Bedeutung für das Erreichen persönlicher Ziele³³ durch Erweiterung der Surveyinstrumente systematischer als bislang auch mit Zielen, die Eltern für ihre Kinder verfolgen,

³² Für die Entwicklung der Kinder- und Jugendfragebögen wurde gezielt Kompetenz von außen geholt (vgl. Schupp et al. 2008).

³³ Nicht nur in Bezug auf Einkommen, sondern auch in Bezug auf Lebenszufriedenheit.

kombiniert werden können. Werte und Einstellungen, die wenig mit der individuellen Zeitverwendung zu tun haben wie z. B. politische Einstellungen und Präferenzen bezüglich Umwelt und Umweltschutz sollen nach wie vor am Rande erhoben werden.

Da die individuelle bzw. intergenerationelle Lebenslaufperspektive immer mehr in den Vordergrund der Analysen tritt, die mit den Daten des SOEP unternommen werden, wird auch immer deutlicher, dass Beginn und Ende von beobachteten Lebensläufen genauer erhoben werden sollten. Es geht also nicht mehr nur darum, dass das Leben von der „Wiege bis zu Bahre“ beobachtet wird. Neuere theoretische, gesellschaftliche und forschungsstrategische Entwicklungen machen es notwendig, dass Lebensläufe von der Schwangerschaft bis zum Sterbeprozess und schließlich den Erinnerungen an Verstorbene erhoben werden (von der „Zeugung bis zur Hinterbliebenenrente“).

Für die verschiedenen Lebensbereiche sollen internationalen Entwicklungen folgend systematischer als bisher die biologischen Grundlagen einerseits („Bio-Marker“ und Persönlichkeitsmerkmale) und andererseits Netzwerke, in denen sich einzelne Personen, Familien und Haushalte befinden, erhoben werden. Diese systematischere Betrachtung wird stark von neuen technologischen Möglichkeiten der Messung und Analyse getrieben. Dazu gehören die Möglichkeit Erhebungen über Internet und Mobiltelefon durchzuführen sowie Auswertungsmöglichkeiten für Biomarker. Ein derart umfassender „Mess“-Anspruch kann offensichtlich leicht zu einer Überforderung von Befragten und zu sinkenden, und vor allem selektiven Ausschöpfungsquoten führen. Insofern werden wir neue surveymethodische Ansätze wie einen konsequenten „Multi-Method-Approach“ und „Matrix Sampling“ prüfen. Beim Matrix-Sampling werden sozusagen bewusst fehlende Werte erzeugt, die später durch Imputationen aufgefüllt werden, indem zufällig gesteuert bestimmte Fragen nur bei Subsamples gestellt werden, um so die Befragungslast zu reduzieren.

Letztlich wird nichts anderes angestrebt als jenes umfassende Erhebungsprogramm zu verwirklichen, das bereits in der klassischen sozialwissenschaftlichen Erhebung bei den „Arbeitslosen von Marienthal“ entwickelt und ansatzweise umgesetzt wurde (vgl. Jahoda et al. 1933). Die Methodik dieser Studie wurde jedoch im Zuge des Erfolgs der standardisierten Survey-Forschung seit den 70er Jahren zunehmend vergessen.

Im Folgenden werden ausgewählte Themenbereiche angesprochen, für die die Weiterentwicklung der SOEP-Erhebungsinstrumente im Hinblick auf eine genauere Abbildung des Lebensverlaufs besonders relevant ist.

Schwangerschaft und Kindheit: Haushaltspanels sind konzeptionell besonders gut geeignet, Lebensläufe von Geburtskohorten nachzuzeichnen, da in einem laufenden Haushaltspanel die Stichprobenmitgliedschaft nicht erst mit der Geburt beginnt (wie das bei konventionellen Kohortenstudien der Fall ist), sondern die „virtuelle“ Stichprobenmitgliedschaft aus der vorherigen Teilnahme eines oder beider Elternteile resultiert.

