



SOEP

The German
Socio-Economic
Panel Study

20 Years of Household Panel Data for Germany

**The SOEP survey and organization between
continuity and innovation**

Jürgen Schupp



Contents

SOEP

- Some Basic Facts: Continuity
- Recent Innovations: Data and Methodology
- Prospects



The SOEP

- Facts (1): Institutional setting
 - SOEP = independent department at DIW Berlin
 - “institutionalized” funding since 2003
 - member of the *Leibniz*-Association (WGL)
 - cooperation with other WGL-service institutions (RWI, WZB)



The SOEP

- Facts (2)
 - Interdisciplinary approach (Economics, Sociology, Demography, Political Sciences, Psychology, Geography, ...)
 - Currently 21 waves of data (1984-2004) for West Germany
 - Currently 15 waves of data (1990-2004) for East Germany
 - Over-sampling of immigrants from the very beginning on !
 - Interviews with all adult household members (no proxies)
 - Objective and subjective Indicators
 - Theory-based and pre-tested questions
 - Most experienced field-work agency (*TNS Infratest Sozialforschung*)



SOEP – Special topical modules

	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	
<i>Employment biography</i>	A+B	R	R	R	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	
<i>Marriage and family biography</i>		A+B	R	R	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	
<i>Social origin, first job</i>			A+B	R	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	BIO	
<i>Neighborhood</i>			A+B							A-C						A-E					A-G			
<i>Social security and its perception</i>				A+B					A-C					A-D						A-G				
<i>Assets, inheritance</i>					A+B													A-F		A-G				
<i>Further education or training and qualification</i>						A+B				A-C							A-F				A-G			
<i>Use of time and preferences</i>							A+B					A-D					A-F						A-G	
<i>Family and social services</i>								A-C					A-D					A-F						A-G
<i>Working (job) conditions</i>		A+B	A+B	A+B								A-D						A-F						A-G
<i>Future expectations</i>											A-C						A-E							A-G
<i>Ecology and environmental behaviour</i>															A-E						A-G			

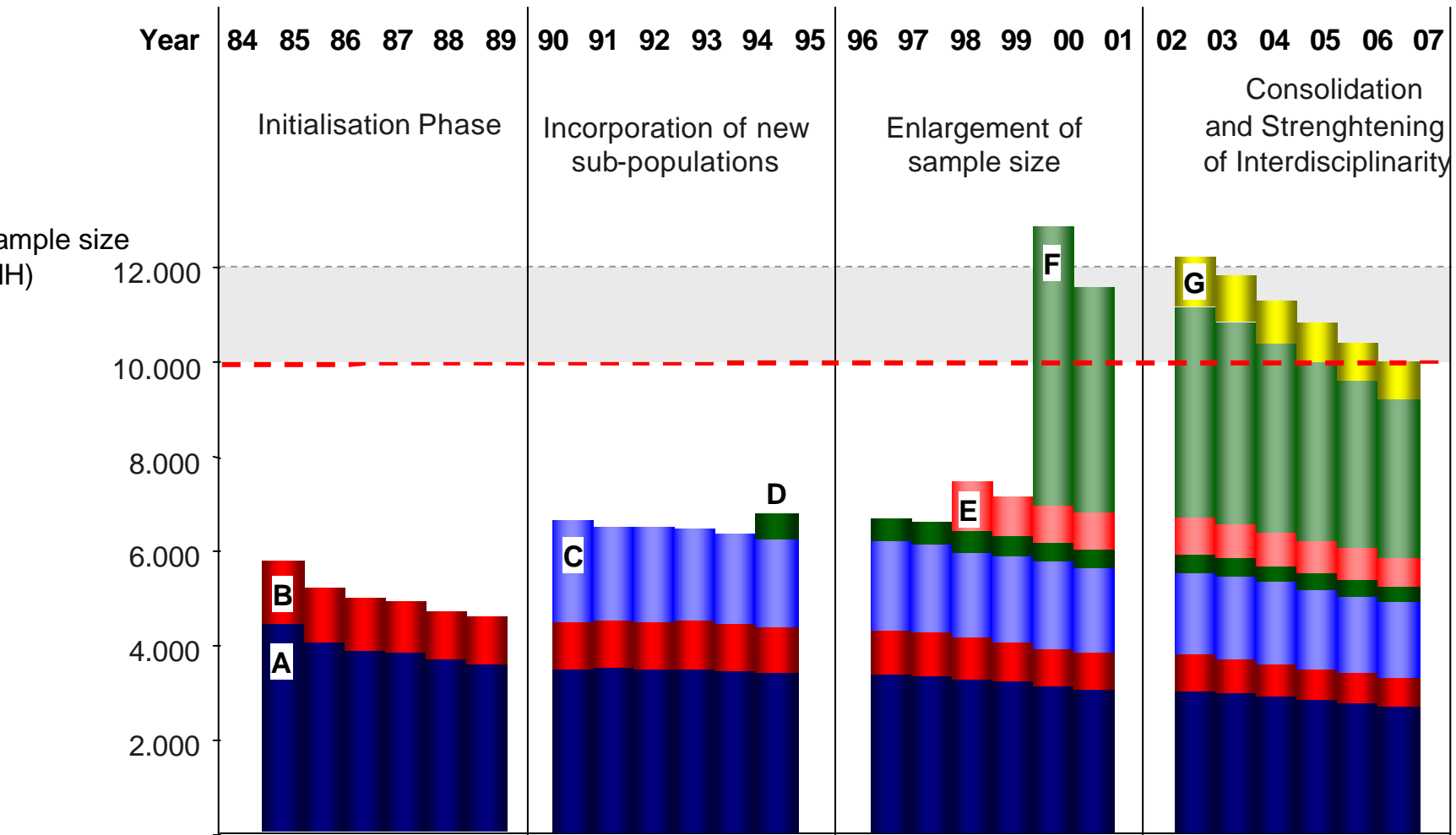
Legend:

