

Steffen Hillmert, Ralf Künster, Petra Spengemann, Karl Ulrich Mayer

Projekt „Ausbildungs- und Berufsverläufe der
Geburtskohorten 1964 und 1971 in Westdeutschland“.
Dokumentation

Teil I Projektüberblick und Projektbeschreibung

Berlin 2004



Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland ist Teil des Forschungsprogramms 'Lebensverläufe und gesellschaftlicher Wandel' am Max-Planck-Institut für Bildungsforschung, Berlin, und wurde seit 1998 in Kooperation mit dem Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg, durchgeführt.

The project *Education, Training, and Occupation: Life Courses of the 1964 and 1971 Birth Cohorts in West Germany* is the most recent part of the German Life History Study (GLHS) and has been conducted since 1998 by the Max Planck Institute for Human Development, Center for Sociology and the Study of the Life Course (Berlin) in cooperation with the Institute for Employment Research (Institut für Arbeitsmarkt- und Berufsforschung – IAB, Nürnberg).

Inhalt der Dokumentation

- Teil I **Projektüberblick und Projektbeschreibung**
- Überblick über das Projekt (mit einer Zusammenfassung, auch in Englisch)
 - Beschreibung der Befragungsstruktur (Module)
 - Beschreibung der Datenstruktur
 - Beschreibung der Materialien
 - Beschreibung des Scientific Use Files
 - Beschreibung der einzelnen Projektphasen (Chronologie): Design, Erhebung, Edition, Vercodung, Datenprüfung, Fallselektion
 - Liste der am Projekt Beteiligten

- Teil II **Fragebogen**
- Fragen
 - Variablennamen (einschließlich hinzugefügter Variablen)
 - Anweisungen für die Interviewer
 - Antwortvorgaben
 - Fragensteuerung

- Teil III **Variablenliste**
- Liste aller Variablennamen, Variablen- und Value-Labels

- Teil IV **Editionshandbuch**
- Grundlagen der Edition
 - Allgemeine und Zeitanpassungsregeln
 - Editionsverlauf
 - Spezifische Editionsregeln zu den einzelnen Modulen
 - Variablenindex
 - Stichwortindex

- Teil V **Nachrecherche**
- Nachrechercheregeln, Ablaufdiagramme und Anschreiben

- Teil VI **Einzelfallentscheidungen**
- Liste der getroffenen Einzelfallentscheidungen

- Teil VII **Vercodung**
- Beschreibung der Vercodungen

- Teil VIII **Programmdokumentation**
- Beschreibung des Programms LDEX

- Teil IX **Infas-Methodenberichte**
- Bericht zum Methodentest
 - Methodenbericht zur Hauptstudie

Alle Teile, mit Ausnahme der Infas-Methodenberichte, sind einzeln seitennummeriert und für den beidseitigen Ausdruck vorbereitet.

Überblick

Das Projekt "Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland" ist die jüngste Teilstudie des Forschungsprogramms "Lebensverläufe und gesellschaftlicher Wandel". Dieses Forschungsprogramm wird seit 1983 unter der Leitung von Karl Ulrich Mayer am Max-Planck-Institut für Bildungsforschung (Berlin), Forschungsbereich "Bildung, Arbeit und gesellschaftliche Entwicklung" durchgeführt.

Das hier vorgestellte Projekt wurde in Kooperation mit dem Institut für Arbeitsmarkt- und Berufsforschung (IAB)*, Nürnberg, entwickelt und durchgeführt. Die Erhebung der Daten oblag dem Markt- und Meinungsforschungsinstitut Infas in Bonn und wurde in dem Zeitraum 1998/99 durchgeführt.

Die Individualdaten enthalten Informationen über 2909 Lebensverläufe und erlauben detaillierte Längsschnittanalysen insbesondere zum Ausbildungsverhalten, zu Arbeitsmarkteinstieg und Erwerbsverläufen sowie zu Prozessen der Familienbildung in den achtziger und neunziger Jahren.

Nach einer Phase der umfangreichen Datenedition liegt mit diesen Bänden nun eine ausführliche Dokumentation vor. Die gedruckte Version und die CD-Version sind identisch.

Overview

The project "Education, Training, and Occupation: Life Courses of the 1964 and 1971 Birth Cohorts in West Germany" is the latest project within the research program "Sociology and the Study of the Life Course". It is an ongoing research program headed since 1983 by Karl Ulrich Mayer at the Center for Sociology and the Study of the Life Course at the Max Planck Institute for Human Development in Berlin.

The project introduced here was developed and conducted in cooperation with the Institute for Employment Research (IAB) in Nürnberg. The Marketing and Opinion Research Institute Infas was responsible for collecting the data conducted during the 1998/99 time period.

The survey data contain information from over 2,909 life courses and allow detailed longitudinal analyses particularly of training behavior, labor market entry and occupational careers as well as processes of family formation in the 1980s and 1990s.

After an extensive edition of the data, a full documentation of the study conducted can now be made available in these volumes. The printed version is identical with the version on CD-ROM.

* In diesem Zusammenhang wurde das Projekt mit Mitteln des Europäischen Sozialfonds gefördert.

Einige Hinweise zur Orientierung

Einen schnellen Überblick über das Projekt "LV-West 64/71" vermittelt das Kapitel 1.

Von besonderer Bedeutung für die Nutzerinnen und Nutzer der Daten dürften die Kapitel 2 in Teil I und die Teile II (Fragebogen) und III (Variablenliste) sein. Für spezifischere Fragen ist außerdem das Editionshandbuch (Teil IV) eine wichtige Informationsquelle. In diesem Zusammenhang sei nachdrücklich empfohlen, sich vor Beginn der Datenaufbereitung und -analyse gründlich mit der Dokumentation und der Struktur der Daten vertraut zu machen. Die Lebensverlaufsdaten sind aufgrund ihrer Detailliertheit komplex. Dies bedeutet, dass jeweils Informationen aus mehreren Modulen benötigt werden, um zum Beispiel die vollständigen Bildungs- oder Erwerbsgeschichten zu bekommen.

Die Daten stehen ab Januar 2004 beim Zentralarchiv für Empirische Sozialforschung an der Universität zu Köln unter der ZA-Studiennummer 3927 als Scientific Use File zur Verfügung. Diese Dokumentation ist sowohl für interne wie auch externe Datennutzerinnen und Datennutzer gedacht. Aus Datenschutzgründen gelten aber für den Scientific Use File einige Einschränkungen, die in Kapitel 2.7 (Seite 19) aufgeführt sind.

Wir gehen davon aus, dass die Nutzerinnen und Nutzer in ihren Arbeiten auf die Herkunft der Daten in geeigneter Form hinweisen. Für die Datenbeschreibung ist diese Dokumentation die geeignete Referenz.

Dank

An dieser Stelle möchten wir uns bei den zahlreichen Mitarbeiterinnen und Mitarbeitern bedanken, die das Gelingen dieses Projekts sichergestellt haben, insbesondere den Kooperationspartnern am IAB, Stefan Bender und Hans Dietrich. Eine Aufstellung findet sich auf Seite 31.

Teil I: Inhalt

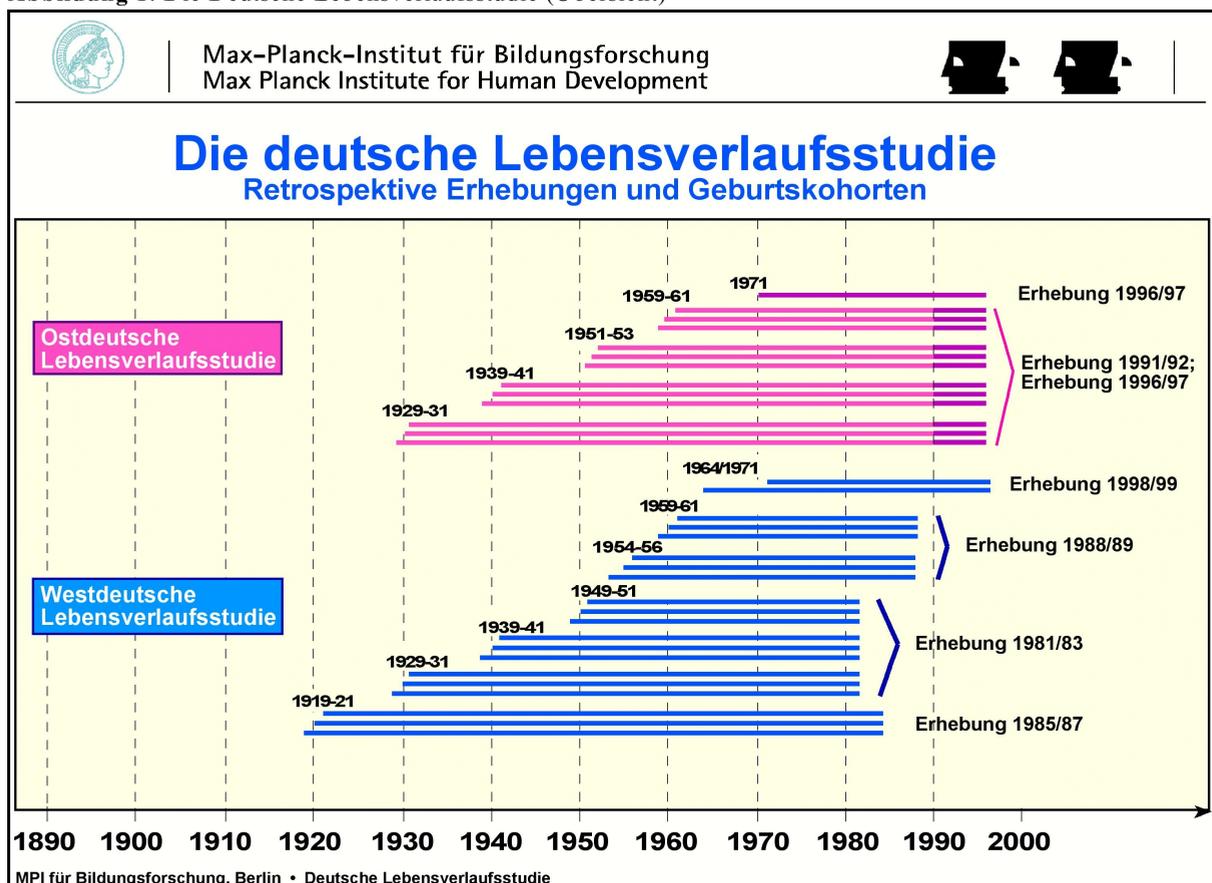
1	Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland	1
1.1	Die deutsche Lebensverlaufsstudie und ihre jüngste Teilstudie	1
1.2	Ablauf der Lebensverlaufsstudie 1964/71	3
1.3	Charakterisierung der Stichprobe	4
1.4	Datenedition	5
1.5	Datenstruktur und bisherige Arbeiten	5
1.6	Fazit	6
1.7	Literatur	7
2	Datenbeschreibung	8
2.1	Datenstruktur	8
2.2	Erläuterungen zu den Modulen	12
2.2.1	Angaben zur Zielperson (Record 1)	12
2.2.2	Eltern (Record 2)	12
2.2.3	Geschwister (Record 3)	12
2.2.4	Wohngeschichte (Record 4)	12
2.2.5	Schulische Ausbildung (Record 5)	12
2.2.6	Berufliche Ausbildung und Weiterbildung (Records 6 und 11)	13
2.2.7	Erwerbstätigkeiten (Record 7/8)	15
2.2.8	Arbeitslosigkeiten (Record 9)	15
2.2.9	Nebentätigkeiten (Record 10)	16
2.2.10	Lückenmodul	16
2.2.11	Berufskontrolle, Politik, Religion (Record 15)	16
2.2.12	Partnerschaften (Record 16)	17
2.2.13	Kinder (Record 17)	17
2.2.14	Wohnung, Haushaltszusammensetzung und -einkommen (Record 18)	17
2.2.15	Interviewende (Record 19)	17
2.3	Erläuterungen zur Datenstruktur	18
2.4	Fragebogen	18
2.5	Variablen	19
2.6	Edierte und codierte Daten	19
2.7	Anonymisierung der Daten für den Scientific Use File	19
2.8	Statistik der Edition und der Nachrecherche	20
3	Chronologie des Projekts "Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland"	21
3.1	Design und Ablauf	21
3.2	Datenerhebung	21
3.2.1	Erhebungsinstrument	21
3.2.2	Fragebogen	21
3.3	Daten	22
3.4	Datenedition	22
3.4.1	Editionshandbuch	22
3.4.2	Fallmaterialien	23
3.4.3	Nachrecherche	23
3.4.4	Verlaufsdokumentation und -statistik von Edition und Nachrecherche	23
3.5	Vercodung	24
3.6	Einzelfallentscheidungen	25
3.7	Datenprüfung	25
3.7.1	Wertebereichsprüfung	26
3.7.2	Zeitliche Konsistenzprüfung	27
3.7.3	Filterführungsprüfung	29
3.7.4	Datenkorrektur	30
3.8	Auswahl der Fälle für den Analysedatensatz	30
4	Beteiligte Personen	31
5	Programm LDEX	31

1 Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland

1.1 Die deutsche Lebensverlaufsstudie und ihre jüngste Teilstudie

Die *Deutsche Lebensverlaufsstudie* (*German Life History Study/GLHS*) ist aus Arbeiten am Sonderforschungsbereich 3 'Mikroanalytische Grundlagen der Gesellschaftspolitik' an den Universitäten Mannheim und Frankfurt/M. hervorgegangen und wird unter der Leitung von Karl Ulrich Mayer seit 1979 durchgeführt, ab 1983 am Max-Planck-Institut für Bildungsforschung. Die Datenbasis setzt sich – mit Ausnahme einer Panelbefragung 1996/97 im Rahmen der ostdeutschen Teilstudien – aus einer Reihe einmaliger, teilweise auch mehrmaliger Retrospektivbefragungen von Personen aus ausgewählten Geburtsjahrgängen zusammen (siehe Abbildung 1; für Westdeutschland vgl. auch Tabelle 1 und Wagner 1996). Die betreffenden Geburtskohorten wurden aufgrund von Vorinformationen zu ihrer kollektiven Lebenssituation (nach Zensusdaten) ausgewählt. In ihrer Gesamtheit erlauben die Lebensverlaufsdaten für Deutschland detaillierte und historisch weiter zurückreichende Analysen, die das gesamte 20. Jahrhundert erfassen können¹. Die früheren Teilstudien sind bereits seit längerer Zeit beim ZA verfügbar (ZA-Studiennummern 2645-2648 für Westdeutschland und 2644 für Ostdeutschland).