Einkommen, Sparen und Vermögen: Die Erfahrung mit reichhaltigen Längsschnittsdaten in den Bereichen Erwerbstätigkeit und Einkommen hat gezeigt, dass konzeptionelle Aussagekraft und empirische Messbarkeit dieser vermeintlich klar messbaren „objektiven“ Merkmale durchaus problematisch sind. Internationalen Ent-

wicklungen folgend soll in den nächsten Jahren geprüft und getestet werden wie die Erhebung von Einkommen, Sparen (und eventuell Konsum) sowie Vermögen verbessert werden kann.

Kompetenzen: Im SOEP – wie in fast allen anderen einschlägigen Surveys – beschränkte sich die Operationalisierung von „Humankapital“ lange Zeit auf die Erhebung von Bildungsabschlüssen wie des höchsten erworbenen schulischen wie beruflichen Bildungszertifikates. Seit der Einführung des SOEP-Jugendfragebogens sowie einer im Jahr 2001 erfolgten Revision des Biographiefragebogens wird systematisch versucht, anhand standardisierter Messverfahren bei Befragten kognitive Kompetenzen in Form von Tests zu bestimmen. Zudem erfolgt eine schrittweise Erweiterung bei der Erhebung sogenannter nicht-kognitiver Fähigkeiten, also von Kompetenzen, die auch außerhalb von Bildungsinstitutionen erworben werden, also insbesondere in der frühen Kindheit im Elternhaus. Auch in diesem Bereich wurde das SOEP-Erhebungsprogramm in den letzten Jahren schrittweise erweitert. Diese Erweiterungen sollen jetzt in einem Netzwerk der Leibniz-Gemeinschaft (WGL) weiter vorangebracht werden.³⁴

Gesundheit und biologische Grundlagen sozialen- und ökonomischen Verhaltens: Trotz des zunehmenden Interesses Surveys mit „Bio-Markern“ zu verknüpfen, erscheint ein von medizinischen Fragestellungen getriebener schrankenloser Ausbau der Erhebung von so genannten Biomarkern für das SOEP nicht sinnvoll. Unter Biomarker werden dabei individuelle biologische Merkmale eines Befragten wie etwa seine Handgreifkraft oder sein Blutdruck verstanden. Biomarker, die die sozial- und verhaltenswissenschaftlichen Analysen ergänzen und zum Teil enorm vertiefen, sind sicherlich grundsätzlich sinnvoll. Das SOEP kann und soll aber nicht einen eigenständigen Gesundheits-Survey ersetzen. Dies würde zudem die Befragten mit großer Sicherheit inhaltlich und zeitlich überfordern und auch die sinnvolle Arbeitsteilung verschiedener methodischer Ansätze und Surveys behindern.³⁵

Sterben und Erinnerungen: Über Geburten stellt das SOEP seit 2002 differenzierte Informationen zur Verfügung; nicht jedoch über den Tod von Stichprobenmitgliedern. Für Analysen sind aber beide Enden von Lebensläufen gleichermaßen von großer Bedeutung. Deswegen wurde in einem Pretest im Jahr 2007 erstmals getestet, Informationen über den Tod von Angehörigen zu erheben. Dabei wurde beispielsweise erfragt, ob in den letzten zwölf Monaten jemand verstorben ist und wer und ob die Person im selben Haushalt gelebt hat. Darüber hinaus wurden Fragen hinsichtlich der Enge der Beziehung zu der verstorbenen Person gestellt und die Todesursache erfragt. Sofern die verstorbene Person in der Vergangenheit selbst SOEP Befragte war, wird es aufgrund dieser Befragungstechnik erstmals möglich sein, Längsschnittsinformationen von Befragten mit Informationen zu ihrem Tod zu verknüpfen. Auf diese Art und Weise werden ab dem Erhebungsjahr 2009 wichtige

³⁴ Vgl. Forschungsnetzwerk „Nicht-kognitive Fähigkeiten: Erwerb und ökonomische Konsequenzen“. Für nähere Informationen siehe <http://www.zew.de>.