A = Sample A; B = Sample B; C = Sample C; D = Sample D; E = Sample E; F=Sample F, G=Sample.

BIO = Biography questionnaire to be answered by first time respondents

R = Retrospectively asked biographical information

The SOEP sample 1984 - 2007



Source: TNS-Infratest Sozialforschung

Longitudinal Data:

Examples based on balanced panels

- 20 waves balanced panel of respondents (\$NETTO==1) **n=3,332**
- 20 wave balanced panel of participants (child or respondent; \$NETTO==1 | 2) **n=4,240**
- 20 waves panel incl. temporary drop-outs (file \$PLUECKE, \$NETTO==1 | 2 | 4) **n=4,738**
- 14 waves balanced panel of respondents in East German sample (PSAMPLE==3 & \$NETTO==1) **n=1,976**

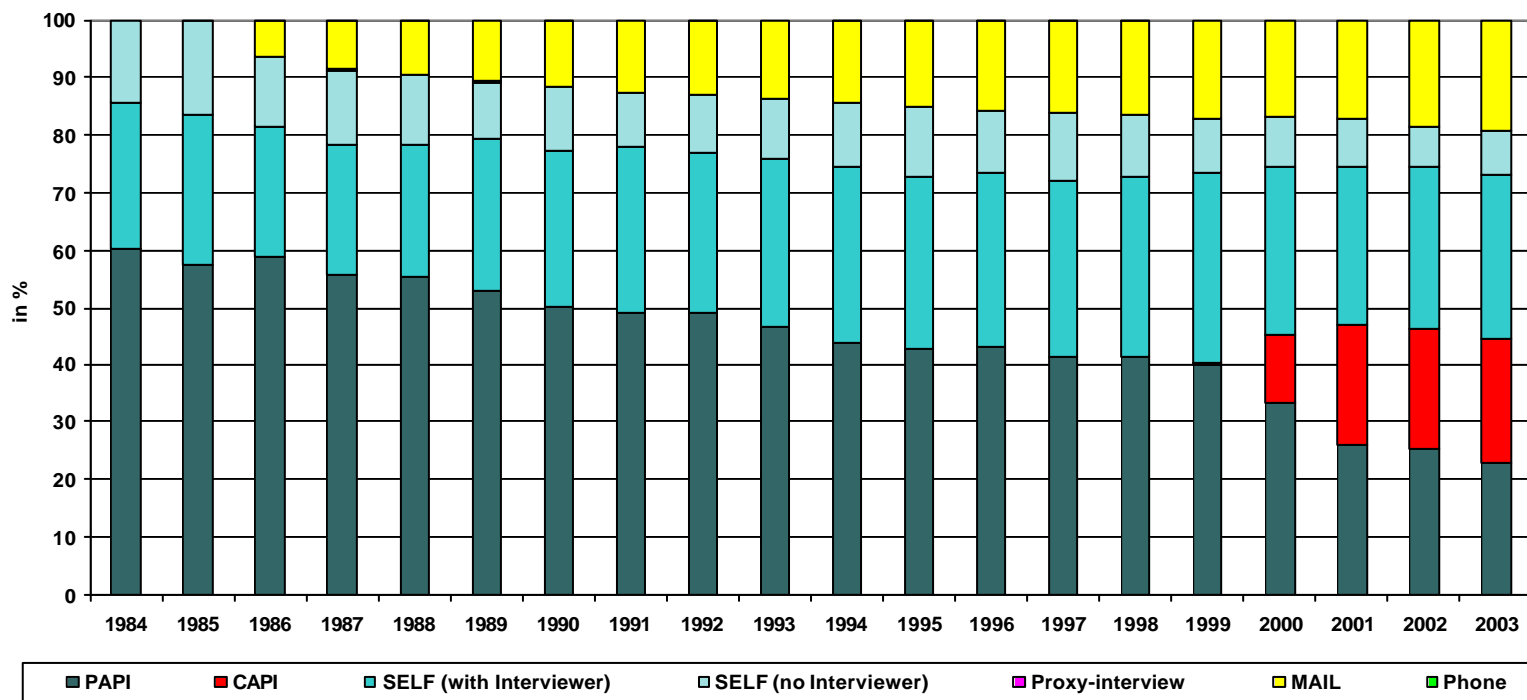


Continuity and Innovation – Interview Mode

- Methodology-Mix using standardized instruments
 - Face-to-face is default interview mode for new SOEP-respondents (PAPI)
 - Self-administered interview with Interviewer contact (SELF)
 - self-administered with interviewer present
 - completely self-administered; Interviewer picks up questionnaire after completion
 - *completely* in written form and mailed back to Infratest (MAIL)
 - (almost) no SOEP-interviews by telephone
 - no Proxy-Interviews
- Since 1998 stepwise implementation of CAPI
- prospective: Self-administered interviewing via Internet



Development of Interview Mode; Sample A, 1984-2003





SOEP-Instruments used in 2004

- Standard Instruments
 - Address protocol (filled in by interviewer)
 - Household questionnaire
 - Individual questionnaire for each HH member aged 16 and over
- Instruments for specific Target Groups
 - Add-on questionnaire „Life History“ (first time respondents aged 18 and over)
 - Add-on questionnaire „Youth/Adolescence“ (first time respondents aged 16/17)
 - Add-on questionnaire „Mother & Child“ (New born infants)
 - Add-on questionnaire „Gap“ (Temporary Drop-outs)
 - Experiment on „Trust and Trustworthiness“ (Face-to-Face Sample F)



Recent Innovations and further plans (I)

- **2002/03:** Integration of results from the additional tracking of drop-outs
 - migration
