

Introduction to the German Socio-Economic Panel (SOEP)

Jan Goebel

IZA Red Cube Seminar
May 20, 2009

<http://www.diw.de/soep>

- Introduction
- Sub-samples and Survey Related Issues
- Sample Size Development
- Content
- Data Access
- Innovations

Household Panel Data

- **The Data**

- Survey of private households and persons in Germany (official title „Living in Germany“)
- Annual Interviews since 1984; 25th (!) wave in 2008
- Started with 5,921 Households in 1984 (2007: 11,689) and 12,245 persons (2007: 21,234)
- Samples added: East Germans 1990, Migrants 1994, High-Income 2002, General Refreshments in 1998, 2000, and 2006

- **Institutional Facts**

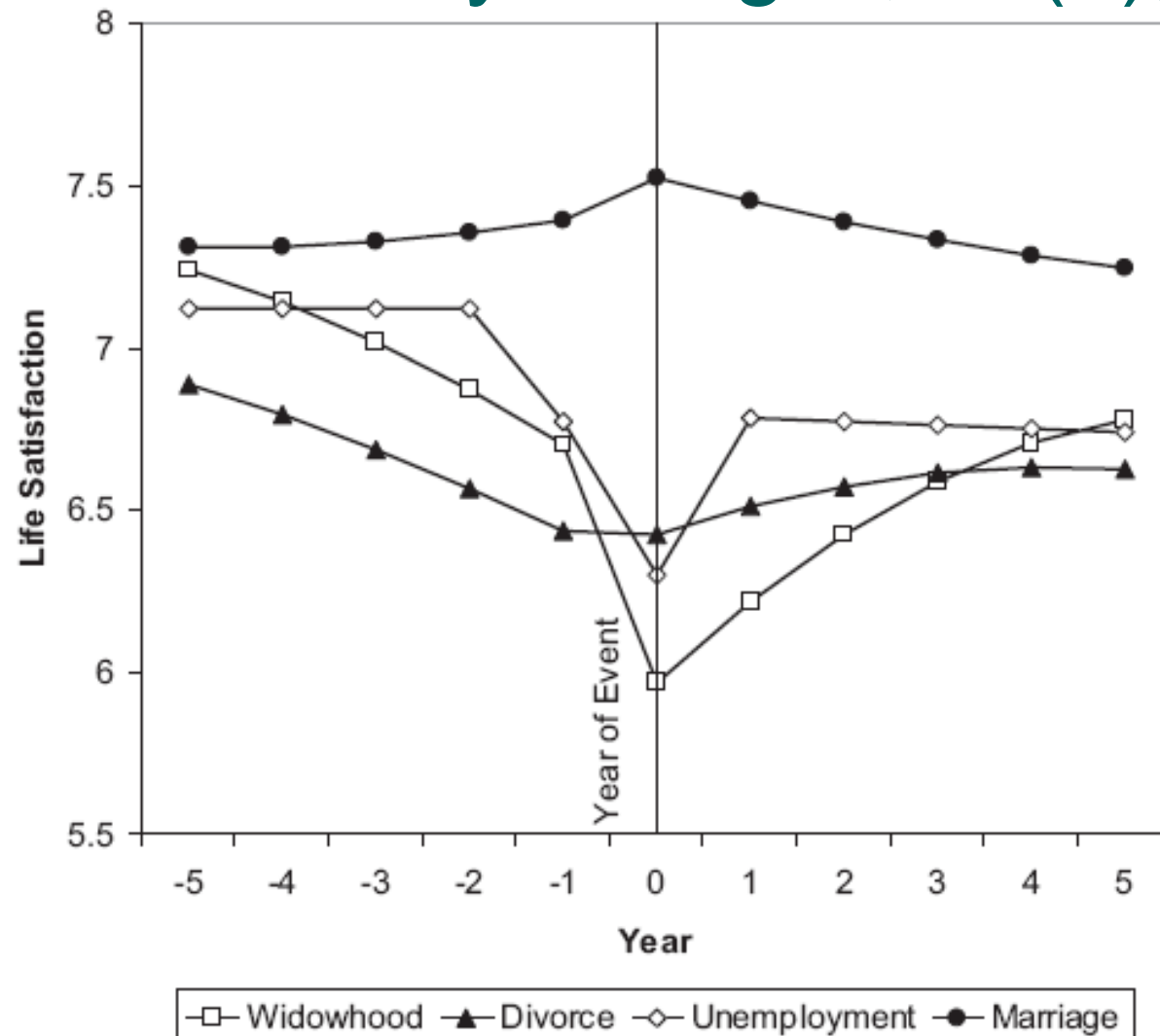
- Run by social scientists, totally research driven!
- Located: department SOEP at DIW Berlin; institutionalized funding since 2003 (WGL institute)
- Field-work agency: TNS Infratest Sozialforschung

- Representative micro-data on persons, **households**, families
- **Objective** (e.g. income) and **Subjective** (e.g. satisfaction) indicators
- Measure **Stability** and **Change** in Living Conditions
- **Coverage of all life-stages**
 - Observation unit “household” guarantees appropriate coverage from birth to death
- **Effects on third persons**
 - Effect of divorce not only on separated partners but also children.
- **Joint decisions**
 - Occupation of partner affects labor supply
 - Family diet and individual health.
- **Dyadic analysis**
 - Intergenerational research (mobility)
 - Sibling estimates, twin studies

- **Heterogeneity**
 - Analysis of intra-individual change vs inter-individual difference
 - Controlling for (inter-individual) unobserved heterogeneity.
- **Causality**
 - Controlling for temporal order of events and behavior better suited for the identification of causal effects than cross-sectional data.

Diener et al. 2006, American Psychologist, 61(4), 310

Figure 2
Adaptation to Good and Bad Events



- Cumulation of rare events by pooling (mortality, divorce, ...)
- Increasing coverage of changes in institutional settings and the potential impact on individual behavior
- Comparison of “intentions” and actual behavior (how relevant are expectations at the individual level?)
- Intergenerational analysis (linking parents and children)
- Increasing potential for cohort analysis
- Improved efficiency in statistical modeling due to increased number of observations per individual

→ increasing with panel duration!