³⁵ Des weiteren besteht grundsätzlich die Gefahr, dass Befragte bei der Erhebung von zu sensitiven „Bio-Markern“ – wie beispielsweise bei Blutentnahmen – nicht mehr weiter bereit sind, an der Wiederholungsbefragung teilzunehmen. Da es von außerordentlicher Wichtigkeit ist, die „Panelmortalität“ so gering wie möglich zu halten, bedarf es deshalb eines sehr vorsichtigen Vorgehens hinsichtlich der Verknüpfung von individuellen Befragungsdaten mit „Bio-Markern“.

neue Analysen hinsichtlich des Zusammenhangs von Lebensführung, sozialem Umfeld und Todesumständen möglich werden.

Schluss

Die wissenschaftsgetragene Panel-Studie SOEP wird nach 25 Erhebungswellen (seit 1984) so intensiv und breit genutzt wie noch nie. Inzwischen hat sich auch weltweit die Erkenntnis durchgesetzt, dass Längsschnittstudien Kausalanalysen möglich machen. Dadurch ergeben sich für die Infrastruktureinrichtung SOEP (mindestens) zwei Herausforderungen.

Zum ersten stellt sich die Frage, ob und wie sozial- und wirtschaftswissenschaftliche Interventionsstudien mit dem SOEP verbunden werden können bzw. ob und wie Interventionsstudien das SOEP als „Kontrollstichprobe“ nutzen können.

Zum zweiten stellt sich die Frage, wie sich Infrastruktureinrichtungen wie das SOEP, die langfristig und für eine breite (interdisziplinäre) Nutzung angelegt sind, stärker spezifischen (theoriegeleiteten) Fragestellungen öffnen können. Bislang wird mit Studien wie dem SOEP (durchaus erfolgreich) versucht, möglichst vielen potentiellen Nutzern interessante Daten zur Verfügung zu stellen. Darunter leidet naturgemäß die theoretische und methodische Spezifität der Fragestellungen und der Datenerhebung. Die Längsschnittstudie SOEP wird sich in den nächsten Jahren speziellen Fragestellungen ausdrücklich öffnen und die Governance des Surveys so organisieren, dass gezielte Fragestellungen in die verschiedenen Teilstichproben des SOEP effektiv eingebracht werden können. Auch soll ein methodischer Service für andere Studien angeboten werden, die das SOEP als „Kontrollstichprobe“ nutzen wollen.

Danksagung Wir bedanken uns bei unseren Kolleginnen und Kollegen der SOEP-Survey-Gruppe am DIW Berlin für die gute Zusammenarbeit und für viele Kommentare zu früheren Fassungen dieses Aufsatzes. Einem anonymen Referee und insbesondere dem Herausgeber dieser Zeitschrift danken wir für außergewöhnlich konstruktive Kritik und konkrete Verbesserungsvorschläge.

Open Access Dieser Artikel wird zu den Bedingungen der “Creative Commons Attribution Noncommercial License” zur Verfügung gestellt. Damit ist eine nichtkommerzielle Nutzung, Verbreitung und Vervielfältigung erlaubt, sofern die Autoren des Artikels und die genaue Quelle angegeben sind.

Literatur

- Andreß H-J, Bröckel M (2007) Income and Life Satisfaction After Marital Disruption in Germany. *J Marriage Fam* 69(2):500–512
- Andreß H-J, Borgloh B, Güllner M, Wilking K (2003) Wenn aus Liebe rote Zahlen werden – Über die wirtschaftlichen Folgen von Trennung und Scheidung. Wiesbaden
- Anger S et al (2008) Zur Weiterentwicklung von SOEPsurvey und SOEPservice. *Vierteljahrsh Wirtschaftsforsch* 77(3):157–177
- Beaudry P, Green DA (2003) Wages and Employment in the US and Germany: What Explains the Differences? *Am Econ Rev* 93(3):573–602
- Boersch-Supan A, Jürges H (2006) Early Retirement, Social Security and Well-Being in Germany. NBER Working Paper 12303, Cambridge, National Bureau of Economic Research (NBER)
- Burkhauser RV, Giles P, Lillard DR, Schwarze J (2005) Until Death Do Us Part: An Analysis of the Economic Well-Being of Widows in Four Countries. *J Gerontol Ser B Soc Sci* 60(5):238–246