Abbildung 1: Die Deutsche Lebensverlaufsstudie (Übersicht)



¹ Vgl. dazu <http://www.mpib-berlin.mpg.de/de/forschung/bag/index.htm>; siehe auch Brückner/Mayer 1998.

In der neuesten Teilstudie (zur Programmatik des Projekts vgl. auch Corsten/Hillmert 2001) wurden die bis Ende der achtziger Jahre in der alten Bundesrepublik durchgeführten Befragungen durch Daten ergänzt, die die Bildungs- und Erwerbsverläufe der Geburtsjahrgänge 1964 und 1971 in Westdeutschland bis Ende der neunziger Jahre verfolgen.

Inhaltliche Ausgangspunkte dieser Teilstudie waren die phasenweise sehr schwierige Lage auf dem Ausbildungs- und Arbeitsmarkt der achtziger und neunziger Jahre. Hinzu kommt die besondere demographische Situation des Jahrgangs 1964. Da es sich hierbei um die größte Kohorte seit dem Bestehen der Bundesrepublik handelt, ist eine besondere Konkurrenzsituation innerhalb der Bildungsinstitutionen und beim Einstieg in den Arbeitsmarkt zu vermuten (vgl. auch Hillmert 2001). Im Gegensatz dazu ist der Jahrgang 1971 bereits deutlich kleiner gewesen. Schließlich erlaubt die Wahl der Geburtskohorte 1971 den direkten Vergleich mit der entsprechenden Teilstudie in Ostdeutschland.

Gegenüber den bisherigen Lebensverlaufsstudien weist die jüngste Erhebung eine Reihe von Neuerungen auf. Den Schwerpunkt der Studie bildet die genaue Rekonstruktion der Ausbildungs- und Erwerbsphasen einschließlich der vielfältigen Übergänge zwischen Ausbildungen und Erwerbstätigkeiten. Stichproben- und Befragungsdesign wurden leicht verändert. So wurden im Gegensatz zu den bisherigen Lebensverlaufsstudien auch in Deutschland lebende (Deutsch sprechende) Ausländer in die Untersuchung einbezogen. Außerdem wurden auch zeitlich parallele Verläufe von Ausbildung und Erwerbstätigkeit monatsgenau aufgenommen. Schließlich wurde mit dieser Studie auch in methodischer Hinsicht Neuland betreten. Dies galt zum einen für die automatisierte Datenprüfung während des Interviews. Zum anderen wurde für die Möglichkeit einer Verknüpfung der erhobenen Daten mit Informationen aus der Beschäftigtenstatistik die Zustimmung der Befragten eingeholt. Ansonsten hat sich das Fragenprogramm aber an den bisherigen Teilstudien orientiert, und dies erlaubt langfristige Vergleiche zwischen den einzelnen Geburtskohorten.

Tabelle 1: Die westdeutschen Teilstudien der Lebensverlaufsstudie

Teilstudie	Lebensverläufe und Wohlfahrtsentwicklung (LV I)	Die Zwischenkriegskohorte im Übergang zum Ruhestand (LV II)	Berufszugang in der Beschäftigungskrise (LV III)	Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 (LV-West 64/71)
ZA-Studiennummer	2645	2646/2647	2648	3927
Erhebungszeitraum	1981-83	LV IIA: 1985/86 LV II/TEL: 1987/88	1989	1998-99
Geburtskohorten	1929-31 1939-41 1949-51	1919-21	1954-56 1959-61	1964 1971
Realisierte Fälle	1929-31: 709 1939-41: 733 1949-51: 729	LV IIA: 407 LV II/TEL: 1005	1954-56: 1008 1959-61: 1000	1964: 1474 1971: 1435
Erhebung	Persönliche Interviews	LV IIA: Persönliche LV II/TEL: telefonische Interviews	Telefonische Interviews	Größtenteils telefonische Interviews
Grundgesamtheit	Deutsche Wohnbevölkerung der entsprechenden Geburtsjahrgänge in Privathaushalten in der Bundesrepublik Deutschland			Deutschsprachige Wohnbevölkerung dieser Geburtsjahrgänge in Privathaushalten in den alten Bundesländern

Die Erhebung für die Teilstudie der 1964 und 1971 Geborenen in Westdeutschland wurde gemeinsam vom Institut für Arbeitsmarkt- und Berufsforschung in Nürnberg und dem Max-Planck-Institut für Bildungsforschung in Berlin in Auftrag gegeben. Die Haupterhebung der Daten erfolgte in den Jahren 1998 und 1999 in Kooperation mit dem Umfrageinstitut Infas (siehe Infas 1999). Der Anhang enthält eine Übersicht über die am Projekt beteiligten Personen.

Angesichts ihrer Detailliertheit bedurfte die Erhebung der Lebensverlaufsdaten einer längeren Vorbereitung. Nach der Erhebung, ab Anfang 1999, wurden die Daten im Max-Planck-Institut für Bildungsforschung genau geprüft und ediert. Falls nötig, wurden die Zielpersonen noch einmal telefonisch kontaktiert und um ergänzende Informationen gebeten. Ende 2002 war diese Phase dann weitgehend abgeschlossen.

1.2 Ablauf der Lebensverlaufsstudie 1964/71

Auf der Grundlage der am Max-Planck-Institut für Bildungsforschung bereits durchgeführten Lebensverlaufsstudien wurde im Februar 1997 mit der Fragebogenentwicklung und der Konzeption der Studie begonnen.

Zunächst wurde ein Methodentest durchgeführt, um die Ausschöpfung bei der Einwilligung zur Zuspiegelung der Sozialversicherungsdaten verbessern zu helfen; das Stichprobendesign wurde gleichzeitig für den Methodentest und die Hauptstudie festgelegt. Nach einer mehrwöchigen Testphase wurde eine Kurzfassung des geplanten Erhebungsinstruments mit den wichtigsten Fragemodulen von Mai bis Juni 1997 getestet (siehe im Einzelnen: Infas 1997). Im ersten großen Pretest wurden dann im September und Oktober 1997 insgesamt 104 Interviews durchgeführt, davon 98 vollständig. Anschließend wurden die Tonbandaufzeichnungen der Interviews am Max-Planck-Institut ausgewertet und kommentiert. Außerdem wurden die bereits vorliegenden Daten und das Erhebungsinstrument (Computerprogramm) geprüft, um Fehler und Probleme an das Erhebungsinstitut zurückmelden zu können. Nach zahlreichen weiteren Tests und daraus resultierenden Änderungen des Erhebungsinstruments – einschließlich der Einführung einer automatischen Prüfung der chronologischen Konsistenz – wurde ein zweiter Pretest durchgeführt. Hierbei wurden noch einmal 43 Interviews, davon 39 vollständig, durchgeführt. Von März bis Mai 1998 wurden erneute Überarbeitungen vorgenommen und dann ab Juni 1998 die ersten Telefoninterviews für die Hauptstudie mit der aktualisierten Version des Erhebungsinstruments durchgeführt. Auch nach Feldbeginn waren noch Änderungen nötig, die gesondert dokumentiert wurden. Die meisten Interviews erfolgten telefonisch. Darüber hinaus wurde aus Vergleichsgründen eine Reihe persönlicher Interviews geführt. Zu diesem Zweck wurden im Juli zusätzliche Interviewer geschult. Insgesamt wurden im Rahmen der Erhebung 2.935 Interviews von 53 Interviewerinnen und Interviewern durchgeführt. Um die Ausschöpfung zu erhöhen, wurde im September 1998 zusätzlich mit einer Non-Response-Studie begonnen (siehe dazu im Einzelnen: Infas 1999: 41 ff.).

Die Tabelle 2 stellt wichtige Merkmale der Haupterhebung zusammen.

Tabelle 2: Grunddaten zur Lebensverlaufsstudie 1964/71 (ZA-Studiennummer: 3927)

Grundgesamtheit	Deutschsprachige Wohnbevölkerung der Geburtsjahrgänge 1964 und 1971 in Privathaushalten in den alten Bundesländern (inkl. West-Berlin)
Stichprobe	Einwohnermeldestichprobe (n = 6519) in 97 ausgewählten Gemeinden, gezogen Juni 1997 - Dezember 1997, Schichtung nach Bundesland, Regionstyp und Geburtskohorte
Erhebungszeitraum	Mai 1998 - Januar 1999
Erhebungsmethode	Mündliche, computerunterstützte Befragung (CATI, bei 249 von 2911 auswertbaren Interviews CAPI)
Durchschnittliche Interviewdauer (netto)	69 Minuten
Verfügbare Interviews (nach Abschluss der Edition)	N = 2909
Ausschöpfungsquote	Bezogen auf Bruttostichprobe (n = 6519): Nach Erhebung 45%, nach Prüfung und Edition 45% Bezogen auf bereinigte Stichprobe* (n = 4404): Nach Erhebung 67%, nach Prüfung und Edition 66%
* Abgezogen sind hier u.a. die Ausfallgründe 'falsche Zielperson' (nicht Deutsch sprechende Ausländer) und '(neue) Adresse nicht ermittelbar' (vgl. Infas 1999: 36); verbleibende Ausfallgründe sind insbesondere Nichterreichbarkeit und Verweigerungen.	

1.3 Charakterisierung der Stichprobe

Mit einer Ausschöpfungsquote von ca. 45% (brutto) bzw. 66% (netto) liegt die Studie über den Quoten vergleichbarer Befragungen. Durch die spätere Nachbefragung konnten darüber hinaus die Lücken in einzelnen Fragen überwiegend gefüllt werden.

Um herauszufinden, wie repräsentativ die Stichprobe ist, wurden in einem ersten Schritt ausgewählte, auf Basis der Lebensverlaufsstudie gewonnene, jeweils altersspezifische Verteilungen mit verschiedenen Jahrgängen des Mikrozensus verglichen (verwendet wurden die Scientific Use Files; vgl. Hillmert/Kröhnert 2001). Angesichts der dort vorhandenen Fallzahlen (der Mikrozensus ist eine 1%-Stichprobe der Haushalte in der Bundesrepublik Deutschland) und der Pflicht zur Teilnahme an dieser Befragung steht eine umfassende Vergleichsgrundlage zur Verfügung. Bei der Gegenüberstellung der einfachen Merkmalsverteilungen (Nutzerdatensätze des Mikrozensus) finden sich einige – im übrigen für Surveys typische – Abweichungen zwischen Mikrozensus und Lebensverlaufsstudie. So sind in den Lebensverlaufsdaten Personen ohne oder mit nur geringer Schul- und Berufsausbildung offenbar unterrepräsentiert, Personen mit mittleren Bildungsabschlüssen hingegen etwas überrepräsentiert. Auch sind die jeweils errechneten Arbeitslosenquoten in den Lebensverlaufsdaten durchweg geringer. Allerdings können auch große Datenbasen wie der Mikrozensus nicht in jedem Fall als objektive Referenz gelten. So sind die resultierenden Merkmalsverteilungen jeweils nicht nur von der Art der Stichprobe, sondern auch vom Design und der Qualität der Erhebung abhängig. Hinsichtlich der beiden letzteren Gesichtspunkte ist die Lebensverlaufsstudie mit ihrem differenzierten Fragenprogramm und der einzelfallorientierten Datenprüfung klar im Vorteil.

In den Fällen finden sich insgesamt 158 Personen (=5,4%) mit einer ausländischen Staatsangehörigkeit (1964: 3,9%; 1971: 7,0%)². Dies sind deutlich weniger Ausländer, als eigentlich zu erwarten wäre. Eigene Berechnungen mit dem Mikrozensus 1998 (alte Bundesländer einschließlich Gesamtberlin; hochgerechnet) weisen für die Kohorten 1964 und 1971 Anteile von 12,3% bzw. 18,3% aus (vgl. Seibert 2003: 12). Allerdings sind auch die Zielpopulationen unterschiedlich: die Lebensverlaufsstudie hat von vornherein nur Deutsch sprechende Ausländer eingeschlossen, mit denen die relativ komplexen Telefoninterviews geführt werden konnten.

1.4 Datenedition

Im Anschluss an die Erhebung wurden die Daten einem aufwändigen Editionsprozess unterzogen. Hierbei wurden alle Fälle einzeln im Hinblick auf fehlende oder unplausible Angaben geprüft³. Bereits im Juli 1998 wurde mit einer Voredition der bereits verfügbaren Daten begonnen. Die eigentliche Edition der Daten wurde am Max-Planck-Institut für Bildungsforschung Anfang Januar 1999 begonnen. Sie dauerte einschließlich Nachrecherche, Eingabe der Korrekturdaten und Datenprüfung bis Ende 2002. An allen Editionsstufen waren insgesamt 21 Mitarbeiterinnen und Mitarbeiter beteiligt. Ferner wurden diverse offen erhobene Daten (Texte) codiert. Im Januar 2000 wurde mit der Codierung der Berufsangaben begonnen, weitere Vercodungen folgten in unregelmäßigen Abständen.

