

EDITION VON LEBENSVERLAUFSDATEN: ZUR RELEVANZ EINER SYSTEMATISCHEN EINZELFALLBEARBEITUNG BEI STANDARDISIERTEN BEFRAGUNGEN

STEFFEN HILLMERT

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, die Erhebung, Erfassung und Analyse der Daten bis zur theoretischen Reflexion und Darstellung der Forschungsergebnisse reichen. Auf jeder dieser Stufen können Fehler auftreten, die die Gültigkeit der Ergebnisse einschränken, und es sind jeweils 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 Datensätzen am Beispiel 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 eventuelle 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 Lebensverlaufsstudien handelt es sich bei der Edition um explizit regelgeleitete, meist mehrjährige Prozesse, in denen die Daten fallweise überprüft werden. Hierbei stellen sich mitunter auch grundsätzliche theoretische Fragen. Anhand von Beispielen aus der jüngsten Teilstudie des

Lebensverlaufsprojektes soll im Folgenden auf zentrale Aspekte der Datenedition eingegangen werden.¹

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 im Wesentlichen aus einmaligen (nur in einer Teilstudie auch mehrmaligen), standardisierten Retrospektivbefragungen von Personen aus neun ausgewählten Geburtsjahrgängen zusammen, die systematisch über ihren bisherigen Lebensverlauf Auskunft geben (vgl. Brückner/Mayer 1998). In ihrer Gesamtheit erlauben diese Lebensverlaufsdaten für Deutschland detaillierte und historisch weit zurückreichende Analysen, die fast das gesamte 20. Jahrhundert erfassen.²

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 die Lebensverläufe, 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 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.³ Die Haupterhebung der Daten erfolgte in den Jahren 1998 und 1999 in Kooperation mit dem Umfrageinstitut Infas (Infas 1999). Erfolgreich befragt werden konnte eine Stichprobe von rund 2.900 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.

2. Grundprobleme der Datenedition

Die Daten der einzelnen Lebensverlaufsstudien sind nach Lebensbereichen in verschiedenen Modulen abgelegt (zum Beispiel Schulgeschichte; Ausbildungsgeschichte; Erwerbsgeschichte; Partnerschaftsgeschichte). Dabei ist zwischen Personendatensätzen und Episodendatensätzen zu unterscheiden. Erstere enthalten im Wesentlichen Querschnittsanga-

1 Der Verfasser hat die Edition dieser Teilstudie seit Anfang 2000 geleitet. Für hilfreiche Kommentare zu diesem Beitrag danke ich Karl Ulrich Mayer, Heike Trappe sowie den anonymen Gutachtern.

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

3 In diesem Zusammenhang gefördert durch Mittel des Europäischen Sozialfonds.

ben (über die Zeit konstante Personenmerkmale), letztere nehmen für jede Person einzelne Episoden (zum Beispiel Ausbildungen oder Arbeitsverhältnisse) und deren Merkmale gesondert auf. Ziel der Lebensverlaufsstudien ist es, für jeden Lebensbereich möglichst präzise (monatsgenau) und vollständig die Sequenz einzelner lebensgeschichtlicher Abschnitte zu rekonstruieren.

Derart komplexe Daten sind fehleranfällig. Sie werden daher nach der Erhebung einem aufwändigen Editionsprozess unterzogen (vgl. Tölke 1982; Brückner et al. 1983). Hierbei werden alle Fälle im Hinblick auf fehlende oder unplausible Angaben geprüft. Um die Nachvollziehbarkeit des Editionsprozesses zu sichern, werden eine Reihe von Prinzipien zugrunde gelegt (siehe auch die Dokumentationen zur Edition in verschiedenen Teilstudien des Lebensverlaufsjprojekts). Grundsätzlich gilt, dass nach Möglichkeit alle Unplausibilitäten aufzuklären und alle verfügbaren Informationen aus den Interviews auszuschöpfen sind.

Ein großer Teil des auftretenden Korrekturbedarfs ist nicht inhaltlich gelagert: hier handelt es sich um die Folgen von Steuerungsfehlern im Instrument, um Tipp- oder Hörfehler der Interviewer und ähnliches. Diese können oft einfach und mit hoher Sicherheit korrigiert werden, wie auch im nächsten Abschnitt deutlich werden wird. In einigen Fällen stellen sich aber stärker inhaltliche und auch theoretisch bedeutsame Probleme. Auf die wichtigsten soll hier kurz eingegangen werden.

(1) Ist eine inhaltlich orientierte Edition, die über die Korrektur erhebungstechnischer Fehler hinausgeht, überhaupt sinnvoll? Mögliche Fehler in den Daten fallen in der Prüfung auf, wenn die Angaben logisch inkonsistent sind oder nicht im Rahmen des Erwartbaren liegen. Prinzipiell besteht damit die Möglichkeit, dass durch Korrekturen an den Daten eine künstliche Konsistenz in den Angaben hergestellt, also ungewöhnliche Lebensverläufe ‚geglättet‘ werden könnten. Angaben der Befragten können unplausibel erscheinen, obwohl sie korrekt sind. Die Daten würden fehlerhaft, wenn hier eine Korrektur erfolgte.

In der Edition kann zum Beispiel das Verhältnis von Ausbildung und Erwerbstätigkeit unplausibel erscheinen, d.h. erworbene Qualifikationen und Tätigkeit einer Person passen offenbar nicht zusammen. Eine einfache Korrektur der Angaben in einer der beiden Variablen wäre ein heikler Eingriff, zumal es sich bei diesem Vergleich um einen Zusammenhang handelt, der selbst Gegenstand soziologischer Forschung ist (man denke an Themen wie ‚Bildungsrenditen‘ und ‚ausbildungs(in)adäquate Beschäftigung‘). Für die Ergebnisse bestünde die Gefahr von Artefakten, wenn bereits im Prozess der Edition Ausbildungs- und Beschäftigungsniveau einander angepasst werden und dann in den späteren Analysen die hochgradige Übereinstimmung als Forschungsergebnis präsentiert wird.