- Butz WP, Boyle Torrey B (2006) Some Frontiers in Social Science. *Science* 312:1898–1900
- Casey T, Dustmann C (2007) Intergenerational Transmission of Language Capital and Economic Outcomes. In: IZA DP No. 3074, Bonn, Institute for the Study of Labor (IZA)
- Cawley J, Spieß CK (2008) Obesity and skill attainment in early childhood. *Econ Hum Biol*, in Druck
- Clark AE, Diener E, Georgellis Y, Lucas RE (2008) Lags and Leads in Life Satisfaction: A Test of the Baseline Hypothesis. *Econ J* 118(529):F222–F243
- Clark AE, Frijters P, Shields M (2008) Relative Income, Happiness and Utility: An Explanation for the Easterlin Paradox and Other Puzzles. *J Econ Lit* 46(1):95–144
- Cooke LP (2007) Persistent Policy Effects on the Division of Domestic Tasks in Reunified Germany. *J Marriage Fam* 69(4):930–950
- Diener E, Lucas RE, Scollon CN (2006) Beyond the hedonic treadmill: Revising the adaptation theory of well-being. *Am Psychol* 61(4):305–314
- Drever AI, Hoffmeister O (2008) Immigrants and Social Networks in a Job-Scarce Environment: The Case of Germany. *Int Migr Rev* 42(2):425–444
- Erlinghagen M (2007) Soziales Engagement im Ruhestand: Erfahrung wichtiger als frei verfügbare Zeit. *DIW Wochenber* 74(39):565–570
- Fertig M, Tamm M (2008) Die Verweildauer von Kindern in prekären Lebenslagen. In: Bertram H (Hrsg) *Der UNICEF-Bericht zur Lage der Kinder in Deutschland*. CH Beck, München, S 152–166
- Frick JR, Grabka MM (2003) Imputed Rent and Income Inequality: A Decomposition Analysis for the Great Britain, West Germany and the US. *Rev Income Wealth* 49(4):503–537
- Frick JR, Goebel J, Schechtman E, Wagner GG, Yitzhaki S (2006) Using Analysis of Gini (ANoGi) for detecting whether two sub-samples represent the same universe: The German Socio-Economic Panel Study (SOEP) Experience. *Sociol Method Res* 34(4):427–468
- Frick JR, Jenkins JP, Lillard DR, Lipps O, Wooden M (2008) Die internationale Einbettung des Sozio-oekonomischen Panels (SOEP) im Rahmen des Cross-National Equivalent File (CNEF). *Vierteljahrsh Wirtschaftsforsch* 77(3):110–129
- Galler HP (1987) Zur Längsschnittgewichtung des Sozio-oekonomischen Panels. In: Krupp H-J, Hanefeld U (Hrsg) *Lebenslagen im Wandel: Analysen 1987*, Band 2 der Reihe: Sozio-oekonomische Daten und Analysen für die Bundesrepublik Deutschland. Campus Verlag, Frankfurt, S 295–317