- Bias due to selective attrition
- Panel effects (incl. *positive* learning effects !!!)
- Representative coverage of changes in underlying population due to immigration
- “Compensation rules” for long-term shrinking number of observations
- Continuity of questioning (phrasing, modes, ...)
 - inter-temporal comparability of indicators vs.
 - need for adjustments to cover institutional changes

- Introduction
- **Sub-samples and Survey Related Issues**
- Sample Size Development
- Content
- Data Access
- Innovations

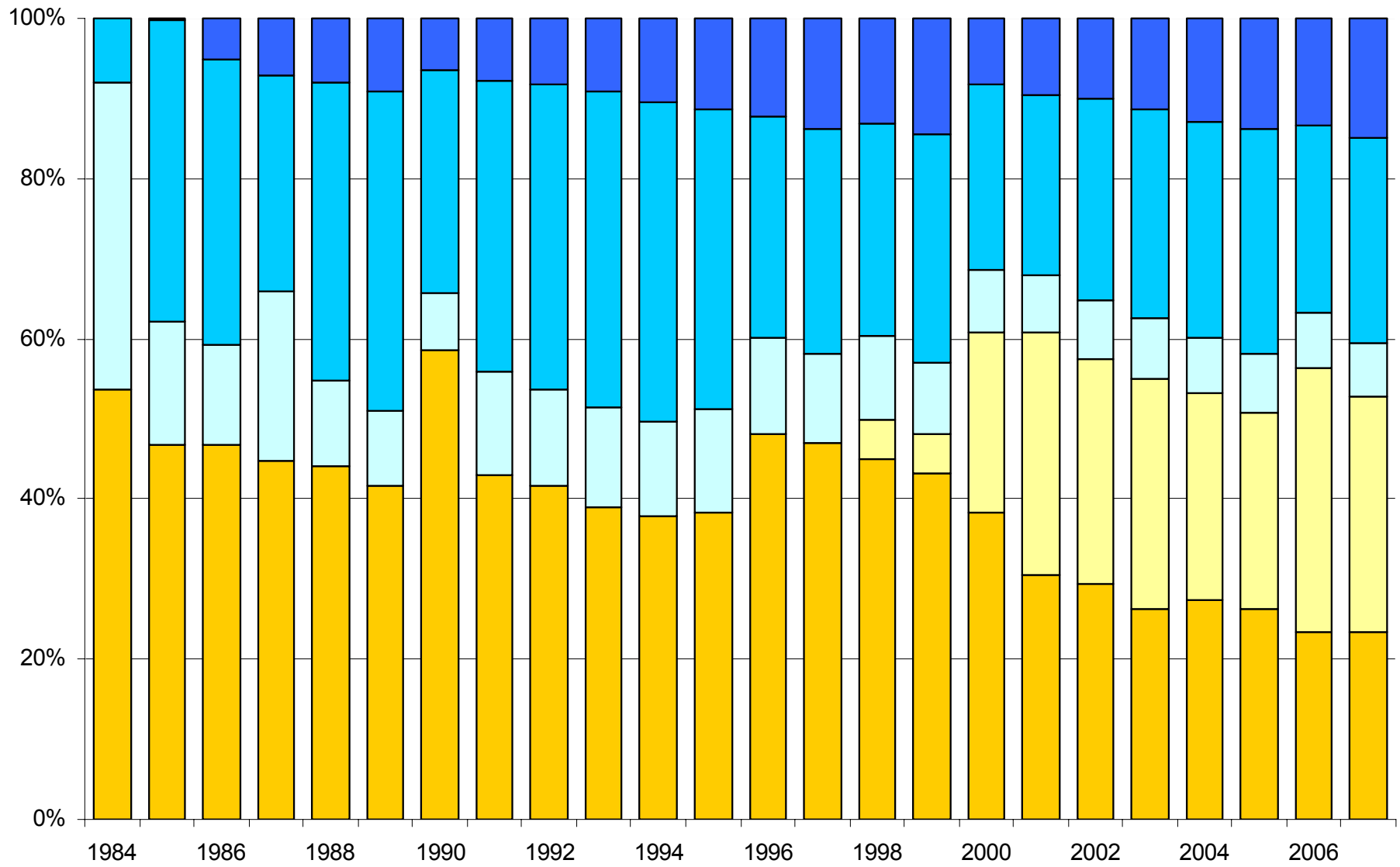
- . **A "West-German" residents**
 - . started in **1984**, n=4528 households
 - . Head is either German or other nationality than those in Sample B.
- . **B "Foreigners"**
 - . started in **1984**, n=1393 households
 - . Head is either Turkish, Italian, Spanish, Greek, or Yugoslavian (oversampling)
- . **C "East-Germans"**
 - . started in **1990**, n=2179 households.
 - . Head was a citizen of the GDR. (expansion of survey territory, oversampling)
- . **D "Immigrants"**
 - . started in **1994/95**, n=522 households
 - . At least one HH member has moved to Germany after 1984. (expansion of survey population)
- . **E "Refreshment sample"**
 - . started in **1998**, n=1067 households
 - . Random sample covering all existing subsamples (total population).
- . **F "Innovation sample"**
 - . started in **2000**, n=6052 households. Random sample covering all existing subsamples (total population).
- . **G "High Income sample"**
 - . started in **2002**, n=1224 households.
 - . Monthly net Household income > 7.500 DM (oversampling)
- . **H "Refreshment sample"**
 - . started in **2006**, n=1506 households.
 - . Random sample covering all existing subsamples (total population)

Methodology-Mix with standardized instruments

- ⇒ Face-to-face individual interviews with all HH-members 16+
 - Household interview with "head of household"
 - Paper-and-pencil interviews (samples A through D, E1, F, G)
 - No proxy interviews / (almost) no phone interviews
 - Self-Completers: Data Agency resolves inconsistencies

- ⇒ Since 1998 stepwise implementation of CAPI

- ⇒ prospective: self-admin. Interviews via Internet (Test in 2004)



Containing general information (filled in by interviewer)

- ⇒ on households (e.g. size, housing area, regional information); on individuals (e.g. sex, year of birth, relation to head)
- ⇒ on the process of field work (e.g. number of contacts, reason for drop-outs, interview mode)

