

Methodenverbund

„Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“

Arbeitspapier Nr.: 12

Bernhard Schimpl-Neimanns
Zentrum für Umfragen, Methoden und Analysen (ZUMA)

Filekonzept zum Mikrozensuspanel

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Deutsche
Forschungsgemeinschaft

DFG

Methodenverbund
„Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“

Arbeitspapier Nr. 12

Filekonzept zum Mikrozensuspanel

Bernhard Schimpl-Neimanns

Februar 2006

ZUMA
Quadrat B2, 1
Postfach 12 21 55
68072 Mannheim
Telefon: +49 621-1246-263
Telefax: +49 621-1246-100
E-mail: schimpl-neimanns@zuma-mannheim.de

1 Einleitung¹

Mit Abschluss des vom Bundesministerium für Bildung und Forschung (BMBF) und der Deutschen Forschungsgemeinschaft (DFG) geförderten Verbundprojektes „Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“² wird der Mikrozensus erstmals der Forschung als Scientific Use File analog zu den Mikrozensus-Querschnittsfiles zur Verfügung stehen. Selbst bei der Auswahl des nur von 1996 bis 1999 befragten Rotationsviertels des Mikrozensus liegen, insbesondere im Vergleich zu den von der Forschung selbst erhobenen Paneldaten, für Analysen sehr große Fallzahlen vor. Damit ist es möglich, ökonomische und soziale Strukturen und ihre Veränderungen in tiefer sachlicher Gliederung zu analysieren, auch für kleine Bevölkerungsgruppen verlässliche Aussagen zu machen und den sozialen und wirtschaftlichen Wandel differenzierter als bislang zu beobachten. Damit diese Potenziale von der Forschung umgesetzt werden können ist eine möglichst einfache Datennutzung nötig. Voraussetzungen hierfür sind informationsreiche und aufbereitete Datenfiles sowie eine gute Dokumentation, die zeigt, welche Daten für die Bearbeitung verschiedener Fragestellungen zur Verfügung stehen. Für Paneldaten, die in der Regel komplexer organisiert sind als Querschnittsdaten, kommt hinzu, dass Zeitvergleiche bzw. Panelauswertungen für alle Untersuchungseinheiten (z. B. Haushalte, Personen) mittels leicht handhabbarer Identifikatoren möglich sind. Eventuelle Fragen- und Variablenänderungen im Zeitverlauf sollen klar dokumentiert werden. Eine gute Dokumentation gibt auch Hinweise auf die bei Paneldaten spezifischen Aspekte der Datenqualität, damit bei Auswertungen zwischen „echter“ Merkmalsveränderung und zeitlich inkonsistenten Angaben unterschieden werden kann.

Das für die Bereitstellung des Mikrozensuspanels 1996-1999 als Scientific Use File zu entwickelnde Filekonzept soll deshalb wie bei den Mikrozensus-Grundfiles (faktisch anonymisierte 70 %-Querschnitts-Substichproben) flexible Auswertungsmöglichkeiten bieten. Hierbei geht es um Fragen der Substichprobenziehung, der Variablenauswahl, und die Anzahl der Erhebungszeitpunkte eines bzw. mehrerer möglicher Panelfiles sowie allgemein um die Verknüpfbarkeit der Einheiten einzelner Erhebungen.

Vor dem Hintergrund, dass die statistischen Ämter im Zuge der seit 1996 gewonnenen Erfahrungen die bisherigen Probleme zeitlich inkonsistenter Ordnungsnummern lösen können, der Mikrozensus ab 2005 unterjährig mit über das ganze Jahr verteilten Befragungen durchgeführt

¹ Die hier zusammengefassten Überlegungen zum Filekonzept wurden im Verbundprojekt diskutiert und geben damit den Diskussionsstand des Projekts wieder. Für konstruktive Hinweise zu einer früheren Fassung danke ich Wolf Bihler, Robert Herter-Eschweiler, Michael Konold, Ulrich Rendtel und Joachim Wackerow.

² Siehe hierzu die WWW-Seite des Projektes unter www.destatis.de/mv/mzpanel_start.htm.

wird und sich ab 2005 weitere Designänderungen hinsichtlich der Unterstichproben ergeben haben, werden für die mittel- bis längerfristige Datenweitergabe Überlegungen zu einem Filekonzept vorgestellt, welche auch die regelmäßige Aktualisierung der Daten berücksichtigen.