Für die Edition wurde ein differenziertes Regelwerk⁴ konstruiert, das sukzessive angepasst wurde. Auftretende Inkonsistenzen und neue Probleme wurden in der Arbeitsgruppe diskutiert und entschieden, und die einzelnen Bearbeiter erhielten Supervision. Die Datenedition hat sich hier in mehreren Phasen über einen Zeitraum von etwa dreieinhalb Jahren erstreckt, was den zeitlichen Abstand zwischen Erhebung und Verfügbarkeit der Daten erklärt. Eine wichtige Informationsmöglichkeit war ferner die telefonische Nachrecherche bei ca. 1.000 Zielpersonen. Diese wurden noch einmal um klärende oder ergänzende Informationen gebeten.

Durch die systematische Edition sind qualitative Verbesserungen der Datenbasis auch und gerade in einer Phase möglich, in der die eigentliche Datenerhebung bereits abgeschlossen ist. Dies trägt dazu bei, den Gesamtfehler im Forschungsprozess zu minimieren. Gleichzeitig wird dabei von der linearen Struktur des Forschungsprozesses, die durch das idealtypische Modell nahegelegt wird, zugunsten einer Wiederholung beziehungsweise Korrektur einzelner Projektschritte und -entscheidungen abgewichen.

1.5 Datenstruktur und bisherige Arbeiten

Ähnlich wie bei den vorangegangenen Teilstudien wurden die Daten in nach Lebensbereichen differenzierten Modulen erhoben (vgl. die Übersicht in Abschnitt 2.1, Seite 10f.). Die Informationen wurden jeweils in *quer-* und *längsschnitorientierten* Teildatensätzen abgelegt, die jeweils durch die Extension -q bzw. -sp gekennzeichnet sind.

Während erstere zeitlich unveränderliche Merkmale oder zusammenfassende Informationen über individuelle Verläufe enthalten, sind letztere nach einzelnen Abschnitten ('Spells' oder 'Episoden') gegliedert.

² Weitere 114 Personen mit der deutschen Staatsangehörigkeit wurden im Ausland geboren.

³ Zum Prozess der Edition in dieser Teilstudie und Konsequenzen für inhaltliche Analysen vgl. ausführlich Hillmert 2002a.

⁴ Vgl. Editionshandbuch, Teil IV dieser Dokumentation.

Für jede dieser Episoden wurden jeweils gleichartige, umfassende Informationen erhoben. Dadurch ist es sowohl möglich, Lebensverläufe in ihrer zeitlichen und inhaltlichen Komplexität genau zu beschreiben. Da beispielsweise für jeden Ausbildungsabschnitt Dauer, Fachrichtung und Ausbildungseinrichtung erfasst wurden, lassen sich vollständige Ausbildungsverläufe hinsichtlich ihrer fachlichen oder institutionellen Kontinuität beschreiben (vgl. Hillmert/Jacob 2003a). Darüber hinaus können empirische Bezüge zwischen Entwicklungen in verschiedenen Lebensbereichen hergestellt werden.

Die bisherigen Analysen haben sich konzentriert auf Fragen von Bildung und Ausbildung (Hillmert/Jacob 2003b, Jacob 2003, Seibert 2003) sowie Muster des Arbeitsmarkteinstiegs und frühere Erwerbskarrieren (Bender/Dietrich 2001, Hillmert 2002b), insbesondere auch unter dem Aspekt unterwertiger Beschäftigung (Büchel/Pollmann-Schult 2002, Pollmann-Schult/Büchel 2002). Als Vorteil der Daten erweist sich dabei die Möglichkeit zu innerdeutschen West-Ost-Vergleichen (Trappe/Goedicke 2003) und längerfristigen Kohortenvergleichen (Mayer/Hillmert 2003).

Einen Überblick über zentrale Projektergebnisse gibt das Buch "Geboren 1964 und 1971: Neuere Untersuchungen zu Ausbildungs- und Erwerbschancen in Westdeutschland" von Hillmert/Mayer (im Erscheinen).

Die Daten stehen ab sofort interessierten Wissenschaftler(inne)n für Analysezwecke zur Verfügung (ZA-Studiennummer 3927). Im Zuge der notwendigen Anonymisierung wurden hierfür einige Veränderungen vorgenommen. So wurden etwa offene Textangaben (Ausbildungs- und Arbeitsstätten, Wohnorte etc.) entfernt, die Fallnummern geändert und bei verschiedenen Variablen Kategorien zusammengefasst.

Um auch externen Wissenschaftler(inne)n die Nutzung der Daten zu erleichtern, ist eine vollständige Dokumentation vorhanden, die zusammen mit den Daten an die Nutzer/-innen abgegeben wird. Sie enthält ausführliche Informationen zur Datenstruktur und zu den einzelnen Variablen (einschließlich generierter Variablen) sowie den detaillierten Fragebogen. Die in die Dokumentation integrierten Methodenberichte liefern Informationen zur Stichprobe, das Editionshandbuch enthält die Regeln und Entscheidungen der Datenedition. Darüber hinaus wird der Prozess von Erhebung, Prüfung und Edition der Daten dargestellt.

1.6 Fazit

Zusammenfassend lässt sich festhalten, dass die vorliegende Studie ein spezifisches Informationspotenzial hat. Es gibt eine Reihe von immanenten Beschränkungen des gewählten Designs der Studie, wie etwa die Konzentration auf zwei große Kohorten, um für die Analysen hinreichend große Fallzahlen in altershomogenen Gruppen zu bekommen. Zu den Besonderheiten bzw. Stärken dieser Studie zählt die besondere Datenqualität, die durch aufwändige Vorbereitung, Erhebung und Edition der Daten gesichert wurde. Hinzu kommt die Detailliertheit der Angaben, die für eine Vielzahl von Variablen Längsschnittanalysen ermöglicht. Die zentralen verlaufsrelevanten Merkmale wurden jeweils gesondert für jede einzelne (Teil-)Episode erhoben. Darüber hinaus ermöglicht es der modulweise Aufbau der Erhebung, Querbezüge zwischen parallelen Aktivitäten bzw. unterschiedlichen Lebensbereichen herzustellen. Schließlich besteht noch die Möglichkeit, die Daten auf Individualebene mit prozessproduzierten Daten zu verknüpfen, um sie einerseits zu validieren⁵ und andererseits mit zusätzlichen Informationen anzureichern. Diese Möglichkeit steht externen Nutzer(inne)n allerdings nicht zur Verfügung.

⁵ Im Rahmen eines Dissertationsprojekts wurde versucht, diesen Vergleich auch für Aussagen über mögliche Gedächtniseffekte bei Retrospektivbefragungen zu nutzen (siehe auch Reimer 2003).

Die Daten des Projekts *Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland* sind die derzeit aktuellsten, differenziert und repräsentativ erhobenen Informationen über Lebensverläufe in (West-)Deutschland.

1.7 Literatur

- Bender, Stefan/Dietrich, Hans 2001: Unterschiedliche Startbedingungen haben langfristige Folgen. Der Einmündungsverlauf der Geburtskohorten 1964 und 1971 in Ausbildung und Beschäftigung. IAB-Werkstattbericht 11/2001. Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung.
- Brückner, Erika/Mayer, Karl Ulrich 1998: Collecting life history data: Experiences from the German life history study. In: Janet Z. Giele/Glen H. Elder Jr. (Hg.): *Methods of life course research: Qualitative and quantitative approaches*. Thousand Oaks: Sage, 152-181.
- Büchel, Felix/Pollmann-Schult, Matthias 2002: Overcoming a period of overeducated work – does the quality of apprenticeship matter? *Applied Economics Quarterly – Konjunkturpolitik* 48: 304-316.
- Corsten, Michael/Hillmert, Steffen 2001: Qualifikation, Berufseinstieg und Arbeitsmarktverhalten unter Bedingungen erhöhter Konkurrenz. Was prägt Bildungs- und Erwerbsverläufe in den achtziger und neunziger Jahren? Arbeitspapier Nr. 1 des Projektes *Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland*. Berlin: Max-Planck-Institut für Bildungsforschung (<http://www.mpib-berlin.mpg.de/dok/full/Corsten/corste11.htm>).
- Hillmert, Steffen 2001: Kohortendynamik und Konkurrenz an den zwei Schwellen des dualen Ausbildungssystems. Arbeitspapier Nr. 2 des Projektes *Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland*. Berlin: Max-Planck-Institut für Bildungsforschung (<http://www.mpib-berlin.mpg.de/dok/full/Hillmert/kohorten/index.htm>).
- Hillmert, Steffen 2002a: Edition von Lebensverlaufsdaten: zur Relevanz einer systematischen Einzelfallbearbeitung bei standardisierten Befragungen. *ZUMA-Nachrichten* 26/H. 51, 120-140.
- Hillmert, Steffen 2002b: Stabilität und Wandel des 'deutschen Modells': Lebensverläufe im Übergang zwischen Schule und Beruf. In: Wingens, Matthias/Sackmann, Reinhold (Hg.): *Bildung und Beruf. Ausbildung und berufsstruktureller Wandel in der Wissensgesellschaft*. Weinheim: Juventa, 65-81.
- Hillmert, Steffen/Jacob, Marita 2003a: Bildungsprozesse zwischen Diskontinuität und Karriere: das Phänomen der Mehrfachausbildungen. *Zeitschrift für Soziologie* 32, 4/2003, 325-345.
- Hillmert, Steffen/Jacob, Marita 2003b: Social inequality in higher education: is vocational training a pathway leading to or away from university? *European Sociological Review* 19, 3/2003, 319-334.
- Hillmert, Steffen/Kröhnert, Steffen 2001: Vergleich der Lebensverlaufsstudie 64/71-West mit Erhebungen des Mikrozensus auf Basis ausgewählter Randverteilungen. Technischer Arbeitsbericht. Berlin: Max-Planck-Institut für Bildungsforschung (<http://www.mpib-berlin.mpg.de/dok/full/Hillmert/vergleic/index.htm>).
- Hillmert, Steffen/Mayer, Karl Ulrich, im Erscheinen: *Geboren 1964 und 1971: Neuere Untersuchungen zu Ausbildungs- und Erwerbchancen in Westdeutschland*. Wiesbaden: Westdeutscher Verlag.
- Infas 1997: *Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland*. Methodentest (Manuskript). Bonn: Institut für angewandte Sozialwissenschaft (auch abgedruckt in Teil IX dieser Dokumentation).
- Infas 1999: *Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland*. Methodenbericht zur Hauptstudie. Bonn: Institut für angewandte Sozialwissenschaft (auch abgedruckt in Teil IX dieser Dokumentation).
- Jacob, Marita 2003: *Ausmaß, Strukturen und Ursachen von Mehrfachausbildungen. Eine Analyse von Ausbildungsverläufen in den achtziger und neunziger Jahren in Westdeutschland*. Dissertation, Freie Universität Berlin.
- Mayer, Karl Ulrich/Hillmert, Steffen 2003: New ways of life or old rigidities? Recent changes in social structures and life courses in Germany and their political impacts. In: Kitschelt, Herbert/Streeck, Wolfgang (Hg.): *Germany beyond the stable state. Special issue of West European Politics*.
- Pollmann-Schult, Matthias/Büchel, Felix 2002: Ausbildungsinadäquate Erwerbstätigkeit: eine berufliche Sackgasse? Eine Analyse für jüngere Nicht-Akademiker in Westdeutschland. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung* 35: 371-384.
- Reimer, Maike 2001: *Autobiographisches Gedächtnis und retrospektive Datenerhebung: Die Rekonstruktion und Validität von Lebensverläufen*. Dissertation, Freie Universität Berlin.
- Seibert, Holger 2003: "Wer zu spät kommt..." Schulausbildung und der Erwerbseinstieg von Ausbildungsabsolventen ausländischer Herkunft in Deutschland. Arbeitspapier Nr. 6 des Projektes *Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland*. Berlin: Max-Planck-Institut für Bildungsforschung (<http://www.mpib-berlin.mpg.de/dok/full/seibert/werzuspa/index.htm>).

Trappe, Heike/Goedicke, Anne 2003: Ein Land – zwei Erwerbssysteme? Die Folgen geschlechtstypischer Berufsentscheidungen in Ost- und Westdeutschland. In: Allmendinger, Jutta (Hg.): Entstaatlichung und soziale Sicherheit: Verhandlungen des 31. Kongresses der Deutschen Gesellschaft für Soziologie in Leipzig 2002. Opladen: Leske + Budrich.

Wagner, Michael 1996: Lebensverläufe und gesellschaftlicher Wandel: Die westdeutschen Teilstudien. ZA-Information 38, 20-27.

2 Datenbeschreibung

2.1 Datenstruktur

Wie in den vorangegangenen Lebensverlaufsstudien, erfolgte die Befragung auch in der jüngsten Teilstudie modulweise, d.h. es wurden nacheinander die Fragen zu bestimmten Lebensbereichen gestellt. Jedes Modul war in sich abgeschlossen. Zwischen den Modulen konnte während des Interviews nicht gewechselt werden. Aus diesem Grunde waren zum einen die Lückenerfassung und zum anderen besondere Steuerungsvariablen notwendig.

Die Daten sind zeilenweise in Records abgelegt, die inhaltlich in Bezug zu den Modulen stehen (vgl. Abschnitt 2.3). Einige Module bzw. Records enthalten ausschließlich *Querschnittsdaten*, einige ausschließlich sogenannte *Episoden- oder Spelldaten*, die meisten jedoch beides.

Zu den *Querschnittsdaten* gehören unveränderliche Daten (wie Geburtsort und -datum der Zielperson, die leiblichen Eltern), Daten, die nur für einen bestimmten Zeitpunkt erhoben wurden (wie beispielsweise das Datum der ersten Haushaltsgründung, der Berufswunsch am Ende der Schulzeit), und Daten, die ausschließlich zur Fragensteuerung erfasst wurden. Für diese Art von Daten existiert in der entsprechenden Datei immer genau eine Zeile pro Person.

