

Arbeitsberichte

Working Papers

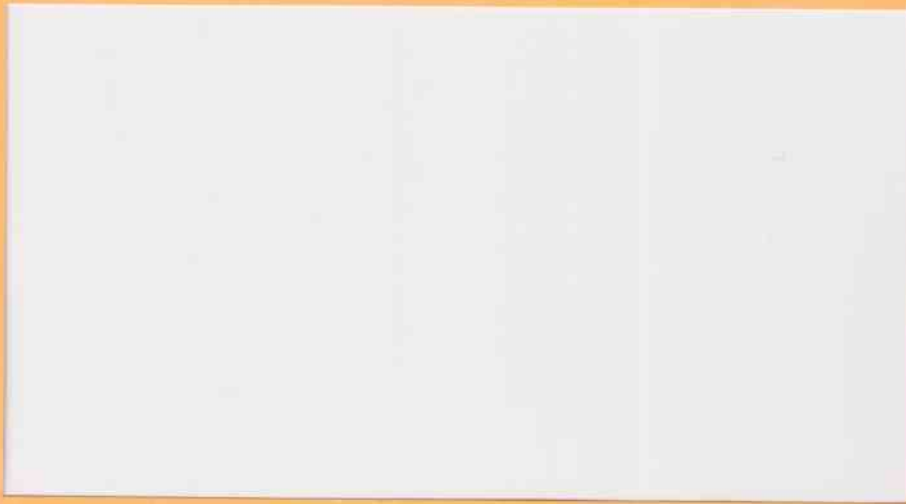
Arbeitspapier Nr. 4 des Projekts Ausbildungs- und Berufsverläufe  
der Geburtskohorten 1964 und 1971 in Westdeutschland

Steffen Hillmert

Die Edition von Lebensverlaufsdaten  
Einzelfallprüfungen, Korrekturentscheidungen und ihre Relevanz

G 2002/1237-2002,7





MPI für Bildungsforschung  
  
11021586

Projekt

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

---

Arbeitspapier Nr. 4 des Projekts Ausbildungs- und Berufsverläufe  
der Geburtskohorten 1964 und 1971 in Westdeutschland

Steffen Hillmert

Die Edition von Lebensverlaufsdaten  
Einzelfallprüfungen, Korrekturentscheidungen und ihre Relevanz

7/2002



g 2002/1237-2002,7

Kontaktadresse:

Dr. Steffen Hillmert  
Max-Planck-Institut für Bildungsforschung  
Lentzeallee 94  
14195 Berlin

hillmert@mpib-berlin.mpg.de

<http://www.mpib-berlin.mpg.de>

*Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland* ist seit 1998 ein Projekt am Max-Planck-Institut für Bildungsforschung, Berlin, Forschungsbereich *Bildung, Arbeit und gesellschaftliche Entwicklung* (Leiter: Prof. Dr. K.U. Mayer).

Die wissenschaftlichen Mitarbeiterinnen und Mitarbeiter des Projektes sind gegenwärtig (7/2002): Dr. Steffen Hillmert, Dipl.-Soz. Marita Jacob (Doktorandin), Prof. Dr. Karl Ulrich Mayer, Dr. Antje Mertens, Holger Seibert, M.A. (Doktorand) und Dr. Heike Trappe. Kooperationspartner am Institut für Arbeitsmarkt- und Berufsforschung (IAB) sind Dipl.-Soz. Stefan Bender und Dr. Hans Dietrich.

## Inhalt

<b>1. Einleitung</b> .....	<b>4</b>
<b>2. Editionsphasen und Gesamtaufwand in der LV64/71-Studie</b> .....	<b>5</b>
<b>3. Typische Probleme der Datenedition</b> .....	<b>7</b>
3.1 Grundlegendes Vorgehen und Regeln bei der Datenedition.....	7
3.2 Inkonsistenzen und ihre Bearbeitung.....	9
3.3 Uni- und bivariate Verteilungen in der Gesamtstichprobe: das Beispiel ‚Ausbildung und berufliche Stellung‘.....	12
3.4 Qualitätsverbesserung durch die Edition.....	15
<b>4. Fazit: Effekte und Effizienz der Edition</b> .....	<b>16</b>
<b>Literatur</b> .....	<b>18</b>
<b>Anhang (ergänzende Übersicht)</b> .....	<b>19</b>

## Zusammenfassung

In der empirischen Sozialforschung wird großes Gewicht auf die Datenanalyse und ihre Methoden gelegt. Stufen des Forschungsprozesses, die der eigentlichen Analyse vorgelagert sind, erhalten hingegen geringere Aufmerksamkeit, obwohl auch sie für die Qualität der Ergebnisse entscheidend sind. Der Beitrag beleuchtet mit der fallbezogenen Edition weitgehend standardisierter Lebensverlaufsdaten einen solchen Teilschritt. Anhand von Beispielen aus der jüngsten Teilstudie des Lebensverlaufsprojektes, die knapp 3000 Lebensverläufe der Geburtsjahrgänge 1964 und 1971 in Westdeutschland umfasst, wird auf verschiedene theoretische und praktische Aspekte der Datenedition eingegangen: typische Datenprobleme, unterschiedliche Informationsquellen und Möglichkeiten der Datenprüfung, grundlegende Annahmen, Regeln und Entscheidungen, Aufwand der Datenedition, Veränderungen im Zuge der Datenkorrektur auf der Ebene des Einzelfalls und der Gesamtstichprobe sowie mögliche theoretische Konsequenzen.

## 1. Einleitung

Ein empirisches Forschungsprojekt lässt sich idealtypisch in Phasen gliedern (etwa Schnell et al. 1995: 7-13), die von der Problemstellung und Konzeptionalisierung über die Auswahl der Untersuchungseinheiten, der Erhebung, Erfassung und Analyse der Daten bis zur theoretischen Reflexion und Darstellung der Ergebnisse reichen. Auf jeder dieser Stufen können Fehler auftreten, die die Gültigkeit der Ergebnisse einschränken, und auf jeder Stufe sind spezifische Vorkehrungen dagegen möglich. Ziel sollte es aber bleiben, den *Gesamtfehler* möglichst klein zu halten. Während in der Methodologie der Forschung großes Gewicht auf die Datenanalyse und ihre Methoden gelegt wird, erhalten vorgelagerte Stufen des Forschungsprozesses eher geringe Aufmerksamkeit, obwohl auch sie für die Qualität der Ergebnisse entscheidend sind. Dieser Beitrag beschäftigt sich näher mit einem solchen Teilschritt: der fallbezogenen Edition von Lebensverlaufsdaten.

Individuelle Lebensverlaufsdaten sind weitgehend standardisiert erhobene, aber in ihrer Struktur besonders komplexe empirische Daten. Daher ist hier der Aufwand für die Prüfung und ggf. Korrektur der Daten auch nach der Erhebung besonders groß, und erst nach der Datenedition steht der eigentliche Analysedatensatz zur Verfügung. Im Falle der am Max-Planck-Institut für Bildungsforschung durchgeführten, retrospektiven Lebensverlaufsstudien handelt es sich bei der Edition um explizit regelgeleitete, meist mehrjährige Prozesse, in denen die Daten einzelfallweise überprüft werden.

Ein Großteil der Unklarheiten in den Daten erweist sich in der Edition als eindeutige Fehler, die relativ einfach zu bereinigen sind. Allerdings bleiben verschiedene Unklarheiten bestehen. Für die Aufklärung stehen verschiedene Methoden und Hilfsmittel zur Verfügung, wie z.B. inhaltliche Überlegungen oder gezielte Nachbefragungen. Hierbei stellen sich mitunter auch grundsätzliche theoretische Probleme.

Anhand von Beispielen aus der jüngsten Teilstudie des Lebensverlaufsprojektes soll im Folgenden auf zentrale Aspekte der Datenedition eingegangen werden<sup>1</sup>: Zunächst werden Organisation und die einzelnen Phasen der Edition kurz beschrieben (Abschnitt 2). Anschließend wird auf typische Probleme der Datenedition eingegangen (Abschnitt 3). Dabei werden zunächst grundlegende Annahmen und Regeln dargestellt. An Einzelfällen wird das Vorgehen der Edition konkret erläutert. Schließlich werden an einem Beispiel Veränderungen in ganzen Merkmalsverteilungen im Zuge der Edition betrachtet. Die zentrale Frage bei einer Bewertung der Edition ist natürlich, wie stark sich die Ergebnisse hierdurch verändern. Kommt man mit uneditierten Daten zu den gleichen Schlüssen wie mit edierten Daten? Im abschließenden Abschnitt 4 wird die generelle Frage nach dem Sinn und Nutzen einer umfangreichen Edition wieder aufgenommen.

Auch wenn im Lebensverlaufsprojekt inzwischen verschiedene Erfahrungen mit umfangreichen Dateneditionen vorliegen, ist ein solches Unternehmen nicht von vorneherein bis ins Detail planbar, und selbstverständlich stellt auch der hier beschriebene Editionsprozess eine Entwicklung dar; darüber hinaus wurde er nicht primär unter methodologischen Gesichtspunkten konzipiert. Insofern sind die folgenden Ausführungen eher als Schilderung des Geschehenen zu lesen, weniger als das Programm, das am Anfang stand und dann exakt umgesetzt wurde.