Im Folgenden wird zunächst das Arbeitsfile des Panelprojekts als Ausgangsbasis des künftigen Scientific Use Files dargestellt. Aufbauend auf den damit im Projekt gemachten Erfahrungen werden Modifikationen des Arbeitsfiles für eine Weitergabe des Mikrozensuspanels als Scientific Use File diskutiert. Im Hinblick auf eine zu verstetigende Bereitstellung von Mikrozensuspaneldaten wird abschließend ein mittel- und längerfristiges Filekonzept skizziert.

2 Das Arbeitsfile Mikrozensuspanel 1996-1999 des Projekts

Das im Projekt verwendete Arbeitsfile basiert auf den vom Statistischen Bundesamt zusammengeführten Querschnittserhebungen eines Rotationsviertels und umfasst den Zeitraum von 1996 bis 1999. Bei den ersten Versuchen der Zusammenführung zeigten sich Probleme aufgrund von im Zeitverlauf inkonsistenten Ordnungsnummern. Mithilfe der Merkmale Geburtsjahr und Geschlecht, die zusätzlich zu den für die Zusammenführung benutzten Identifikatoren (Bundesland, Regierungsbezirk, Auswahlbezirk, Haushalt) verwendet wurden, konnten auf Personenebene – zum Teil aber erst nach umfangreichen manuellen Korrekturen – circa 97 Prozent der Sätze verknüpft werden, wobei für rund 70 Prozent räumlich immobile Personen vollständige Verlaufsangaben von 1996 bis 1999 vorliegen. Auf Haushaltsebene konnten rund 90 Prozent der Haushalte zusammengeführt werden (Heidenreich 2002; Herter-Eschweiler 2003a: 50-53). Nicht zusammengeführt werden konnten Personen in Gemeinschaftsunterkünften (u. a. Alten- oder Studentenwohnheime), da das Geburtsjahr und das Geschlecht in diesen Einrichtungen aufgrund zu starker Klumpungen nicht zur Identifikation der zu einer Person gehörenden Querschnittsangaben geeignet sind. Das Arbeitsfile des Methodenverbundprojekts, mit dem die inhaltlichen und methodischen Fragen der Stichprobenselektivität untersucht werden, ist eine faktisch anonymisierte 60-Prozent-Substichprobe von Auswahlbezirken, wobei alle Haushalte und Personen eines ausgewählten Auswahlbezirks unabhängig vom Zusammenführungserfolg enthalten sind. Bezogen auf Privathaushalte entspricht das Arbeitsfile zirka einer 65-Prozent-Haushaltssubstichprobe. Der Stichprobenumfang beträgt rund 54.500 Haushalte und 120.000 Personen pro Erhebungszeitpunkt.

3 Weitergabe des Mikrozensuspanels 1996-1999 als Scientific Use File

Im Zuge der Projektarbeiten wurden mit dem Arbeitsfile eine Reihe von Erfahrungen gesammelt, die in die Konzeption des Scientific Use Files des Mikrozensuspanels einfließen. Dies betrifft die folgenden Aspekte: Substichprobenziehung, Panelzeitpunkte, Erhebungs- und Auswertungseinheiten, Merkmalsumfang und Datendokumentation.

3.1 Substichprobenziehung

Im Arbeitsfile sind wegen den Problemen zeitlich konsistenter Ordnungsnummern Personen in Gemeinschaftsunterkünften nicht enthalten. Unabhängig vom Zusammenführungserfolg umfasst das File jedoch alle Personen und Haushalte der für die Substichprobe ausgewählten Auswahlbezirke, d. h. auch die nicht zusammenführbaren Einheiten. Die grundsätzlichen Vorteile dieses Vorgehens bestehen einerseits darin, dass Forscher eigenständige methodische Analysen, beispielsweise zum Ausfall, vornehmen können. Andererseits bietet dieses Längsschnittfile auch die Option von Querschnittsanalysen. Damit in Verbindung steht der grundsätzliche Vorteil, dass aufgrund des im Mikrozensus realisierten Prinzips der Flächenstichprobe in den Auswahlbezirken „automatisch“ alle Populationsveränderungen erfasst werden. Die Ziehung einer Substichprobe von Auswahlbezirken bietet deshalb vielfältige Möglichkeiten in Bezug auf Längsschnittanalysen sowie hinsichtlich methodischer Fragestellungen zur räumlichen Mobilität bzw. der damit potenziell verbundenen Stichprobenselektivität.