Beispiel: Querschnittsdaten eines Falls aus der Wohnortgeschichte (Datei REC4Q)

CASEID	Anzahl Spells in der Datei REC4SP	Gründung des ersten eigenen Haushalts		erster Auszug aus dem Elternhaus zum selben Zeitpunkt?	Falls WG13 = 2 (nein): Erster Auszug aus dem Elternhaus		RECID	HH1DAT	AUS1DAT
	NWG	Monat	Jahr	WG13	Monat	Jahr			
xxxxxx	7	5	86	2	5	82			

Zu den *Spelldaten* zählen wiederkehrende Daten (z.B. Angaben zu den Wohnorten, den besuchten allgemeinbildenden Schulen, den beruflichen Ausbildungen, ausgeübten Erwerbstätigkeiten, eventuellen Phasen der Arbeitslosigkeit, ausgeübten Nebentätigkeiten, besuchten Lehrgängen/Kursen zur beruflichen Weiterbildung, Phasen anderer Aktivitäten oder nicht berichteter Zeiträume) und Angaben zu Eltern/Stiefeltern, Geschwistern, Kindern und Partnern, zu denen eine Zielperson mehrfach in Beziehung stehen kann. Ein Spelldatenfile enthält gegebenenfalls mehrere Datenzeilen pro Person.

Beispiel: Verlaufsdaten eines Falls aus der Wohnortgeschichte (Datei REC4SP)

CASEID	SORT-ID	WG1B	WG2	Wohnort von ... bis				MAR-KE12	Wechselgrund	WG10	RECID	START	END	...
				WG3MA	WG3JA	WG3ME	WG3JE							
yyyyyy	1	Ort A	Bundesland A	5	64	48	83		Studium		773	1004		
yyyyyy	2	Ort B	Bundesland B	9	83	9	85		Studienortwechsel	4	1005	1029		
yyyyyy	3	Ort C	Bundesland A	50	85	6	86		Studienwechsel	4,2	1030	1038		
yyyyyy	4	Ort D	Bundesland A	47	86	6	87		Studienortwechsel	4,2	1039	1050		
yyyyyy	5	Ort E	Bundesland A	47	87	8	91		berufliche Versetzung	4,2	1051	1100		
yyyyyy	6	Ort F	Bundesland A	9	91	8	96		Arbeitsstellenwechsel	4,2	1101	1160		
yyyyyy	7	Ort A	Bundesland A	49	96	11	98			4,2	1161	1187		

Übersicht über Fragenbereiche und Module

Querschnittsdaten	Spelldaten	Querschnittsdaten	Spelldaten	Querschnittsdaten	Spelldaten	Querschnittsdaten	Spelldaten
-------------------	------------	-------------------	------------	-------------------	------------	-------------------	------------

SCHUL- UND BERUFS-AUSBILDUNG	Modul	SCHUL-AUSBILDUNG		BERUFS-AUSBILDUNG		AUS- UND WEITERBILDUNG	
	Record-nummer	5		6		11	
	Beispiele	erste Schule im Ausland, nachgeholte Schulabschlüsse, Berufsvorbereitungsjahr, Berufswunsch und Steuerungsvariablen	Schulbesuchsdauer, Schultyp, -abschluss, Abgangsklasse, Abschlussnote	Anzahl Ausbildungen, Adäquatheit der Ausbildung und Steuerungsvariablen	Ausbildungsberuf, Dauer, Ausbildungsstätte, -abschluss, Finanzierung, Gründe für Abbruch/ Unterbrechung, zusätzliche Abschlüsse	Anzahl Weiterbildungen, Editions-kommentar	besuchte Lehrgänge/ Kurse, Dauer, Kostenträger, Zeitaufwand, Zertifikat
	Datei-namen	rec5q	rec5sp	rec6q	rec6sp	rec11q	rec11sp

ERWERBS-TÄTIGKEITEN UND PHASEN DER NICHT-ERWERBS-TÄTIGKEIT	Modul	ERWERBSTÄTIGKEITEN		ARBEITSLOSIGKEIT		NEBENTÄTIGKEITEN		LÜCKEN	
	Record-nummer	7/8		9		10			
	Beispiele	Anzahl Erwerbstätigkeiten, Wunsch nach Erwerbstätigkeit / Wehr-/Zivildienst und -dauer	Beruf/Tätigkeit, Dauer, berufliche Stellung, Stellenfindung, Rentenversicherungsbeiträge, Betrieb, Arbeitsvertrag, Verdienst, Arbeitszeit, Angaben zur Tätigkeitsveränderung	Anzahl Arbeitslosigkeiten	Dauer, Leistungen und Nutzung des Arbeitsamts, Bezug von Arbeitslosengeld, Dauer des Leistungsbezugs, Gründe für Zahlungsunterbrechung	Anzahl Nebentätigkeiten und Editions-kommentar	Neben- oder geringfügige Tätigkeit, Dauer, Zeitaufwand, Arbeitsvertrag, Verdienst	Anzahl Lücken	Lückenaktivität (mit keinem anderen Modul zu erfassen), Dauer
	Datei-namen	rec7q / rec8	rec7sp	rec9q	rec9sp	rec10q	rec10sp	lueq	luesp

ALLGEMEINE ANGABEN ZUR ZIELPERSON	Modul	ZIELPERSON		BERUFSKONTROLLE, POLITIK, RELIGION		INTERVIEWENDE	
	Recordnummer	1		15		19	
	Beispiele	Geschlecht, Geburtsdatum, Staatsangehörigkeit etc.		Kontrollüberzeugungen, Mitgliedschaft in Parteien/Verbänden, Religionszugehörigkeit		Rentenversicherung, Einverständnis zur Zuspiegelung der Sozialversicherungsdaten, Panelbereitschaft	
	Dateinamen	rec1		rec15		rec19	

FAMILIE	Modul	ELTERN		GESCHWISTER		PARTNERSCHAFTEN		KINDER	
	Recordnummer	2		3		16		17	
	Beispiele	leibliche Eltern: Geburtsjahr, Herkunft, Schul- und Ausbildungsabschluss, Erwerbstätigkeit	Stief-/Pflegermütter / Stief-/Pflegerväter (wie Eltern + Dauer des Zusammenlebens)	Anzahl Geschwister	Vornamen der Geschwister, Geschlecht, Geburtsjahr, Schul- und Ausbildungsabschluss	Familienstand, Anzahl Partnerschaften	Dauer der Partnerschaft, Zusammenzugs-/Heiratsdatum, Trennung, Geburtsjahr, Herkunft, Schul- und Ausbildungsabschluss, Erwerbstätigkeit	Anzahl Kinder	Vornamen der Kinder, Geschlecht, Geburtsdatum, Kindschaftsverhältnis, Sterbedatum, Schulbesuch, Betreuungseinrichtungen und -personen
	Dateinamen	rec2	rec2sm / rec2sv	rec3q	rec3sp	rec16q	rec16sp	rec17q	rec17sp

WOHNEN	Modul	WOHNORTGESCHICHTE		EINKOMMEN UND HAUSHALT	
	Recordnummer	4		18	
	Beispiele	Anzahl Wohnorte, Haushaltsgründung, Auszug aus dem Elternhaus	Wohnorte, Dauer	Angaben zur aktuellen Wohnung, zur Haushaltszusammensetzung und zu den Einkommen des Haushalts, Lebenszufriedenheit	
	Dateinamen	rec4q	rec4sp	rec18	

2.2 Erläuterungen zu den Modulen

2.2.1 Angaben zur Zielperson (Record 1)

In diesem Record gibt es für jeden Fall genau einen Datensatz, der allgemeine Angaben zur Zielperson wie Geburtsdatum, Geburtsort, erstes Zuzugsdatum nach Westdeutschland, Staatsangehörigkeit(en) und Angaben zum Interview wie Interviewnummer, Interviewdatum, Interviewunterbrechung (Record und Datum) und Interviewdauer enthält.

2.2.2 Eltern (Record 2)

Zu diesem Modul finden sich in der Datei REC2 für die leiblichen Eltern Angaben wie Geburts- und Sterbejahr, höchster allgemeinbildender Schulabschluss, beruflicher Ausbildungsabschluss, Erwerbsstatus, Erwerbstätigkeit, berufliche Stellung, Ehe der Eltern, Herkunft der Eltern sowie Anzahl der Stief-/Pflegetöchter und Stief-/Pflegetäter.

Die Dateien REC2SM und REC2SV enthalten in Spellform jeweils die Angaben Dauer des Zusammenlebens, Geburts- und Sterbejahr, höchster allgemeinbildender Schulabschluss, beruflicher Ausbildungsabschluss, Erwerbsstatus, Erwerbstätigkeit, berufliche Stellung.

2.2.3 Geschwister (Record 3)

Hier ist in der Querschnittsdatei (REC3Q) die Anzahl der Geschwister abgelegt. In der Datei REC3SP finden sich in Spellform die Angaben Geschwistervorname, Geburtsdatum, Schulabschluss und höchster beruflicher Ausbildungsabschluss.

2.2.4 Wohngeschichte (Record 4)

Mit der "Wohngeschichte" wurden die verschiedenen Wohnorte der Zielperson seit Geburt erfasst. Die Datei REC4Q enthält die Querschnittsangaben Anzahl der Wohnorte, Zeitpunkt der ersten eigenen Haushaltsgründung und Zeitpunkt des ersten Auszugs aus dem Elternhaus.

Die Datei REC4SP enthält die Spelldaten für jeden Wohnort der Zielperson, an dem diese länger als 3 Monate lebte. Für jeden Spell sind hier die Daten Wohnort (mit zugehörigem Landkreis, Bundesland oder Land) und dessen Vercodung (Korrigierter Ortsname, Gemeindekennziffer, Postleitzahl, Status der Wohnortvercodung, neue Bezeichnung des eingemeindeten Ortes, Bundesland, Regierungsbezirk, Kreis und Gemeinde), Wohndauer und Gründe für Wohnortwechsel zu finden.

2.2.5 Schulische Ausbildung (Record 5)

Aufgrund der Einbeziehung Deutsch sprechender Ausländer in die Studie wurde über das Zuzugsdatum bzw. das Alter der Zielperson beim ersten Zuzug nach Westdeutschland entweder in die Fragen zu Schulen im Ausland oder in den Fragenteil der deutschen Schulen gesteuert. Zunächst wurden diejenigen, die nicht in Westdeutschland oder der DDR geboren und beim ersten Zuzug nach Westdeutschland nicht älter als 7 Jahre waren, nach dem Zeitraum des Schulbesuchs im Ausland gefragt, und eventuelle Schulbesuche in Westdeutschland wurden nur mit den Fragen zum nachgeholt Schulabschluss erfasst.

Da auf diese Weise jedoch keine Wechsel innerhalb westdeutscher Schultypen oder Lücken zwischen einzelnen Schulphasen sichtbar wurden, mussten diese Angaben in normale SchulsPELLS umgetragen werden.

Neben den Steuerungsvariablen für den Übergang zum nächsten Modul und der Anzahl der (in der Datei REC5SP vorhandenen) SPELLS enthält die Datei REC5Q die Daten zum Auslandschulbesuch, zu nachgeholt oder zuerkannten Schulabschlüssen, zum Berufsvorbereitungsjahr, zum Berufswunsch und, falls keine Berufsausbildung gemacht wurde, die Begründung hierfür. Zu beachten ist bei den Variablen zum nachgeholt Schulabschluss, dass hier nur erreichte Abschlüsse vorliegen. Die Zeit, die mit dem Anstreben eines nachzuholenden Schulabschlusses verbracht wurde, ist nur aus der Lückenaktivität "allgemeinbildende Schule" (Lückenmodul, Datei LUESP) ersichtlich, und auch hier nur, wenn es die einzige Aktivität der Zielperson in diesem Zeitraum war.

In der Datei REC5SP liegen die Daten zum Schulbesuch in Form einzelner SPELLS vor. Kriterium für einen SPELLwechsel war lediglich der Wechsel von einem Schultyp auf einen anderen. Ein Wechsel von Schulen begründete keinen neuen SPELL, und zwar auch dann nicht, wenn der Schulwechsel mit einem Ortswechsel verbunden war. Schulbesuche im Ausland können auch hier als SPELL vorliegen, da mit den oben beschriebenen "Auslandsschulvariablen" nur die Zeit erfasst wurde, die eine Zielperson im Ausland von der Einschulung bis zum ersten Zuzug nach Westdeutschland auf einer Schule (oder mehreren Schulen) verbrachte. Wurde die erste Schule in Westdeutschland und später eine Schule im Ausland besucht, so gibt es hierfür einen SPELL mit dem Schultyp "Sonstige Schule (im Ausland)".

2.2.6 Berufliche Ausbildung und Weiterbildung (Records 6 und 11)

Die Datei REC6Q enthält neben der Anzahl der SPELLS (in der Datei REC6SP) und den Steuerungsvariablen zum nächsten Modul die Variable zur Adäquatheit der Ausbildung.

Die Datei REC6SP enthält die Variablen der AusbildungssPELLS. Hier sind sowohl abgeschlossene als auch unterbrochene und abgebrochene Ausbildungen erfasst. SPELLwechsel gründen in: Ende einer Ausbildung, Unterbrechung einer Ausbildung, Ende einer Ausbildungsphase (wie z.B. Studium und Referendariat oder Fachschulausbildung und Anerkennungsjahr) oder Wechsel der Ausbildungsstätte (nicht im Falle von Referendariaten und Vikariaten; im Falle eines Studiums nur, wenn ein Wechsel zwischen In- und Ausland vorlag).

Die Datei REC11Q enthält die Anzahl der WeiterbildungssPELLS (NAWB) sowie die Korrekturvariable zur Anzahl der SPELLS (NAWBK) und die Kommentare der Edition (AWKOM) zu nicht erfassten Weiterbildungen (siehe dazu im Einzelnen den Abschnitt 3.10.10 des Editionshandbuchs in Teil IV dieser Dokumentation).

Zu den Verlaufsdaten in der Datei REC11SP gehören die Bezeichnung des Kurses/Lehrgangs, der Kostenträger, die Art der Weiterbildung, der Zeitpunkt des Beginns, die Dauer, der Zeitaufwand und das Jahr des Zertifikats/Abschlusszeugnisses, falls der Kurs mit Zertifikat/Abschlusszeugnis beendet wurde.