- Gerhards J, Hans S (2006) Zur Erklärung der Assimilation von Migranten an die Einwanderungsgesellschaft am Beispiel der Vergabe von Vornamen. *DIW Discussion Paper No. 583*. German Institute for Economic Research, Berlin (DIW Berlin)
- Gerstorff D, Ram N, Estabrook R, Schupp J, Wagner GG, Lindenberger U (2008) Life Satisfaction Shows Terminal Decline in Old Age: Longitudinal Evidence from the German Socioeconomic Panel Study. *Dev Psychol* 44(4):1148–1159
- Goebel J (2005) Job submission instructions for the SOEPremote System at DIW Berlin. www.diw.de/documents/dokumentenarchiv/17/44069/soepremote2005.pdf
- Goebel J et al (2008) Mikrodaten, Gewichtung und Datenstruktur der Längsschnittstudie Sozio-oekonomisches Panel (SOEP). *Vierteljahrsh Wirtschaftsforsch* 77(3):77–109
- Glowsky D (2007) Staatsbürgerschaft als Ressource bei der Heirat ausländischer Frauen – eine Analyse mit Daten des Sozio-oekonomischen Panel. *Z Soziol (ZfS)* 36(4):282–330
- Grabka MM (2007) Codebook for the \$PEQUIV File 1984–2006 – CNEF Variables with Extended Income Information for the SOEP. *DIW Berlin Data Documentation No 21*
- Haisken-DeNew J, Frick JR (2005) DTC, Desktop Companion to the German Socio-Economic Panel (SOEP). www.diw.de/deutsch/sop/service/dtc/dtc.pdf
- Headey B (2008) Life Goals Matter to Happiness: A Revision of Set-Point Theory. *Soc Indic Res* 86(2):213–231
- Horwitz D, Thompson D (1952) A Generalisation of Sampling without Replacement From a Finite Universe. *J Am Stat Assoc* 47:663–685
- Jahoda M, Lazarsfeld PF, Zeisel H (1933) *Die Arbeitslosen von Marienthal*. S Hirzel, Leipzig
- Jürges H (2006) Gender Ideology, Division of Housework, and the Geographic Mobility of Families. *Rev Econ Househ* 4(4):299–323
- Jürges H (2007) Unemployment, life satisfaction and retrospective error. *J R Stat Soc Ser A Stat Soc* 170(1):43–61
- Kroh M, Spieß M (2008) Documentation of Sample Sizes and Panel Attrition in the German Socio Economic Panel (SOEP) (1984 until 2007). *DIW Berlin Data Documentation No 39*
- Krupp H-J (2008) Die Anfänge: Zur Entstehungsgeschichte des SOEP. *Vierteljahrsh Wirtschaftsforsch* 77(3):15–26