Die Übernahme aller Einheiten ausgewählter Auswahlbezirke ist allerdings gegenüber der Ziehung einer Haushaltssubstichprobe, nach der die Scientific Use Files der Mikrozensus Querschnitterhebungen gebildet werden, mit einem größeren Klumpeneffekt und damit einhergehend einem höheren Stichprobenfehler verbunden. Da jedoch die genannten Vorteile der Substichprobenziehung von Auswahlbezirken überwiegen, spricht viel dafür, dieses Ziehungsverfahren vorläufig beizubehalten. Alternativen und eventuelle Modifikationen werden in den nächsten Abschnitten angesprochen. Zunächst sind jedoch das Verfahren der beim Arbeitsfile angewandten Substichprobenziehung und die auf Praxiserfahrungen zurückgehenden Änderungsvorschläge genauer zu beschreiben.

Für das Arbeitsfile wurden die Auswahlbezirke der Originaldaten nach den folgenden Stichprobenmerkmalen sortiert: Bundesland, Regierungsbezirk, regionale Anpassungsschicht, regionale Schicht, regionale Untergruppe, Gemeindegrößenklasse, Gebäudegrößenklasse und Auswahlbezirksnummer. Auf Grundlage der nach dieser Sortierung nummerierten Haushalte wurden in die systematische Auswahl von 60 Prozent der Auswahlbezirke jene Auswahlbezirke übernommen, deren letzten beiden Platzziffern von 40 ganzzahligen Zufallszahlen ver-

schieden waren (Herter-Eschweiler 2003b). Die Berücksichtigung des Stichprobendesigns bei der Ziehung der Substichprobe gewährleistet eine hohe Ergebnisgenauigkeit. Näherungsweise können durch die systematische Ziehung mit Zufallsstart Effekte einer geschichteten Auswahl erreicht werden. Allerdings gibt es für systematische Ziehungen keine erwartungstreue Schätzung des Stichprobenfehlers. In der Regel wird die Varianz im Vergleich zu einer geschichteten Stichprobe überschätzt (Wolter 1985: 248ff.).

In Bezug auf eine höhere Ergebnisgenauigkeit und eine bessere Varianzschätzung im Scientific Use File des Mikrozensuspanels ist die von Bihler (2005) entwickelte Alternative einer echten geschichteten Auswahl von Auswahlbezirken oder Wohnungen von entscheidender Bedeutung. Die Schichtung lehnt sich an die Sortierfolge an, die bei der Ziehung des Arbeitsfiles verwendet wurde. Zusätzlich wird aber nach der gegenwärtigen Größe der Auswahlbezirke³ geschichtet, da die bei der Grundauswahl von 1987 im früheren Bundesgebiet bzw. 1991 in den neuen Bundesländern verwendete Gebäudegrößenklasse als Richtgröße der Auswahlbezirke nicht den aktuellen Stand abbildet. Das Merkmal Gebäudegrößenklasse bleibt vergrößert als Schichtmerkmal erhalten, um zwischen den bei der Grundauswahl bereits bestehenden Auswahlbezirken einerseits und den seither hinzu gekommenen Neubaubezirken andererseits zu unterscheiden.

Neu ist ein Schichtmerkmal zur Unterscheidung von Auswahlbezirken der Unterstichprobe, das bei der Ziehung des Arbeitsfiles nicht berücksichtigt wurde. Die Auswahlsätze sind auf Regierungsbereichsebene unterschiedlich groß: 0,4, 0,6, 0,8 oder auch 1 Prozent. Die Teilnahme eines Auswahlbezirks an den Unterstichproben hängt vom Auswahlatz und einem besonderen Rotationsschema ab. Der Grund für die variablen Auswahlätze besteht in der Umsetzung von Genauigkeitsanforderungen von Eurostat⁴ beim Erhebungsprogramm der in den Mikrozensus integrierten EU-Arbeitskräftestichprobe. Die Anforderungen sind für die regionale Ebene von Regierungsbezirken festgelegt. Im Bundesdurchschnitt ergibt sich im Mittel ein Auswahlatz von 0,45 Prozent. Zur Erhöhung der Ergebnisgenauigkeit der

³ Die 10 Größenklassen beziehen sich auf die durchschnittliche Zahl der Personen in Privathaushalten im Auswahlbezirk.

⁴ Siehe Artikel 3 der Verordnung (EG) Nr. 577/98 des Rates vom 9.3.1998 zur Durchführung einer Stichprobenerhebung über Arbeitskräfte in der Gemeinschaft. URL: http://www.bundesstatistik.de///download/d/stat_ges/erwerb/085.pdf

Unterstichprobenmerkmale bei Querschnittsauswertungen sollte diese Eigenschaft bei der Stichprobenziehung berücksichtigt werden.⁵

Schließlich soll insgesamt eine höhere Ergebnisgenauigkeit angestrebt werden. Bis auf Weiteres (s. u.) wird deshalb das Mikrozensuspanel als eine geschichtete Substichprobe von Auswahlbezirken gezogen. Das Ergebnis bildet dann eine 70 %-Substichprobe von Haushalten und Personen eines Rotationsviertels (70 % * 25 %). Der Auswahlatz entspricht der Ziehungswahrscheinlichkeit faktisch anonymisierter Querschnittsdaten des Mikrozensus (Grundfile; 70 %).

