

Einführung in das GSOEP: Grundlagen, Fragebogen und Item Correspondence

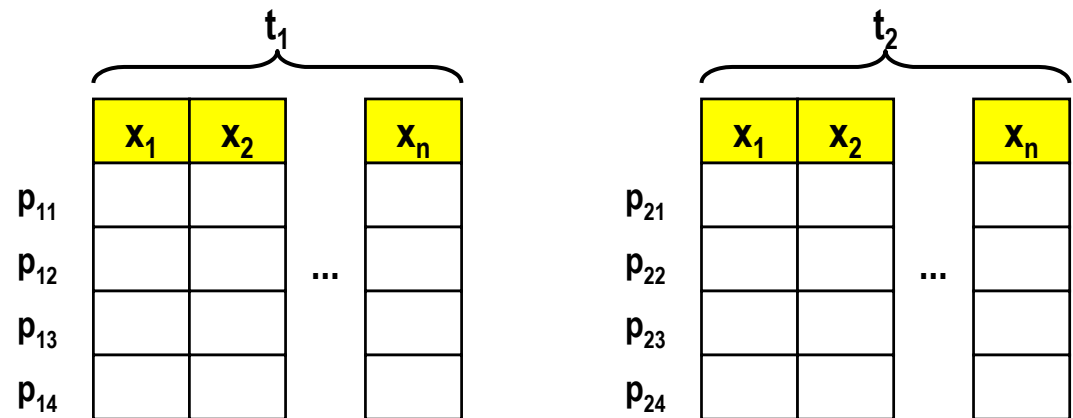
08. Dezember 2003

Wiederholung: Datensätze: Querschnitt ↔ Längsschnitt I

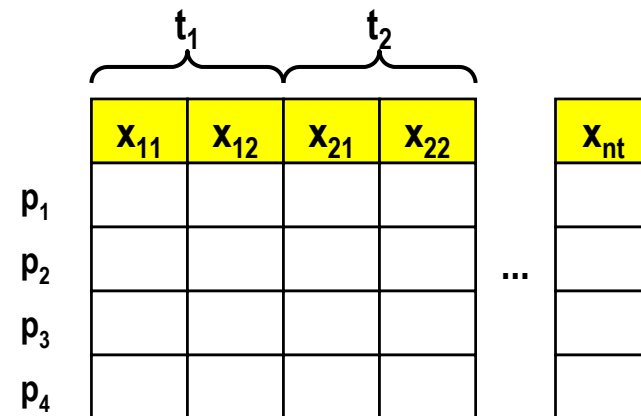
- > **Querschnittsdaten** = nur zu einem Zeitpunkt erhoben oder keine zeitspezifischen Informationen
- > **Längsschnittdaten** = zu mehreren Zeitpunkten erhoben oder zeitspezifische Informationen
 - **Trendstudie**: mind. drei Querschnittserhebungen, gleiche Instrumente, verschiedene Personen
 - **Panel**: mehrere Erhebungszeitpunkte mit gleichem Abstand, gleiche Instrumente, gleiche Personen
- > Kausale Schlussfolgerungen empirisch nur mit Längsschnittdaten zu ziehen (Ursache-Wirkungs-Zusammenhang ist zeitabhängig)
- > theoretische Kausalitäten auch mit Querschnittsdaten möglich, aber nicht empirisch testbar

Wiederholung: Datensätze: Querschnitt ↔ Längsschnitt II

> Trendstudie:



> Panelstudie:



Wiederholung: Längsschnitt

spezifische Probleme von Panelstudien:

> **Panelmortalität:**

- stetige Verringerung des Panel-Bestands (ca. 5-15% pro Welle)
- Ursachen: Umzug, Tod, Verweigerung, „drop out“
- Gegenmaßnahme: Panelpflege

> **Paneleffekte:**

- Anpassung von Antworten an vorhergehende Befragung

> Beispiele für Panelstudien: SOEP, BHPS, PSID, ECHP, IAB-Betriebspanel

Was ist das GSOEP?