---

<sup>1</sup> Der Verfasser hat den Editionsprozess für diese Teilstudie seit Anfang 2000 geleitet. Ich danke Petra Spengemann und Anette Veauthier für ihre Unterstützung bei der Ausarbeitung dieses Beitrages. Als weitere langjährige Mitglieder der Editionsgruppe seien an dieser Stelle Christiane Hertwig und Jan Stöckigt erwähnt. Für hilfreiche Kommentare danke ich zudem Karl Ulrich Mayer und Heike Trappe.

Das Projekt 'Lebensverläufe und gesellschaftlicher Wandel' wird unter der Leitung von Karl Ulrich Mayer seit 1979 durchgeführt, ab 1983 am Forschungsbereich Bildung, Arbeit und gesellschaftliche Entwicklung des Max-Planck-Instituts für Bildungsforschung. Die Datenbasis setzt sich aus im Wesentlichen aus einmaligen (nur in einem Fall auch mehrmaligen), standardisierten Retrospektivbefragungen von Personen aus neun ausgewählten Geburtsjahrgänge zusammen, die systematisch über ihren bisherigen Lebensverlauf Auskunft geben (vgl. auch Brückner/Mayer 1998). In ihrer Gesamtheit erlauben diese Lebensverlaufsdaten für Deutschland detaillierte und historisch weit zurückreichende Analysen, die fast das gesamte 20. Jahrhundert erfassen<sup>2</sup>.

In der neuesten Teilstudie (vgl. Corsten/Hillmert 2001) werden die bis Ende der achtziger Jahre in der alten Bundesrepublik durchgeführten Befragungen durch Daten ergänzt, die insbesondere die Bildungs- und Erwerbsverläufe der Geburtsjahrgänge 1964 und 1971 in Westdeutschland bis Ende der neunziger Jahre verfolgen. Im Gegensatz zu den bisherigen Lebensverlaufsstudien wurden auch in Deutschland lebende (deutschsprachende) Ausländer in die Untersuchung einbezogen sowie bewusst zeitlich parallele Ausbildungs- und Erwerbsverläufe aufgenommen. Das Projekt wird in Zusammenarbeit mit dem Institut für Arbeitsmarkt- und Berufsforschung (IAB) in Nürnberg durchgeführt<sup>3</sup>. Die Haupterhebung der Daten erfolgte in den Jahren 1998 und 1999 in Kooperation mit dem Umfrageinstitut Infas (s. Infas 1999). Erfolgreich befragt werden konnte eine Stichprobe von ca. 2900 Personen. Die Interviews erfolgten zumeist in der CATI (computer-assisted telephone interviewing)-Technik. Seit 1999 werden die Daten im Max-Planck Institut für Bildungsforschung geprüft und ediert.

Die Daten sind, je nach Lebensbereich, in verschiedenen Modulen abgelegt (z.B. Schulgeschichte; Ausbildungsgeschichte; Erwerbsgeschichte; Partnerschaftsgeschichte). Dabei ist zwischen Personendatensätzen und Episodendatensätzen zu unterscheiden. Erstere enthalten im Wesentlichen Querschnittsangaben (Personenmerkmale), letztere nehmen für jede Person einzelne Episoden (z.B. Ausbildungsabschnitte, Arbeitsverhältnisse) und deren Merkmale gesondert auf. Die Edition hat sich dabei bisher auf die Module konzentriert, die auch im Zentrum der Forschungen am Institut stehen, nämlich Ausbildungen und Erwerbstätigkeiten. Ziel der Lebensverlaufsstudien ist es, für jeden Lebensbereich möglichst präzise und vollständig die Sequenz einzelner lebensgeschichtlicher Abschnitte zu rekonstruieren.

## 2. Editionsphasen und Gesamtaufwand in der LV64/71-Studie

Derart komplexe Daten sind fehleranfällig (vgl. Brückner et al. 1983). Die Edition solcher Daten ist eine umfangreiche Aufgabe, die im Projekt nur arbeitsteilig bewältigt werden kann. Mit den unmittelbaren Prüfungen und Korrekturen ist dabei vor allem ein Team von bis zu zwölf studentischen Mitarbeiterinnen und Mitarbeitern betraut gewesen. Die Datenedition in der jüngsten Lebensverlaufsstudie hat sich in mehreren Phasen über einen Zeitraum von etwa dreieinhalb Jahren erstreckt. Die folgenden Absätze geben eine kurze Beschreibung dieser Editionsphasen:

Die **Voredition (VorED)** begann kurz nach Beginn der Erhebung, jedoch bereits vor der ersten Datenlieferung. Der Schwerpunkt lag dabei am Anfang in erster Linie auf der Beurteilung der Interviewerinnen und Interviewer und erst in zweiter Linie auf der Edition. Der Voredition standen noch keine Einzelfallprotokolle (also die Angaben einer

<sup>2</sup> Vgl. dazu <http://www.mpib-berlin.mpg.de/de/forschung/bag/index.htm>

<sup>3</sup> Dort betreut von Hans Dietrich und Stefan Bender; gefördert durch Mittel des Europäischen Sozialfonds.

bestimmten Person in Form eines Papierausdrucks) zur Verfügung. Grundlage waren die Tonbänder, Biographieschemata (Auflistung der Lebensverlaufsepisoden in standardisierter Form) und Interviewprotokolle. Die Tonbänder wurden anfangs komplett, später nur noch in ausgewählten Modulen abgehört. Die Mitglieder der Editionsgruppe sollten dabei Fehler und problematische Stellen des Interviews so genau wie möglich protokollieren. Eine Erstfassung des Editionshandbuchs stand ab Januar 1999 zur Verfügung.

Die Erstedition (ErstED) begann im Januar 1999, nachdem die aus studentischen Hilfskräften bestehende Editionsgruppe an einer zweitägigen Schulung teilgenommen hatte. In der Schulung wurden die Ziele und der Stand der Arbeiten des Projekts vorgestellt. Das Erhebungsinstrument und die für die Interviewer schwierigen Stellen wurden erläutert und durch ein nachgestelltes Interview verdeutlicht. Die Editionsregeln wurden an einem Fall demonstriert, danach bearbeiteten die Editeure einen Beispielfall einzeln und besprachen ihre Ergebnisse zunächst Gruppen und dann im Plenum. Als Arbeitsmaterialien waren vorhanden: das Editionshandbuch in der damaligen Fassung, eine Variablenliste, das Computerprogramm für die Erhebung (um die Steuerung nachvollziehen zu können) sowie vereinfachte Biographieschemata, das Einzelfallprotokoll und, falls vorhanden, das Interviewprotokoll und die Ergebnisse der Voredition. Die Editeure waren anschließend gehalten, keine eigenmächtigen Entscheidungen zu treffen, sondern Probleme zuerst mit der Supervision zu besprechen, die entweder auf die passende Editionsregel verwies und diese noch einmal erläuterte oder entschied, den Fall in einer Editionssitzung vorzustellen oder der Einzelfallentscheidung seitens der Projektleitung vorzulegen. Aus Kapazitätsgründen wurde nach kurzer Zeit entschieden, nur noch die Hauptmodule (im Wesentlichen Schul- und Berufsausbildungen, Arbeitslosigkeit und Erwerbstätigkeiten sowie persönliche Angaben) mit den dazwischen liegenden Lücken zu edieren, dabei aber die Informationen aus den anderen Modulen zu nutzen.

Die Zweitedition (ZweitED) wurde hauptsächlich zum Zweck der Überprüfung, Korrektur und Anpassung der bis dahin ersteditierten Fälle an die inzwischen deutlich veränderten Editionsregeln durchgeführt. Alle Fälle, die zwischen Januar und Juli 1999 ersteditiert wurden, wurden erneut bearbeitet. Dies betraf 911 Fälle, wovon der letzte im Januar 2000 zweiteditiert wurde.

Die Korrektur eingabe nach Erst- bzw. Zweitedition erfolgte für alle Fälle, die entweder zweiteditiert waren oder bei denen die Erstedition im August 1999 oder später stattgefunden hatte und die von Erst- oder Zweitedition nicht als Nachrecherche-Fälle deklariert waren. Sie erfolgte zu diesem Zeitpunkt, weil ursprünglich nicht mehr als die Editionsschritte Erst-, Zweitedition und Nachrecherche vorgesehen waren.

Die Ausbildungsvercodung sollte in einem separaten Arbeitsschritt zunächst alle Fälle bearbeiten, die in der Variable Ausbildungsabschluss den Code ‚Sonstiges‘ enthielten oder die von den Editeuren als inkonsistent (z.B. zwischen Ausbildungsstätte und Abschluss oder zwischen Abschluss und prüfender Stelle) erachtet wurden. Einige der angegebenen Ausbildungsabschlüsse konnten ohne Weiteres einer der vorhandenen Kategorien zugeordnet werden. In vielen Fällen mussten jedoch aufwändige Recherchen angestellt werden, um die Angaben der Befragten zu verifizieren oder korrigieren zu können. Neben der Korrektur machten es diese Recherchen möglich, differenziertere Editionsregeln zu erstellen; sie erforderten aber auch neue, nachträglich einzuführende Kategorien oder Variablen.