Die folgende Übersicht und die daran anschließenden Beispiele verdeutlichen grob die unterschiedlichen Arten von Daten der beiden Records. Näheres erläutert das Editionshandbuch (Teil IV dieser Dokumentation).

Modul AB	Modul AB - Zusätzlicher Abschluss	Modul AWB
<p>In der Variable AB15 findet sich der mit dieser Ausbildung erreichte Abschluss.</p> <p>Betriebliche oder überbetriebliche Ausbildung wie Lehre, Facharbeit, Weiterbildung</p> <p>Meisterlehrgang (Vollzeit) z.B. auf einer Fachschule</p> <p>Berufsfachschulausbildung inkl. Ausbildung an einer Schule des Gesundheitswesens (med. Hilfsberufe, Krankenpfleger, MTA etc.)</p> <p>Fachschulausbildung (z.B. Erzieherin, Techniker)</p> <p>Fachhochschulstudium</p> <p>Studium an einer Berufsakademie oder einer Verwaltungs- und Wirtschaftsakademie</p> <p>Hochschul-/Universitätsstudium</p> <p>Postgraduales Studium</p> <p>Praktikum, Volontariat, praktisches Jahr, Anerkennungsjahr, Referendariat, Vikariat</p> <p>Berufsgrundbildungsjahr</p> <p>Verwaltungsausbildung des öffentlichen Dienstes</p> <p>Vorbereitungsdienst für Beamte</p>	<p>In den Variablen ABZ27-ABZ29 stehen nur Abschlüsse, wenn sie neben einer anderen Aktivität erreicht wurden und wenn sie auf die Ausbildung (dieses Ausbildungsspell) aufbauen.</p> <p>Ausbildereignung</p> <p>Meisterbrief, Techniker (wenn der zugehörige Lehrgang berufsbegleitend absolviert wurde)</p> <p>Zusatzqualifikation (wie Fach..., staatl. gepr. ..., ...betriebswirt u.a.)</p> <p>Promotion, Habilitation</p> <p>Facharztqualifikation</p> <p>Lizenzwerb</p> <p>Laufbahnprüfung (Aufstieg für Beamte)</p>	<p>Dieses Modul enthält Angaben zu Lehrgängen/Kursen der beruflichen Fort- und Weiterbildung (Dauer, Kostenträger, Zeitaufwand, Zertifikat).</p> <p>Meisterlehrgang (berufsbegleitend)</p> <p>Softwareschulung</p> <p>Geschäftsendenglisch</p> <p>Mitarbeiterführung</p> <p>Verkaufsschulung</p> <p>Seminar für Betriebsräte</p> <p>Facharztweiterbildung</p>

Beispiele:

Modul AB	Modul AB - Zusätzlicher Abschluss	Modul AWB
AB-Spell 2: Bauzeichner	AB-Spell 2: staatl. gepr. Tiefbautechniker	
AB-Spell 1: Bankkaufmann	AB-Spell 1: ABZ28A Bankfachwirt ABZ28B Bankbetriebswirt	AWB-Spell 1: Einlagengeschäft AWB-Spell 2: Verkaufsseminar ... AWB-Spell 10: Wertpapiergeschäft
AB-Spell 1: Kinderpflegerin		AWB-Spell 1: Ausbilderschein
AB-Spell 1: Umweltschutztechniker	AB-Spell 1: Technischer Betriebswirt	AWB-Spell 1: Ausbildereignungsprüfung AWB-Spell 2: Technischer Betriebswirt
AB-Spell 1: Medizinstudium mit Abschluss "3. Staatsexamen"	AB-Spell 1: Promotion	AWB-Spell 1: Facharzt Augenarzt
AB-Spell 2: Arzt im Praktikum mit Abschluss "Staatl. Anerkennung eines vorangegangenen Abschlusses"	AB-Spell 2: Facharzt Augenarzt	

2.2.7 Erwerbstätigkeiten (Record 7/8)

Die Querschnittsdaten dieses Moduls finden sich in den Dateien REC7Q (Anzahl der Spells, Steuerungsvariablen und die Frage an Zielpersonen, die nicht bis zum Interviewzeitpunkt erwerbstätig waren, ob sie sich wünschen, wieder erwerbstätig zu werden) und REC8 (Daten zu Wehr-/Zivildienst).

In der Datei REC7SP sind die Daten der Erwerbstätigkeitsspell abgelegt. Als Erwerbstätigkeit waren alle bezahlten Beschäftigungen, die in Vollzeit oder Teilzeit (mindestens 15 Stunden pro Woche) ausgeübt wurden, zu erfassen. Spellwechsel gründen in: Berufs- oder Tätigkeitsveränderung, Wechsel der beruflichen Stellung, Wechsel des Betriebes, Wechsel von Voll- zu Teilzeit oder umgekehrt.

2.2.8 Arbeitslosigkeiten (Record 9)

Die Datei REC9Q enthält lediglich die Zahl der Arbeitslosigkeitsspell. Die Spells selbst finden sich in der Datei REC9SP. Sie enthalten Daten wie Dauer der Arbeitslosigkeit, Angaben zu Arbeitslosengeldbezug und anderen Leistungen des Arbeitsamts sowie Gründe für Zahlungsunterbrechung.

2.2.9 Nebentätigkeiten (Record 10)

Die Datei REC10Q enthält die Anzahl der NebentätigkeitssPELLS, die Variable NT1 (die Frage, ob es Nebentätigkeiten gab), die Korrekturvariable der SPELLzahl und den Kommentar zu nicht erfassten Nebentätigkeiten.

Die NebentätigkeitssPELLS sind in der Datei REC10SP enthalten. Zu jedem SPELL gehören die Art der Tätigkeit, die Dauer, die durchschnittliche wöchentliche Arbeitszeit, Fragen zum Arbeitsvertrag, zu Sozialversicherungsbeiträgen und zum Verdienst.

Zur Unterscheidung von Haupt- und Nebentätigkeit siehe die Abschnitte 3.7.13 (Seite 83), 3.9.1 und 3.9.2 (Seite 88f.) des Editionshandbuchs (Teil IV dieser Dokumentation).

2.2.10 Lückenmodul

Mit diesem Modul, das ursprünglich aus den Records 12 bis 14 und Teilinformationen aus den Records 5, 6, 7 und 8 bestand, wurden während des Interviews die Aktivitäten der Zielpersonen zwischen Modulen oder zwischen zwei Episoden innerhalb des gleichen Moduls erfasst. Das heißt, nach der Erfassung einer bestimmten Phase wurde, wenn sich die entsprechende nächste nicht unmittelbar anschloss, gefragt, was die Zielperson bis zum Beginn der nächsten getan hat. Wenn die befragte Person beispielsweise im Modul "Berufsausbildung, Record 6" angab, zwischen zwei Ausbildungsepisoden einen Lückenzeitraum von mehr als 3 Monaten zu haben, dann wurde sie nach ihrem Status zwischen diesen beiden Ausbildungen befragt und diese Informationen im Record 6 separat gespeichert. Es konnten verschiedene Lückenaktivitäten angegeben werden, für die zunächst nur die Dauer erfasst wurde. Die so angelegten LückensPELLS unterschieden sich in solche, die als "Nebenaktivität" galten und solche, die als "Hauptaktivität" mit einem der folgenden Module noch genauer zu erfassen waren (etwa eine zwischen zwei Ausbildungen verbrachte Erwerbsepisode). Im letzteren Fall wurden die LückensPELLS insgesamt überflüssig, da sie einerseits als Lückenaktivität und andererseits noch einmal als Hauptaktivität erfasst waren. Sie wurden durch die automatische oder die manuelle Edition gelöscht. Durch dieses Verfahren wurden nicht nur größere Datenmengen gelöscht, sondern auch inhaltsleer gewordene Variablen bzw. einzelne Ausprägungen von Variablen.

Von den ursprünglichen Lückenaktivitäten blieben übrig: "Erziehungsurlaub/Mutterschutz", "krank/in Rehabilitation/Kur", "in allgemeinbildender Schule" und "etwas anderes gemacht". Nachträglich hinzugefügt wurden die Lückenaktivitäten: "nicht berichteter Zeitraum", "Phase mit Weiterbildungsaktivität unbestimmter Dauer", "Deutschkurs", "Arbeitsamtsmaßnahme", "Sozialamtsmaßnahme" und "Editionsmissing".

Die Daten des Lückenmoduls finden sich in den Dateien LUEQ (Anzahl der LückensPELLS) und LUESP (Lückentyp und Dauer).

2.2.11 Berufskontrolle, Politik, Religion (Record 15)

Dieses Modul enthält ausschließlich Querschnittsdaten zu Berufskontrolle, Politik (Mitgliedschaft in Organisationen/Partei) und Religionszugehörigkeit, die in der Datei REC15 abgelegt sind.

2.2.12 Partnerschaften (Record 16)

Die Querschnittsdaten zu Partnerschaften (Anzahl und Familienstand) enthält die Datei REC16Q.

Die Datei REC16SP enthält zu jeder erhobenen Partnerschaft (das sind solche, die mindestens ein Jahr gedauert haben) die Angaben zur Dauer, zum Zusammenleben, zur Heirat, zur Trennung/Beendigung der Partnerschaft, zu Herkunft, Geburtsjahr, höchstem Schulabschluss und höchstem Ausbildungsabschluss der Partnerin/des Partners sowie zur Erwerbstätigkeit der Partnerin/des Partners zu Beginn und am Ende der Partnerschaft.

2.2.13 Kinder (Record 17)

In der Datei REC17Q sind die Angaben zur Anzahl der Kinder enthalten. Erfasst wurden leibliche, Stief-, Pflege- Adoptivkinder und Kinder der Partnerin/des Partners, die im Haushalt der Zielperson aufgewachsen sind.

Die Datei REC17SP enthält zu jedem Kind: Vorname, Geschlecht, Geburts- und eventuell Sterbedatum, Kindschaftsverhältnis, Schulbesuch, Betreuung durch Institutionen und Personen sowie angestrebter Schulabschluss.

2.2.14 Wohnung, Haushaltszusammensetzung und -einkommen (Record 18)

Dieses Modul enthält wieder ausschließlich Querschnittsdaten (Datei REC18), und zwar: Angaben zur aktuellen Wohnung (Art und Kosten), zur Haushaltszusammensetzung, zu den Einkünften der einzelnen Haushaltsmitglieder und des gesamten Haushalts sowie die Fragen zur Lebenszufriedenheit.

2.2.15 Interviewende (Record 19)

Die Datei REC19 enthält die Daten (nur Querschnitt) des letzten Moduls wie Rentenversicherungsbeiträge, Zustimmung zum Zusammenspielen der Daten mit den Sozialversicherungsdaten, Panelbereitschaft und die Einschätzungen des Interviews durch Zielperson und Interviewer/-in.

2.3 Erläuterungen zur Datenstruktur

Aus technischen Gründen bestand die Notwendigkeit, während eines Interviews mehrfach ganze Datenblocks aus dem Arbeitsspeicher zu entfernen, um dort Platz für neue Informationen zu schaffen. Die entfernten Datenblocks wurden auf der Festplatte des Rechners in einer so genannten Rohdatendatei im ASCII-Format zeilenweise abgelegt. Diese Datenzeilen, die gleichzeitig *Quer-* und *Längsschnittinformationen* enthalten konnten, werden mit dem Begriff "Record" bezeichnet und sind in der Abfolge ihrer Abspeicherung von 1-19 nummeriert. Die Records wurden so konzipiert, dass sie inhaltlich zusammengehörige Informationen zusammenfassen und diese in gleicher Abfolge strukturieren, wie die Frageabfolge im Erhebungsinstrument. Daher kann für jeden Record dessen Bezug zu einem inhaltlich abgegrenzten Modul des Erhebungsinstruments angegeben werden (z.B. Record 2 = Eltern, Record 5 = Schulausbildung u.s.w.). Bei der späteren Konstruktion von auswertbaren Datenfiles (SPSS-Datenfiles) aus den Rohdaten wurden einige strukturelle Veränderungen vorgenommen:

- Es wurden *Querschnitts-* und *Episodendaten* in voneinander getrennten Datenfiles gespeichert, um auch datentechnisch dem speziellen Charakter von *Spelldaten* Rechnung zu tragen. Dabei wurde im Falle der *Episodendaten* eine zusätzliche Episodenzählervariable SORTID angelegt, sodass jede Episode durch die Fallnummervariable CASEID und den Episodenzähler SORTID eindeutig gegenüber jeder anderen Episode gleichen Typs definiert ist.
- Es wurden sogenannte Lückenepisoden, die immer in gleicher Weise Lückenzeiträume zwischen Hauptaktivitäten beschreiben und deren Erfassung auf unterschiedliche Records verteilt war, in einem neuen Record 20 zusammengefasst, um das "Handling" von Lückenepisoden zu vereinfachen und Redundanzen zu beseitigen. Hierdurch wurden auch einige Records (Record 12-14) nicht mehr benötigt (siehe auch Teil IV dieser Dokumentation: Editionshandbuch, Seite 107) und entfielen.
- Es wurden über unterschiedliche Records verteilte Informationen zu Erwerbstätigkeiten im Record 7 und zu Arbeitslosigkeitsepisoden in Record 9 gebündelt (siehe auch Teil IV dieser Dokumentation: Editionshandbuch, Seite 107).
- Es wurden einige wenige Variablen aus verwaltungspraktischen Gründen zwischen Records umverlagert.
- Die Records wurden um Korrektur- und Nachvercodungsvariablen erweitert.