- > Eine repräsentative Längsschnittstudie privater Haushalte in der Bundesrepublik Deutschland
- > Seit 1984 jährliche Wiederholungsbefragung von Deutschen, Ausländern und Zuwanderern in den alten und neuen Bundesländern
- > Over-Sampling von Migrant/innen
- > Stichprobe 2001: ca. 12.000 Haushalte mit mehr als 22.000 Personen
- > Themenschwerpunkte:
 - Haushaltszusammensetzung
 - Erwerbs- und Familienbiographie
 - Erwerbsbeteiligung und berufliche Mobilität
 - Einkommensverläufe
 - Gesundheit und Lebenszufriedenheit
- > Objektive (Einkommen etc.) und subjektive (Zufriedenheiten etc.) Indikatoren
- > Steht universitären Forschungseinrichtungen für Forschung und Lehre in SPSS, SAS, TDA, STATA und ASCII Format zur Verfügung

Stärken des GSOEP: Besondere Analysemöglichkeiten

- > Längsschnittdesign (Panelcharakter)
- > Individualbasis, Haushalts- und Familienkontext (Befragung aller HH-Mitglieder über 16 Jahre)
- > Möglichkeit innerdeutscher Vergleiche
- > Überproportionale Ausländerstichprobe: gegenwärtig größte Wiederholungsbefragung bei Ausländern in der Bundesrepublik Deutschland („Gastarbeiter“-Nationalitäten)
 - Erhebung von Zuwanderung: die einzige methodisch zuverlässige Stichprobe von Zuwanderern, die von 1984 bis 1995 nach Westdeutschland gekommen sind
- > Hohe Kontinuität: 1984 beteiligten sich im SOEP-West 5921 Haushalte mit 12290 Personen an der Befragung; nach 17 Wellen im Jahre 2000 sind es noch 4060 Haushalte mit 7623 Personen
- > Ergänzungsstichproben: Ostdeutsche, Refreshment Sample, Hohe Einkommensgruppen
- > Umfangreiche Online-Dokumentation (<http://www.diw.de/deutsch/sop/>)

Die wichtigsten Themengebiete

- > Bevölkerung und Demographie
- > Aus- und Weiterbildung, Qualifikation
- > Arbeitsmarkt und berufsbezogene Dynamiken
- > Verdienst, Einkommen und soziale Sicherung
- > Wohnen
- > Gesundheit
- > Haushaltsproduktion
- > Grundlegende Orientierung (Präferenzen, Werte, etc.)
- > Allgemeine und besondere Zufriedenheit

Spezielle Themen-Module I

- > A – 1984: Erwerbsbiographie seit dem 15. Lebensjahr
- > B – 1985: Ehe- und Familienbiographie
- > C – 1986: Soziale Wurzeln, erster Job, Nachbarschaft
- > D – 1987: Soziale Sicherung, Frühverrentung, Behinderungen
- > E – 1988: Vermögen
- > F – 1989: Weiterbildung und Qualifikation
- > G – 1990: Zeiteinteilung und –nutzung
- > H – 1991: Familien- und soziale Dienstleistungen
- > I – 1992: Soziale Sicherheit und Armut
- > J – 1993: Weiterbildung und Qualifikation
- > K – 1994: Nachbarschaft, Werte, Erwartungen

Spezielle Themen-Module II

- > L – 1995: Zeiteinteilung und –nutzung, Einkommen
- > M – 1996: Familien- und soziale Netzwerke
- > N – 1997: Soziale Sicherheit und Armut
- > O – 1998: Ökologie und Umweltverhalten, indirekte Besteuerung
- > P – 1999: Zeiteinteilung und –nutzung, Erwartungen
- > Q – 2000: Weiterbildung und Arbeitsmarkt
- > R – 2001: Soziale Netzwerke, Arbeitsbedingungen
- > S – 2002: Soziale Sicherheit, Vermögen
- > T – 2003: Ökologie und Umweltverhalten