Die Nachrecherche (NR) startete nach einer Schulung der damit betrauten Mitarbeiterinnen und Mitarbeiter im April 2000. Zu diesem Zeitpunkt waren bereits durch Erst- oder Zweitedition knapp 700 Fälle dafür vorgesehen. Aus der Endedition bzw. der Ausbildungsvercodung kamen später ca. 500 hinzu, so dass schließlich 1272 Fälle als NR-Fälle deklariert waren. Davon wurden 110 Fälle nicht oder nicht bis zum festgelegten Datum (Dezember 2001) erreicht, bei drei Fällen schienen weitere Kontaktversuche nicht sinnvoll. Ein Teil der NR-Fälle musste lediglich wegen der fehlenden Note des Ausbildungsabschlusses nachbefragt werden (Steuerungsfehler im Instrument; sog. Nur-Noten-Fälle). Alle Personen sollten zunächst telefonisch kontaktiert werden, und das Nachrecherche-Interview sollte möglichst telefonisch geführt werden. Für die Fälle, für die bei der vom Erhebungsinstitut zur Verfügung gestellten Adresse keine Telefonnummer vorhanden war, oder bei denen die angegebene Telefonnummer nicht mehr korrekt war, mussten vorab Recherchen durchgeführt werden. War dies über keines der verwendeten Telefonnummernverzeichnisse möglich, wurden die Personen angeschrieben (Terminanfrage). Sie wurden gebeten, ihre Telefonnummer und einen günstigen Termin für das Telefonat zu nennen. Es wurde ihnen mit diesem Schreiben allerdings auch die Möglichkeit einer schriftlichen Befragung angeboten. Die Anschreiben waren für NR-Fälle und Nur-Noten-Fälle leicht unterschiedlich. In 141 Fällen wurden weitere Auskünfte explizit verweigert oder Terminanfragen nicht beantwortet, so dass insgesamt 1011 Nachrecherchen erfolgreich durchgeführt werden konnten.

Die Endedition (EndED) wurde für alle Fälle durchgeführt. NR-Fälle mussten sowohl vorbereitend bearbeitet als auch nach der Nachrecherche endeditiert werden. Zu den vorhandenen Materialien standen jetzt auch Fragebögen (in einer ausformulierten Papierversion) zur Verfügung, der aus dem Quellfile des Erhebungsinstruments erstellt



worden war. Diese Fragebögen wurden dann jeweils mit den Ergänzungen und Änderungen, die im Editionsverlauf beschlossen wurden, versehen. Das Editionshandbuch war speziell für die Endedition ein weiteres Mal überarbeitet worden. Besonders umfangreich war die Überarbeitung des Abschnitts zum Modul Berufsausbildung. Aber auch nach Beginn der Endedition wurden die Editionsregeln weiter angepasst und vervollständigt.

Abgeschlossen wurde die Edition durch die automatisierte, modulweise Datenprüfung im Anschluss an die Einzelfalledition. Daneben gab es zu jedem Zeitpunkt der Edition Einzelfallentscheidungen. Diese wurden getrennt für die Phasen Erst-, Zweit- und Endedition dokumentiert, da im Zuge der Endedition die eine oder andere Einzelfallentscheidung aus einer früheren Phase rückgängig gemacht werden musste. Nach Abschluss aller Fälle wird die Dokumentation nur noch die Einzelfallentscheidungen aufweisen, die während der Endedition getroffen wurden oder die erhalten blieben. Ferner waren nachträgliche Korrekturen und Überprüfungen aus verschiedenen Gründen notwendig. Wurde beispielsweise während einer bestimmten Phase eine Editionsregel mit weitreichenden Folgen neu erstellt, so mussten alle Fälle, die bis zu diesem Zeitpunkt bereits bearbeitet waren, dementsprechend korrigiert werden. Dies zeigt, dass sich die Edition im fortgeschrittenen Stadium nicht nur mit den Rohdaten, sondern immer wieder mit eigenen früheren Entscheidungen auseinandersetzen musste.

Für die Endedition mit telefonischer Nachrecherche, die die letzte und bei weitem umfangreichste Phase der Einzelfalledition war, wurde eine Statistik geführt, die den Zeitaufwand bei der konkreten Fallarbeit aufzeichnet. Bei einem mittleren Zeitaufwand von etwa 100 Minuten und etwa 2.900 Fällen ergibt sich ein Gesamtvolumen von etwa 5.000 Arbeitsstunden. Dabei handelte es sich in zwei Drittel der Fälle um fehlende Angaben in ein oder zwei Modulen bzw. Unklarheiten in maximal drei Episoden. Der Arbeitsaufwand der Edition insgesamt dürfte ein Mehrfaches dieser Zeit beansprucht haben.

### 3. Typische Probleme der Datenedition

#### 3.1 Grundlegendes Vorgehen und Regeln bei der Datenedition

Im Zuge der Edition werden alle Fälle einzeln geprüft. Dafür stehen eine Reihe von Informationsquellen zur Verfügung:

- der Rohdatensatz aus der Befragung
- zusätzliche handschriftliche Notizen der Interviewer
- in den meisten Fällen Tonbandaufzeichnungen des Interviews (mit Einverständnis der Befragten).

Hinzu kommt als weitere Informationsmöglichkeit insbesondere die telefonische Nachrecherche bei den Zielpersonen selbst (bis Ende 2001). Falls nötig, wurden diese noch einmal telefonisch kontaktiert und um klärende bzw. ergänzende Informationen gebeten. Da dies in einem zeitlichen Abstand von bis zu drei Jahren gegenüber der Hauptbefragung erfolgte, kann die Nachrecherche allerdings ihrerseits mit Konsistenz- und Erinnerungsproblemen (Reimer 2001) behaftet sein kann, auf die hier nicht weiter eingegangen werden kann. Schließlich wurden Informationen aus externen Quellen (über Institutionen etc.) genutzt.

Um die Nachvollziehbarkeit des Editionsprozesses zu sichern, werden eine Reihe von Prinzipien zugrunde gelegt (siehe bereits Tölke 1982 und Dokumentationen zur Edition in verschiedenen Teilstudien des Lebensverlaufsprojekts). Grundsätzlich gilt, dass alle Unplausibilitäten aufzuklären und alle verfügbaren Informationen aus den Interviews auszuschöpfen sind. Individuelle Entscheidungen der Bearbeiter sollen möglichst unterbleiben. Vielmehr wird ein differenziertes Regelwerk konstruiert, das sukzessive angepasst wird. Auftretende Inkonsistenzen und neue Probleme werden in der Arbeitsgruppe diskutiert. Die einzelnen Bearbeiter erhalten

Supervision. Schließlich werden alle Regeln und auch alle (Einzelfall-)Entscheidungen ausführlich dokumentiert (vgl. Spengemann/Hillmert 2002).

Entgegen der Kritik an der (quantitativen) empirischen Sozialforschung, die durch ‚Kategorienfehler‘ vermeintlich entscheidende Tendenzen des sozialen Wandels übersieht, geht es bei der Edition nicht darum, lückenhafte Angaben um jeden Preis zu füllen, unplausible Angaben anzupassen oder Antworten auf jeden Fall in die festgelegten Kategorien einzuordnen. Über derart „blinde Empirie“ schreiben etwa Beck/Beck-Gernsheim (1994: 32; 39): „Quantitative Methoden setzen Kategorisierungen, Gruppenbegriffsbildungen voraus (...). Eine sich individualisierende Gesellschaft entzieht sich aber diesen untersuchungstechnischen Standardisierungszwängen (...). Deshalb ist es für eine auf ihre technische Brillanz stolze Soziologie schwer, sich über ihren eigenen Schatten hinweg für die Fragen der sich individualisierenden Gesellschaft zu öffnen.“

Für die einzelfallorientiert arbeitende Edition spielen zwar Motive der *Praktikabilität* durchaus eine Rolle - etwa das grundsätzliche Ziel, die weitgehende Vollständigkeit der Daten für Analyse Zwecke herzustellen. Durch die Einzelfallprüfung kommt die Edition aber bewusst dem Anspruch nach kategorialer Differenzierung entgegen (siehe unten), und die Hauptleitlinie bleibt die der *Qualität* der Daten im Sinne der Übereinstimmung mit der objektiven Realität bzw. der tatsächlichen Wahrnehmung der Zielpersonen.

Neben der Frage nach der richtigen kategorialen Erfassung ist auch die Warnung vor falscher Konsistenz prinzipiell berechtigt. Wie das folgende Beispiel zeigt, ist ein Fehler aber immer in beide Richtungen - durch Korrektur und durch Unterlassen - möglich.

Zunächst einmal ist aber festzuhalten, dass ein großer Teil des auftretenden Korrekturbedarfs nicht inhaltlich gelagert ist: hier handelt es sich um die Folgen von Steuerungsfehlern im Instrument, um Tipp- oder Hörfehler der Interviewer etc. Diese können oft einfach und mit hoher Sicherheit korrigiert werden, wenn etwa aus der Tonbandaufzeichnung eindeutig hervorgeht, dass der Interviewer zwei Zahlen vertauscht hat. Es ist daher durchaus legitim, Unterschiede zwischen edierten und nicht-edierten Daten im Wesentlichen als korrigierte Fehler zu interpretieren und diese dann auf Eigenschaften von Personen oder Interviewer zurückzuführen (vgl. Brückner 1995).