 - death (improved potential for analysis on mortality)
- **2002:** Introduction of a sample of High Income Households (subsample G)
- **2004:** Pointers on twins BIOTWIN (118 twins, 3 triple)
- 2007: Refresher sample (subsample H)
- 2012: Refresher sample



Recent Innovations and further plans (II)

- Improving data for cohort analyses using age-specific instruments
 - 2001: Introduction of age-specific data on youth/adolescence for first-time respondents aged 16/17 (*questionnaire “Youth”*)
 - 2003: Introduction of age-specific data on new born children (about 0-15 months) around childbirth (*questionnaire „Mother and Child“*)
 - 2005: Introduction of age-specific data on infants (about 3 years) (*questionnaire „Infant“*) (*2004: pretest in field*)
 - 2007/2012: building up on the 2003 birth cohort further age-specific instruments will collect information on children (*questionnaire „Children“*)
 - aged 6-7 in 2007
 - age 12-13 in 2012



Recent Innovations and further plans (IV)

➤ Survey Methodology

- Tracking emigrants – A pilot in 2003
- 2004: Improved Geo-Code data (ZIP-code level)
 - to be analyzed at DIW Berlin and via LISSY)
- 2004/05: Testing Internet Based Data Collection (“SOEP-Online”)
- 2003ff: Integration of experimental data (“trust and trustworthiness”) as „add-on“ to sample F collected in 2003 to 2005
- 2005: pre-testing physical health measures
- 2006: provision of “Experimental Slots” in subsamples ???
- 200x: Testing Links to Institutions/ Organizations ???
- 200x: Re-Interviewing Drop-Outs ???



Pilot - Experimental Economics – The SOEP Behavioral Experiment

"Trust and Trustworthiness"

SOEP-Cooperation with:
Ernst Fehr, University of Zurich.
Data available to all SOEP users.

Idea:

- to generate behavioral data in representative samples of the population
- to combine behavioral data with survey data
- to observe longitudinal effects

SOEP program

- 2002: Pretests
- 2003: First wave, n = 1,432
(random subsample,
95% participation rate)
- 2004: Second wave
- 2005: Third wave

The experiment is conducted at the end of the individual interview.