Wie dieses Beispiel zeigt, ist ein Fehler aber immer in beide Richtungen - durch Korrektur und durch Unterlassen - möglich. Da die Informationen in den meisten Variablen der Studie als objektive Aussagen über die Realität (etwa formale Bildungsabschlüsse) und nicht als rein subjektive Wahrnehmungen verstanden werden, führen auch unterbleibende Korrekturen zu Fehlern in den Ergebnissen. Dies ist der Fall, wenn die in Analysen festgestellte Inkonsistenz eher auf lückenhaften Angaben als auf realen Abweichungen beruht, wenn also etwa vorhandene Ausbildungen nicht genannt oder nicht aufgenommen wurden. Ähnliches gilt, wenn sich die Befragten selbst (objektiv) falsch einer Kategorie zuordnen. In diesen Fällen würde es zu unrealistisch hoher Inkonsistenz kommen. 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 (mit Kategorien wie ‚Angestellte mit einfacher Tätigkeit‘ oder ‚Angestellte mit qualifizierter Tätigkeit‘) überdurchschnittlich betreffen. Auch wenn den Befragten für die einzelnen Kategorien jeweils Beispiele von Tätigkeiten vorliegen, dürfte ihnen diese eher abstrakte Skala weniger vertraut sein als etwa die Bezeichnung ihrer konkreten Ausbildungsabschlüsse.

Jede einzelne Editionsentscheidung stellt also eine Risikoabwägung dar: fehlerhafte Angaben sollen möglichst korrigiert, fehlerhafte Korrekturen möglichst vermieden werden.

(2) Wie lässt sich eine fehlerhafte Edition verhindern? Verschiedene Vorkehrungen gegen Fehler dieser Art sind möglich. Basis der Plausibilitätsprüfung (Wie lange dauerte eine bestimmte Ausbildung? Ist das Verhältnis von Ausbildung und Erwerbstätigkeit plausibel?) sollten nicht nur die empirischen Mittelwerte, sondern entsprechende Streuungsbereiche und auch mögliche Ausnahmen sein. Bei einer detaillierten Befragung wie der Lebensverlaufsstudie stehen zudem meistens Informationen aus anderen Variablen zur Verfügung, im vorliegenden Beispiel etwa nähere Angaben zur betreffenden Erwerbstätigkeit. Dies ist übrigens auch der Grund für die Entscheidung, die Edition *einzelfallweise* durchzuführen: ein solches Vorgehen erlaubt es, nicht nur alle Variablen innerhalb einer Episode zu berücksichtigen, sondern diese sowohl in einen zeitlichen Verlauf als auch in den Kontext verschiedener Lebensbereiche einzuordnen. Daneben sollten auch Sekundärinformationen wie institutionelle Vorgaben herangezogen werden (etwa gesetzliche Zulassungsregeln). Bei der Beurteilung der Glaubwürdigkeit einzelner Angaben können im Einzelfall Kriterien wie die Menge und Detailliertheit von Angaben, Sicherheit im Antwortverhalten oder die zu erwartende Kenntnis bestimmter Kategorisierungen berücksichtigt werden.

Grundsätzlich sollten Entscheidungen nicht ad hoc getroffen, sondern in der Editionsgruppe argumentativ abgewogen werden. Individuelle Entscheidungen der Bearbeiter sollen unterbleiben. Die Plausibilitätskontrolle dient zunächst auch nur dazu, bestimmte Fälle oder Fragen noch einmal genauer zu prüfen; aus ihr sollte keine automatische Korrekturvorschrift folgen, Eingriffe in die Daten sollten vorsichtig gehandhabt werden. Gibt es keinen hinreichenden Beleg für eine veränderte Kategorisierung, sollte die Angabe unverändert belassen werden. Bestehen Zweifel an der Richtigkeit der Angaben, lässt sich dies aber dadurch markieren, dass zusätzliche Angaben neben die Originalangaben treten (explizite Korrekturvariablen, als ungefähr kenntlich gemachte Zeitangaben und ähnliches). Auch eine eventuelle Unsicherheit der Informationen kann also ausgewiesen werden. Die Lebensverlaufsstudie 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) Schließlich könnte die Gefahr des ‚Kategorienfitting‘ bestehen: Werden die Angaben auch und gerade in der Editionsphase dem starren Erhebungsinstrument angepasst? Eine Kritik an der (quantitativen) empirischen Sozialforschung lautet, dass sie durch vorab definierte Kategoriensysteme entscheidende Tendenzen des sozialen Wandels übersieht. 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.“

Bei der Edition sollte es aber nicht darum gehen, lückenhafte Angaben um jeden Preis zu füllen, unplausible Angaben anzupassen oder Antworten auf jeden Fall in die festgelegten Kategorien einzuordnen. Für die einzelfallorientiert arbeitende Edition spielen zwar Motive der *Praktikabilität* durchaus eine Rolle - etwa das grundsätzliche Ziel, die weitgehende Vollständigkeit der Daten für Analysezwecke herzustellen: eine sehr lückenhafte Datenbasis wäre nur bedingt verwendungsfähig. Dies gilt aber nicht absolut. Durch die Einzelfallprüfung kommt die Edition zudem bewusst dem Anspruch nach möglicher kategorialer Differenzierung entgegen, und die Hauptleitlinie bleibt die der *Qualität* der Daten im Sinne der Übereinstimmung mit der objektiven Realität bzw. der tatsächlichen Wahrnehmung der Zielpersonen. So kann die Tatsache, dass Informationen beim Interview nicht richtig zugeordnet werden konnten, gerade bei der Edition auffallen; solche Probleme zeigen sich insbesondere in der starken Besetzung von Restkategorien oder

auch in Rückfragen der Interviewten. Zum anderen können im Zuge der Edition Kategorienschemata verändert werden, um letztlich möglichst alle Antworten präzise zu erfassen.

3. Edition in der Lebensverlaufsstudie 1964/71

3.1 Editionsverlauf

Gerade unter Berücksichtigung der eben genannten Anforderungen ist die Edition von detaillierten Lebensverlaufsdaten bei großen Stichproben eine umfangreiche Aufgabe, die im Projekt nur arbeitsteilig bewältigt werden kann. Mit den unmittelbaren Prüfungen und Korrekturen ist in der jüngsten Lebensverlaufsstudie insbesondere ein Team von bis zu zwölf studentischen Mitarbeiterinnen und Mitarbeitern betraut gewesen. Für die Edition wurde ein differenziertes Regelwerk konstruiert, das sukzessive angepasst wurde. Auftretende Inkonsistenzen und neue Probleme wurden in der Arbeitsgruppe diskutiert und entschieden. Die einzelnen Bearbeiter erhielten Supervision. Schließlich wurden alle Regeln und auch alle (Einzelfall-)Entscheidungen ausführlich dokumentiert (Spengemann/Hillmert 2002).