In einigen Fällen stellen sich aber stärker inhaltliche Probleme. Der Edition kann zum Beispiel das Verhältnis der von einer Person genannten Ausbildung und Erwerbstätigkeit unplausibel erscheinen, und sämtliche verfügbaren Informationen sind bereits ausgeschöpft. Eine Korrektur der Angaben in einer der beiden Variablen wäre ein heikler Eingriff, zumal es sich hier um einen Zusammenhang handelt, der Thema soziologischer Forschung ist (vgl. Themen wie ‚Bildungsrenditen‘ und ‚ausbildungs(in)adäquate Beschäftigung‘). Für die Ergebnisse bestünde tatsächlich die Gefahr von Artefakten, wenn etwa bereits im Prozess der Edition Ausbildungs- und Beschäftigungsniveau einander angepasst werden und dann in den späteren Analysen hochgradige Übereinstimmung als Forschungsergebnis präsentiert wird.

Da die Informationen in den meisten Variablen der Studie aber als objektive Aussagen über die Realität und nicht als rein subjektive Wahrnehmungen verstanden werden (etwa die Nennung formaler Bildungsabschlüsse), gilt auch das umgekehrte Problem: fehlende Korrekturen führen zu Fehlern in den Ergebnissen, wenn die Inkonsistenz eher auf lückenhaften Angaben als auf realen Abweichungen beruht (was vermutlich etwa dann der Fall ist, wenn ein praktizierender Arzt angeblich keinen Ausbildungsabschluss hat). Ähnliches gilt für den Fall, dass die Befragten

zwar Angaben machen, sie sich selbst aber (objektiv) falsch einer Kategorie zuordnen – sei es aus Unkenntnis, sei es auf Grund von Verständigungsschwierigkeiten. In diesem Fall könnte es zu unrealistisch hoher Inkonsistenz kommen – natürlich auch zu falscher Konsistenz, diese ist aber so nicht erkennbar.

Diese Art von Fehlern scheint besonders dann möglich, wenn die Befragten mit Konzepten konfrontiert werden, die nicht unbedingt Teil ihrer Lebenswelt sind – von Problemen sozialer Erwünschtheit beim Antwortverhalten einmal ganz abgesehen. Im konkreten Fall dürfte dies beispielsweise die Variable der beruflichen Stellung („Angestellte mit einfacher Tätigkeit“, „Angestellte mit qualifizierter Tätigkeit“ etc.) überdurchschnittlich betreffen. Diese eher abstrakte Skala dürfte den Befragten weniger vertraut sein als etwa die Bezeichnung ihrer konkreten Ausbildungsabschlüsse, wenngleich ihnen für die einzelnen Kategorien der beruflichen Stellung jeweils Beispiele von Tätigkeiten genannt wurden.

Die Edition prüft u.a. die Plausibilität des Verhältnisses von Ausbildung und Erwerbstätigkeit. Gleichwohl werden Eingriffe in die Daten vorsichtig gehandhabt, und es gibt keine generelle Bereinigungsvorschrift. Im Einzelfall werden bestimmte Plausibilitätsüberlegungen angestellt. So erscheint Überqualifizierung („taxifahrende Akademiker“) – jedenfalls prinzipiell – plausibel, während Unterqualifizierung zumindest dann bezweifelt wird, wenn sie institutionellen gesetzlichen Regelungen widerspricht. Dabei kann z.B. auch berücksichtigt werden, dass besonders ungewöhnliche, aber dennoch wahre Konstellationen Zusatzerklärungen des Befragten nahe legen würden, die auf dem Tonband festgehalten worden sind. Bei der Beurteilung der Glaubwürdigkeit einzelner Angaben werden im Einzelfall Kriterien wie die Menge und Detailliertheit von Angaben, Sicherheit im Antwortverhalten oder die allgemeine Kenntnis über bestimmte soziologische Kategorien berücksichtigt.

Gibt es keinen hinreichenden Beleg für eine veränderte Kategorisierung, bleibt die Angabe unverändert. Bestehen große Zweifel an der Richtigkeit der Angaben, wird dies markiert. Generell gilt, dass später nicht mehr erkennbare Korrekturen allein wegen Unplausibilitäten i.d.R. nicht erfolgen. Vielmehr wird dieser Korrekturschritt in solchen Fällen als Information erhalten, indem zusätzliche Angaben neben die Originalangaben treten (explizite Korrekturvariablen, als ungefähr kenntlich gemachte Zeitangaben u.ä.). Die Befragung enthält auch explizit subjektive Variablen (etwa die Frage nach den Gründen für den Wechsel eines Arbeitsverhältnisses), bei denen keine objektiven Plausibilitätskriterien angelegt werden. Auch hier können die Daten aber zum Beispiel im Hinblick auf Eingabe- und Übertragungsfehler geprüft werden.

### *3.2 Inkonsistenzen und ihre Bearbeitung*

Die Konsistenzprüfung ist nicht die einzige Aufgabe der Edition, aber doch ein wesentlicher Teil. Mögliche Inkonsistenzen wurden bereits bei der Erhebung reduziert, da während des Interviews eine automatische (programmgesteuerte) Konsistenzprüfung zwischen bestimmten Angaben stattgefunden hat. Diese Prüfung betraf insbesondere zeitliche Unplausibilitäten (das Ende einer Episode lag zeitlich angeblich vor ihrem Beginn o.ä.) und Lücken. In solchen Fällen wurde der Interviewer entweder unmittelbar zu einer Nachfrage aufgefordert, oder am Ende des Interviews wurde noch einmal gezielt nach Aktivitäten in den Lückenzeiträumen gefragt.

Ursachen für verbleibende Inkonsistenzen in den Daten können sein:

- Instrumentfehler wie eine fehlerhafte Steuerung oder eine unterbliebene Präzisierung bzw. Differenzierung von Kategorien

- Fehler der Interviewer
- im Sinne der vorgegebenen Kategorien falsche Selbsteinschätzung der Befragten, Erinnerungsfehler oder bewusst falsche Angaben
- sowie reale Abweichungen.

Die folgende Übersicht zeigt wesentliche Arten von Inkonsistenzen im Hinblick auf die modulare Struktur der Daten:

Tab. 1: Mögliche Inkonsistenzen in den Daten (Beispiele)

	Innerhalb von Modulen	Zwischen Modulen
Synchron	(a) Variablen innerhalb einer Episode; z.B. Ausbildungsstätte (Fachschule) entspricht nicht dem Abschluss (Fachhochschuldiplom)  (b) Parallele Episoden z.B. mehrere parallele Ausbildungen in Vollzeit	Parallelität von Zuständen; z.B. längere Ausbildung im Ausland, obwohl im Inland gewohnt
Diachron	Aufbauqualifikationen; z.B. Meisterqualifikation ohne vorherige Berufsausbildung	Reihenfolge (und Niveau) von Aktivitäten; z.B. hochqualifizierte Tätigkeit ohne Ausbildung

Das Vorgehen soll zunächst durch die ausführliche Darstellung einiger konkreter Fälle erläutert werden, in denen für Datenedition typische Aufgaben auftreten:

#### Fall A. (Jahrgang 1971)

##### *Daten vor Edition*

Frau A., die 1977 eingeschult wurde, beendet im Januar 1985 die Sonderschule nach der achten Klasse mit dem einfachen Hauptschulabschluss. Sie absolviert von September 1986 bis September 1987 ein Berufsvorbereitungsjahr und beginnt 1986 eine Bäckerlehre, an deren genauen Anfangszeitpunkt sie sich nicht erinnert. Diese Ausbildung sagt ihr aber nicht zu, zudem hat sie gesundheitliche Probleme. Sie bricht die Bäckerlehre ab, da sie eine bessere Ausbildungsstelle als Hauswirtschaftshelferin findet. Der fehlende Zeitraum zwischen dem Ende der allgemeinbildenden Schule im Januar 1985 und dem Beginn des Berufsvorbereitungsjahres im September 1986 kann nur durch die Informationen aus einer Lückenabfrage ergänzt werden: Frau A. ist von Januar 1985 bis Juni 1986 krank (oder in Rehabilitation) und besucht von Mai 1986 bis September 1987 nochmals eine allgemeinbildende Schule.