Die resultierenden SPSS-Datenfiles sind nach ihrer Record-Herkunft mit Labels versehen (z.B. REC1.SAV). Waren sowohl *Querschnitts-* als auch *Episodendaten* im Ursprungsrecord vorhanden, dann existieren nun zwei SPSS-Datenfiles, deren Namen entweder mit "q" (*Querschnittsdesign*) oder "sp" (*Spelldesign*) enden (z.B. REC5Q.SAV und REC5SP.SAV). Die oben beschriebenen Daten der umstrukturierten Lückenepisoden machen eine Ausnahme von dieser Benennungsregel (LUEQ.SAV und LUESP.SAV). Eine weitere Ausnahme bilden die aus den Elterndaten (Record 2) ausgekoppelten und im *Spelldesign* gespeicherten Informationen zu Stiefvätern (REC2SV.SAV) und Stiefmüttern (REC2SM.SAV). Zusätzlich zu den Bezügen der Dateinamen der SPSS-Files auf den jeweiligen Ursprungsrecord existiert in jedem Datenfile die Variable RECID, die noch einmal explizit den Record-Bezug kenntlich macht.

2.4 Fragebogen

Der bereinigte Fragebogen (Teil II dieser Dokumentation) wurde aus dem Quellfile des Erhebungsprogramms nachträglich erstellt (zu den weiteren Versionen des Fragebogens vgl. Abschnitt 3.2.2, Seite 21). Er enthält zu allen in den Datenfiles vorhandenen Variablen die gestellten Fragen bzw. vergebenen Variablen-Labels, die Intervieweranweisungen, die Antwortvorgaben sowie alle Ergänzungen und Änderungen. Auch ist die Fragensteuerung (Filter-

führung) nachvollziehbar. Durch die Bereinigung um programminterne, technische Routinen spiegelt er nur noch bedingt das ursprüngliche Erhebungsinstrument "ODIN" wider, ist jedoch unverzichtbares Hilfsmittel zum Verständnis der nutzbaren SPSS-Datenfiles.

2.5 Variablen

Die Variablen- und Valuelabel-Liste (Teil III dieser Dokumentation) gibt eine vollständige Übersicht der Variablen jedes Moduls, geordnet nach Records.

2.6 Edierte und codierte Daten

Wie die Daten durch die Edition bearbeitet wurden, ist dem Editionshandbuch zu entnehmen (Teil IV dieser Dokumentation). Ein speziell für diese Dokumentation erstellter Index erlaubt das rasche Auffinden einzelner Editionsregeln in dem umfangreichen Handbuch.

Ebenso finden sich im Editionshandbuch die Regeln zu Nachvercodungen, die von den Mitarbeiter(inne)n der Edition auszuführen waren.

Daneben gab es aber auch Vercodungen, die nicht im Rahmen der Edition, sondern in einem separaten Arbeitsschritt durchgeführt wurden. Eine Übersicht über die Vercodungen enthält Abschnitt 3.5. Die Beschreibungen im Einzelnen sind dem Teil VII dieser Dokumentation zu entnehmen.

2.7 Anonymisierung der Daten für den Scientific Use File

Folgende Regeln wurden für die Bearbeitung der Lebensverlaufdaten zur Weitergabe an das "Zentralarchiv für Empirische Sozialforschung" an der Universität Köln bzw. zur direkten Weitergabe an Nutzer/-innen im Sinne eines "faktisch anonymisierten Datensatzes" festgelegt:

1. Alle Klartextvariablen wurden aus den Daten entfernt.
2. Die originalen Fallnummern wurden durch zufallsgenerierte Fallnummern ersetzt.
3. Die Angaben zu Geburtstagen wurden aus den Daten entfernt. Geburtsmonate und -jahre verblieben.
4. Außer den Ortsgrößen- und den Bundeslandvariablen wurden alle ortsbeschreibenden Variablen, wie Postleitzahl und Gemeindekennziffer, entfernt.
5. Alle Berufsvercodungen (Vercodung nach Bundesanstalt für Arbeit 1988 und Statistisches Bundesamt 1992) wurden nur als Dreisteller in die Daten übernommen.
6. Die Nationalität der Eltern und der Partner wurde, soweit erfragt, nur durch klassifizierte Variablen einbezogen.

2.8 Statistik der Edition und der Nachrecherche

Durchgeführte Interviews		2935
- unvollständige Interviews		-23
- fehlende Datenlieferung		-1
- nachträglich ausgeschlossene Fälle ("falsche" ZP)		-2
Anzahl der Fälle im Analysedatensatz		2909
Vollständige Interviews		2911
- telefonische Interviews (CATI)		2662
- persönliche Interviews (CAPI)		249
Fälle mit Einverständnis zur Interviewaufzeichnung auf Tonband		2521
- Fälle, für die eine Tonbandaufzeichnung des Interviews vorliegt		2482
- Fälle, für die keine Tonbandaufzeichnung vorliegt		39
Voredierte Fälle		246
Erstedierte Fälle		2911
Zweitedierte Fälle		911
- Einzelfallentscheidungen nach Erst- bzw. Zweitedition		118
Nachträgliche Überprüfung der Einkommensvariablen bzgl. Währung (W-Code)		1245
Korrektureingabe nach Erst- bzw. Zweitedition		2239
Abschlussedition (EndED) mit Nachrecherche (NR)		2911
- Einzelfallentscheidungen nach Abschlussedition		266
> nur edierte Fälle		1639
> NR-Fälle		1272
- allgemeine NR-Fälle		1071
- "Nur-Noten"-Fälle (nur eine oder mehrere Noten eines Ausbildungsabschlusses waren nachzuerheben)		201
- erfolgreich/teilweise erfolgreich durchgeführte Nachrecherchen		1016
- erfolgreich durchgeführte "Nur-Noten"-NR		174
- schriftliche (teilw.) erfolgreiche NR		17
-- schriftliche erfolgreiche "Nur-Noten"-NR		12
- "falsche" ZP (erreichte Person war nicht ZP)		2
- verweigerte NR		141
- ZP verweigert selbst am Telefon		51
- andere Person verweigert am Telefon		2
- ZP verweigert durch Nichtbeantwortung einer Terminanfrage, einer Erinnerung oder einer schriftl.		85
- nicht durchgeführte NR		113
- Projektleitung erachtete NR als nicht sinnvoll		3
- ZP war nicht erreichbar bzw. Adressprüfung wurde nach Datum nicht mehr durchgeführt		48
- NR wurde nach Datum nicht mehr durchgeführt		62
Nachrecherche in	inkl. Nur-Noten-Fälle	ohne Nur-Noten-Fälle
1 Modul	440	266
2 Modulen	289	289
3 Modulen	165	165
4 Modulen	74	74
5 Modulen	40	40
6 Modulen	6	6
7 Modulen	2	2
mehr als 7 Modulen	0	0
Summe	1016	842

3 Chronologie des Projekts "Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland"

3.1 Design und Ablauf

Stichprobendesign, Methodenstudie und Pretests sind in den Berichten

- Ausbildungs- und Berufsverläufe der Geburtskohorten 1971 und 1964. Bericht zum Methodentest. infas Sozialforschung, Juni 1997, Nr. 2582.7, Bonn.

und

- Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland. Methodenbericht zur Hauptstudie. infas, September 1999, Nr. 2582, Bonn.

beschrieben. Beide Berichte sind in Teil IX dieser Dokumentation zu finden.

3.2 Datenerhebung

Die Details zu Vorbereitung (INT-Schulung, INT-Handbuch) und Durchführung der Interviews sind ebenfalls in

- Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland. Methodenbericht zur Hauptstudie. infas, September 1999, Nr. 2582, Bonn.

beschrieben.

3.2.1 Erhebungsinstrument

Für die Hauptstudie kamen eine Telefon- und eine persönliche Version des Programms "ODIN" zum Einsatz. Die Versionen unterschieden sich darin, dass bei der CATI-Version die Adressen und zugehörigen Case-IDs automatisch eingeblendet wurden, während bei der CAPI-Version die Case-ID von den Interviewern anhand einer Adress-/Case-ID-Liste eingegeben werden musste.

Die Quell-Files der verschiedenen Versionen des Erhebungsprogramms (auch alle Pretest- und anderen Vorläuferversionen) sind archiviert, werden jedoch wegen ihres Umfangs hier nicht dokumentiert.

3.2.2 Fragebogen

Der Fragebogen wurde aus dem Quellfile des Erhebungsinstruments erstellt, um die Lesbarkeit von Fragen, Filterprüfungen und Fragensteuerung zu erleichtern. Er liegt in den folgenden Fassungen vor:

- Originalversion (ohne Editionsergänzungen, mit allen Prüfungen, Steuerungen),
- Editionsversion (mit allen Prüfungen, Steuerungen, hervorgehobene Korrekturen fehlerhafter Steuerungen, hervorgehobene Ergänzungen und Änderungen durch die Edition, sowohl Antwortkategorien, Variablen als auch Filterführung betreffend),

- bereinigte Version (ohne programminterne Prüfungen, mit korrigierter Rechtschreibung, mit allen Ergänzungen und Änderungen, jedoch ohne Hervorhebung).

Den aktuellen Stand enthält die bereinigte Version (Teil II dieser Dokumentation), der auf die vorliegenden Daten abgestimmt ist.

3.3 Daten

Die Daten wurden von Infas während der Erhebungsphase in mehreren Teillieferungen in Form von Rohdaten im ASCII-Format geliefert. In einem ersten Bearbeitungsschritt wurden SPSS-Syntaxfiles entwickelt, mit deren Hilfe die Rohdaten, getrennt nach Records und *Längs-* und *Querschnittsdaten*, in SPSS-Datenfiles eingelesen und mit Variablen- und Value-Labels versehen wurden. Daher war es bereits zu einem sehr frühen Zeitpunkt der Datenerhebung möglich, auf Probleme, die das sehr komplexe Erhebungsinstrument "ODIN" verursachte, zu reagieren. Jede neue Datenteillieferung wurde sukzessive an bereits bestehende Datenfiles angefügt. Auswertungen der Datenfiles gaben bereits erste wertvolle Hinweise für die spätere Edition der Daten.

Die Belange der Datenedition machten es erforderlich, die SPSS-Daten in einem zweiten Schritt in eine MS-ACCESS-Datenbank zu übernehmen. In der Datenbank wurden Report-Masken generiert, die es möglich machten, die Daten jedes Falles in strukturierter Weise als sogenannte Einzelfallprotokolle auszudrucken. Die Einzelfallprotokolle dienten der Edition als Korrekturvorlage. Die Korrekturen konnten nach Abschluss der Edition mit Hilfe eines für diesen Zweck entwickelten Eingabeprogramms menügeführt über Eingabemasken in die Datenbank übertragen werden.

Die Datenbank erwies sich nach Abschluss der Editionsarbeiten auch als außerordentlich nützliches Werkzeug für die Durchführung der abschließenden Datenprüfung (siehe auch Abschnitt 3.7).

Die Entwicklung weiterer SPSS-Syntaxfiles war notwendig, um die korrigierten Daten der Datenbank in analysierbare SPSS-Datenfiles zu verwandeln und aus diesen wiederum die vorliegenden Public-Use-Files zu erzeugen.

3.4 Datenedition

3.4.1 Editionshandbuch

Auf der Grundlage des ständig aktualisierten Editionshandbuchs erfolgte die Edition der einzelnen Fälle. Das Handbuch diente den Mitarbeiter(inne)n der Edition als Regelwerk. Anhand der letzten Version dieses Handbuchs (Teil IV dieser Dokumentation) lassen sich somit sehr genau sowohl der Prozess der Edition als Ganzes als auch die Gründe für einzelne Korrekturen/Veränderungen an den Daten nachvollziehen.

Nicht alle Module wurden in die Edition einbezogen. Die Hauptmodule 'Angaben zur Zielperson', 'Schulausbildung', 'Berufsausbildung', 'Erwerbstätigkeiten', 'Arbeitslosigkeit', 'Nebentätigkeiten', 'Aus- und Weiterbildung' sowie das 'Lückenmodul' waren durch die Edition vollständig zu prüfen und zu korrigieren. Im Modul 'Berufskontrolle, Politik, Religion' fand keine Edition statt, für alle anderen gab es einige ausgewählte Editionsanweisungen.

3.4.2 Fallmaterialien

Der Edition standen zu den einzelnen Fällen folgende Materialien zur Verfügung:

- Tonband (falls vorhanden)
- Protokoll der Voredition (sofern der Fall voreditiert wurde)
- Interviewprotokoll (sofern vorhanden)
- Biographieschema
- Einzelfallprotokoll
- Variablenliste
- Fragebogen (ergänzte Version)

Diese Materialien sind im Kapitel 1.4 (Seite 6ff.) des Editionshandbuchs (Teil IV dieser Dokumentation) detailliert beschrieben.

3.4.3 Nachrecherche

Nach bestimmten Kriterien, die in Abschnitt 2.3.3 auf Seite 28 des Editionshandbuchs (Teil IV dieser Dokumentation) beschrieben sind, wurden 1272 Fälle von der Edition für eine Nachbefragung vorgesehen. In 1016 Fällen konnte die Nachrecherche erfolgreich oder teilweise erfolgreich durchgeführt werden. Mit der Durchführung waren Mitarbeiter/-innen der Edition betraut, die dabei bestimmte Nachrecherche-Regeln zu befolgen hatten. Diese Regeln enthalten Anweisungen zur Vorbereitung, zur Kontaktaufnahme mit der zu befragenden Person, zum Verhalten während einer telefonischen Nachrecherche sowie zur Dokumentation der Nachrecherche in den Daten selbst und im Verlauf. Sie sind im Teil V dieser Dokumentation abgedruckt).