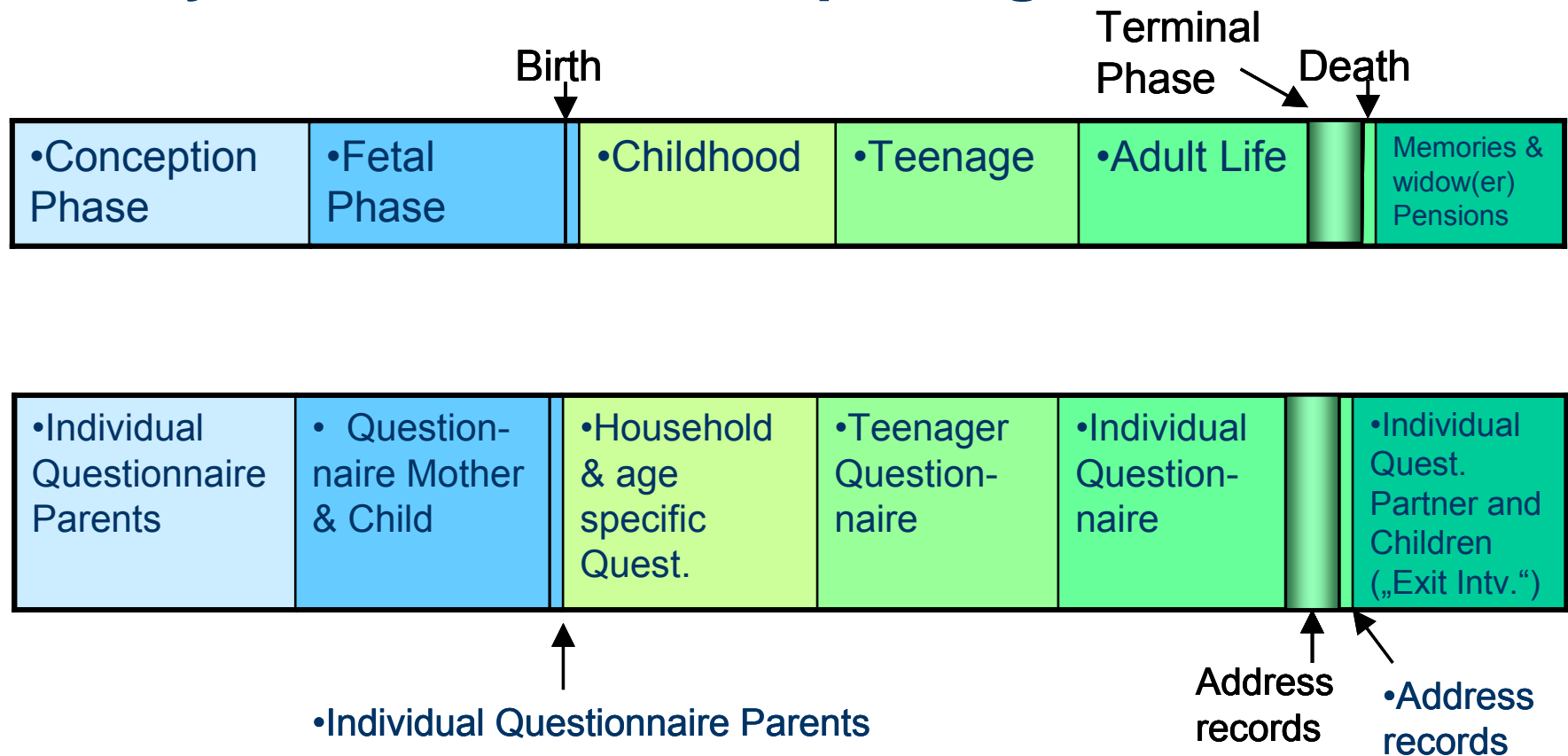
The Standard Instruments

- Household questionnaire (annual)
- Personal questionnaire (annual, all 17+)
- Personal „gap“ questionnaire (temp. drop-outs)
- Biography questionnaire (second interview only)

Special Populations

- Sub-sample-specific questionnaires for “Germans”, “Foreigners”, and “East Germans” (until 1995)
- Age-specific questionnaires: Age 1 (since 2003), Age 2 (since 2005), Age 5 (since 2008), Age 12 (prospective 2013), Age 17 (since 2001)
- Interviewer questionnaire (since 2006)
- Event-triggered questionnaires planned

Survey Instruments for capturing the life course



- Introduction
- Sub-samples and Survey Related Issues
- **Sample Size Development**
- Content
- Data Access
- Innovations

Sample	Starting Year	Households	Respondents
A and B	1984	5,921	12,245
C	1990	2,179	4,453
D1/D2	1995	522	1,078
E	1998	1,067	1,923
F	2000	6,052	10,890
G	2002	1,224	2,671

- All persons in HH are to be surveyed also the following years.
- At same address as well as after a residential move within Germany (→ regional mobility)
- Personal interviews at age of 16 (→ demographic development)
- Persons moving into an existing SOEP household.
- Since 1989 these (new) persons are also followed in case of leaving the household. This had not been the case up to wave 5 (1988).

- Principle: follow until two consecutive temporary drop-outs of all household members or a final refusal.
- Gaps: small questionnaire including questions on central information which is missing for the year of the drop-out (e.g. employment status).

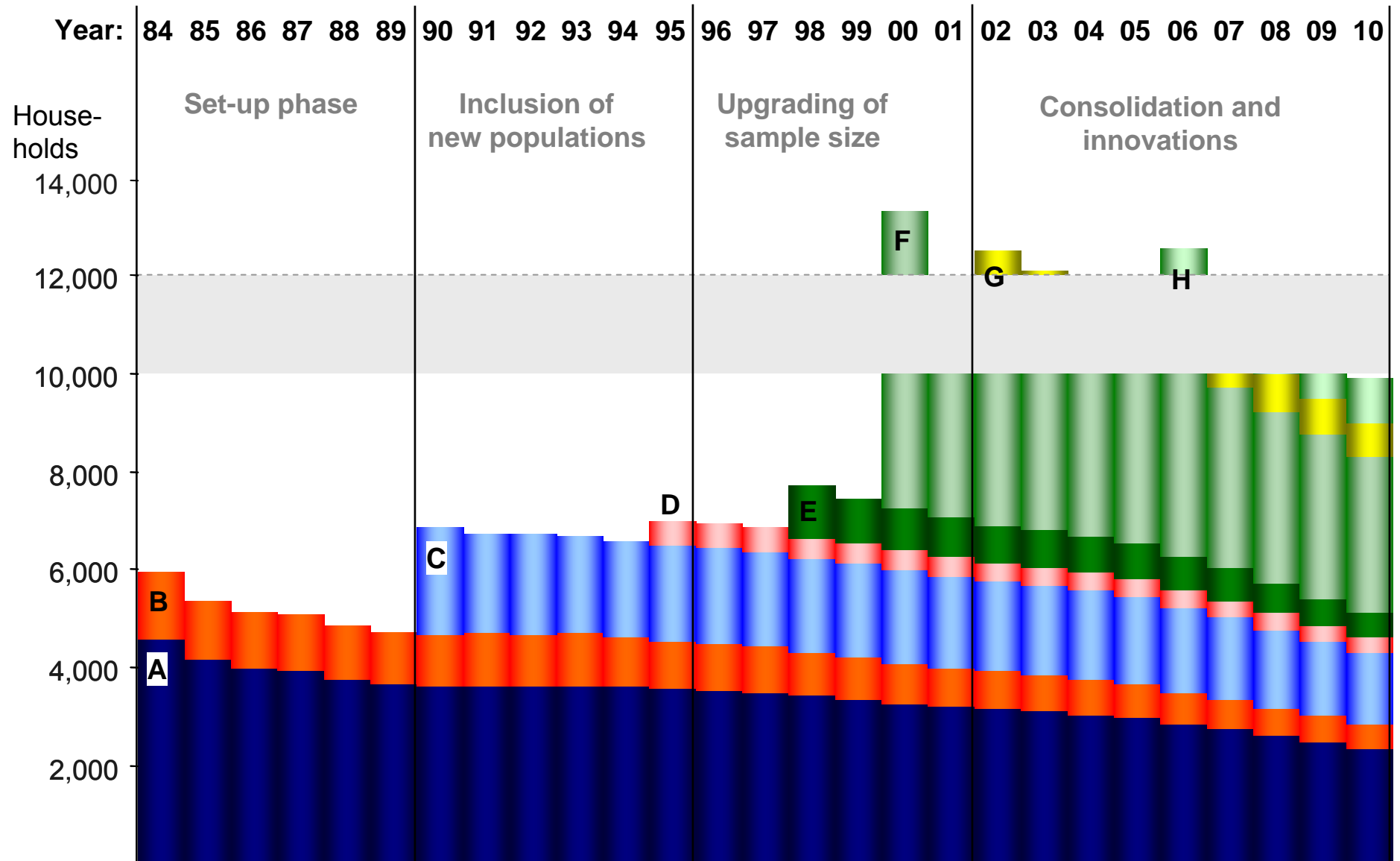
		<i>Households</i>	
		Old	New
<i>Persons</i>	Old	<ul style="list-style-type: none"> • “classic case” without change of address • entire household moves 	<ul style="list-style-type: none"> • Move-out
	New	<ul style="list-style-type: none"> • Birth • Move-in 	<ul style="list-style-type: none"> • Birth • Caused by splitt-offs of old persons from old households*