Subsamples des SOEP

- > A – (West-)Deutsche Einwohner/innen; $N_{1984} = 4528$ HH
- > B – Migrant/innen; $N_{1984} = 1393$ HH [Oversampling]
- > C – (Ost-)Deutsche Einwohner/innen; $N_{1990} = 2179$ HH
- > D – Immigrant/innen; $N_{1994} = 522$ HH
- > E – Refreshment; $N_{1998} = 1067$ HH
- > F – Innovation; $N_{2000} = 6052$ HH
- > G – Hohe Einkommen; $N_{2002} = 1224$

Follow-Up-Konzept

- > Alle Personen eines HH über 16 Jahre werden im folgenden Jahr ebenfalls befragt
- > Auch bei Umzug werden die Personen weiterhin befragt
- > Auch in einen SOEP-Haushalt einziehende werden miteinbezogen (seit 1989)
- > Auswirkungen:
 - regionale Mobilität ist beobachtbar
 - demographische Entwicklungen sind im Sample
 - Panel-Mortalität ist eingeschränkt

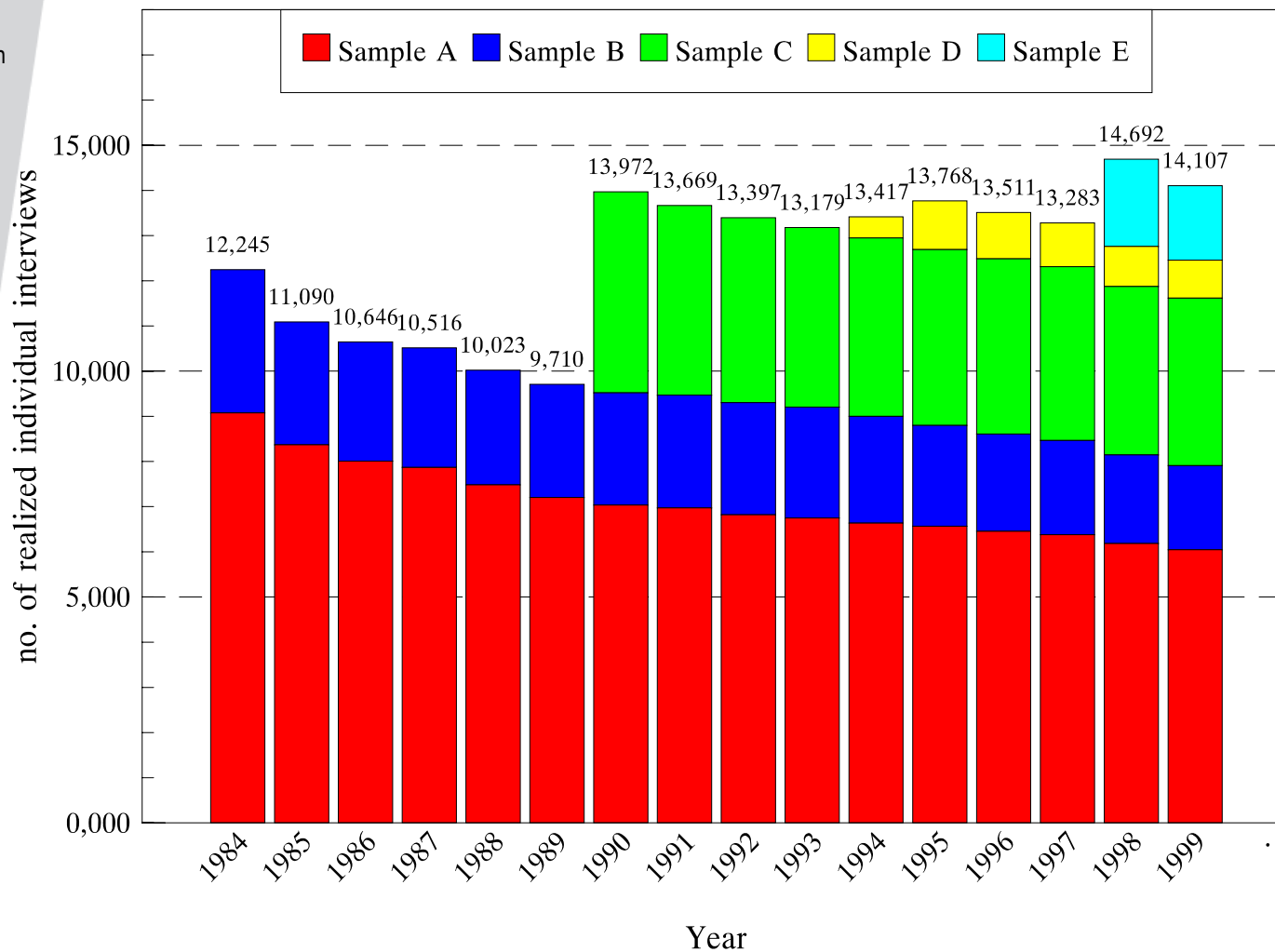
Alte und neue Haushalte bzw. Personen

Haushalte

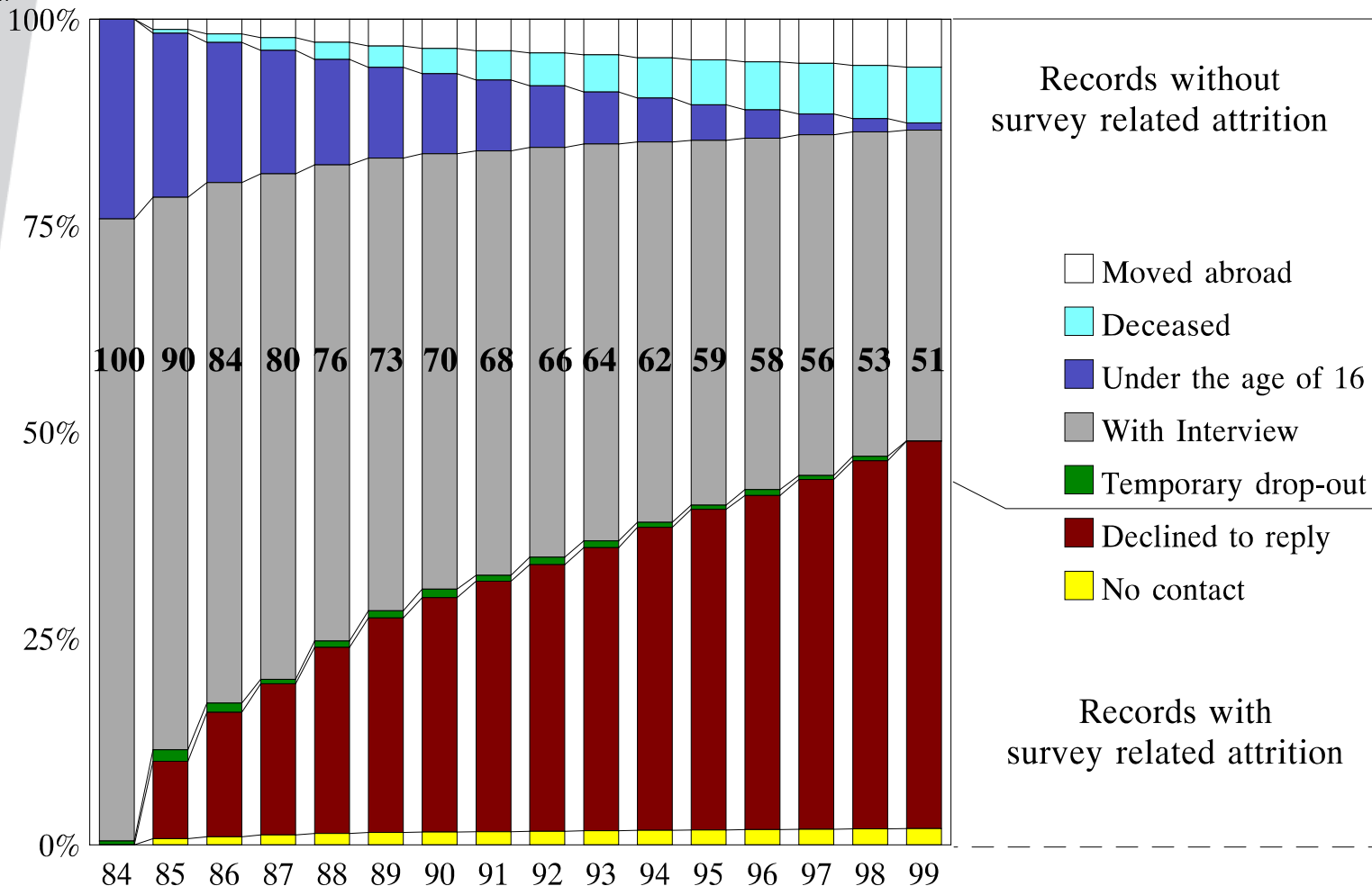
Personen

	alt	neu
alt	klass. Fall (keine Adressenänderung) ganzer HH zieht um	Auszug
neu	Geburt Einzug	Geburt Auszug alter Personen aus alten HH (Auflösung)

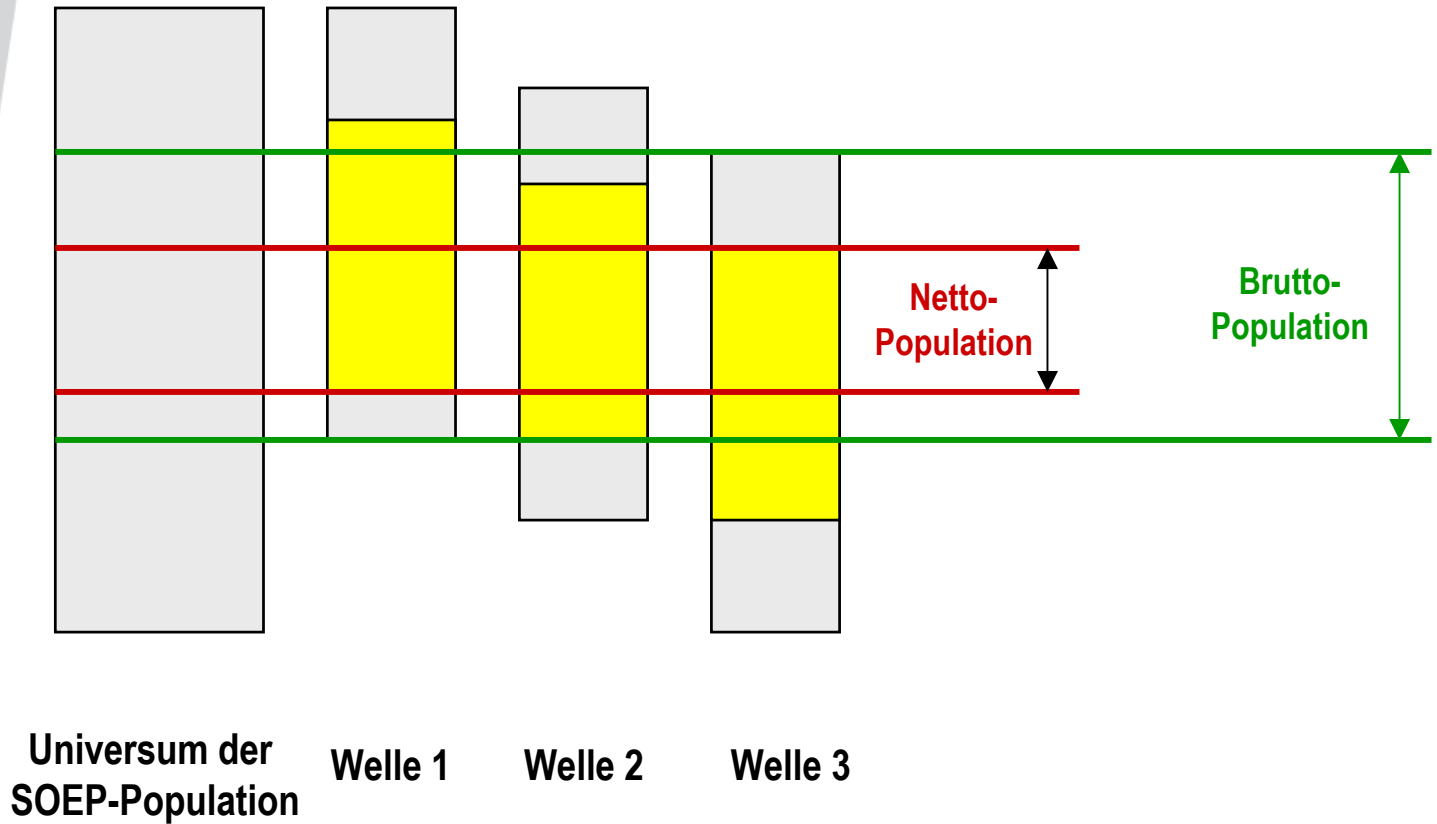
Erfolgreich interviewte Personen 1984-1999