*Aufgabenstellung für die Edition: Der Abschluss der Sonderschule mit dem einfachen Hauptschulabschluss nach acht Schuljahren ist zu überprüfen. Die Überschneidung zwischen der Krankheitsphase von Januar 1985 bis Juni 1986 und dem ebenfalls nur aus der Lückenabfrage bekannten weiteren Besuch einer allgemeinbildenden Schule von Mai 1986 bis September 1987 ist ebenso unplausibel wie die Parallelität dieses Schulbesuchs zum Berufsvorbereitungsjahr.*

##### *Korrektur durch Edition und Nachrecherche*

Die Edition zieht zunächst den Interviewmitschnitt zur Klärung dieser Inkonsistenzen heran, kann aber nur feststellen, dass Frau A. im Zeitraum von Mai 1986 bis September 1987 keine weitere allgemeinbildende Schule, sondern von September 1986 bis September 1987 tatsächlich ein Berufsvorbereitungsjahr absolviert hat. Somit kann der als Lückenaktivität erfasste weitere Schulbesuch gestrichen werden. Zur Klärung der zeitlichen Überschneidung von Krankheitsphase und Berufsvorbereitungsjahr und der fragwürdigen Angabe zum Schulabschluss muss eine Nachbefragung durchgeführt werden: Frau A. beendet die Sonderschule im April 1986 nach der neunten Klasse nicht mit einem Hauptschulabschluss, sondern mit dem Sonderschulabschluss. Den einfachen Hauptschulabschluss bekommt sie erst im Juli 1989 mit dem Abschluss der Ausbildung zur

Hauswirtschaftshelferin zuerkannt. Vor dem Beginn des Berufsvorbereitungsjahres im September 1986 absolviert Frau A. ab Mai 1986 ein zweimonatiges Betriebspraktikum in einer Bäckerei 'um diesen Beruf kennen zu lernen'. Direkt im Anschluss beginnt sie im Juli 1986 die Bäckerlehre. Sie bricht diese Lehre nach zwei Monaten ab und absolviert anschließend ein Berufsvorbereitungsjahr. Die in der Lückenabfrage erhobene Krankheitsphase von Januar 1985 bis Juni 1986 ist nun wegen ihrer Parallelität zum Besuch der Sonderschule und zum Betriebspraktikum zu streichen. Frau A. bestätigt in der Nachrecherche, dass es eine solche Phase nicht gab.

#### **Fall B. (Jahrgang 1964)**

##### *Daten vor Edition (I)*

Herr B. schließt im April 1983 eine Maurerlehre ab und wird im Anschluss von seinem Ausbildungsbetrieb als Maurer übernommen. Er arbeitet bis Mai 1992 (ohne Unterbrechung) als Facharbeiter in diesem Betrieb. In den Jahren 1983/84, 1984/85 und 1985/86 ist Herr B. jeweils in den Monaten Dezember bis April arbeitslos. Für den Zeitraum zwischen der im April 1983 beendeten Maurerausbildung und dem Beginn einer weiteren Ausbildung, die Herr B. von April 1986 bis Juni 1986 (in Teilzeit) absolviert, gibt er in der Lückenabfrage drei Erwerbstätigkeitszeiträume an: von April bis Dezember 1983, von April bis Dezember 1984 und von April bis Dezember 1985.

*Die Edition hat Folgendes zu klären: Die Phasen der Arbeitslosigkeit in den Monaten Dezember bis April der Jahre 1983/84, 1984/85 und 1985/86 liegen innerhalb des über neun Jahre andauernden Erwerbstätigkeitszeitraumes als Maurer von 1983 bis 1992. Gleiches gilt für die nur aus der Lückenabfrage bekannten Erwerbstätigkeitsphasen April bis Dezember 1983, April bis Dezember 1984 und April bis Dezember 1985. Besteht der neunjährige Erwerbszeitraum aus mehreren von Arbeitslosigkeit unterbrochenen Abschnitten?*

##### *Korrektur durch Edition und Nachrecherche*

Zur Klärung kann die Edition Interviewprotokoll und Tonband nutzen: Die im Interviewprotokoll festgehaltenen Phasen der saisonbedingten Arbeitslosigkeit in den Wintermonaten von 1983 bis 1992 werden durch das Abhören des Tonbandes aber nur teilweise bestätigt: Herr B. ist bis einschließlich 1985/86 jeweils von Dezember bis März arbeitslos. Er muss die neunjährige Erwerbstätigkeit als Maurer in diesen Zeiträumen unterbrechen. Die Erwerbstätigkeit muss also zwischen Mai 1983 und Mai 1992 in vier Phasen aufgeteilt werden: April bis Dezember 1983, April bis Dezember 1984 und April bis Dezember 1985 und April 1986 bis Mai 1992. Das Abhören des Tonbandes bestätigt, dass die drei in der Lückenabfrage für die Monate April bis Dezember 1983-85 aufgenommenen Erwerbstätigkeitszeiträume mit den ersten drei dieser Erwerbstätigkeitsphasen (April bis Dezember in den Jahren 1983-85) identisch und deshalb zu streichen sind. Genauere Informationen zu den vier Erwerbstätigkeitsphasen müssen durch die Nachbefragung ergänzt bzw. komplett nacherhoben werden: Herr B. ist von Mai 1983 bis November 1983, von April 1984 bis Dezember 1984, von April 1985 bis Dezember 1985 und von April 1986 bis Mai 1992 jeweils als Maurer in der gleichen Firma beschäftigt.

##### *Daten vor Edition (II)*

Herr B. hat von April 1986 bis Juni 1986 neben seiner Erwerbstätigkeit als Maurer eine weitere Ausbildung zum Sprengmeister absolviert. Sein ‚Vater war in diesem Beruf selbstständig‘, und um dessen ‚Betrieb übernehmen zu können‘, benötigte Herr B. eine ‚Lizenz‘. Diese Ausbildung fand in Teilzeitform an einer Berufsfachschule statt und Herr B. beendete sie mit dem Meistertitel.

*Die Zuordnung dieser Qualifikation als Ausbildung ist zu prüfen. Die Berufsfachschule ist hier als Ausbildungsstätte ebenso in Frage zu stellen wie die Meisterqualifikation.*

##### *Korrektur durch Edition (Ausbildungsvercodung)*

Mit Hilfe der berufskundlichen Informationen des Arbeitsamtes kann festgestellt werden, dass es sich beim Sprengmeister um einen Anlernberuf handelt, für den ein staatlich anerkannter Lehrgang oder eine erfolgreiche Sachkundeprüfung zu absolvieren ist. Eine Meisterqualifikation im Sinne der Edition wird nicht erreicht, sondern lediglich die Berechtigung zur Ausübung bestimmter Tätigkeiten. Die Ausbildung ist zu streichen und als Weiterbildung zu erfassen.

## Fall C. (Jahrgang 1964)

### *Daten vor Edition*

Frau C., die 1970 eingeschult wurde, hat nach einer sechsjährigen Grundschulzeit von August 1976 bis Juli 1981 die Realschule besucht und diese nach der zehnten Klasse mit dem qualifizierenden Hauptschulabschluss beendet. Nach dem Schulabschluss nimmt Frau C. eine Erwerbstätigkeit als Serviererin auf und übt bis Juni 1985 verschiedene Haupterwerbs- und Nebentätigkeiten aus. Ab August 1984 bis Juli 1985 holt sie den Realschulabschluss nach. Von August 1986 bis August 1988 absolviert sie als erste Ausbildung die Fachoberschule für Sozialpädagogik, da sie ‚weiter will‘ und beendet diese mit einem Berufsfachschulabschluss. Im Anschluss an die Fachoberschule beginnt Frau C. ein Architekturstudium an einer Hochschule oder Universität und beendet dieses im Juli 1990 mit dem Abschluss ‚Vordiplom‘. Ab Oktober 1990 studiert Frau C. Geschichte, Kulturwissenschaften und Kunstgeschichte an einer Hochschule oder Universität. 1996 beendet sie das Studium ohne Abschluss. In dem nur durch eine Lückenabfrage definierten Zeitraum zwischen dem Ende des nachgeholtten Schulabschlusses im Juli 1985 und dem Beginn der Ausbildung an der Fachoberschule für Sozialpädagogik im August 1986 hat Frau C. ‚etwas Anderes gemacht‘.

*Der Besuch einer Fachoberschule für Sozialpädagogik erscheint als Ausbildungsphase fragwürdig, ebenso der Abschluss derselben mit einem Berufsfachschulabschluss. Obwohl Frau C. als höchsten allgemeinbildenden Schulabschluss nur den Realschulabschluss hat, beginnt sie nach Beendigung der Fachoberschule für Sozialpädagogik mit einem Berufsfachschulabschluss ein Architekturstudium an einer Hochschule oder Universität und nimmt anschließend ein weiteres Hochschulstudium auf.*

### *Korrektur durch Edition*

Beim Abhören des Interviewmitschnittes wird festgestellt, dass die Benennung der Ausbildungsstätte als Berufsfachschule und des Abschlusses als Berufsfachschulabschluss keine Angaben von Frau C. sind, sondern auf Interpretation bzw. eigenmächtiger Eingabe durch den Interviewer beruhen. Frau C. hat als Ausbildungsstätte die Fachoberschule für Sozialpädagogik genannt und dort keinen beruflichen Abschluss erworben, sondern das Fachabitur. Die Ausbildung wird gestrichen und für den Zeitraum August 1986 bis August 1988 wird ein zweiter nachgeholtter Schulabschluss für das Fachabitur eingefügt. Aus der Tonbandaufzeichnung geht auch hervor, dass Frau C. das Architekturstudium begann, weil sie ‚das Vordiplom brauchte, um an die Uni zu gehen‘. Sie hat als Ausbildungsstätte die ‚Hochschule Bremen‘ angegeben, die vom Interviewer, der keine weiteren Möglichkeiten nannte, als Hochschule/Universität aufgenommen wurde. Bei der betreffenden Hochschule in Bremen handelt es sich aber um eine Fachhochschule, so dass die Ausbildungsstätte entsprechend korrigiert wird. Dieses Studium hat Frau C. ohne Abschluss beendet, da es sich beim Vordiplom nicht um einen Ausbildungsabschluss im Sinne der Edition handelt. Mit dem an der Fachoberschule erworbenen Schulabschluss hatte Frau C. nur die Zugangsberechtigung für die Fachhochschule und bestimmte Universitätsstudiengänge. Das an der Fachhochschule erworbene Vordiplom ermöglichte Frau C. den Zugang zur Universität, an der sie im Anschluss an das Architekturstudium Geschichte, Kulturwissenschaften und Kunstgeschichte studiert.