* Remember that households *new* to the SOEP may already have existed before contacting the survey

- Persons exit by:
 - Death
 - Moving abroad
- Persons enter by:
 - Birth
 - Moving into a SOEP household from somewhere else in Germany or from abroad
 - Reaching age of 16 years (minimum respondents age is given by the calendar year, in which a person turns 17 years of age)
 - Split-offs of at least one *old* person from an *old* household

1. Making a successful contact to a given household
 2. Realizing a successful interview
- Social groups typically associated with problems in respect to re-contacting and re-interviewing:
 - Single person households
 - Mobile households and persons
 - Young adults leaving parental home

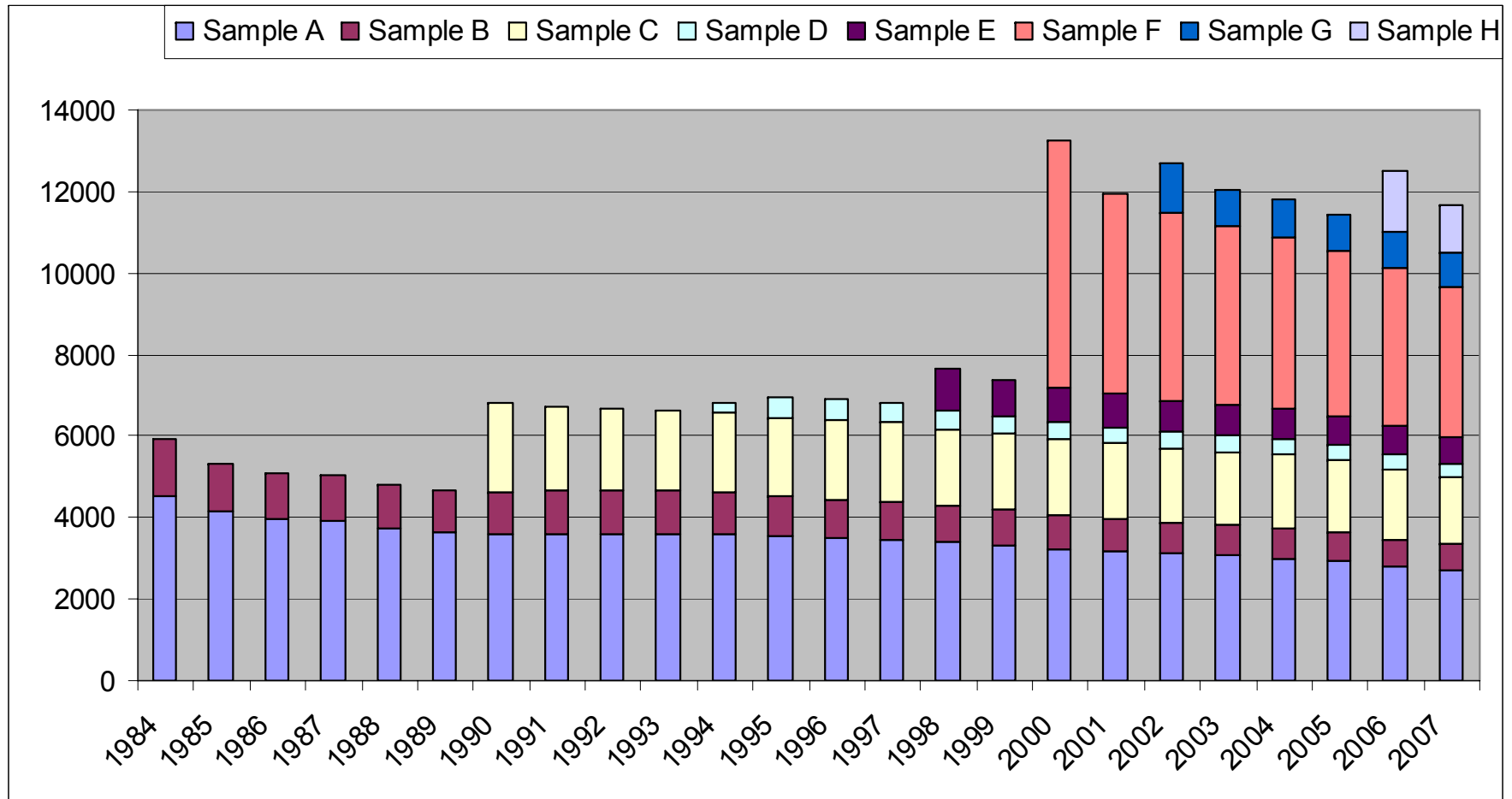
- ⇒ Announcement by mail 2 weeks prior to interview
- ⇒ For each successful interview, any respondent
 - receives a gift related to the yearly topical module (until 2007)
 - ticket for monthly nationwide lottery
 - since 2008: 5 € ex-ante incentive (until 2007: 1,50 € ex-post)
- ⇒ Addresses are kept up to date by the field work agency
- ⇒ Households receive the brochure „Living in Germany“ and information about data privacy regulations
- ⇒ After interview (during summer): thank you-letter and „porto-card“
- ⇒ Special treatments:
 - contact via phone
 - central case-by-case treatment for “problematic” households
 - mailing more information on request
 - website “Leben in Deutschland”
- ⇒ The interview situation (face-to-face) ensures a personal
 - relationship, which makes it harder to withdraw from the survey.
 - Thus, the stability of the interviewer over time is very crucial.