Verbleib der interviewten Personen von 1984 (Subsample A + B)



Die SOEP-Datenstruktur (Längs- und Querschnitt)



Haushalts-Dateien

- > HPFAD HH-Grundinformationen für alle Wellen
- > HHRF HH-Gewichtungsvariablen
- > \$HBRUTTO HH-Adressenregister (jede Welle)
- > \$H / \$HOST HH-Variablen (jede Welle)
- > \$HGEN generierte HH-Variablen (jede Welle)

Matching-Variablen: HHNR, HHNRAKT (\$HHNR)

Personen-Dateien

- > PPFAD Grundinformationen (für alle Wellen)
- > PHRF Gewichtungsvariablen
- > \$PBRUTTO Register (jede Welle)
- > \$P / \$POST Variablen (jede Welle)
- > \$PGEN generierte Variablen (jede Welle)
- > \$PKAL Kalendarien (jede Welle)
- > \$PLUECKE Daten zum temporären Drop-Out (jede Welle)
- > \$KIND Kinder-Daten (jede Welle)

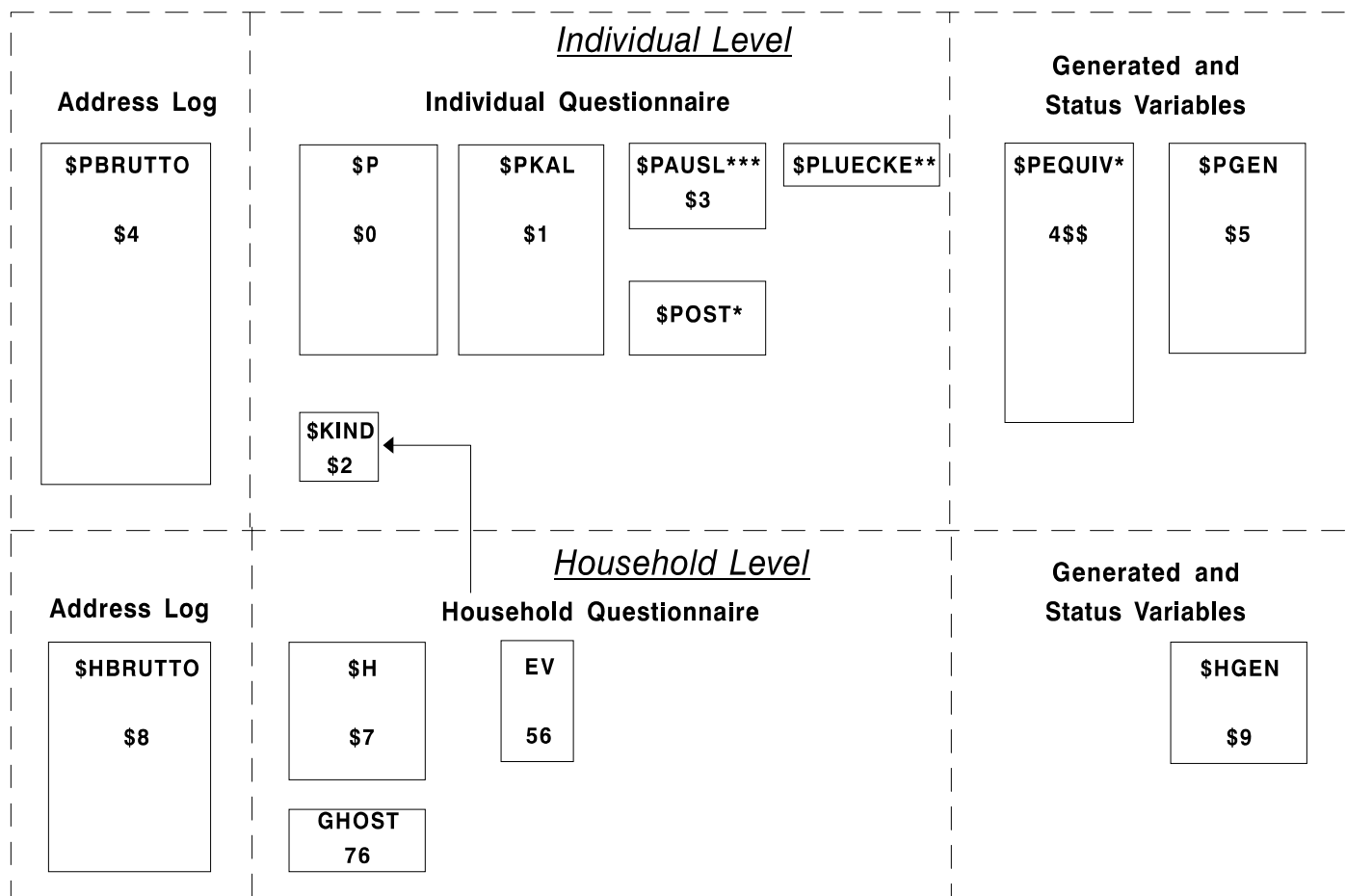
Matching-Variablen: HHNR, HHNRAKT, PERSNR