Die Datenedition hat sich hier in mehreren Phasen über einen Zeitraum von etwa dreieinhalb Jahren erstreckt. Die folgenden Absätze geben eine kurze, vereinfachte Beschreibung dieser Editionsphasen.

Eine *Voredition* ausgewählter Interviews begann bereits kurz nach Beginn der Erhebung. Der Schwerpunkt lag dabei in erster Linie auf der Beurteilung der Interviewer und erst in zweiter Linie auf der Edition der Daten. Fehler und problematische Stellen des Interviews sollten so genau wie möglich protokolliert werden. Grundlage waren die Biographieschemata (Auflistung der Lebensverlaufsepisoden in standardisierter Form) und Interviewerprotokolle (zusätzliche handschriftliche Notizen der Interviewer). Außerdem standen in den meisten Fällen Tonbandaufzeichnungen der Interviews (mit Einverständnis der Befragten) zur Verfügung.

Die *Erstedition* der 2911 Fälle begann im Januar 1999 nach einer Schulung der Editionsgruppe. Die Arbeitsmaterialien wurden erweitert: es zählten hierzu eine erste Version des Editionshandbuchs, die Variablenliste, das Computerprogramm für die Erhebung (um die Steuerung nachvollziehen zu können) sowie der Rohdatensatz in der Form von Einzelfallprotokollen (also alle Angaben eines Falles als Papierausdruck) und, falls vorhanden, die Ergebnisse der *Voredition*. Die Editeure waren gehalten, keine eigenmächtigen Entscheidungen zu treffen, sondern Probleme zuerst mit der Supervision zu besprechen. Die passenden Editionsregeln wurden dann noch einmal erläutert, oder der Fall wurde in einer Editionsitzung vorgestellt beziehungsweise der Projektleitung zur Einzelfallentscheidung

vorgelegt. Aus Gründen der inhaltlichen Fokussierung und der verfügbaren Arbeitskapazität konzentrierte sich die Edition auf die Hauptmodule (im Wesentlichen Schul- und Berufsausbildungen, Arbeitslosigkeit, Erwerbstätigkeiten und die dazwischen liegenden Lücken, sowie persönliche Angaben), dabei wurden aber auch die Informationen aus den anderen Modulen genutzt. Eine Reihe von Fällen wurde zur Überprüfung, Korrektur und Anpassung an die inzwischen deutlich veränderten Editionsregeln wiederholt ediert.

Eine wichtige Informationsmöglichkeit war anschließend die telefonische *Nachrecherche* bei den Zielpersonen. Falls nötig, wurden sie noch einmal telefonisch kontaktiert und um klärende oder ergänzende Informationen gebeten. Die telefonische Nachrecherche startete nach einer Schulung der damit betrauten Mitarbeiter im April 2000. Zu diesem Zeitpunkt waren bereits durch die Erstedition rund 700 Fälle dafür vorgesehen. Im Rahmen der Endedition kamen weitere Fälle hinzu, so dass schließlich 1.272 Fälle als Nachrecherchefälle deklariert waren. Davon wurden 110 Fälle nicht bis zum festgelegten Ende der Nachrecherche (Dezember 2001) erreicht, bei einigen wenigen Fällen schienen weitere Kontaktversuche nicht sinnvoll. Ein Teil der Fälle musste lediglich wegen der fehlenden Note eines Ausbildungsabschlusses nachbefragt werden (Steuerungsfehler im Befragungsinstrument der Haupterhebung). Alle Personen sollten zunächst telefonisch kontaktiert werden, und das Nachrecherche-Interview sollte möglichst telefonisch geführt werden. Vom Erhebungsinstitut wurde die für diesen spezifischen Zweck notwendige Kontaktinformation zur Verfügung gestellt. War keine Telefonnummer vorhanden oder diese nicht mehr korrekt, mussten vorab Nachforschungen durchgeführt werden; war dies über keines der verwendeten Telefonnummernverzeichnisse möglich, wurden die Personen zwecks Terminanfrage angeschrieben. In 141 Fällen wurden weitere Auskünfte explizit verweigert oder Terminanfragen nicht beantwortet, so dass insgesamt 1.011 Nachrecherchen erfolgreich durchgeführt werden konnten. Bei der Nachrecherche ging es in zwei Drittel der Fälle um fehlende Angaben in ein oder zwei Modulen bzw. Unklarheiten in maximal drei Episoden, es handelte sich also im Wesentlichen um gezielte Ergänzungen.

Die *End- bzw. Hauptedition* wurde zusammen mit der telefonischen Nachrecherche begonnen und für alle Fälle durchgeführt. Zusätzlich zu den bislang vorhandenen Materialien stand jetzt auch der Fragebogen in einer ausformulierten Papierversion zur Verfügung, die aus dem Programmcode des Erhebungsinstruments erstellt worden war. Das Editionshandbuch wurde speziell für die Endedition ein weiteres Mal überarbeitet, insbesondere im Abschnitt zum Modul Berufsausbildung. Aber auch nach Beginn der Endedition wurden Editionsregeln weiter angepasst und vervollständigt.

Ende 2002 wurde die Edition durch die automatisierte, modulweise Datenprüfung abgeschlossen. Daneben gab es zu jedem Zeitpunkt der Edition Einzelfallentscheidungen. Sie

wurden getrennt für die verschiedenen Phasen dokumentiert, da im Zuge der Endedition die eine oder andere Entscheidung aus einer früheren Phase rückgängig gemacht werden musste. Wurde beispielsweise eine Editionsregel mit weitreichenden Folgen neu erstellt, so mussten alle Fälle, die bis zu diesem Zeitpunkt bereits bearbeitet waren, entsprechend 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. Nach Abschluss der Edition weist die Dokumentation nur noch die Einzelfallentscheidungen auf, die erhalten blieben oder während der Endedition neu getroffen wurden.