An die Betrachtung einzelner Fälle schließt sich unmittelbar die Frage an, inwieweit sich Merkmalsverteilungen und Zusammenhänge in der gesamten Stichprobe durch ein solches Editionsverfahren verändern. Das Ausmaß dieser Veränderungen variiert natürlich je nach Variable bzw. inhaltlicher Problemstellung. Im Folgenden sollen solche Veränderungen an einem für Fragen der Bildungs- und Arbeitsmarktforschung relevanten Beispiel dargestellt werden.

### *3.3 Uni- und bivariate Verteilungen in der Gesamtstichprobe: das Beispiel ‚Ausbildung und berufliche Stellung‘*

Die folgende Übersicht zeigt für die Variable ‚Berufliche Stellung‘ Veränderungen im Laufe der Edition. Die Informationen, die zu diesen Veränderungen geführt haben, stammen aus verschiedenen Quellen (insbes. zusätzliche Angaben des Befragten, Ergebnisse der Nachrecherche oder Plausibilitätsüberlegungen). Die folgenden Auswertungen beziehen sich auf eine Teilstichprobe des Datensatzes, nämlich jene 85% der Personen, deren Daten zum Stichtag 1.11.2001 einzelfallweise ediert waren. Vergleich ist der Datenstand vor Beginn der Edition. In der folgenden Tabelle werden besonders häufige Kategorien der Variablen ‚Berufliche Stellung‘ (Arbeiter und Angestellte) präsentiert:

Tab. 2: Veränderung von Angaben der beruflichen Stellung im Zuge der Edition (inklusive Veränderungen in der Episodenstruktur)

Berufliche Stellung	Vor Edition			Nach Edition		
	Fälle (in % aller Episoden)	davon Anteil mit abgeschl. Berufsausbildung	davon Anteil mit abgeschl. (Fach-) Hochschulstudium	Fälle (in % aller Episoden)	davon Anteil mit abgeschl. Berufsausbildung	davon Anteil mit abgeschl. (Fach-) Hochschulstudium
Ungelernte Arbeiter	353 (6%)	37%	3%	360 (6%)	41%	3%
Angelernte Arbeiter	557 (9%)	59%	1%	551 (9%)	62%	1%
Facharbeiter	1231 (20%)	85%	0%	1342 (21%)	96%	0%
Vorarbeiter	91 (2%)	88%	0%	98 (2%)	93%	0%
Angestellte mit einfacher Tätigkeit	498 (8%)	52%	2%	486 (8%)	63%	3%
Angestellte mit qualifizierter Tätigkeit	2213 (37%)	70%	5%	2252 (35%)	90%	3%
Angestellte mit hochqualifizierter Tätigkeit	521 (9%)	41%	45%	561 (9%)	42%	53%
Angestellte mit höheren Leitungsfunktionen	94 (2%)	57%	26%	61 (1%)	53%	41%
...						
N (Erwerbsepisoden mit benennbarer beruflicher Stellung)	6063 (100%)	64%	9%	6481 (100%)	76%	9%

Westdeutsche Lebensverlaufsstudie 64/71, eigene Berechnungen.

Die Verteilungen von Bildung und beruflicher Stellung beziehen sich auf Erwerbsepisoden; diese erlauben keinen unmittelbaren Rückschluss auf die Verteilung nach Personen.

Im Vergleich der beiden Datensätze unterschiedlichen Editionsstandes lassen sich folgende Unterschiede erkennen:

- Insgesamt zeigen sich nur geringe Verschiebungen in der relativen Besetzung der Positionskategorien (eindimensionale Verteilung/Prozentwerte). Deutlich stärker verändert sich die Bildungsverteilung.
- In einzelnen Kategorien gibt es allerdings auch bei den Positionen zum Teil größere Veränderungen (so nimmt die Anzahl der Episoden, die höheren Leitungsfunktionen zuzurechnen sind, deutlich ab).
- Die Veränderung der Qualifikationsverteilung betrifft nicht alle Positionskategorien gleichermaßen. Somit gibt es teilweise beträchtliche Abweichungen in den Qualifikationsverteilungen innerhalb dieser Positionen; sie liegen in Größenordnungen zwischen 10 und 20 Prozentpunkten, etwa beim Anteil abgeschlossener Ausbildungen bei Erwerbsepisoden von Facharbeitern (96% statt 85%) oder qualifizierten Angestellten (90% statt 70%). Ähnliches gilt für den Anteil des Hochschulstudium bei höheren Angestellten.

- Somit dürften sich auch Unterschiede für Folgeanalysen zum Zusammenhang von beruflicher Stellung und Ausbildung ergeben, etwa wenn der Anteil der - dann im Einzelnen als solche zu definierenden - (in)adäquat Beschäftigten numerisch ausgewiesen wird. Die genannte überproportionale Zunahme des Anteils von Qualifizierten in ausgewählten Positionen (etwa Hochschulabsolventen in höheren Leitungspositionen) bedeutet auch, dass Angestellte ohne Hochschulstudium tendenziell aus dieser höheren Kategorie herausfallen - insbesondere deshalb, weil es aufgrund der Tätigkeitsbeschreibungen unplausibel ist, dass es sich tatsächlich um höhere Leitungspositionen handelte.

Dies führt zwar in der Gesamtverteilung zu keiner qualitativen Veränderung des klaren (und i.d.R. monotonen) Zusammenhangs zwischen Position und Qualifikationsniveau. Offensichtlich verändert sich aber seine Stärke, und er wird eher deutlicher. Definiert man adäquate Ausbildungsniveaus wie folgt: ‚keine abgeschlossene Ausbildung‘ für un- und angelernte Arbeiter sowie einfache Angestellte; ‚abgeschlossene Ausbildung‘ für Fach- und Vorarbeiter sowie qualifizierte Angestellte; ‚abgeschlossenes (Fach-)Hochschulstudium‘ für hochqualifizierte Angestellte; ‚abgeschlossene Berufsausbildung oder Studium‘ für Angestellte mit höheren Leitungsfunktionen, so erhält man vor Edition einen Anteil von 66% (status-)adäquater Erwerbsepisoden und nach Edition einen Anteil von 76%. Unterschiede dieser Größenordnung könnten sich gerade bei der Gegenüberstellung mit Referenzdaten inhaltlich auswirken, etwa bei der Frage, ob es im Vergleich mit älteren Geburtsjahrgängen einen historischen Trend im Ausmaß ausbildungsadäquater Beschäftigung gibt oder nicht.

In den Vergleich des Datensatzes zu zwei Zeitpunkten gehen mehrere Veränderungsdimensionen ein. Im Zuge der Edition wurden nicht nur Angaben in allen einzelnen relevanten Variablen, insbesondere berufliche Stellung und Ausbildungsniveau, verändert, sondern auch die Struktur der Ausbildungs- und Erwerbsgeschichten (Episodenanzahl). Die folgende Tabelle 3 zeigt numerische Veränderungen in der Episodenanzahl auf Einzelfallebene im Laufe der Edition. Bei diesen Episoden handelt es sich nicht in jedem Fall um ganze Ausbildungen oder Erwerbsverhältnisse; vielmehr handelt es sich häufig um Teilabschnitte, in denen sich bestimmte Merkmale verändert haben. Tendenziell zeigt sich eine leichte Zunahme der Episodenanzahlen, also eine noch etwas stärkere Differenzierung der Daten.

Tab. 3: Veränderung der Episodenanzahl (in den Modulen Ausbildung und Erwerbstätigkeit) im Zuge der Edition

	Modul Ausbildung	Modul Erwerbstätigkeit
Reduktion um 2 und mehr	1 %	2 %
Reduktion um 1	5 %	6 %
Konstant	80 %	77 %
Erweiterung um 1	11 %	12 %
Erweiterung um 2 und mehr	3 %	3%
<i>N (Personen)</i>	<i>2397 (100 %)</i>	<i>2369 (100 %)</i>

Wird bei den inhaltlichen Vergleichen jeweils die für den betreffenden Editionsstand geltende Struktur der Daten zugrunde gelegt, können die allein dadurch bedingten Unterschiede potentiell relativ groß sein. Hingegen sind sie naturgemäß kleiner, wenn die beiden Datenstände nicht insgesamt, sondern nur im Hinblick auf spezifische Veränderungen verglichen werden.



Deshalb im Folgenden (Tabelle 4 im Anhang) zusätzlich als Beispiel die Gegenüberstellung bei weitgehend identischen Episoden. Berücksichtigt werden hier also nur Veränderungen in einer Variable (berufliche Stellung), die jeweils in ‚ursprünglicher‘ und ‚korrigierter‘ Form vorhanden ist<sup>4</sup>, wobei aber die Episodenstruktur die gleiche ist. Die Unterschiede zwischen den beiden Datenbasen sind nun tatsächlich geringer, allerdings ändert sich wenig am Muster dieser Unterschiede. Wohlgedacht handelt es sich hier nicht um einen globalen Vergleich, der ergebnisrelevante Veränderungen der Daten insgesamt zeigt, sondern um ein Beispiel für eine bedingte Veränderung der Variablen, die im Hinblick auf ein spezifisches Forschungsinteresse (‚Verkodungsqualität‘) relevant sein kann.