Biographie- / Spell-Dateien

- > BIOIMMIG Migrations-Biographie
- > BIOJOB Erwerbsbiographie (Arbeitseinstieg)
- > PBIOSPE Biographie-Kalender
- > ARTKALEN Aktivitäts-Kalender
- > EINKALEN Einkommens-Kalender
- > PFLEGE Pflegebedürftigkeits-Kalender
- > SOZKALEN Sozialunterstützungs-Kalender
- > BIOBIRTH Geburtsbiographie
- > BIOMARSM Familienstands-Biographie (monatlich)
- > BIOMARSY Familienstands-Biographie (jährlich)
- > BIOPAREN Eltern-Informationen

Matching-Variablen: HHNR, HHNRAKT, PERSNR

Die SOEP Datenstruktur (Querschnitt)



\$: Wave specification: A,B,C ... N for file names; 1,2,3 ... 14 for file numbers.

* Waves G and H only; ** Waves B through M only; *** Waves A through L only

Die SOEP-Datenstruktur (Längsschnitt)

<i>Individual/ Household</i>				<i>Cumulative Individual Data</i>	<i>Spell</i>					<i>Individual</i>		
					<i>House- hold</i>	<i>Individual</i>						
META- DATA		WEIGHTING FACTORS		DROP- OUTS	SOCIAL ASSIST. (month)	CALENDAR (month)	OCCUP. BIO (year)	MARITAL STATUS (year) (month)		BIRTHS (women only)	PARENTAL INFO	
PPFAD 01	HPFAD 02	PHRF 03	HHRF 07	YBBRUTTO 260	SOZ- KALEN 289	Occup. ART- KALEN 81	Income EIN- KALEN 82	PBIO- SPE 80	BIO- MARSY 282	BIO- MARSM 281	BIO- BIRTH 280	BIO- PAREN 283
				GENERATED INCOME								
				RBEINK 375								
				PERSONS NEEDING CARE								
				PFLEGE 268								
											MIGRATION (migrants only)	FIRST JOB
											BIO- IMMIG 284	BIO- JOB 285

Variablen-Benennung

Ziffer

Bedeutung

1

Welle

2

Analyseebene (Person, HH)

3-4

Fragennummer im Fragebogen

5-6

Item-Nummer im Fragebogen

5 / 7

Differenzierung je nach Sample

(A = Ausländer, O = Ostdeutsche)

2-8

ganze Namen (in \$BRUTTO und \$P)