Für die Endedition mit telefonischer Nachrecherche, die die bei weitem umfangreichste Phase der fallorientierten Edition war, wurde eine Statistik zum Zeitaufwand bei der konkreten Fallarbeit geführt. Bei einem mittleren Zeitaufwand von etwa 100 Minuten und etwa 2.900 Fällen ergibt sich daraus ein Gesamtvolumen von etwa 5.000 Arbeitsstunden. Der Arbeitsaufwand für die Edition insgesamt - inklusive Schulungen, Regelkonzeption, Interviewvorbereitung, Einzelfallentscheidungen, Vercodungen, Organisation usw. - hat ein Mehrfaches dieser Zeit beansprucht.

3.2 Vorgehen bei der Datenedition: einige Fallbeschreibungen

Im Zuge der Edition wurden also alle Fälle einzeln auf Basis verschiedener Informationsquellen wie dem Rohdatensatz aus der Befragung, Interviewerprotokollen und zumeist der Tonbandaufzeichnung des Interviews geprüft. Gerade mit Hilfe des Tonbandes konnte eine Reihe von Unplausibilitäten relativ einfach geklärt werden, wenn daraus beispielsweise eindeutig hervorging, dass der Interviewer zwei Zahlen vertauscht hatte. Daneben wurden auf dem Tonband auch zusätzliche Informationen wie Interaktionen zwischen Interviewer und Befragtem (Erläuterungen, Rückfragen etc.) festgehalten. Hinzu kam als wichtige Informationsmöglichkeit die telefonische Nachrecherche. Da diese in einem zeitlichen Abstand von bis zu drei Jahren gegenüber der Hauptbefragung erfolgte, kann die Nachrecherche ihrerseits mit einigen Konsistenz- und Erinnerungsproblemen (Reimer 2001) behaftet gewesen sein, auf die hier nicht weiter eingegangen werden kann. Schließlich wurden Informationen aus externen Quellen (etwa aus Handbüchern zur Berufskunde oder dem Internet) genutzt.

Die Konsistenzprüfung war 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 stattfand. Diese Prüfung betraf insbesondere zeitliche Unplausibilitäten (Beispiel: das Ende einer Episode lag zeitlich angeblich vor ihrem Beginn) und Lücken. In solchen Fällen wurde der Interviewer entweder unmittelbar zu einer Nach-

frage aufgefordert, oder es wurde am Ende des Interviews noch einmal gezielt nach Aktivitäten in den Lückenzeiträumen gefragt.

Ursachen für verbleibende Inkonsistenzen in den Daten können Instrumentfehler wie eine fehlerhafte Steuerung oder eine unterbliebene Präzisierung bzw. Differenzierung von Kategorien, Fehler der Interviewer, eine im Sinne der vorgegebenen Kategorien falsche Selbsteinschätzung der Befragten, Erinnerungsfehler oder bewusst falsche Angaben sowie reale Abweichungen gewesen sein.

Die folgende Übersicht (Tabelle 1) zeigt wesentliche Arten von Inkonsistenzen im Hinblick auf die modulare Struktur der Daten:

Tabelle 1: Dimensionen der Konsistenzprüfung und 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	Mehrfachausbildungen und Aufbauqualifikationen; z.B. Meisterqualifikation ohne vorherige Berufsausbildung	Reihenfolge (und Niveau) von Aktivitäten; z.B. hochqualifizierte Tätigkeit ohne Ausbildung

Die Edition hat sich dabei auf die Module konzentriert, die auch im Zentrum der Forschungen am Institut stehen, nämlich Ausbildungen und Erwerbstätigkeiten. Das Vorgehen soll zunächst durch die Darstellung zweier konkreter Fälle erläutert werden. Diese Lebensverläufe können nicht inhaltlich ‚repräsentativ‘ für die gesamte Stichprobe sein, und sie sind nicht in ihrer ganzen Komplexität dargestellt; es kommen hier aber für die Datenedition typische Anforderungen zum Ausdruck:

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 am Ende des Interviews ergänzt werden: Frau A. ist von Januar 1985 bis Juni 1986 krank 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.

Bearbeitung durch Edition und Nachrecherche

Die Edition zieht zunächst den Interviewmitschnitt zur Klärung dieser Inkonsistenzen heran und kann 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 sogenannte Lückenaktivität erfasste weitere Schulbesuch gestrichen werden. Zur Klärung der anderen offenen Fragen 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 macht 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 beginnt anschließend das Berufsvorbereitungsjahr. Die in der Lückenabfrage erhobene Krankheit war lt. Frau A. keine dauerhafte Phase.

Fall B. (Jahrgang 1964)

Daten vor Edition

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. Andererseits ist Herr B. nach Angaben aus dem Interviewerprotokoll aber in den Jahren 1983/84, 1984/85 und 1985/86 jeweils in den Monaten Dezember bis April arbeitslos. Für den Zeitraum zwischen der im April 1983 beendeten Maurerausbildung und April 1986 gibt er in der Lückenabfrage drei Erwerbstätigkeitszeiträume an: von April bis Dezember 1983, von April bis Dezember 1984 und von April bis Dezember 1985. 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.

Aufgabenstellung für die Edition

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 also aus mehreren von Arbeitslosigkeit unterbrochenen Abschnitten oder handelt es sich um unterschiedliche Erwerbsverhältnisse? Die Zuordnung der Qualifikation zum Sprengmeister als Ausbildung ist zu prüfen. Die Berufsfachschule ist hier als Ausbildungsstätte ebenso in Frage zu stellen wie die Meisterqualifikation.