3.4.4 Verlaufsdocumentation und -statistik von Edition und Nachrecherche

Zu jedem einzelnen Fall wurde in einer Datei (EDVerlGES.XLS) der Gesamtverlauf dokumentiert, sie enthält die folgenden Angaben:

- Anzahl der Tonbänder, falls vorhanden,
- Bemerkungen zum Tonband und sonstige Notizen,
- Bearbeiter/-in und Datum der jeweiligen Bearbeitungsschritte (Voredition, Erstedition, Zweitedition, Einzelfallentscheidung während Erstedition, W-Code, Korrekturingabe nach Erst- oder Zweitedition, Nachrecherche/Endedition, Einzelfallentscheidung während Endedition, AB-Code, Korrektur/Überprüfung/Nachtrag, letzte Korrekturingabe),
- Dauer des Tonbandabhörens im Zuge der Erstedition,
- von Tonband gehörte Module im Zuge der Erstedition,
- ob es sich um einen NR-Fall handelt,
- Dauer der Bearbeitungsschritte Nachrecherche/Endedition, Einzelfallentscheidung (Endedition), AB-Code, Korrektur/Überprüfung/Nachtrag.

Eine zweite Datei (NR_Verlauf.XLS) enthält ausschließlich zu den Nachrecherchefällen die Informationen:

- Kontaktversuche (Anzahl und Editionsmitarbeiter/-in),
- Ergebnis der Adressprüfung,
- Datum und Stand von Terminanfragen,
- Bemerkungen zur Reaktion der befragten Person,
- Bemerkungen zur Nachrecherche

- Abschluss der Nachrecherche (Bearbeiter/-in, Datum, erfolgreich/nicht erfolgreich/verweigert, ggf. Art der Verweigerung),
- Art der Nachrecherche (Nur-Noten-Nachrecherche/allgemeine Nachrecherche),
- Modul, in dem eine oder mehrere Angaben nacherhoben wurden,
- Anzahl der Spells, die teilweise oder komplett nacherhoben wurden.

Das Führen dieser Dateien diene der internen Überprüfung des Editions- und Nachrecherchestands. Die einzelnen Dateien werden lediglich archiviert.

3.5 Vercodung

Freitextvariablen wurden nachträglich codiert. Die folgende Übersicht enthält alle vercodeten Variablen. Die Vercodungsvariablen und die verwendeten Klassifikationen bzw. Kategorienschemata sind der jeweiligen Beschreibung in Teil VII dieser Dokumentation zu entnehmen.

	Modul	Ausgangsvariable/ zu vercodender Text
Ausbildungsberuf des Partners/der Partnerin	Rec16sp	FP6A
Erwerbsberuf des Partners/der Partnerin zu Beginn der Partnerschaft		FPET2
Erwerbsberuf des Partners/der Partnerin am Ende der Partnerschaft		FPET5
Beruf der Mutter bis ZP 16	Rec2	ELMU4
Beruf der Mutter vor Geburt der ZP		ELMU8
Beruf der Stiefmutter bis ZP 16	Rec2sm	ELMU4
Beruf des Vaters	Rec2	ELVA5
Beruf des Vaters als ZP 15		ELVA6
Beruf des Stiefvaters	Rec2sv	ELVA5
Beruf des Stiefvaters als ZP 15		ELVA6
Ausbildungsberuf der ZP	Rec6sp	AB1
Erwerbsberuf der ZP	Rec7sp	BG1
Nebentätigkeit	Rec10sp	NT2
Branche	Rec7sp	BG13
Berufswechselgründe	Rec7sp	BG32 BG32A(1-11) BG32B(1-9)
Angestrebter Ausbildungsabschluss	Rec6sp	AB1
Zusätzlicher beruflicher Abschluss	Rec6sp	ABZ28 ABZ28B ABZ28C
Wohnort	Rec4sp	WG1B
Standort des Ausbildungsbetriebs	Rec6sp	ABL4S
Standort des Betriebs/der Arbeitsstätte	Rec7sp	BG11B

Des Weiteren gab es

- nachträgliche Zuordnungen zu bestehenden und ergänzten Kategorien (vor allem Freitextangaben unter der Kategorie "Sonstiges"), die im Rahmen der Edition erfolgten. Dies betraf Schulabschlüsse der Eltern, Geschwister, Partner und Kinder, Ausbildungsabschlüsse der Eltern, Geschwister und Partner sowie Ausbildungsabschlüsse der Zielperson (AB-Code) und
- Vercodungen von Sachverhalten durch das Einfügen neuer Variablen, und zwar wurde bei Einkommensangaben die Währung codiert und bei den Ausbildungen der Zielperson wurde codiert, ob sich die Ausbildungsstätte im Ausland oder der DDR befand (ABL2Z).

Diese "Vercodungen" wurden im Rahmen der Edition durchgeführt und sind im Einzelnen im Editionshandbuch (Teil IV dieser Dokumentation) beschrieben.

3.6 Einzelfallentscheidungen

Während jeder Phase der Edition war es möglich, dass Einzelfallentscheidungen getroffen wurden. Stellte die Edition in einem Fall eine Unklarheit fest, die mit den bestehenden Regeln nicht lösbar war und die sich auch nicht verallgemeinern ließ, sodass eine neue Regel hätte aufgestellt werden können, wurde zusammen mit der Projektleitung eine Einzelfallentscheidung getroffen. Diese Entscheidungen wurden dokumentiert und begründet, und zwar für die Phasen Erst- und Zweitedition in der Datei EinzelfERST.XLS und für die Endedition in der Datei EinzelfallEND.XLS. Im Zuge der Endedition kam es vor, dass Einzelfallentscheidungen, die in früheren Phasen getroffen wurden, rückgängig gemacht oder geändert werden mussten, weil es mittlerweile passende bzw. anders lautende Regeln gab. In diesen Fällen wurde auch die entsprechende Einzelfall-Datei korrigiert.

Die Datei EinzelfallEND.XLS ist in gekürzter und anonymisierter Form in Teil VI dieser Dokumentation zu finden.

3.7 Datenprüfung

Die Datenprüfung nach Abschluss der Edition verfolgte das Ziel, Datenfehler und Plausibilitätsprobleme, die von der Edition nicht entdeckt oder durch die Datenkorrektur eingabe erzeugt wurden, ausfindig zu machen. Der Datenprüfung lagen einerseits die Filterführungsrichtlinien des Erhebungsinstrumentes und das formalisierte Regelwerk der Edition zugrunde. Ziel der Datenprüfung war es, Verstöße gegen diese Regelwerke ausfindig zu machen, in diesem Sinne fehlerhafte Daten zu entdecken und einer Korrektur zuzuführen. Es ließen sich bei der Datenprüfung drei generelle Aufgabenbereiche differenzieren

- die Wertebereichsprüfung der einzelnen Variablen,
- die Prüfung der zeitlichen Konsistenz von Episoden und
- die Filterführungsprüfung,

die im Folgenden genauer beschrieben werden.

3.7.1 Wertebereichsprüfung

Ausgangsbasis der Wertebereichsprüfung waren die Merkmalsausprägungen, die bei der Konstruktion der einzelnen standardisierten Variablen, festgelegt worden waren. Die erste Wertebereichsprüfung wurde bereits mittels des Datenerhebungsinstrumentes durchgeführt, indem nur bestimmte Werte in die entsprechenden Felder der Datenerfassungsmasken eintragbar waren. Auf eine abschließende Prüfung konnte jedoch nicht verzichtet werden, da eine Vielzahl von Einflüssen (Fehler im Erhebungsinstrument, Editionsveränderungen, Nacherhebung von Daten, nachträgliche standardisierte Vercodung offen erhobener Angaben) dazu führten, dass Abweichungen der Daten von den vorgegebenen Wertebereichen auftreten konnten.

Der erste Schritt zur Wertebereichsprüfung bestand in der Aufbereitung der Daten in mit Variablen- und Value-Labels versehenen SPSS-Datenfiles. Mit Hilfe dieser Datenfiles wurden dann Häufigkeitsauszählungen aller numerischen und aller standardisierten nicht-numerischen Variablen gemacht.

Die Prüfung konzentrierte sich bei den kategorialen Variablen auf die Kontrolle, ob für alle aufgeführten Werte auch Wertelabels vorhanden sind. War dies nicht der Fall, dann wurde damit entweder auf einen Wert außerhalb des Wertebereichs oder auf ein fehlendes Label eines Werts hingewiesen.

Abbildung 2: Wertebereichsprüfung am Beispiel der kategorialen Variable ABL2

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
ABL2 In welcher Ausbildungsstätte haben Sie Ihre Ausbildung absolviert?	Gültig				
	-9 Edition: k.A.	63	1,3	1,3	1,3
	-8 weiss nicht	2	,0	,0	1,4
	-5 Edition: Tnz.	3	,1	,1	1,4
	1 in einem privaten Betrieb	2153	44,9	44,9	46,3
	10 im oeffentlichen Dienst	409	8,5	8,5	54,8
	15 in einer Berufs-Fachschule	418	8,7	8,7	63,5
	16 an einer Universitaet, Hochschule, Akademie	742	15,5	15,5	79,0
	17	256	5,3	5,3	84,3
	18 Fachhochschule der öffentlichen Verwaltung	37	,8	,8	85,1
	22 an einer Fachschule	157	3,3	3,3	88,3
	23	255	5,3	5,3	93,6
	24 in einer anderen schulischen Einrichtung	305	6,4	6,4	100,0
Gesamt	4800	100,0	100,0		

Werte ohne Label → (pointing to rows 17 and 23)

Bei metrischen bzw. intervallskalierten Variablen richtete sich die Kontrolle, neben den auch hierfür teilweise vorhandenen standardisierten Werten, auf den Beginn und das Ende des Wertebereichs der entsprechenden Variablen, um sogenannte "Werteausreißer" zu entdecken. Die Einschätzung, ob diese sich noch innerhalb eines plausiblen Rahmens bewegen, war dann wieder Aufgabe der Edition.

Abbildung 3: Wertebereichsprüfung am Beispiel der metrischen Variable BG19A

BG19A Wie viele Stunden pro Woche zu Beginn im Arbeitsvertrag fest- gelegt?	43,40	1	,0	,0	92,5
	43,50	6	,1	,1	92,6
	43,70	2	,0	,0	92,6
	43,80	3	,0	,0	92,7
	44,00	7	,1	,1	92,8
	44,50	1	,0	,0	92,8
	45,00	66	,9	,9	93,7
	45,20	1	,0	,0	93,8
	46,00	10	,1	,1	93,9
	46,50	4	,1	,1	94,0
	47,00	4	,1	,1	94,0
	48,00	31	,4	,4	94,5
	48,20	1	,0	,0	94,5
	48,50	4	,1	,1	94,5
	50,00	34	,5	,5	95,0
	54,00	4	,1	,1	95,1
	55,00	4	,1	,1	95,1
	56,00	3	,0	,0	95,2
	58,00	2	,0	,0	95,2
	60,00	12	,2	,2	95,4
	63,00	1	,0	,0	95,4
	70,00	3	,0	,0	95,4
	72,00	1	,0	,0	95,4
	80,00	4	,1	,1	95,5
	95,00 keine festgelegte Arbeitszeit	312	4,2	4,5	100,0
	98,00	1	,0	,0	100,0
	Gesamt	6964	92,8	100,0	
	Fehlend System	540	7,2		
	Gesamt	7504	100,0		

Arbeitszeit von mehr als 80 Stunden/Woche möglich?

3.7.2 Zeitliche Konsistenzprüfung

Bei der Datenerhebung wurden die einzelnen Lebensbereiche wie Schulausbildung, Berufsausbildung, Erwerbstätigkeiten und Erwerbslücken, Wehr-/Zivildienstzeiten, Nebentätigkeiten und Weiterbildungen unabhängig voneinander in separaten thematischen Blöcken erfragt. Faktisch sind diese Lebensbereiche jedoch nicht unabhängig voneinander. So kann eine Person, von der aufgewendeten Zeit her gesehen, nicht gleichzeitig mehrere Vollzeitaktivitäten wie Erwerbstätigkeiten und Ausbildungen parallel betrieben haben. Einige Aktivitäten schließen sich gegenseitig inhaltlich-logisch aus, wie etwa Arbeitslosigkeit und Erwerbstätigkeit (Schwarzarbeitsphasen natürlich ausgenommen) und dürfen daher nicht parallel vorhanden sein. Bestimmte Aktivitäten wurden in dem oben beschriebenen Erhebungsdesign mehrfach in unterschiedlichen Modulen erfasst. So wurden beispielsweise Ausbildungen, die innerhalb der Erwerbsgeschichte absolviert wurden, sowohl im Ausbildungsmodul als auch im Lückenmodul festgehalten. Zu all diesen Problembereichen entwickelte die Edition Regeln und Bestimmungen der Datenbereinigung, die dann ebenfalls Voraussetzungen für die technische Datenprüfung waren.

Grundlage für die Prüfung der zeitlichen Konsistenz der Daten war ein so genannter Spellsplitting-File. Hierfür wurden die Episoden aus den unterschiedlichen Lebensbereichen in der Weise in einen Datenfile zusammengeführt, dass zeitliche Parallelitäten in und zwischen den Lebensbereichen, aber auch bislang undefinierte Zeiträume, erkennbar und damit prüfbar wurden.

Abbildung 4: Spellsplitting-File

	caseid	sid	start	end	as	ab	bg	lue
1	XXXXXX	1	861	944	.	.	.	2
2	XXXXXX	2	945	992	1	.	.	.
3	XXXXXX	3	993	1086	2	.	.	.
4	XXXXXX	4	1087	1097	2	.	1	.
5	XXXXXX	5	1098	1112	.	.	1	.
6	XXXXXX	6	1113	1183	.	1	1	.
7	YYYYYY	1	858	930	.	.	.	2
8	YYYYYY	2	931	979	1	.	.	.
9	YYYYYY	3	980	1051	2	.	.	.
10	YYYYYY	4	1052	1085	.	1	.	.
11	YYYYYY	5	1086	1086	.	.	.	1
12	YYYYYY	6	1087	1096	.	.	1	.
13	YYYYYY	7	1097	1119	.	2	1	.
14	YYYYYY	8	1120	1125	.	2	2	.
15	YYYYYY	9	1126	1180	.	.	2	.
16	YYYYYY	10	1181	1183	.	.	.	1
17								

Das obige Beispiel zeigt in vereinfachter Form einen Spellsplitting-File, der die Schul-, Ausbildungs-, Erwerbs- und Aktivitätslückengeschichte von zwei Befragten beschreibt, beginnend mit deren Geburtsdatum und endend mit dem Interviewzeitpunkt. Die Zeitangaben sind gemessen in Monaten seit Januar 1900. AS, AB, BG und LUE sind Statusvariablen für Schul-, Ausbildungs-, Erwerbs- und Lückenepisoden. Die Werte in AS, AB und BG stellen die Episodenummer im jeweiligen Lebensbereich dar. LUE hat die beiden Ausprägungen 1 "undefinierter Lückenzeitraum" und 2 "Zeitraum zwischen Geburt und erstem Schulbesuch". Überschneidungen zwischen Aktivitäten und fehlende Informationen zu bestimmten Zeiträumen sind, wie an diesem Beispiel deutlich wird, mittels solcherart aufbereiteter Daten sofort erkennbar.