Wie oben bereits erwähnt, ist die Ziehung von Auswahlbezirken mit der Übernahme aller Haushalte und Personen dieser Einheiten aufgrund des Klumpeneffekts mit einem im Vergleich zur uneingeschränkten Zufallsauswahl höheren Stichprobenfehler verbunden. Analog zum Grundfile des Mikrozensus könnte zwar der Klumpeneffekt durch die Ziehung einer Haushaltssubstichprobe reduziert werden, jedoch kommt für das Mikrozensuspanel eine Haushaltssubstichprobe hauptsächlich deshalb nicht in Frage, weil die Einheit Haushalt zeitlich nicht stabil ist. Haushalte können somit nicht bzw. nicht direkt für die Auswahl verwendet werden. Falls man dennoch eine Haushaltsauswahl vornehmen wollte, wären sehr komplexe und für die Auswertungspraxis schwer handhabbare Inklusionsregeln zum Vorgehen bei Zu- und Wegzügen sowie temporären Ausfällen sowohl für Haushalte als auch Personen zu entwickeln. Ersatzweise zu Haushalten kommt jedoch die Wohnung als Einheit der Substichprobenziehung in Frage, zumal für die Mikrozensusquerschnitte, in denen Fragen zur Wohnung gestellt werden (1998, 2002) Wohnungssubstichproben gezogen werden. Für das Mikrozensuspanel eröffnen sich mit einer Wohnungssubstichprobe Verknüpfungsmöglichkeiten zum Grundfile (siehe Abschnitt 4).

Hinsichtlich des Stichprobenplans sind Wohnungen hierarchisch zwischen dem Auswahlbezirk und den darin zu befragenden Haushalten angeordnet. Aus der Panelperspektive sind Wohnungen weitgehend zeitlich stabile Einheiten, insofern die aus einem Auswahlbezirk wegziehenden Haushalte und Personen durch die nachziehenden Einheiten ersetzt werden. Da Wohnungen aber auch geteilt oder zusammengelegt werden, können im Zeitverlauf Veränderungen auftreten. Um durch eine Wohnungssubstichprobe den Klumpeneffekt reduzieren zu können, sind jedoch eindeutige, zeitlich konsistente Wohnungsidentifikatoren unabdingbar.

⁵ Längsschnittanalysen der Unterstichprobenmerkmale sind nur eingeschränkt möglich. Aufgrund der regional unterschiedlichen Auswahlätze und aufgrund von Datenschutzregeln zur Umsetzung der faktischen Anonymität, insbesondere in Bezug auf regionale Kennzeichnungen, ist keine Information über die Inklusionswahrscheinlichkeiten für die Unterstichprobenmerkmale im Längsschnitt möglich (siehe dazu auch Abschnitt 3.4).

Wohnungsteilungen und –zusammenlegungen treten nur in begrenztem Ausmaß auf und können voraussichtlich bei Dateninspektionen der vorliegenden Rotationsviertel erkannt und dann im Längsschnittfile entsprechend gekennzeichnet werden. Der Aufwand für manuelle Prüfungen erscheint vertretbar. Von größerer Bedeutung ist, dass aufgrund zeitlich inkonsistenter Ordnungsnummern die Grundvoraussetzung zeitstabiler Wohnungsidentifikatoren bis zum Mikrozensus 2004 nicht gegeben ist. Bis dahin kann das Mikrozensuspanel deshalb nur auf Basis einer Substichprobe von Auswahlbezirken gebildet werden.

3.2 Panelzeitpunkte

Abbildung 1 zeigt, wie sich die Mikrozensusquerschnitte aus den Rotationsgruppen der einzelnen Stichproben zusammensetzen. Beispielsweise wurden 1996 aus der dritten Stichprobe die Rotationsgruppen drei und vier und aus der vierten Stichprobe die Rotationsgruppen eins und zwei befragt. Durch Kombination verschiedener Stichproben und Rotationsgruppen können also Längsschnittdaten konstruiert werden, die Zeiträume von zwei, drei oder vier Jahren abdecken (siehe Abb. 2).