Sample Size Development

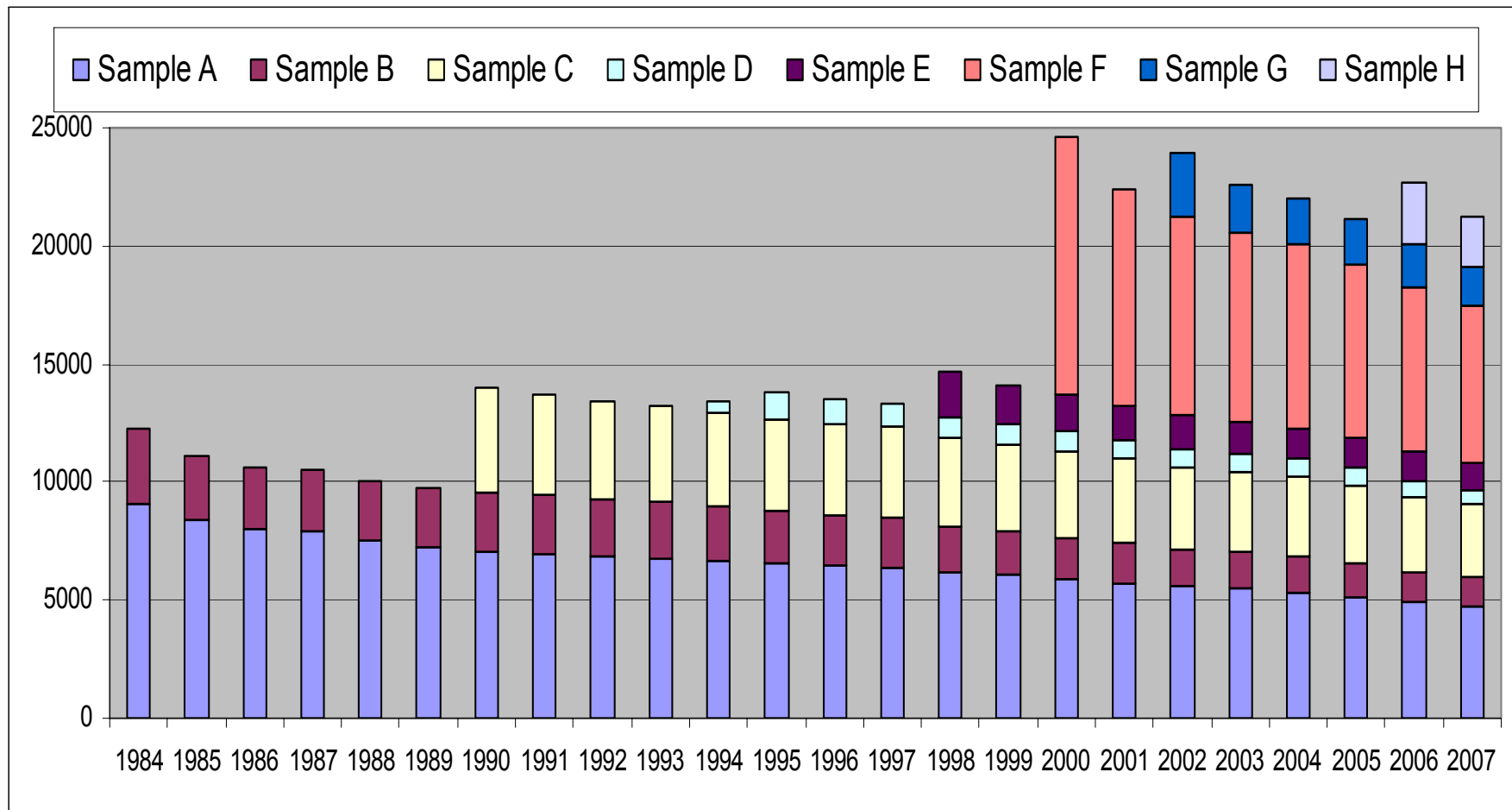
Cross-sectional perspective

⇒ Number of successfully interviewed households by sample

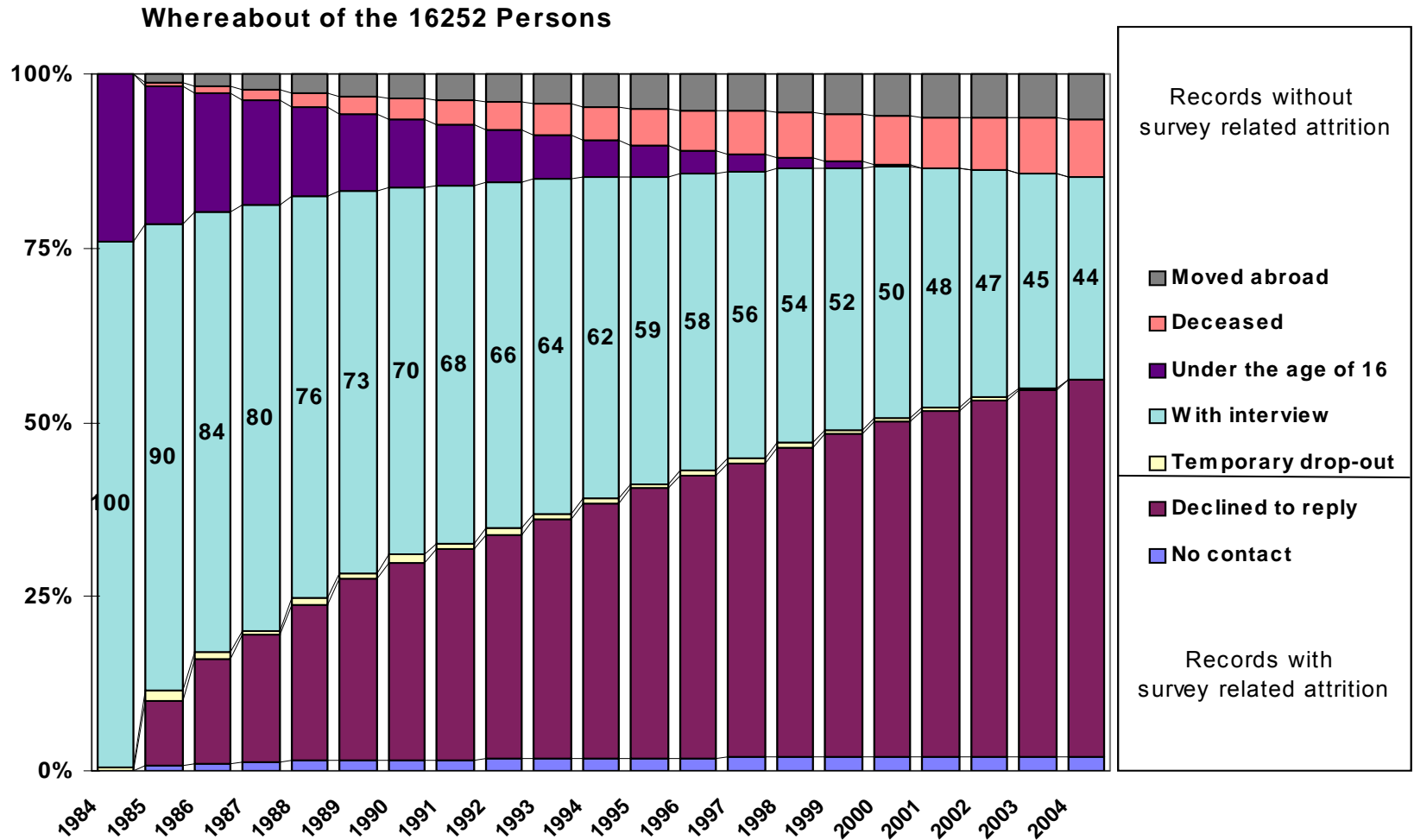


Cross-sectional perspective

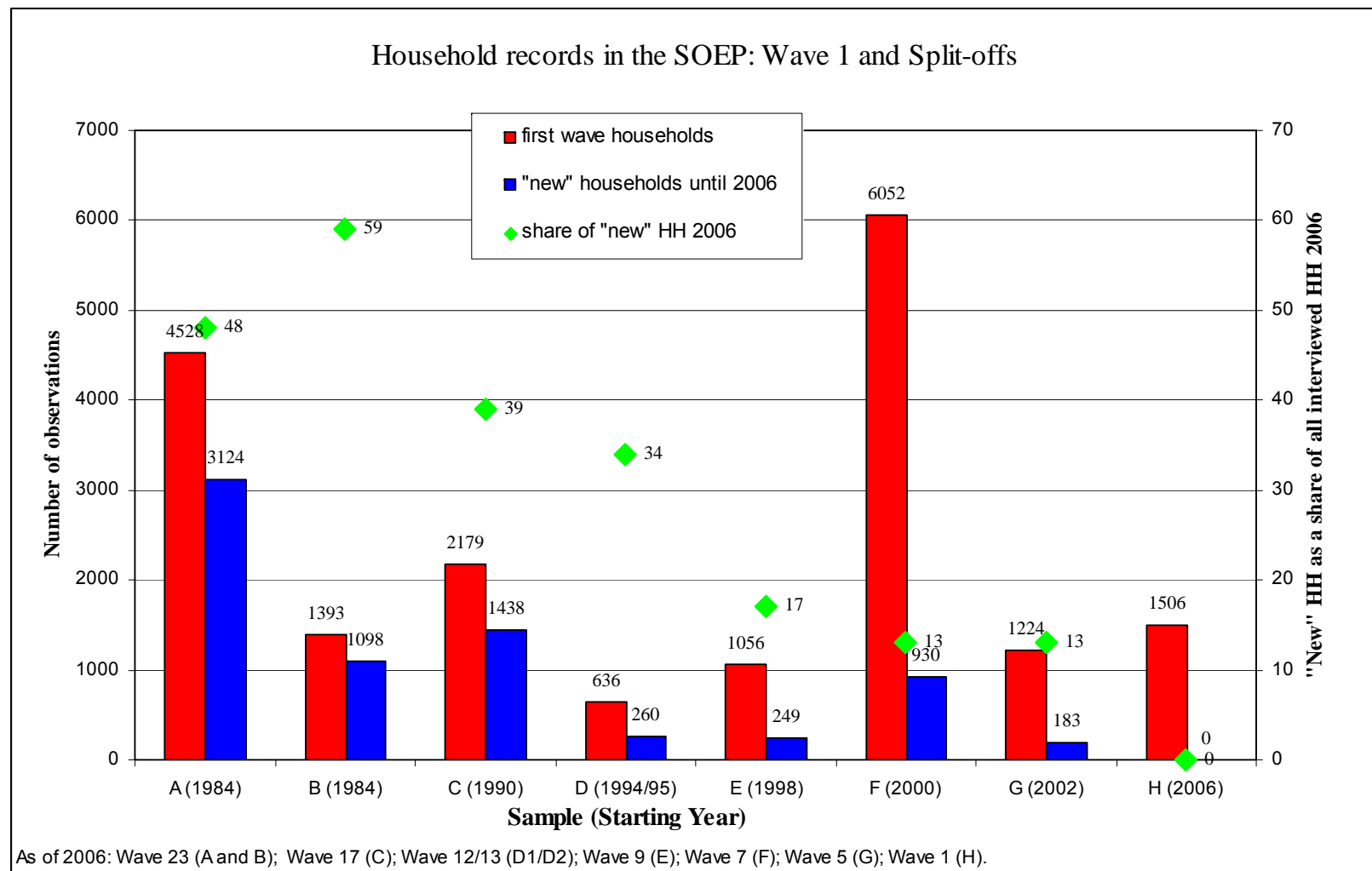
⇒ Number of successfully interviewed persons by sample



Longitudinal perspective: Samples A+B since 1984



Incidence of New households after Wave 1



- Introduction
- Sub-samples and Survey Related Issues
- Sample Size Development
- **Content**
- Data Access
- Innovations

- Questions about a point of time (present)
e.g. current employment status or current levels of satisfaction
- Retrospective questions on certain events (past)