Bearbeitung durch Edition und Nachrecherche

Zur Klärung kann die Edition Interviewprotokoll und Tonband nutzen: Die im Interviewprotokoll festgehaltenen Phasen saisonbedingter Arbeitslosigkeit in den Wintermonaten von 1983 bis 1992 werden durch das Abhören des Tonbandes 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 ist also zwischen Mai 1983 und Mai 1992 in vier Phasen aufzuteilen: April bis Dezember 1983, April bis Dezember 1984 und April bis Dezember 1985 und April 1986 bis Mai 1992. Das Abhören des Tonbandes bestätigt auch, 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 beziehungsweise 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. Für diese Phasen können jeweils die fehlenden Informationen ergänzt werden. Mit Hilfe der berufskundlichen Informationen des Arbeitsamtes kann ferner festgestellt werden, dass es sich beim Sprengmeister um einen Anlernberuf handelt, für den ein staatlich anerkannter Lehrgang oder eine erfolgreiche Sachkundeprüfung vorzuweisen ist. Eine Meisterqualifikation im Sinne der Edition wird nicht erreicht, sondern lediglich die Berechtigung zur Ausübung bestimmter Tätigkeiten. Die Ausbildung wird als berufliche Weiterbildung erfasst.

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 Veränderun-

gen anhand von Beispielen aus den beiden für die Studie wohl wichtigsten Modulen, den Ausbildungs- und Erwerbsgeschichten, dargestellt werden.

3.3 Veränderungen von uni- und bivariaten Merkmalsverteilungen im Zuge der Edition

In die Unterschiede zwischen Merkmalsverteilungen, die auf einem unterschiedlichen Datenstand beruhen, gehen mehrere Veränderungsdimensionen ein. Im Zuge der Edition wurden nicht nur Angaben in allen einzelnen relevanten Variablen, sondern auch die Struktur der Ausbildungs- und Erwerbsgeschichten (Episodenanzahl) verändert. Bei diesen Episoden handelt es sich nicht in jedem Fall um ganze Ausbildungen oder Erwerbsverhältnisse; vielmehr sind es häufig Teilabschnitte. Auch bei gleicher Episodenstruktur können sich aber die Werte in den einzelnen Variablen verändern.

Die Tabellen im Anhang geben anhand einiger Beispiele Auskunft über Verteilungsänderungen im Zuge der Edition. Verglichen wird jeweils der Datenstand nach dem Abschluss der Einzelfalldition mit jenem vor Beginn der Edition. Tabelle 2 bezieht sich auf die Variable ‚Ausbildungsabschluss‘ im Modul Ausbildungsgeschichte, Tabelle 3 auf die Variable ‚differenzierte berufliche Stellung‘ im Modul Erwerbsgeschichte, wobei jeweils nur ausgewählte (stärker besetzte) Kategorien ausgewiesen werden. Die Werte für die Spalte ‚nach Edition‘ beziehen sich bei der Variable ‚berufliche Stellung‘ auf die im Zuge der Edition eingeführte Korrekturvariable - neben diesem Vorschlag der Edition blieben hier also die Originalangaben im Datensatz erhalten. In beiden Tabellen wird ferner unterschieden zwischen einer Betrachtung aller Episoden einerseits und nur jeweils der ersten genannten Episode andererseits. Letztere lässt sich auch als Verteilung von Personen interpretieren. Schließlich werden Veränderungen in der Episodenstruktur auf der Individualebene angegeben. Es zeigt sich, dass sich bei mehr als einem Fünftel der Befragten bereits die Anzahl der angegebenen Ausbildungs- oder Erwerbsepisoden im Zuge der Edition verändert hat.

Im Vergleich der Datensätze mit unterschiedlichem Editionsstand lassen sich für die Verteilungen der beiden Variablen folgende Unterschiede erkennen:

(1) Variablen unterscheiden sich im Ausmaß der Veränderung. Für die Variable Ausbildungsabschluss zeigen sich insgesamt eher größere, für die beruflichen Positionskategorien eher geringere Verschiebungen in der relativen Besetzung. Dies dürfte auch auf die bessere Abgrenzbarkeit der einzelnen Ausbildungskategorien zurückgehen.

(2) Nicht alle Kategorien sind gleich betroffen. Gerade bei den Ausbildungsabschlüssen geht ein Teil der Veränderungen auf die Präzisierung einzelner Kategorien zurück. So waren in der Haupterhebung als Ausbildungskategorien sowohl ‚Meister/Techniker‘ als

auch ‚Fachschulabschluss‘ vorhanden. Dies führte zu Überschneidungen. Mit der Präzisierung der Kategorie ‚Fachschulabschluss‘ (,... ohne Meister/Techniker‘) wurde dieses Problem im Zuge der Edition bereinigt. Auch bei den beruflichen Positionen gibt es bei einzelnen Kategorien zum Teil große relative Veränderungen: der Anteil der Episoden, die höheren Leitungsfunktionen zuzurechnen sind, nimmt beispielsweise deutlich ab.

(3) Grobe Kategorien (Zusammenfassungen) und einfache Maßzahlen sind offenbar prozentual weniger betroffen als einzelne Kategorien (ähnliches gilt für Mittelwerte bei anderen Variablen). Hier fallen Abweichungen in einzelnen Kategorien entweder nicht ins Gewicht, oder sie heben sich tendenziell auf. Fasst man beispielsweise alle Lehrabschlüsse zusammen, so erhält man hierfür vor Edition 68,2 und nach Edition 68,0 Prozent, für alle (Fach-)Hochschulabschlüsse 9,6 und 8,8 Prozent (jeweils erste Episode). Ähnliches gilt, wenn man bei der beruflichen Stellung alle Arbeiter (38,4 und 37,3 Prozent) oder alle Angestellten zusammenfasst (55,8 und 53,3 Prozent). Bei einzelnen Kategorien sowie bestimmten Zusammenhängen kann die Abweichung natürlich deutlich größer sein. Extrem große Unterschiede können sich auf der Ebene des Einzelfalles ergeben, wenn etwa einzelne Episoden vollständig gelöscht werden.

In einem weiteren Schritt werden nun Zusammenhänge zwischen diesen beiden Variablen analysiert: betrachtet wird das Qualifikationsniveau der Befragten in Abhängigkeit von ihrer beruflichen Stellung. Selbst dieses einfache Beispiel setzt voraus, die Längsschnittstruktur beider Module zu berücksichtigen. Sowohl das Qualifikationsniveau als auch die berufliche Stellung sind zeitveränderlich, so dass auch die Kombination beider Variablen im Zeitverlauf zu analysieren ist. Basis sind hier die Episoden der Erwerbsgeschichte, für die jeweils das in diesem Zeitraum aktuell vorhandene Qualifikationsniveau bestimmt wird; sollte sich die Qualifikation während einer Erwerbsepisode verändert haben, wird diese entsprechend geteilt. Da die beiden Module in ihrer genauen Zeitstruktur aufeinander bezogen werden, können sich nun zusätzlich Änderungen auswirken, die die Edition an den (monatsgenauen) Zeitangaben sowohl der Ausbildungs- als auch der Erwerbsepisoden vorgenommen hat; auf diese komplexen Zusammenhänge kann hier allerdings nicht näher eingegangen werden.