Abbildung 1: Rotation im Mikrozensus

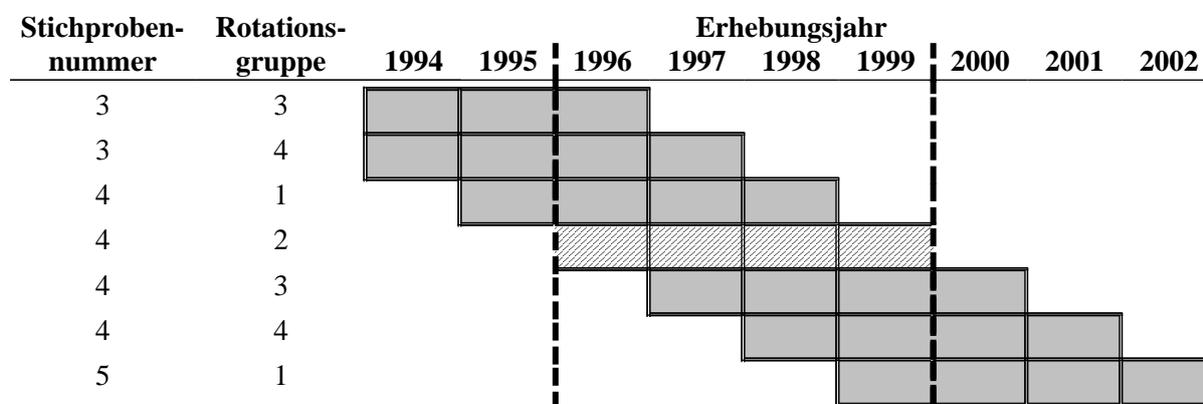


Abbildung 2: Beobachtungsdauer, Zeitraum und Zusammensetzung der Stichprobe

Beobachtungsdauer	Periode	Stichprobe / Rotationsgruppe
4 Jahre	1996 - 1999	4/2
3 Jahre	1996 - 1998	4/2, 4/1
	1997 - 1999	4/2, 4/3
2 Jahre	1996 - 1997	4/2, 3/4, 4/1
	1997 - 1998	4/2, 4/3, 4/1
	1998 - 1999	4/2, 4/3, 4/4

Der mit dem Scientific Use File des Mikrozensuspanels 1996-1999 untersuchbare Zeitraum ist in Abbildung 1 gekennzeichnet. Für substanzwissenschaftliche Verlaufsanalysen steht damit der maximale Informationsgehalt von vier Befragungen zur Verfügung. In methodischer Hinsicht bieten die vier Zeitpunkte bzw. drei Übergänge zugleich Möglichkeiten, mithilfe von statistischen Modellen "echte" Veränderungen von zeitlich inkonsistenten Angaben ("Messfehler") zu unterscheiden (Singer und Willett 2003; Vermunt 1997).

Das Mikrozensuspanel 1996-1999 enthält zwar für vier Zeitpunkte Informationen, allerdings nur für ein Rotationsviertel. Für manche Fragestellungen sind jedoch Längsschnitte über zwei oder drei Zeitpunkte ausreichend. So umfasst beispielsweise das vom Statistischen Bundesamt gebildete Mikrozensuspanel 1996-1997 knapp 500.000 Personen, wobei für rund 440.000 räumlich immobile Personen zu beiden Zeitpunkten Angaben vorliegen (Heidenreich 2002: 218). Da Längsschnittdaten über zwei und drei Erhebungszeitpunkte im Vergleich zum Mikrozensuspanel 1996-1999 deutlich höhere Fallzahlen besitzen, wäre es zu begrüßen, wenn das Statistische Bundesamt die bereits für interne Zwecke gebildeten Zweijahreslängsschnitte ebenfalls als Scientific Use File für die Forschung zur Verfügung stellen könnte.

3.3 Erhebungs- und Auswertungseinheiten

In Bezug auf die Erhebungs- und Auswertungseinheiten sollte das Mikrozensuspanel neben den bei den Querschnittdaten möglichen Auswertungen insbesondere Panelanalysen bieten. Für die Bearbeitung von Forschungsfragen auf Basis der Scientific Use Files werden die Einheiten Person, Haushalt, Familie, Lebensgemeinschaft und Wohnung herangezogen. Für Berechnungen des Stichprobenfehlers werden außer den Informationen zur Schichtung die Identifikatoren bzw. Ordnungsnummern der Auswahlbezirke als Primäreinheiten sowie Identifikatoren der Haushalte und Personen als Sekundäreinheiten benötigt.