Wie beim obigen Globalvergleich gezeigt, gibt es aber auch bei den eindimensionalen Verteilungen (bzw. darauf aufbauender einfacher Maßzahlen) insgesamt eher geringe Veränderungen zwischen den verschiedenen Editionsstufen. Hier fallen einzelne Abweichungen entweder nicht ins Gewicht oder sie heben sich tendenziell auf. Bei einzelnen Kategorien sowie bestimmten Zusammenhängen kann die Abweichung natürlich größer sein. Extrem große Unterschiede können sich auf der Ebene des Einzelfalles ergeben, wenn etwa einzelne Episoden vollständig gelöscht werden.

### *3.4 Qualitätsverbesserung durch die Edition*

Neben Konsistenzprüfungen im engeren Sinn leistet die Edition eine ganze Reihe weiterer Beiträge zur Verbesserung der Datenqualität. Hierzu zählen das Schließen von Lücken in den Angaben, sofern dies möglich ist (etwa durch Nachbefragung), die Korrektur von Systemfehlern in der Ausgangserhebung sowie die Präzision bzw. Entdifferenzierung von Kategorien.

Die Kategorisierungen nach der Edition sind zum Teil erheblich differenzierter als vorher. Genau dies aber kann ‚Kategorienfehler‘, wie sie oben angesprochen wurden, verhindern. Trifft ein veraltetes Kategoriensystem auf eine veränderte soziale Realität, so wird sich dies in der Form von Zuordnungsproblemen ausdrücken. Die Edition hat die Möglichkeit, eben diese Probleme sichtbar zu machen, darauf zu reagieren und adäquatere Kategorien zu finden. Dabei bleibt diese Eingriffsmöglichkeit nicht wie oft üblich auf Projektphasen beschränkt, die der Haupterhebung vorgelagert sind (Vorstudien, Pre-Tests). Eine stillschweigende, verzerrende Anpassung der Antworten an ein unflexibles Kategoriensystem wird so verhindert.

So gab es bei der Variable Ausbildungsabschluss vor Beginn der Endedition 539 mal die Angabe ‚Anderer Ausbildungsabschluss‘ (=18,9% der aktuell gültigen Angaben), nach Edition (bei denselben Personen) nur noch 14 mal (0,4%). Dies bedeutet zum einen, dass ein Teil der zunächst unspezifischen Angaben vorhandenen Kategorien zugeordnet werden konnte. Zum anderen wurden neue (präzise) Kategorien für Ausbildungsabschlüsse geschaffen, die bei der Erhebung offenbar nur unzureichend berücksichtigt worden waren (u.a. verschiedene schulische Abschlüsse, Abschlüsse im Ausland, Berufsakademien). Das Kategoriensystem wurde somit auf Grund empirischer Beobachtungen weiterentwickelt.

Zusammenfassend lassen sich zahlreiche Belege dafür angeben, dass die Edition – allein durch die gezielten Nachbefragungen – die Datenbasis deutlich korrigierte und verbesserte. Eine bisher noch nicht eingesetzte Möglichkeit, dies auch zu quantifizieren, wäre die systematische

---

<sup>4</sup> Die Verteilungen werden jeweils auf Basis der Episoden mit benennbarer beruflicher Stellung ausgewiesen; vor Edition ist diese Anzahl etwas geringer, da im Prozess der Edition u.a. einigen Fällen mit ‚unbekannter‘ Stellung Werte zugeordnet werden konnten.

Gegenüberstellung von Daten unterschiedlichen Editionsstandes mit als valide angesehenen externen Informationen<sup>5</sup>.

Gerade in der Edition zeigen sich auch an verschiedenen Stellen Verbesserungsmöglichkeiten für zukünftige Datenerhebungen. So sollten für inhaltlich wichtige, aber eher abstrakte Konzeptionen und Skalen (wie die berufliche Stellungen) gezielt (weitere) Zusatzinformationen (Tätigkeitsmerkmale) erhoben werden, die eine spätere Plausibilitätsprüfung erleichtern.

#### 4. Fazit: Effekte und Effizienz der Edition

Lohnt sich der Aufwand einer solchen Datenedition? Mit Sicherheit ergibt sich eine deutliche Verbesserung der Datenqualität, nicht nur im Sinne der Praktikabilität, sondern auch der ‚Wahrheitsähnlichkeit‘. Ob der Ressourceneinsatz gerechtfertigt ist, bleibt natürlich eine Wertentscheidung, die sich darauf gründet, was den jeweiligen Forschern x % an Fehlerreduktion wert sind und wie detailliert soziale Prozesse abgebildet werden sollen.

Für die Systematik des Forschungsprozesses bleibt festzuhalten, dass qualitative Verbesserungen der Datenbasis auch und gerade in einer Phase möglich sind, in der die eigentliche Datenerhebung bereits abgeschlossen ist. Dies trägt dazu bei, den Gesamtfehler zu minimieren, solange die Konzentration auf die Edition nicht zu Lasten anderer Teilschritte geht.

Wie groß sind die substantiellen Veränderungen infolge der Edition? Beim Vergleich einfacher Durchschnitte von Variablen zeigen sich eher geringe Veränderungen über die Editionsschritte hinweg als bei einzelnen Kategorien. Man könnte nun argumentieren, dass eine an einfachen Mittelwerten orientierte Analyse der Daten offensichtlich der geringsten ‚Verfälschung‘ unterliege und deshalb vorzuziehen sei. Dies ist aber nur ein Scheinargument, denn dann bräuchte man dann den Aufwand der Datenedition nicht betreiben. Außerdem unterliegen gerade auch Mittelwerte der Stichprobenselektivität bei der Datenerhebung. Nach allem, was wir wissen, hat die Edition die Datenqualität tatsächlich deutlich verbessert und zu erheblich realitätsgerechteren Aussagen führt. Derart aufwendig bearbeitete Daten *sollten* daher auch differenziert ausgewertet werden, sei es mit komplexeren quantitativen Verfahren, sei es auch stärker qualitativ bis hin zu individuellen Fallanalysen. Bilden auf Basis der Lebensverlaufsdaten gewonnene Ergebnisse die reale Struktur von Bildungs- und Erwerbsverläufen zu einfach ab, dürfte dies eher auf das jeweilige Auswertungsverfahren oder die Art der Darstellung als auf die Daten selbst zurückgehen.

Auf jeden Fall sollte aber der Editionsprozess stärker als bisher in die inhaltlichen Analysen einbezogen werden, was im Gegensatz zur Vorstellung eines ‚in jeder Hinsicht korrekten Datensatzes‘ am Ende einer ergebnisneutralen Edition steht. Prinzipiell ist für jede inhaltliche Fragestellung zu fragen, ob und inwieweit sich die Edition der Daten auf den Erklärungsgegenstand (Verteilungen, Zusammenhänge) ausgewirkt hat bzw. haben könnte. Hierfür steht die umfangreiche Dokumentation der Edition zur Verfügung, was insbesondere für externe Datennutzer wichtig sein dürfte.

---

<sup>5</sup> Die Daten der jüngsten Lebensverlaufsstudie erlauben zum Teil die Verknüpfung mit prozessproduzierten amtlichen Längsschnittdaten (Individualinformationen aus der Beschäftigtenstatistik; mit Zustimmung der Befragten). Der Aufbau eines solchen Analysedatensatzes ist gegenwärtig in Arbeit, u.a. als Teil eines Dissertationsprojektes. Während Vergleiche einfacher Querschnittsverteilungen auch mit vorhandenen Daten relativ unproblematisch möglich sind (Hillmert/Kröhnert 2001), sind individuelle Längsschnittvergleiche auf diese zusätzliche Datenquelle angewiesen.

Diese Berücksichtigung der Edition ist besonders dann wichtig, wenn der Forschungsgegenstand zumindest teilweise mit Verfahren und Rationalitäten übereinstimmt, die bereits die Phase der Edition bestimmt haben, d.h. wenn interessierende Zusammenhänge auch als Plausibilitätskriterium gedient haben.

Für solche Zusammenhänge wiederum dürfte die Bestimmung absoluter Niveaus (etwa der Anteil adäquat Beschäftigter) besonders sensibel sein. Trendanalysen oder Gruppenvergleiche scheinen robuster zu sein - allerdings nur dann, wenn das Editionsverfahren konstant bleibt. Insofern ist auch die Konsistenz der Editionsprozesse über verschiedene Teilstudien hinweg zu berücksichtigen. Ideal wäre eine parallele Konzeption und Entwicklung von Editions- und Auswertungsprogramm - dies scheint allerdings in arbeitsteiligen Großprojekten nur schwer möglich.

Außerdem zeigt es sich einmal mehr, dass die statistischen Ergebnisse der empirischen Sozialforschung nicht ‚bis in die dritte Nachkommastelle‘ inhaltlich interpretiert werden sollten, da es - neben der sozialen Realität - auch im gesamten Forschungsprozess eine ganze Reihe von Faktoren gibt, die sie beeinflussen.