e.g. how often did you change your job during the last ten years?
- Retrospective life event history (past)
e.g. employment or marital history
- Monthly calendar on income and labor market (past)
e.g. employment status January through December last year
- Questions concerning future prospects (future)
e.g. satisfaction with life five years from now, or job expectations

→ Replication every year

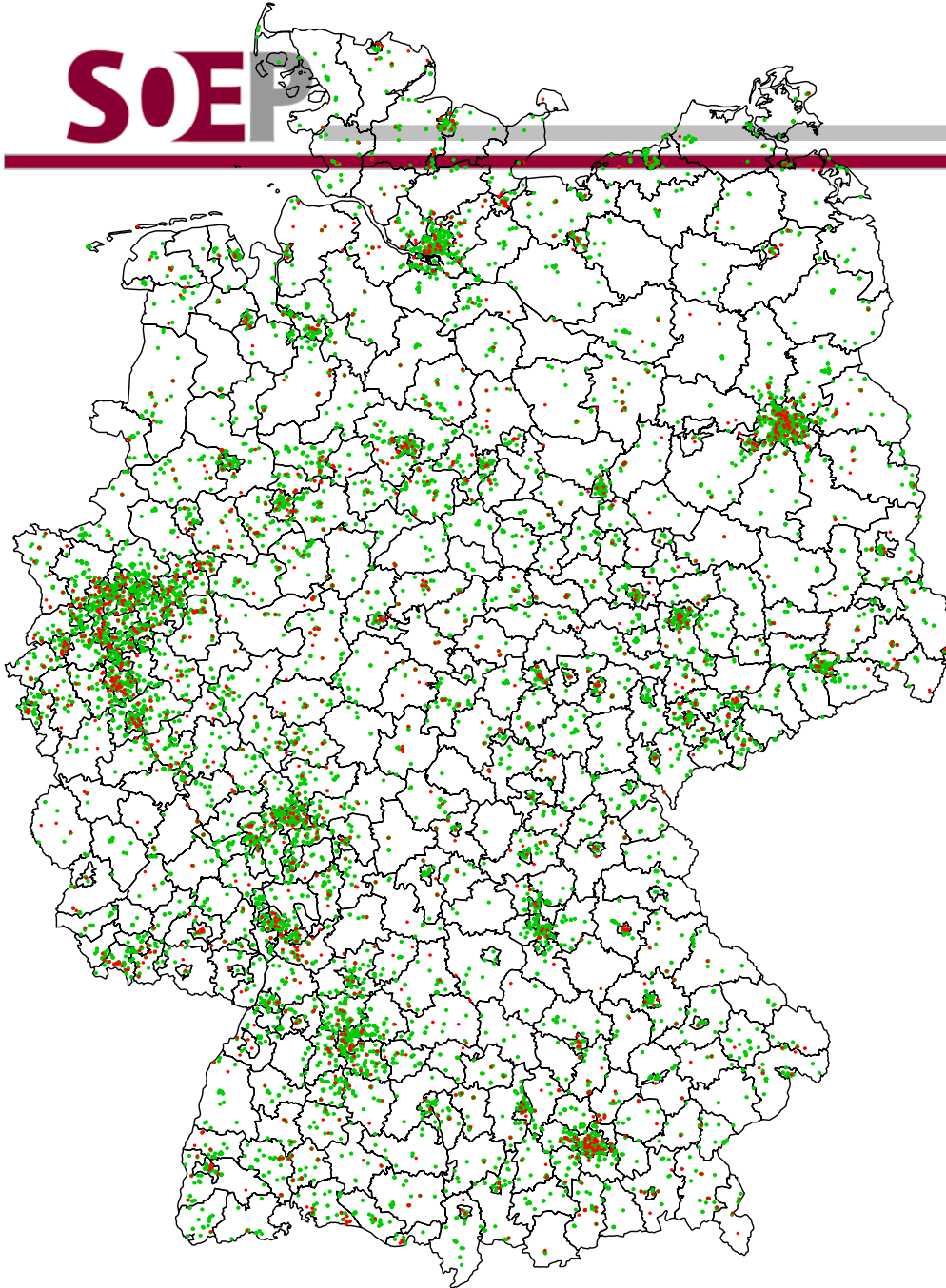
- Population and demography
- Education, training, and qualification
- Labor market and occupational dynamics
- Earnings, income and social security
- Housing
- Health
- Household production
- Basic orientation (preferences, values, etc.)
- Satisfaction with life in general and various aspects