The monetary outcome for each pair of players (0 - 40 €) is then calculated and some weeks later handed over by Infratest (paycheck)



Recent Innovations and further plans (V)

Strengthening Interdisciplinarity

➤ Political Science

- 2004: testing validity of old indicators

➤ Psychology

- 2005: introduction of improved personality indicators



SOEP-Pretesting of next year's programme

Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec
	●	●	●	●	●	●	●	●	●	●	●

- **February - March:** Discussion of pretest design and questionnaire
- **April - May:** Fieldwork
- **May 27:** Data delivery
- **June:** Technical report, recommendations
- **By end of September:** Analysis & conclusions (DIW), finalize questionnaires



The SOEP Pretest for 2005 “Personality & politics”

Design: Multi Trait Multi Method Experiment

Random sample of $n = 750$, with 3 splits of $n = 250$, CAPI interview

Using different scales for measurement of

- life satisfaction
- political positioning of self and political parties

Long item lists for personality testing

(Big Five, justice, trust, reciprocity, altruism, and risk aversion)

Interview duration: 60 minutes!



Basis for decision about a **small set** of questions in SOEP 2005.

Different version of the experiment “trust game”



Questionnaire “Youth” (started 2001)

- Substitutes questionnaire “Life history” for first-time respondents aged 16-17
- New data files: BIOYOUTH
- Themes covered:
 - Relationship to parents
 - leisure time use
 - school performance
 - educational intentions
 - job expectations
 - personality characteristics
 - family expectations
 - Standard indicators on intergenerational mobility



Questionnaire „Mother and Child“ (started 2003)

- 4 page-questionnaire given to mothers with children born “over the last year” (2002 / 2003)
- new data file BIOCHILD
- Themes covered
 - birth related issues (weight, height, timing, etc)
 - problems, etc during pregnancy
 - subjective indicators (mother)
 - Caring for baby and support by father and third parties
- Expected number of observations: approx. **150** per year !!!



Questionnaire „Mother & Child 2 - Infant“ (starting in 2005)

- Referring to the same child 2 years later
- Again 4-page questionnaire
- Themes covered
 - Health of the child, medical treatment
 - Child development
 - Activities of parents and child
 - Caring for baby and support by father and third parties
- Pretest in June/July 2004
 - n=50
 - Questionnaire 8 pages



Questionnaire „Life History“ (changes since 2001)

- now to be filled in by first-time respondents aged 18 and over
- additional topics covered
 - information on adolescence (e.g. school performance)
[new data file BIOSOC]
 - information on childbirth also collected from all men in sample F and new respondents in all other subsamples
[new data file BIOBRTHM]
 - more exact birth information (month of birth)
[new variable GEBMONAT in PPFAD]
 - 2004 data distribution will include information on second residence as given at time of filling in the questionnaire “life history”
[new data file BIORESID]



Recent Innovations and further plans (V)

Research on Data Quality and Service Improvement

- Survey Statistics
 - ongoing: Improvement of Imputations
 - ongoing: Fraud detection
 - 2005/06: Testing Multiple Imputation for Unit- and Item-Non-Response
 - 2005/06: Testing Possibilities of Matrix-Sampling-Design

- Information Management (Service)
 - 2004: LISSY



Conclusion & Prospects (I)

- SOEP stands for theory based data collection – not just “more and better statistics”
- Improved potential for analyses of smaller subpopulations due to increased sample size
- Improved potential for intergenerational analyses based on meanwhile 20 waves of data
- Improved biographical data (e.g., youth and adolescence)
 - background information
 - of self-contained interest
- Improved data on methodological issues (interview modes, imputation of missing values, ...)



Conclusion & Prospects (II)

- international networking improves the cross-national infrastructure for data and analyses
 - cross-sectional databases: LIS, LWS
 - longitudinal databases: CNEF (1984-2003+), CHER (1990-2000+), ECHP (1994-2001)
 - development of user-friendly and x-nationally comparative SOEP-data (ex-post harmonization)
 - education: ISCED, CASMIN (\$PGEN)
 - labor market: ISCO88, NACE (\$PGEN)
 - regional information: NUTS (\$HBRUTTO)
 - income: Canberra Group recommendation (\$PEQUIV)
 - moreover, interaction with other producers of panel data (e.g. BHPS, HILDA) with respect to methodology (e.g. pre-testing) and substantive issues (e.g. timing of special topical modules) simplifies future data harmonization and x-national analyses
 - Common Conferences like the two starting tomorrow!



Comments welcome !!!

<mailto:jschupp@diw.de>