## Literatur

- Beck, Ulrich/Beck-Gernsheim, Elisabeth 1994: Individualisierung in modernen Gesellschaften – Perspektiven und Kontroversen einer subjektorientierten Soziologie. In: dies. (Hg.): Riskante Freiheiten. Individualisierung in modernen Gesellschaften. Frankfurt/M.: Suhrkamp.
- Brückner, Hannah 1995: Surveys don't lie, people do? An analysis of data quality in a retrospective life course study. Materialien aus der Bildungsforschung Nr. 50. Berlin: Max-Planck-Institut für Bildungsforschung.
- Brückner, Erika/Hoffmeyer-Zlotnik, Jürgen/Tölke, Angelika 1983: Die Datenedition als notwendige Ergänzung der Datenerhebung bei retrospektiven Langzeitstudien. ZUMA-Nachrichten 13, S. 73-83.
- Brückner, Erika/Mayer, Karl Ulrich 1998: Collecting life history data: experiences from the German life history study. In: Giele, Janet Z./Elder, Glen H. Jr. (Hg.): Methods of life course research: Qualitative and quantitative approaches. Thousand Oaks: Sage, S. 152-181.
- 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.
- 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.
- Infas 1999: Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland. Methodenbericht zur Hauptstudie. Bonn: Institut für angewandte Sozialwissenschaft.
- Reimer, Maike 2001: Die Zuverlässigkeit des autobiographischen Gedächtnisses und die Validität retrospektiv erhobener Lebensverlaufsdaten. Kognitive und erhebungspragmatische Aspekte. Materialien aus der Bildungsforschung Nr. 71. Berlin: Max-Planck-Institut für Bildungsforschung.
- Schnell, Reiner/Hill, Paul B./Esser, Elke 1995<sup>5</sup>: Methoden der empirischen Sozialforschung. München, Wien: Oldenbourg.
- Spengemann, Petra/Hillmert, Steffen 2002: Editionshandbuch zur Studie ‚Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland‘ (unter Mitarbeit von Henriette Engelhardt, Antje Mertens und Heike Trappe; Stand 4/2002). Berlin: Max-Planck-Institut für Bildungsforschung.
- Tölke, Angelika 1982: Möglichkeiten und Grenzen einer Edition bei retrospektiven Verlaufsdaten: dargestellt an der Lebensverlaufsbefragung des Sfb 3. Arbeitspapier Nr. 90. Frankfurt/M., Mannheim: Universitäten.

## Anhang (ergänzende Übersicht)

Tab. 4: ‚Reine‘ Veränderung von Angaben der beruflichen Stellung im Zuge der Edition (ohne Veränderung auf Grund der veränderten Datenstruktur)

Berufliche Stellung	Ursprünglich			Korrigiert		
	Fälle (in % aller Episoden)	davon Anteil mit abgeschl. Berufsausbildung	davon Anteil mit abgeschl. (Fach-)Hochschulstudium	Fälle (in % aller Episoden)	davon Anteil mit abgeschl. Berufsausbildung	davon Anteil mit abgeschl. (Fach-)Hochschulstudium
Ungelernte Arbeiter	359 (6%)	41%	3%	360 (6%)	41%	3%
Angelernte Arbeiter	559 (9%)	63%	1%	551 (9%)	62%	1%
Facharbeiter	1330 (21%)	95%	0%	1342 (21%)	96%	0%
Vorarbeiter	100 (2%)	93%	0%	98 (2%)	93%	0%
Angestellte mit einfacher Tätigkeit	470 (7%)	65%	2%	486 (8%)	63%	3%
Angestellte mit qualifizierter Tätigkeit	2239 (35%)	88%	5%	2252 (35%)	90%	3%
Angestellte mit hochqualifizierter Tätigkeit	546 (9%)	47%	47%	561 (9%)	42%	53%
Angestellte mit höheren Leitungsfunktionen	94 (2%)	65%	31%	61 (1%)	53%	41%
...						
N (Erwerbsepisoden mit benennbarer beruflicher Stellung)	6376 (100%)			6481 (100%)		

Westdeutsche Lebensverlaufsstudie 64/71, eigene Berechnungen.

Datenbasis: n=6481 Erwerbsepisoden mit benennbarer beruflicher Stellung (nach Edition). Die Bildungsangaben beziehen sich in beiden Fällen auf die Endversion

In der Reihe Arbeitspapiere des Projektes Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland sind bisher folgende Titel erschienen:

2001

---

- Nr. 1 Michael Corsten/ Steffen Hillmert  
Qualifikation, Berufseinstieg und Arbeitsmarktverhalten unter Bedingungen erhöhter Konkurrenz  
Was prägt Bildungs- und Erwerbsverläufe in den achtziger und neunziger Jahren?
- Nr. 2 Steffen Hillmert  
Kohortendynamik und Konkurrenz an den zwei Schwellen des dualen Ausbildungssystems  
Übergänge zwischen Schule und Arbeitsmarkt im Kontext ökonomischen und demographischen Wandels
- Nr. 3 Marita Jacob  
Ausmaß und Strukturen von Mehrfachausbildungen  
Eine Analyse der Ausbildungswege in den achtziger und neunziger Jahren

2002

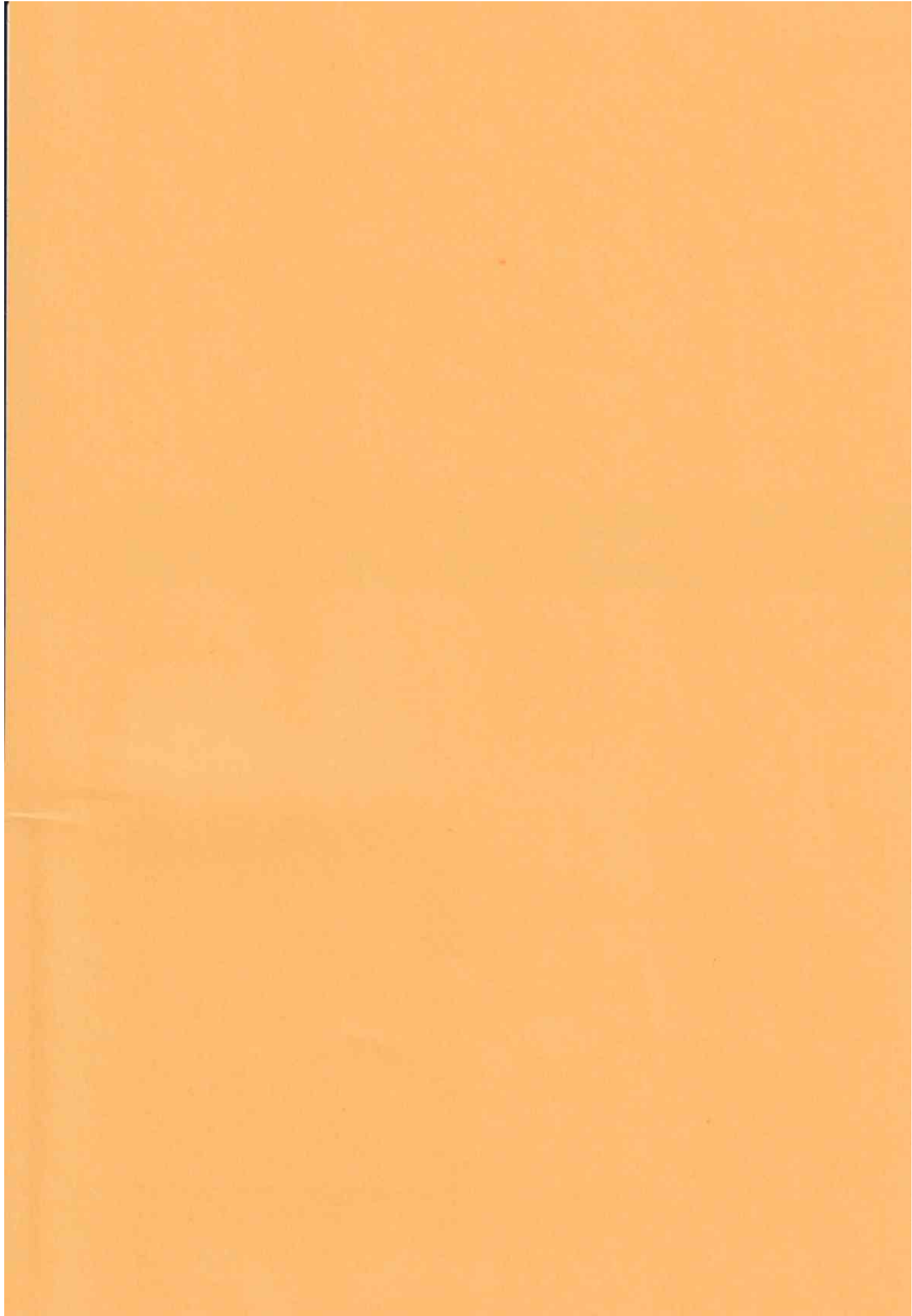
---

- Nr. 4 Steffen Hillmert  
Die Edition von Lebensverlaufsdaten  
Einzelfallprüfungen, Korrekturentscheidungen und ihre Relevanz









Arbeitsberichte

Working Papers

Max-Planck-Institut für Bildungsforschung  
Lentzeallee 94, D-14195 Berlin  
Tel. +49-30-824 06-0  
Fax +49-30-824 99 39  
[www.mpib-berlin.mpg.de](http://www.mpib-berlin.mpg.de)