Tabelle 4 zeigt als Resultat aller Einflüsse die Veränderungen in der Verteilung der Ausbildungsabschlüsse nach beruflicher Stellung, wieder für alle Erwerbsepisoden und die jeweils erste genannte Erwerbsepisode. Erneut wird deutlich, dass die Veränderung nicht alle Kategorien gleichermaßen betrifft. Es gibt teilweise beträchtliche Verschiebungen in den Qualifikationsverteilungen der einzelnen beruflichen Positionen; sie können in der Größenordnung von 10 bis 20 Prozentpunkten liegen, etwa beim Anteil abgeschlossener Ausbildungen bei den qualifizierten Angestellten.

Was bedeutet dies für inhaltliche Analysen, die über die Betrachtung einzelner Kategorien hinausgehen? Definiert man adäquate Ausbildungsniveaus für die einzelnen Positionen wie folgt: ‚keine abgeschlossene Ausbildung‘ für un- und angelernte Arbeiter sowie einfache Angestellte; ‚abgeschlossene Ausbildung‘ für Fach- und Vorarbeiter sowie qualifizierte Angestellte; ‚(Fach-)Hochschulstudium‘ für hochqualifizierte Angestellte; ‚Berufsausbildung oder Studium‘ für Angestellte mit höheren Leitungsfunktionen, kann man den Anteil der (in)adäquat Beschäftigten in diesen Gruppen insgesamt grob bestimmen. Es zeigt sich, dass die Edition zwar in der Gesamtverteilung zu keiner qualitativen Veränderung des klaren (und i.d.R. monotonen) Zusammenhangs zwischen Position und Qualifikationsniveau in diesen Gruppen führt. Offensichtlich verändert sich aber die Stärke dieses Zusammenhangs, und er wird noch deutlicher. So erhält man vor Edition einen Anteil von 71,7 Prozent ausbildungs(niveau)adäquater Erwerbsverhältnisse. Nach Edition ergibt sich einen Anteil von 81,2 Prozent, wobei jeweils die erste Erwerbsepisode zugrunde gelegt wird. Die Anteilsdifferenz beträgt also knapp 10 Prozentpunkte.

Kommt man also mit unedierten Daten zu den gleichen inhaltlichen Schlussfolgerungen wie mit edierten Daten? Das hängt sicherlich von den zugrunde gelegten Hypothesen ab. Für beide Datenstände lässt sich eine einfache Aussage der Art ‚Die Mehrheit der Beschäftigten in den (ausgewählten) Arbeiter- und Angestelltenpositionen verfügt über das entsprechende Qualifikationsniveau‘ treffen. Deutlicher wird die Relevanz der Edition, wenn die Niveauunterschiede genauer betrachtet und möglichen Entwicklungen in der Grundgesamtheit gegenübergestellt werden: eine Aussage wie ‚der Anteil der adäquat Qualifizierten hat in einem (kurzen) Zeitraum um rund 10 Prozentpunkte zu- oder abgenommen‘ wäre sicherlich ein berichtenswertes Ergebnis, wenn er sich auf reale soziale Veränderungen zurückführen ließe. Die Entscheidung, ob und inwieweit eine detaillierte Datenedition durchführt wird, kann also solche Ergebnisse direkt beeinflussen.

3.4 Qualitätsverbesserung durch die Edition

Neben Konsistenzprüfungen im engeren Sinn konnte die Edition eine ganze Reihe weiterer Beiträge zur Verbesserung der Datenqualität in der vorliegenden Studie leisten. Hierzu zählen die Korrektur von Eingabefehlern und das Schließen von Lücken in den Angaben. Solche Lücken gehen nicht nur auf das individuelle Interviewer- bzw. Antwortverhalten, sondern auch auf mögliche Steuerungsfehler im Instrument der Ausgangserhebung zurück. Diese konnten so nachträglich korrigiert werden; ansonsten galt die genaue Fragebogenstruktur der Erhebung aber auch für Edition und Nachbefragung.

Schließlich hat die Edition die notwendige Ausdifferenzierung von Kategorien ermöglicht. In der Ausbildungs-Vercoding etwa wurden in einem separaten Arbeitsschritt unter anderem alle Fälle bearbeitet, die in der Variable Ausbildungsabschluss den Code der

Restkategorie und somit in einer Folgefrage offene Angaben enthielten. Ein Teil dieser Angaben konnte nachträglich einer der vorhandenen Kategorien zugeordnet werden. In vielen Fällen mussten aber aufwändige inhaltliche Recherchen angestellt werden, um die Nennungen präzise erfassen zu können. Bei der Variable Ausbildungsabschluss gab es vor Beginn der Edition 708 mal die Angabe ‚Anderer Ausbildungsabschluss‘ (= 21,0 Prozent der gültigen Angaben; vergleiche Tabelle 2), nach Edition nur noch 9 mal (0,2 Prozent). Dabei wurden neue Kategorien für Ausbildungsabschlüsse geschaffen, die bei der Erhebung nur unzureichend berücksichtigt worden waren (unter anderem verschiedene schulische Abschlüsse, Abschlüsse im Ausland, Berufsakademien). Das Kategoriensystem wurde also auf Basis der empirischen Angaben weiterentwickelt.

Die Kategorisierungen nach der Edition sind somit zum Teil erheblich differenzierter als vorher geworden. Genau dies aber kann ‚Kategorienfehler‘, wie sie in der oben zitierten Kritik angesprochen wurden, entgegenwirken. Trifft ein zu grobes oder 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, Pretests). Eine stillschweigende, verzerrende Anpassung der Antworten an ein unflexibles Kategoriensystem wird so verhindert.