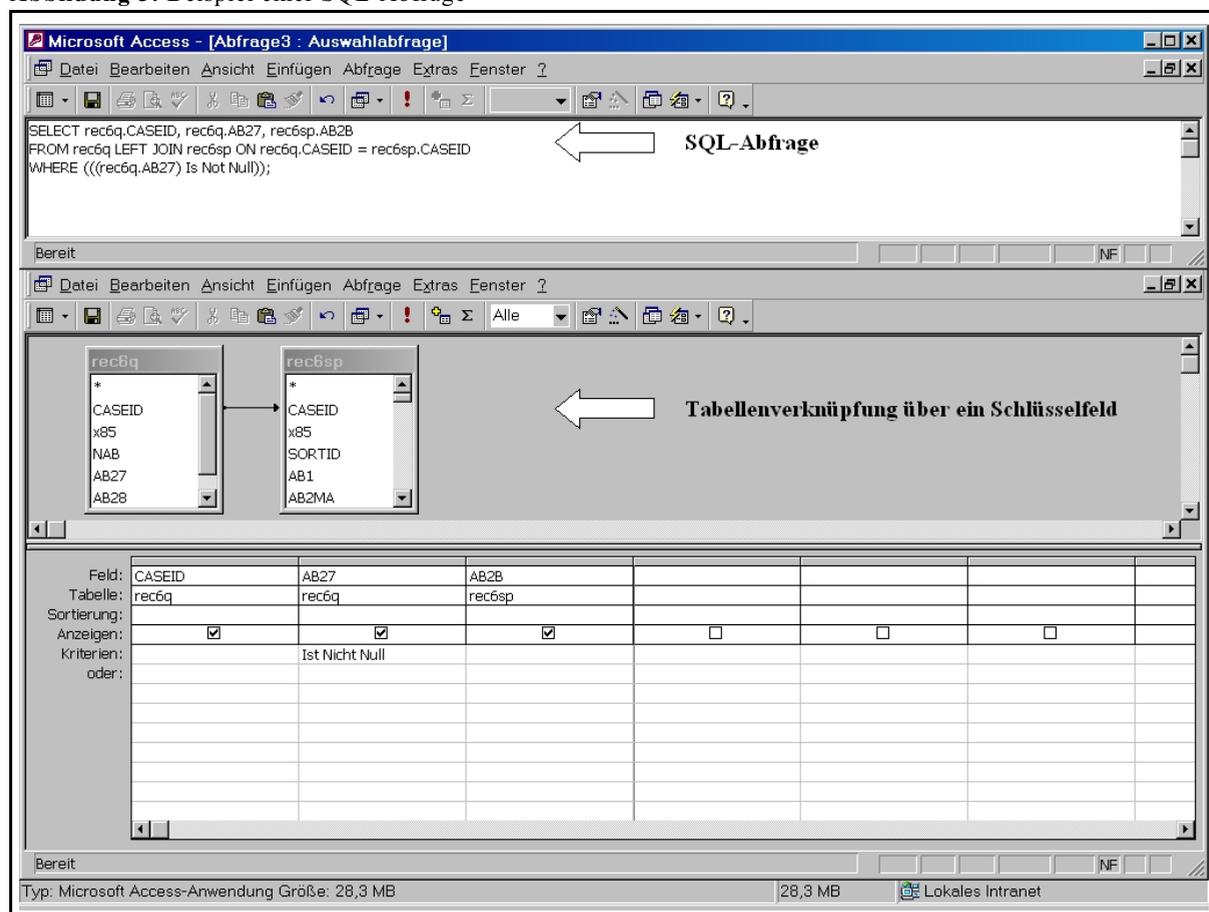
Voraussetzung für die Generierung eines derartigen Spellsplitting-Files waren vor allem valide erhobene Zeitangaben zum Beginn und Ende der Episoden. Fehlende Monatswerte wurden in diesem Zusammenhang näherungsweise geschätzt. Fehlende Jahresangaben wären nur sehr unsicher schätzbar gewesen und führten daher zum Ausschluss dieser Episode aus dem Spellsplitting-File und zu einem Anstieg der Anzahl von undefinierten Lückenzeiträumen. Da die Zeitangaben eine sehr zentrale Rolle bei der Analyse aber auch der Prüfung der Daten spielt, wurde im Editionsprozess sehr viel Zeit und Mühe für die korrekte Rekonstruktion fehlender Angaben aufgewendet.

3.7.3 Filterführungsprüfung

Auch die Filterführung (Steuerung der Abfolge der Fragen des Erhebungsinstrumentes in Abhängigkeit vom Antwortverhalten der Befragten) wurde dem Befragten (bzw. Interviewer) durch das entsprechend programmierte Erhebungsinstrument so vorgegeben, dass es im Erhebungsprozess kaum noch zu Filterführungsfehlern kommen konnte. Je komplexer jedoch das Datenerhebungsinstrument und je umfangreicher die Studie ist, um so eher ist zu erwarten, dass sich entweder Fehler in die Programmsteuerung einschleichen, oder nicht alle Filter durch eine Programmsteuerung eindeutig nachbildbar sind. Zusätzliche Probleme in der korrekten Beachtung der Filter können durch die Veränderungen auftreten, die die Edition in die Daten einbringt. Auch die Nacherhebung von Daten und die Nachvercodung von offen erhobenen Informationen, die u.U. ungeprüft in den Datenbestand integriert werden, sind mögliche Fehlerquellen. Da die Filterregeln nicht nur innerhalb der einzelnen thematischen Module angelegt waren, sondern häufig modulübergreifend wirkten und in vielfältiger Weise *Längsschnitts-* und *Querschnittsdaten* miteinander in Beziehung setzten, war selbst nach Einsatz des mit aufwändig automatisierten Prüfverfahren versehenen Erhebungsinstrumentes die abschließende Filterführungsprüfung unerlässlich.

Voraussetzung für diese Prüfung war eine korrekte, vollständige und im Editionsprozess immer wieder aktualisierte Dokumentation aller Filterführungsregeln. Als optimale Hilfsmittel zur Filterführungsprüfung haben sich PC-gestützte Datenbanksysteme, wie beispielsweise MS-ACCESS, herausgestellt, in denen die Befragungsdaten in separaten Tabellen abgelegt werden. Diese Tabellen können mittels SQL-Abfragen (Structured Query Language) den Prüfbedürfnissen entsprechend angepasst und über Schlüsselfelder miteinander verknüpft werden. Die SQL-Abfragen lassen sich separat speichern und für sehr komplexe Datenbeziehungen wiederum als Ausgangsbasis für weitere SQL-Abfragen nutzen.

Abbildung 5: Beispiel einer SQL-Abfrage



Ebenfalls wichtig sind die sehr einfachen und flexiblen Möglichkeiten, in der Tabellenansicht der Daten Variablen zu Gruppen zusammenzustellen und diese auf einfache Weise nach ihren Werten zu sortieren. Nimmt man als Basis der Gruppierung und Sortierung die Filterbedingungsvariablen, dann lassen sich die filterabhängigen Variablen auf sehr schnelle und sichere Art optisch prüfen. Sind Filterführungsfehler vorhanden, treten diese immer an den Schnittstellen eines Wertewechsels der Filterbedingungsvariablen auf. Bei der optischen Prüfung der Daten brauchen also nicht alle Einzeldaten, sondern nur diese Schnittstellen kontrolliert zu werden. Eine Selektion der fehlerhaften Fälle durch aufwändige Formulierung aller möglichen Fehlerbedingungen ist daher bei dieser Art der Prüfung nicht notwendig.

3.7.4 Datenkorrektur

Die Daten der Lebensverlaufsstudie durchliefen alle drei der oben beschriebenen Stadien der Datenprüfung. In edierten Modulen entdeckte Probleme wurden entweder direkt mit Hilfe des Editionsregelwerks korrigiert oder zur Einzelfallbearbeitung zurück an das Editionsteam verwiesen. Für einige zentrale Module wurde der Datenprüf- und der Datenkorrekturvorgang mehrfach wiederholt, vor allem dann, wenn nach der abschließenden Prüfung noch Korrekturen aus der Edition in die Daten einfließen. Aus arbeitsorganisatorischen Gründen musste in manchen Fällen die logische Abfolge der Arbeitsschritte von Edition zu Datenprüfung durchbrochen werden. So wurde z.B. die Berufsvercodung, mit dem Ziel die Gesamteditionszeit zu verkürzen, separat bereits durchgeführt, bevor die Erwerbsgeschichte abschließend ediert war. Da es aber zwangsläufig zu Veränderungen in den Erwerbsdaten durch den Editionsprozess kam, mussten die separat vercodeten Berufe nachträglich daraufhin überprüft werden, ob sie noch den korrekten Bezug zu den edierten Daten hatten. Bei den Fällen, die durch die Edition verändert worden waren, mussten diese Angaben nachvercodet und nachträglich erneut auf Korrektheit geprüft werden. So gestaltete sich die idealtypische "abschließende Datenprüfung", zumindest für die edierten Module, eher als permanente editionsbegleitende Datenprüfung.

Bei den nicht-edierten Modulen erfolgten dagegen die Datenkorrekturen nach Datenprüfung ausschließlich auf Basis der Filterlogik.

Die hier beschriebenen Maßnahmen, die zur Datenprüfung ergriffen wurden, machen deutlich, dass die Daten der Lebensverlaufsstudie als außerordentlich intensiv und gut geprüfte Daten gelten können. Das bedeutet natürlich nicht, dass sie absolut fehlerfrei sind. Jeder, der sich mit der Prüfung von Daten beschäftigt, kennt das Phänomen, dass selbst nach zigfacher Kontrolle immer noch irgendetwas übersehen wurde. Die Erfahrung zeigt jedoch, dass das Ausmaß der noch vorhandenen Fehler durch die Prüfverfahren auf ein für Auswertungszwecke irrelevantes Maß gesenkt werden konnte.

3.8 Auswahl der Fälle für den Analysedatensatz

In den Editionsprozess wurden 2911 vollständige Fälle einbezogen. Bei zwei Fällen wurde festgestellt, dass die falsche Person befragt wurde. In einigen weiteren Fällen wurde vermutet, dass die Zielpersonen bewusst falsche Angaben machten (DAT1 = 1), Schwierigkeiten bei der Beantwortung der Fragen hatten (DAT1 = 2) oder aufgrund mangelnder Deutschkenntnisse Sprachschwierigkeiten hatten (DAT1 = 3).

Die speziellen Editionsregeln zu den hier genannten Variablen und der erläuternden Textvariable DAT2 sind dem Editionshandbuch (Teil IV dieser Dokumentation, Seite 32) zu entnehmen.

Aus den Daten wurden lediglich die beiden Fälle ausgeschlossen, bei denen festgestellt wurde, dass die falsche Person befragt wurde. Die Datenfiles enthalten somit nur noch 2909 Fälle.

4 Beteiligte Personen

Am Max-Planck-Institut für Bildungsforschung leitet Karl Ulrich Mayer das Gesamtprojekt "Lebensverläufe und gesellschaftlicher Wandel" seit 1983. Die Mitarbeiter/-innen und ihre Aufgaben sind/waren:

Erika Brückner	Fragebogenentwicklung, Pretestbetreuung und -auswertung, Interviewerschulung (1997/1998)
Felix Büchel	Wissenschaftliche Mitarbeit (1998)
Michael Corsten	Wissenschaftliche Mitarbeit (2000/2001)
Henriette Engelhardt	Wissenschaftliche Mitarbeit, Editionsleitung (1998-2000)
Steffen Hillmert	Projektkoordination (ab 2000), Wissenschaftliche Mitarbeit, Editionsleitung
Antje Mertens	Wissenschaftliche Mitarbeit; Betreuung der Edition
Götz Rohwer	Projektplanung (1997)
Heike Trappe	Projektkoordination (bis 2000), Wissenschaftliche Mitarbeit, Betreuung der Edition
Monika Albin	Datenbank (bis 2001)
Anke Höhne	Supervision (1999/2000)
Ralf Künster	Datenbank (ab 2001)
Sonja Menning	Fragebogenentwicklung (1997)
Petra Spengemann	Organisation, Fragebogenentwicklung, Vercodung, Supervision
Anette Veauthier	Vercodung (2001/2002)
Katja Winkler	Pretestauswertung (1998/1999)

Außerdem waren mit Edition und Vercodung zahlreiche studentische Hilfskräfte und mit der Eingabe der Korrekturdaten Mitarbeiterinnen des Zentralen Sekretariats betraut.

Die Projektmitarbeiter des Instituts für Arbeitsmarkt- und Berufsforschung waren Stefan Bender und Hans Dietrich.

Bei infas waren Doris Hess, Bernd Schneider und Jacob Steinwede am Projekt beteiligt, außerdem die zahlreichen Interviewer/-innen.

5 Programm LDEX

Das Computerprogramm zur graphischen Darstellung einzelner Lebensverläufe wurde von Ralf Künster erstellt. Teil VIII dieser Dokumentation enthält die Programmbeschreibung .