Die Grundfiles enthalten Identifikatoren für die folgenden Einheiten: Auswahlbezirk, Wohnung (1998, 2002), Haushalt und Person. Des Weiteren liegen Identifikatoren für die Auswertungseinheiten Familie und Lebensgemeinschaft vor. Diese Einheiten werden nicht im Rahmen der Befragung gebildet, sondern erst anschließend im Zuge der Datenaufbereitung abgegrenzt (Heidenreich und Nöthen 2002).

Das Arbeitsfile des Mikrozensuspanels 1996-1999 enthält die gleichen Identifikatoren wie die Grundfiles sowie Längsschnittidentifikatoren für Auswahlbezirke, Haushalte und Personen, wobei auch plausibilisierte Haushaltsnummern vorliegen, die den Stand nach der manuellen Zusammenführung widerspiegeln. Aufgrund der bis einschließlich Mikrozensus 2004 im

Zeitverlauf inkonsistenten Ordnungsnummern liegen für die Auswertungseinheiten Wohnung, Familie und Lebensgemeinschaft allerdings bislang keine zeitlich konsistenten Identifikatoren vor.

In den als Scientific Use File des Mikrozensus verfügbaren Querschnittserhebungen wird die Wohnung als Untersuchungseinheit nur in den Jahren ausgewiesen, die das Zusatzprogramm zur Wohnsituation enthalten (1998, 2002), gleichwohl enthalten alle Jahrgänge des vollen Mikrozensus Ordnungsnummern zur Wohnung. Da im Mikrozensus nach dem Prinzip der Flächenstichprobe die aus einer Wohnung wegziehenden Haushalte bzw. Personen durch die nachziehenden Haushalte und Personen ersetzt werden, kann der Mikrozensus als Wohnungspanel betrachtet werden. Vor dem Hintergrund, dass in Deutschland keine anderen Wohnungspanels verfügbar sind, bietet das Mikrozensuspanel mit Wohnungsidentifikatoren deshalb innovative Analysemöglichkeiten. Beispielsweise können damit Mieterwechsel beschrieben werden. Zwar lassen sich bis einschließlich 2004 zeitlich konsistente Wohnungsidentifikatoren nur teilweise konstruieren, doch reichen eventuell schon die vorhandenen Daten für experimentelle Auswertungen und den Aufbau von Praxiserfahrungen aus. Es wäre daher zu überlegen, diese Teile der zeitkonsistenten Wohnungsnummern im Scientific Use File bereitzustellen.

In Bezug auf die Einheiten Familie und Lebensgemeinschaft haben erste Plausibilitäts- und Konsistenzprüfungen der Ordnungsnummern einzelner Erhebungszeitpunkte auf einige Probleme im Arbeitsfile aufmerksam gemacht. Beispielsweise zeigten sich bei im Zeitverlauf wechselnden Haushaltsbezugspersonen unplausible Änderungen der Zuordnung von Personen zu Familien. Eine Datenkorrektur setzt allerdings umfangreiche Arbeiten voraus. Diese Arbeiten würden mit der Revision aller abgeleiteten Querschnittsvariablen auf Haushaltsebene (Typisierungen und so genannte Bandsatzergänzungen) beginnen, welche die nach der manuellen Zusammenführung neu gebildeten längsschnittplausiblen Haushalte betreffen. Ergänzend sind entsprechende Änderungen auf Familien- und Lebensgemeinschaftsebene notwendig.⁶

Wohnungen, Haushalte, Familien und Lebensgemeinschaften sind Einheiten, in denen sich die Beziehungen im Zeitverlauf ändern können. (Z. B.: Nach der Heirat eines Paares, das zuvor eine nichteheliche Lebensgemeinschaft bildete.) Dennoch sind aus Sicht der Forschung Längsschnittidentifikatoren für Familien und Lebensgemeinschaften sehr wünschenswert, um entsprechende Längsschnittanalysen zu ermöglichen. Auch wenn die Konstruktion von

⁶ Zu beachten ist, dass die Änderungen zu unterschiedlichen Ergebnissen von Auswertungen der Querschnitts- und Längsschnittsdaten führen können.

Längsschnittidentifikatoren für diese Einheiten nicht einfach ist, wäre zumindest eine Korrektur der oben genannten Probleme anzustreben. Damit kann nicht nur der jeweils bei den Einzelforschern für die Konstruktion von Längsschnittidentifikatoren anfallende Aufwand reduziert werden. Darüber hinaus können nur durch die Bereitstellung einheitlicher Identifikatoren bereits im Scientific Use File die Grundlagen für replizierbare und vergleichbare Forschungsergebnisse geschaffen werden. Doch ist der für die systematische Prüfung und Korrektur notwendige Arbeitsaufwand so groß, dass dieses Vorhaben für das Mikrozensuspanel 1996-1999 mit den gegebenen Ressourcen nicht realisierbar ist. Mittel- und längerfristig sollte das Ziel von Längsschnittidentifikatoren für Familien und Lebensgemeinschaften jedoch weiter verfolgt werden.