→ Replication every ~ 5-10 years

1984	Employment biography	1997	Social security and poverty
1985	Marriage and family biography	1998	Ecology and envir. behavior
1986	Social origins (Bio), neighborh.	1999	Expectations, Use of time
1987	Social security, early retirement	2000	Further education or training
1988	<i>Assets</i>	2001	Social networks, working cond.
1989	Further education or training	2002	Social Security, <i>assets</i>
1990	<i>Time use and preferences</i>	2003	Ecology and envir. behavior
1991	Family and social services	2004	Further education or training
1992	Social security and poverty	2005	<i>Time use and preferences</i>
1993	Further education or training	2006	Family and social networks
1994	Neighborhood, values	2007	<i>Assets</i>
1995	<i>Time use and preferences</i>	2008	Further education or training
1996	Family and social network		

- Locus of control (1996,2005,2009)
- Anomie (1991,1996, 2008)
- Life goals (1990, 1992,1995, 2004, 2008)
- Trust and Fairness (2003,2008)
- Risk aversion (2004, 2006, 2008)
- Big Five Inventory (2005, 2010)
- Reciprocity (2005, 2010)
- Emotions (2007,2008,2009)
- Eagerness (2008)
- Impulsiveness (2008)

	Federal States	Regional Policy Regions	Counties	Communities	Zip-Codes, delivery area
Corresponding NUTS Level	NUTS1	NUTS2	NUTS3	NUTS5/LAU2	-/-
Number of Regions (2005)	16	97	439	12141	17201
Number of Regions with SOEP respondents (2008)	16 (100%)	97 (100%)	439 (100%)	2561 (21,1%)	2420 (14,1%)
Mean number of SOEP Households (Respondents) per region in 2008	684,9 (1246,6)	112,9 (205,6)	24,9 (45,4)	4,3 (7,8)	4,5 (8,2)
Available for SOEP waves since ...	1984	1985	1985	2000	1993
Access policy	Standard-DVD	Extra CD and password (Geocodes)	SOEPremote or at DIW Berlin	Only at DIW Berlin	Only at DIW Berlin (only one computer, in the near future via SOEPgeo at DIW Berlin)



Distribution of SOEP households in 2008 over
Germany

Legend

- Successful HH-Interview in 2008
- No HH-Interview in 2008

- Since 2000 the geographical coordinates for all (gross and net sample) households are at block level available.
- New possibilities for example for mobility studies or segregation.
- Pilot study in cooperation with DLR, to test the integration and possibilities of spatial information of the earth's surface with social scientific questions („Synergies in socioeconomic and remote sensing”).
- Data access for external guest researchers (scheduled to end of 2009) and internal researchers possible through a special server environment at DIW Berlin only.



SOEP Homepage

- ⇒ <http://www.diw.de/soep>
- ⇒ Online Documentation
- ⇒ All questionnaires in German and English



SOEPINFO

- ⇒ Information System: Frequencies, Questionnaires
- ⇒ Getting Started Quickly: Generate Command Files !



Desktop Companion (DTC)

- ⇒ In-depth Description in English
- ⇒ Basic, Extensions, Retrievals, Weighting
- ⇒ Updated Every Year



SOEPLIT

- ⇒ Database of articles, books, papers written using GSOEP
- ⇒ Easy Search Interface

➔ Search for all DP's, Articles, Books using SOEP!

- Search by Author, Year, Title or even over all fields.
- If available links to digital documents are provided.

SOEP
SOEPlit Datenbank

Dies ist die neue SOEPlit-Datenbank
Hier finden Sie [Hilfe](#), mit Informationen zu Suche und Sortierung.
[link zurück](#)

All Details

5456 entries in all: 1 11 21 31 41 51 61 71 81 91 >>

SOEPlitNo	Author	Year	Title	URL
5527	Klein, Thomas	2009	Determinanten der Sportaktivität und der Sportart im Lebenslauf	
5526	Leopold, Thomas	2009	Erbschaften und Schenkungen im Lebenslauf. Eine vergleichende Längsschnittanalyse mit dem Sozio-oekonomischen Panel (Diplomarbeit)	http://www.opus-bayern.de/uni-bamberg/volltexte/2009/179/
5525	Sologon, Denisa Maria; O'Donoghue, Cathal	2009	Policy, Institutional Factors and Earnings Mobility	http://ftp.iza.org/dp4151.pdf
5524	Haan, Peter; Myck, Michal	2009	Dynamics of Poor Health and Non-Employment	http://ftp.iza.org/dp4154.pdf
5523	Wunder, Christoph; Wiencierz, Andrea; Schwarze, Johannes; Küchenhoff, Helmut; Kleyer, Sara; Bleninger, Philipp	2009	Well-Being over the Life Span: Semiparametric Evidence from British and German Longitudinal Data	http://ftp.iza.org/dp4155.pdf
5522	Klick, Jonathan; Stratmann, Thomas	2008	Do Spa Visits Improve Health: Evidence From German Microdata	http://www.qmu.edu/bc/stratmann/Do%20Spa%20Visits%20Improve%20Health%20final.pdf
5521	Bazyl, Monika	2009	Factors influencing tenure choice in European countries	http://iriss.ceps.lu/documents/irissp102.pdf
5520	Schlottter, Martin	2009	Kindergarten attendance and decision makers' political leaning - Difference-in-Difference evidence from German municipalities	http://smvs2009.org/file/235_Schlottter.pdf
5519	Scheubel, Beatrice	2009	Does it pay to be a woman? Labour market effect of maternity-related job protection and replacement incomes	http://smvs2009.org/file/396_Scheubel.pdf
5518	Schmitz, Hendrik	2009	Risk aversion, health behaviour, and adverse selection in the German market for private supplementary health insurance	http://smvs2009.org/file/469_Schmitz.pdf

Search:

SOEPlitNo:

Author:

Year:

Title:

All Details

Search over all fields:

Kommentare, Änderungswünsche und Fehlermeldungen bitte an: [Uta Rahmann](#)

- English or German
- Questionnaires
- Direct Search
- Topics
- Word Search
- Options
- Help

WWW-SOEPinfo (SOEP-DB 2005) - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

[SOEPinfo-Main-Actions] [Basic Information] [Help]

Basket:
0 Vars

[Basket-Actions] Select all Clear Delete [Files]

WWW-SOEPinfo

Intro

WWW-SOEPinfo is an interactive system with extensive search capabilities that provides you with detailed information on the variables in the SOEP dataset. By interactively creating lists of variables in SOEPinfo, the user can then output frequencies information, item correspondence, and even generated SPSS, SAS, and Stata command files.

SOEPinfo should be your first step in starting any new SOEP project!