Zusammenfassend lassen sich zahlreiche Belege dafür angeben, dass die Edition durch systematische Prüfung und gezielte Nachbefragung die Datenbasis ergänzen und deutlich verbessern konnte. 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 Interviewern zurückzuführen (vgl. Brückner 1995). Eine bisher noch nicht eingesetzte Möglichkeit, die Verbesserung der Datenqualität auch zu quantifizieren, wäre die systematische Gegenüberstellung von Daten unterschiedlichen Editionsstandes mit als valide angesehenen externen Informationen.⁴

4 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, unter anderem 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.

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

4. Fazit: Effekte und Effizienz der Edition

Für die eingangs geschilderte Systematik des Forschungsprozesses bleibt festzuhalten, dass durch systematische Edition 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. 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.

Lohnt sich der Aufwand einer solchen Datenedition? Mit Sicherheit verbessert sich die 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 ein bestimmtes Ausmaß an Fehlerreduktion wert ist und wie detailliert soziale Prozesse abgebildet werden sollen. In jedem Fall muss mit Validitätsproblemen gerechnet werden, wenn auf den Schritt der Edition verzichtet wird. In der Lebensverlaufsstudie wird sicherlich überdurchschnittlich viel Wert auf die Datenedition gelegt, allerdings sind die meisten standardisierten Befragungen auch weniger komplex.

Wie groß sind die substantiellen Veränderungen infolge der Edition? Beim Vergleich grober Kategorien von Merkmalen zeigen sich eher geringe Veränderungen über die Editionsschritte hinweg als bei einzelnen, detaillierteren Kategorien. Man könnte nun argumentieren, dass eine an einfachen Anteilen bzw. Mittelwerten orientierte Analyse der Daten offensichtlich der geringsten Verfälschung unterliege und deshalb vorzuziehen sei. Dies ist aber nur ein Scheinargument, denn dann bräuchte man den Aufwand der Datenedition nicht betreiben. Außerdem unterliegen gerade auch einfache Verteilungsparameter der Stichprobenselektivität bei der Datenerhebung. Wenn man davon ausgehen kann, dass die Edition die Datenqualität tatsächlich deutlich verbessert, dann *sollten* derart aufwendig erhobene und bearbeitete Daten auch differenziert ausgewertet werden, sei es mit komplexeren quantitativen Verfahren, sei es stärker qualitativ bis hin zu individuellen Fallanalysen.

Auf jeden Fall sollte aber der Editionsprozess stärker als bisher in die inhaltlichen Analysen einbezogen werden, was die Vorstellung eines ‚in jeder Hinsicht korrekten Datensatzes‘ am Ende einer ergebnisneutralen Edition etwas relativiert. Prinzipiell ist für jede inhaltliche Fragestellung zu klären, ob und inwieweit sich die Edition der Daten auf den Erklärungsgegenstand (Verteilungen, Zusammenhänge) ausgewirkt hat oder haben könnte. Hierfür steht die umfangreiche Dokumentation der Edition zur Verfügung, die insbesondere für spätere externe Datennutzer wichtig sein dürfte. Die 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, das heißt wenn interessierende Zusammenhänge auch als Plausibilitätskriterium für die Prüfung und somit als Basis für Veränderungen gedient haben.

Für solche Zusammenhänge wiederum dürfte die Bestimmung absoluter Niveaus 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 bei der Edition einmal mehr, dass die statistischen Resultate der empirischen Sozialforschung nicht übergenuau, ‚bis in die dritte Nachkommastelle‘, inhaltlich interpretiert werden können. Neben der sozialen Realität gibt es auch im gesamten Forschungsprozess eine ganze Reihe von Faktoren, die die Ergebnisse beeinflussen.

Korrespondenzadresse

Dr. Steffen Hillmert

Max-Planck-Institut für Bildungsforschung

Forschungsbereich Bildung, Arbeit und gesellschaftliche Entwicklung

Lentzeallee 94

D-14195 Berlin

E-mail: hillmert@mpib-berlin.mpg.de

Literatur

Beck, U./Beck-Gernsheim, E., 1994: Individualisierung in modernen Gesellschaften - Perspektiven und Kontroversen einer subjektorientierten Soziologie. S. 10-39 in: dies. (Hrsg.): Riskante Freiheiten. Individualisierung in modernen Gesellschaften. Frankfurt/M.: Suhrkamp.

Brückner, H., 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, E./Hoffmeyer-Zlotnik, J./Tölke, A., 1983: Die Datenedition als notwendige Ergänzung der Datenerhebung bei retrospektiven Langzeitstudien. ZUMA-Nachrichten 13: 73-83.

Brückner, E./Mayer, K. U., 1998: Collecting life history data: Experiences from the German life history study. S. 152-181 in: J. Z. Giele/G. H. Elder Jr. (Hrsg.): Methods of life course research: Qualitative and quantitative approaches. Thousand Oaks: Sage.

Corsten, M./Hillmert, S., 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, S./Kröhnert, S., 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, M., 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, R./Hill, P. B./Esser, E., 1995 (5. Aufl.): Methoden der empirischen Sozialforschung. München: Oldenbourg.

Spengemann, P./Hillmert, S., 2002: Editionshandbuch zur Studie Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland (unter Mitarbeit von H. Engelhardt, A. Mertens und H. Trappe; Stand 9/2002). Berlin: Max-Planck-Institut für Bildungsforschung.

Tölke, A., 1982: Möglichkeiten und Grenzen einer Edition bei retrospektiven Verlaufsdaten: dargestellt an der Lebensverlaufsbefragung des Sfb 3. Arbeitspapier Nr. 90. Frankfurt/M.: Universität.