3.4 Merkmalsumfang

Bis auf wenige Ausnahmen enthält das Arbeitsfile des Mikrozensuspanels 1996-1999 wie die Grundfiles alle erfragten Merkmale des Mikrozensus, einschließlich der Merkmale aus den Zusatz- und Ergänzungsprogrammen. Des Weiteren sind Identifikatoren bzw. Ordnungsnummern der Erhebungseinheiten sowie abgeleitete Variablen (Bandsatzergänzungen, Typisierungen) und methodische Variablen (Filter, Hochrechnungsfaktoren etc.) der Querschnittsdaten nutzbar. Zusätzlich zu den im Grundfile regulär enthaltenen Merkmalen, plausibilisierten Variablen sowie verschiedenen Längsschnittidentifikatoren (siehe Abschnitt 3.3) enthält das Arbeitsfile des Mikrozensuspanels 1996-1999 die im Methodenprojekt gebildeten Längsschnittgewichte und weitere Korrekturfaktoren zur Hochrechnung der Querschnittsangaben, die ebenfalls für Verlaufsanalysen verwendet werden können.

Dieser Merkmalsumfang ermöglicht ergänzend zu Längsschnittanalysen nicht nur Auswertungen des Rotationsviertels im Querschnitt, sondern liefert Forschern auch die Informationen, die für die selbstständige Erzeugung längsschnittplausibler Merkmale bzw. Einheiten notwendig sind (siehe Abschnitt 3.3). Da alle im Arbeitsfile enthaltenen Variablen für die Panelanalyse wichtig sind, sollten sie auch Bestandteil des Scientific Use Files sein.

Im Unterschied zum Grundfile wird das Panelfile keine Unterstichprobenmerkmale für den Zeitpunkt 1999 enthalten. Der Grund hierfür besteht darin, dass in der Unterstichprobe 1999 nur Auswahlbezirke aus Regierungsbezirken vertreten wären, in denen der volle Auswahlatz für die Unterstichprobe von 100 Prozent angewendet wurde. Die Datenschutzregeln zur Umsetzung der faktischen Anonymität beim Mikrozensus-Grundfile sehen jedoch vor, dass Re-

gierungsbezirke als regionale Einheiten nicht erkennbar sein dürfen.⁷ In diesem Zusammenhang ist außerdem anzumerken, dass die statistischen Ämter die Erhebung der Unterstichprobe so organisiert haben, dass einerseits eine möglichst gleichmäßige Belastung der betreffenden Haushalte über die vier Erhebungszeitpunkte und andererseits eine Entlastung bei der ersten und vierten Befragung erreicht wird. Längsschnittanalysen für Unterstichprobenmerkmale sind daher im Mikrozensuspanel nur eingeschränkt möglich.

Infolge von Ausfällen bzw. Weg- und Zuzügen sowie bei nicht zusammenführbaren Haushalten oder Personen liegen nicht für alle Einheiten Angaben für die Erhebungszeitpunkte 1996 bis 1999 vor. Das Mikrozensuspanel ist somit ein unbalanciertes Panel. Für die entsprechende Unterscheidung zwischen vollständigen Verlaufsangaben und Ausfällen infolge räumlicher Mobilität sowie nicht plausibel zusammenführbaren Einheiten liegen im Arbeitsfile bereits hilfreiche methodische Merkmale vor. Für Personen beschreibt die Variable „Kennzeichnung erklärter Personensätze“ (PERKL) den Zusammenführungserfolg, für Haushalte die weitgehend analog konzipierte Variable HHERKL. Daneben existieren Hilfsmerkmale zur Zusammenführung der Querschnittsangaben, mit denen auf Haushaltsebene die Bestandsentwicklungen aufgrund von räumlicher Mobilität und demografischen Veränderungen (Fortzüge, Geburten, Todesfälle) erfasst sind. Diese gute Ausgangslage lässt sich in zwei Punkten verbessern.

(1) Mithilfe von ergänzenden Informationen, die darüber Auskunft geben, ob die Zusammenführungen bereits mittels der hierfür verwendeten Ordnungsnummern oder erst nach manuellen Korrekturen erreicht werden konnte, wären Forscher ohne weiteren Aufwand in der Lage z. B. bei methodischen Analysen zur Antwortstabilität die Übergangswahrscheinlichkeiten nach dem Zusammenführungserfolg zu differenzieren.