Language Version / Spachversion:

German English

[Variables] [Topics] [Questionnaires]

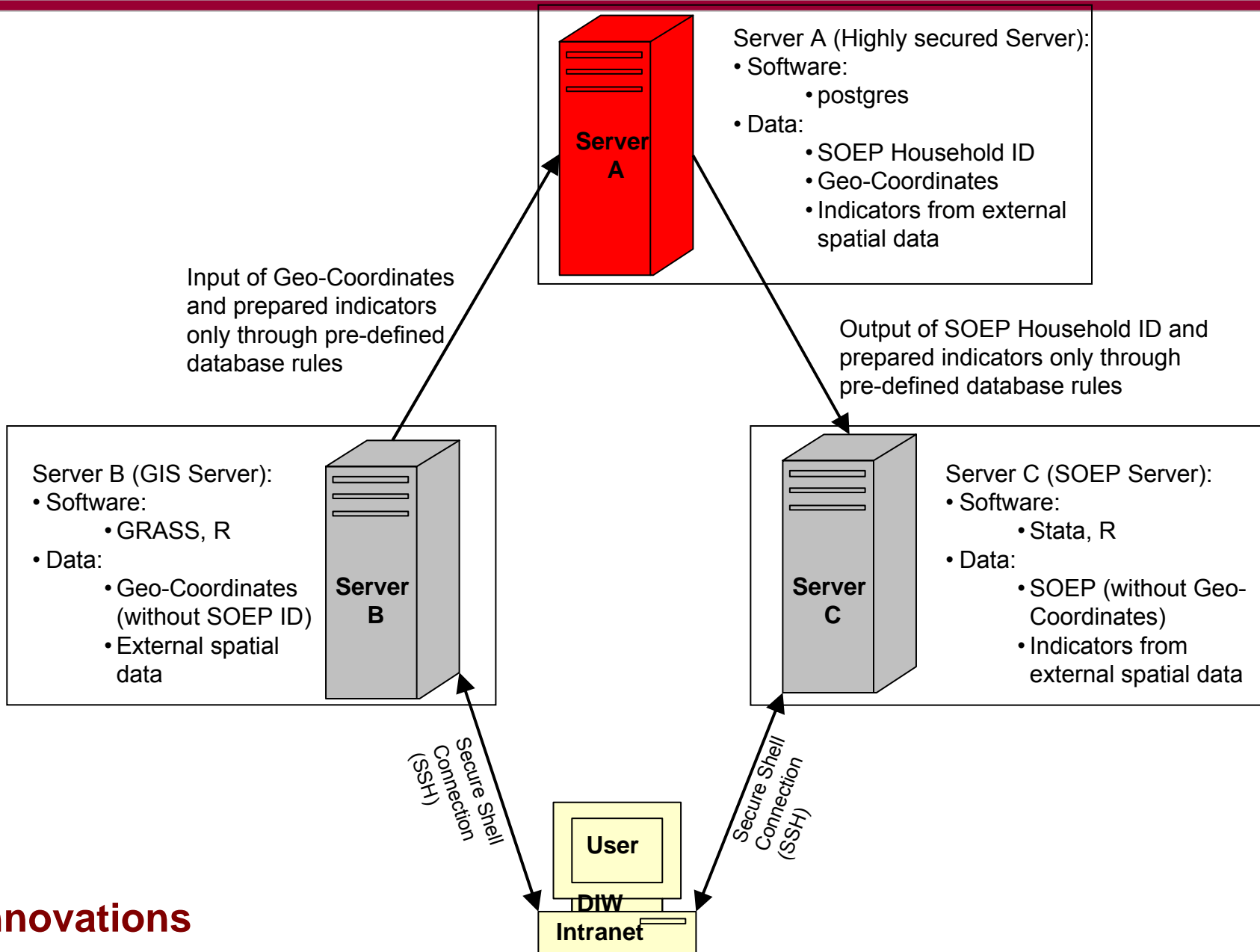
- Introduction
- Sub-samples and Survey Related Issues
- Sample Size Development
- Content
- **Data Access**
- Innovations

- Strict data protection legislation requires ...
 - Data user contract with DIW Berlin
 - Intl. Users 95% sample: Co-operation with Cornell Univ., Ithaca/NY
- Data dissemination currently on CD-Rom only
 - detailed written documentation
 - full German and English labeling (variables and values)
 - within EU: 100% sample
 - outside EU: International Scientific Use Version (95% sub-sample)
 - some variables dropped or recoded (regional info)
- Low nominal data fee (30 EUR/year; 125 USD one-time)
 - yearly updates on CD-Rom

- Access to sensitive information (counties or zip-codes, microm, geo-coordinates, plaintext answers):
 - **Visiting** DIW Berlin, normal workplace (counties, microm and plaintext answers)
 - Remote access via **SOEPremote** (counties)
 - Email with stata syntax, fully automatic, 24/7
 - Visiting DIW Berlin, access to **SOEPgeo** Server (Zip-codes and geo-coordinates)

- Introduction
- Sub-samples and Survey Related Issues
- Sample Size Development
- Content
- **Data Access**
- **Innovations**

- **SOEPgeo** ... allows processing of highly sensitive geo-coordinates (latitude and longitude) and data on zip-code level available on the premises of DIW Berlin. SO: with SOEPgeo, the SOEP department acts as a “Research Data Centre” (FDZ).
- ... will improve current situation, as up to now data on a zip-code level can be analyzed on a single non-networked computer only (for internal as well for external users).
- ... combines geo-coordinates and SOEP IDs on one server only which can not accessed by any user directly.
- It is intended that this server-based infrastructure will be available to all guest researchers in fall 2009.



- Introduction of **event/age-specific questionnaires** (newborns, toddlers, preschoolers, youth, emigrants)
- Identification of twins
- Introduction of questions about **personal traits**
- Introduction of a new module on **individual wealth** holdings
- Two pretests of web-based surveying
- Measuring above and **beyond “survey questions”**
 - Evaluation of health (by taking grip strength)
 - Behavioral experiments
 - Evaluation of cognitive abilities
- Improvements of imputation techniques (MI)
- Improvement of industry (NACE) and ISCO coding – ready for ISCO2008
- Linking **neighborhood** data to SOEP households (microm)
- Linking **geographical codes** to SOEP households
- Establishment of an **interviewer survey**
- (Test) survey of attriters

Innovations

- Link to remote sensing data
- Sampling twins as a related sample (pretest study)
- Test of fully web-based SOEP questionnaires
- Test of mobile-phone-based interviewing and assessments
- Test of the event-triggered questionnaire “The Deceased Person” (in main SOEP wave of 2009)
- Test of statistical matching: SOEP and Social Security Register Data
- Pretests on educational aspirations of parents
- Pretest on factorial design (“vignettes approach”)
- Pretests evaluating the cognitive and non-cognitive skills of preschoolers and links with educational institutions (day care, school)
- Pretests of saliva sample collection

Innovations

- Pretest on the longitudinal impact of incentives
- Pretest on better measurement of incomes, savings and expenditures
- Standardization of personality indicators, mood and emotions
- Questionnaire for children 8-9 years of age

**Thank you for your
attention!**

E-mail:

jgoebel@diw.de

SOEP-Hotline:

soepmail@diw.de

SOEP-Homepage:

www.diw.de/soep/