Anhang

Tabelle 2: Veränderungen im Zuge der Edition: Variable *Ausbildungsabschluss* im Modul *Ausbildungsgeschichte* (ausgewählte Kategorien); Angaben in Prozent

	Vor Edition	Nach Edition
Ausbildungsabschluss (Alle Episoden)		
Praktikum/Volontariat	1,7	10,3
Gewerbliche Lehre	31,6	31,8
Kaufmännische Lehre	22,6	22,6
Berufsfachschulabschluss	3,0	6,5
Meister/Techniker	2,1	2,9
Fachschulabschluss	2,6	2,7**
Fachhochschulabschluss	3,3	4,4
Hochschulabschluss (Diplom)	5,9	3,6
Hochschulabschluss (1. Staatsexamen)	3,3*	2,1
Anderer Abschluss	21,0	0,2
Sonstige Kategorien	2,9 (4 Kat.)	12,9 (15 Kat.)
<i>N (Episoden mit Abschluss) = 100%</i>	3374	3699
Ausbildungsabschluss (Erste Episode)		
Praktikum/Volontariat	1,4	9,8
Gewerbliche Lehre	39,7	39,8
Kaufmännische Lehre	26,2	26,4
Berufsfachschulabschluss	2,9	5,3
Meister/Techniker	0,3	0,0
Fachschulabschluss	1,5	1,2**
Fachhochschulabschluss	1,8	2,4
Hochschulabschluss (Diplom)	4,8	3,4
Hochschulabschluss (1. Staatsexamen)	2,6*	2,3
Anderer Abschluss	15,9	0,1
Sonstige Kategorien	2,9 (4 Kat.)	9,3 (15 Kat.)
<i>N (Personen mit Abschluss in erster Episode) = 100%</i>	2277	2341
Veränderung der Episodenanzahl auf der Individualebene		
Reduktion um 2 und mehr	-	0,8
Reduktion um 1	-	6,1
Konstant	-	78,4
Zunahme um 1	-	11,3
Zunahme um 2 und mehr	-	3,3
<i>N (Personen) = 100%</i>		2904

* alle Staatsexamina; ** ohne Meister/Techniker.

Westdeutsche Lebensverlaufsstudie 64/71, eigene Berechnungen.

Tabelle 3: Veränderungen im Zuge der Edition: Variable Differenzierte berufliche Stellung im Modul Erwerbsgeschichte (ausgewählte Kategorien); Angaben in Prozent

	Vor Edition	Nach Edition
Diff. berufliche Stellung (Alle Episoden)		
Angestellte mit einfacher Tätigkeit	8,2	7,4
Angestellte mit qualifizierter Tätigkeit	37,4	35,9
Angestellte mit hochqualifizierter Tätigkeit	8,7	8,8
Angestellte mit höherer Leitungsfunktion	1,6	0,9
Ungelernte Arbeiter	5,6	5,2
Angelernte Arbeiter	9,2	8,4
Facharbeiter	19,4	19,8
Vorarbeiter	1,4	1,4
Sonstige Kategorien	8,5 (19 Kat.)	12,2 (28 Kat.)
<i>N (Episoden mit benennbarer beruflicher Stellung) = 100%</i>	6826	7359
Diff. berufliche Stellung (Erste Episode)		
Angestellte mit einfacher Tätigkeit	8,3	7,3
Angestellte mit qualifizierter Tätigkeit	40,4	39,5
Angestellte mit hochqualifizierter Tätigkeit	6,5	6,4
Angestellte mit höherer Leitungsfunktion	0,6	0,2
Ungelernte Arbeiter	6,0	5,6
Angelernte Arbeiter	7,3	6,3
Facharbeiter	24,5	24,6
Vorarbeiter	0,6	0,6
Sonstige Kategorien	5,8 (19 Kat.)	9,5 (28 Kat.)
<i>N (Personen mit benennbarer beruflicher Stellung in erster Episode) = 100%</i>	2628	2741
Veränderung der Episodenanzahl auf der Individualebene		
Reduktion um 2 und mehr	-	1,7
Reduktion um 1	-	5,4
Konstant	-	76,2
Zunahme um 1	-	12,7
Zunahme um 2 und mehr	-	3,8
<i>N (Personen) = 100%</i>		2904

Westdeutsche Lebensverlaufsstudie 64/71, eigene Berechnungen.

Tabelle 4: Veränderungen im Zuge der Edition: Ausbildungsverteilung in bestimmten beruflichen Stellungen (ausgewählte Kategorien); Angaben in Prozent

Ausbildungsabschluss*:	Vor Edition			Nach Edition		
	Ohne	Beruf	Studium	Ohne	Beruf	Studium
Berufliche Stellung (Alle Erwerbsepisoden)						
Angestellte mit einfacher Tätigkeit	45,4	51,9	2,6	33,3	64,3	2,4
Angestellte mit qualifizierter Tätigkeit	25,4	69,1	5,5	6,2	88,9	5,0
Angestellte mit hochqualifizierter Tätigkeit	14,0	38,4	47,6	5,9	45,0	49,1
Angestellte mit höherer Leitungsfunktion	17,1	60,4	22,5	4,5	68,2	27,3
Ungelernte Arbeiter	59,2	37,4	3,4	56,3	41,4	2,3
Angelernte Arbeiter	42,3	56,9	0,8	37,5	61,7	0,8
Facharbeiter	14,6	85,2	0,2	4,9	94,9	0,2
Vorarbeiter	11,5	88,5	0,0	7,5	92,5	0,0
...						
<i>N (Episoden mit benennbarer beruflicher Stellung)</i>			6934			7276
Berufliche Stellung (Erste Erwerbsepisode)						
Angestellte mit einfacher Tätigkeit	52,1	44,7	3,2	41,9	55,2	3,0
Angestellte mit qualifizierter Tätigkeit	25,3	69,0	5,7	6,6	88,4	5,0
Angestellte mit hochqualifizierter Tätigkeit	11,8	14,1	74,1	5,2	21,6	73,2
Angestellte mit höherer Leitungsfunktion	11,8	58,8	29,4	0,0	57,1	42,9
Ungelernte Arbeiter	73,9	22,3	3,8	71,6	26,4	2,0
Angelernte Arbeiter	51,6	47,4	1,0	45,9	51,9	2,2
Facharbeiter	12,8	87,1	0,2	4,4	95,6	0,0
Vorarbeiter	20,0	80,0	0,0	5,9	94,1	0,0
...						
<i>N (Personen mit benennbarer beruflicher Stellung in erster Episode)</i>			2624			2595

* Ohne: ohne Ausbildungsabschluss; Beruf: abgeschlossene Berufsausbildung; Studium: abgeschlossenes (Fach-)Hochschulstudium; die Angaben ergänzen sich für jede berufliche Stellung zu 100 Prozent.

Westdeutsche Lebensverlaufsstudie 64/71, eigene Berechnungen.