- Lois D (2008) Einflüsse von Lebensstilmerkmalen auf den Übergang in die erste Ehe. *Z Familienforsch* 20(1):11–33
- Lucas RE (2005) Time Does Not Heal All Wounds: A Longitudinal Study of Reaction and Adaptation to Divorce. *Psychol Sci* 16(12):945–950
- Lucas RE, Clark A, Georgellis Y, Diener E (2004) Unemployment alters the set point for life satisfaction. *Psychol Sci* 15:8–13
- Lucas RE, Clark AE (2006) Do People Really Adapt to Marriage? *J Happiness Stud* 7(4):405–426
- Pannenberg M et al (2005) Sampling and Weighting, chapt 5. In: Haisken-DeNew J, Frick JR (eds) DTC, Desktop Companion to the German Socio-Economic Panel (SOEP). www.diw.de/deutsch/sop/service/dtc/dtc.pdf
- Rehdanz K, Maddison D (2008) Local environmental quality and life-satisfaction in Germany. *Ecol Econ* 64(4):787–797
- Rendtel U (1995) *Lebenslagen im Wandel: Panelausfälle und Panelrepräsentativität*. Campus, Frankfurt/New York
- Rendtel U, Pötter U (1993) Über Sinn und Unsinn von Repräsentativstudien. *Allg Stat Arch (AStA)* 77(3):260–280
- Romeu GL (2006) Effects of short- and long-term unemployment on health satisfaction: evidence from German data. *Appl Econ* 38(20):2335–2350
- Scherger S (2007) Destandardisierung, Differenzierung, Individualisierung – Westdeutsche Lebensläufe im Wandel. VS Verlag für Sozialwissenschaften, Wiesbaden
- Scherger S (2008) Flexibilisierte Lebensläufe? Die Dynamik von Auszug und erster Heirat. In: Szydlík M (Hrsg.) *Flexibilisierung. Folgen für Arbeit und Familie*. VS Verlag, Wiesbaden, S 193–212
- Schimmack U, Schupp J, Wagner GG (2008) The Influence of Environment and Personality on the Affective and Cognitive Component of Subjective Well-Being. *Soc Indic Res*
- Schnitzlein DD (2008) Verbunden über Generationen – Struktur und Ausmaß der intergenerationalen Einkommensmobilität in Deutschland. In: *SOEPPapers* 80. DIW Berlin
- Schräpler J-P, Schupp J, Wagner GG (2008) Who Are the Nonrespondents? An Analysis Based on a New Subsample of the German Socio-Economic Panel (SOEP) including Microgeographic Characteristics and Survey-Based Interviewer Characteristics. DIW Berlin Research Note (in Vorbereitung)
- Schräpler J-P (2004) Respondent Behavior in Panel Studies – A Case Study for Income Nonresponse by Means of the German Socio-Economic Panel (SOEP). *Sociol Method Res* 33(1):118–156
- Schräpler J-P, Wagner GG (2000) Das Verhalten von Interviewern – Darstellung und ausgewählte Analysen am Beispiel des „Interviewer-Panels“ des Sozio-oekonomischen Panels. *Allg Stat Arch (AStA)* 85(1):45–66
- Schräpler J-P, Wagner GG (2005) Characteristics and impact of faked interviews in surveys – An analysis of genuine fakes in the raw data of SOEP. *Allg Stat Arch (AStA)* 89(1):7–20
- Schräpler J-P (2007) A Study of Mode-Effects of a Change from PAPI to CAPI. *Schmollers Jahrb* 127(1):113–125
- Schupp J, Wagner GG (1995) Die Zuwanderer-Stichprobe des sozio-oekonomischen Panels (SOEP). *Vierteljahrsh Wirtschaftsforsch* 64(1):17–26
- Schupp J, Spieß CK, Wagner GG (2008) Die verhaltenswissenschaftliche Weiterentwicklung des Erhebungsprogramms des SOEP. *Vierteljahrsh Wirtschaftsforsch* 77(3):63–76
- Spieß M (2001) Derivation of design weights: The case of the German Socio-Economic Panel (GSOEP). DIW-Materialien/Research Notes No 5
- Spieß M et al (2008) On the Treatment of Non-Original Sample Members in the German Household Panel Study (SOEP) – Tracing and Weighting. Data Documentation des DIW Berlin
- Stutzer A, Frey BS (2003) Does Marriage Make People Happy, Or Do Happy People Get Married? Working Paper, No 143, University of Zurich, Institute for Empirical Research in Economics, Zurich
- Tamm M (2008) Does Money Buy Higher Schooling? Evidence from Secondary School Track Choice in Germany. *Econ Educ Rev* 27(5):536–545
- Tamm M (2005) The Effect of Poverty on the Health of Newborn Children – Evidence from Germany. RWI Discussion Paper, No 33, Essen
- Trommsdorff G (2008) 25 Wellen des Sozio-oekonomischen Paels (SOEP) Gewinn für interdisziplinäre Forschung. *Vierteljahrsh Wirtschaftsforsch* 77(3):195–202
- Wagner GG (2008) Die Längsschnittstudie Sozio-oekonomisches Panel (SOEP) – Die Jahre von der Wende zur Jahrtausendwende. *Vierteljahrsh Wirtschaftsforsch* 77(3):43–62

- Wagner GG, Frick JR, Schupp J (2007) The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. *Schmollers Jahrb* 127(1):139–169
- Zimmermann AC, Easterlin RA (2006) Happily Ever After? Cohabitation, Marriage, Divorce, and Happiness in Germany. *Popul Dev Rev* 32(3):511–528