(2) In den Variablen der einzelnen Erhebungszeitpunkte sind Panelausfälle infolge von Weg- und Zuzügen wie andere Ausfälle durch Nichtzusammenführbarkeit und fehlende Angaben („trifft nicht zu“, „keine Angabe“ etc.) mit „leer“ (bzw. „System Missing“) kodiert. Bei der Unterscheidung zwischen Panelausfällen und anderen Ausfällen incl. Item-Nonresponse muss daher jeweils auf die methodischen Variablen wie z. B. die Variable PERKL zurückgegriffen werden. Zur Erleichterung der Datenauswertungen könnten für Panelausfälle besondere Werte (z. B. -1) vergeben werden.

⁷ Siehe hierzu www.gesis.org/Dauerbeobachtung/GML/Daten/MZ/MZ-Info/Datenweitergabe.htm und insbesondere www.gesis.org/Dauerbeobachtung/GML/Daten/MZ/MZ-Info/Leitfaden.htm.

Die folgenden technischen Punkte runden die obigen, eher inhaltlich orientierten Überlegungen zum Merkmalskatalog ab. Im Arbeitsfile sind die Variablen pro Person, d. h. horizontal in Blöcken nach diesem Schema abgelegt:

- § Identifikatoren/Ordnungsnummern
- § Merkmale 1996 (AF1, ..., AF757)
- § Merkmale 1997 (BF1, ..., BF757)
- § Merkmale 1998 (CF1, ..., CF757)
- § Merkmale 1999 (DF1, ..., DF757)
- § spezielle Panelvariablen

Die erste Stelle des Variablennamens (Präfix) kennzeichnet den Erhebungszeitpunkt (A=1996, ..., D=1999). Präfixe oder Suffixe (z. B. F1_A, ..., F757_A usw.) erleichtern die Arbeit bei wiederholten Variablentransformationen. Die Umbenennung von Variablennamen mit Präfix in Variablennamen mit Suffix ist einfach. Die Benennungen können daher im Scientific Use File beibehalten werden.

Bei Panelanalysen mit gängiger Statistiksoftware (SAS, SPSS, Stata) stellt die horizontale Datenorganisation kein Problem dar. Für bestimmte statistische Modelle bzw. mit anderen Statistikprogrammen ist eventuell eine vertikale Datenorganisation erforderlich, d. h. für jede Untersuchungseinheit stehen die dazu gehörigen Variablen blockweise untereinander. Die in diesem Fall notwendige Reorganisation der Daten ist leicht umzusetzen.⁸ Da für die meisten Forscher ein File mit horizontaler Datenorganisation leichter handhabbar ist, sollte diese beim Scientific Use File beibehalten werden.

3.5 Datendokumentation

Eine gute Datendokumentation ist unabdingbar, damit die Paneldaten sachgerecht und möglichst einfach ausgewertet werden können. Dies gilt umso mehr, als die Scientific Use Files im Rahmen der Sekundärdatenanalyse genutzt werden. D. h., das bei den Datenproduzenten bzw. bei der Primärauswertung vorhandene Hintergrundwissen zu den Daten ist für die Sekundärnutzung aufzubereiten und zu vermitteln. In den obigen Ausführungen zum Filekonzept wurden bereits mit dem Schwerpunkt maschinenlesbarer Dokumentation im File einige zentrale Punkte der Datendokumentation aufgegriffen. Im Folgenden geht es auch um Dokumentationen, die i. d. R. nicht oder nur teilweise im File enthalten sind, sowie um weitere Fragen der Vermittlung von Informationen über die Daten. Die hierbei zu beschreibenden zentralen Inhalte sind (siehe u. a. ICPSR 2005; Lynn 2001; Särndal et al, 1997: 637ff.; Taylor 2000):

⁸ Mit SAS bietet sich hierfür die Prozedur „PROC TRANSPOSE“ an, SPSS stellt den Befehl „VARSTOCASES“ zur Verfügung und mit Stata kann das Kommando „RESHAPE“ verwendet werden.

- § Stichprobenplan
 - Stichprobenplan des vollen Mikrozensus
 - Substichprobenziehung des Scientific Use Files
- § Datenerhebung und Datenaufbereitung
 - Erhebungszeitraum / Berichtszeitraum
 - Befragung durch Interviewer, Selbstausfüller (Methodenmix), Feldarbeit
 - Datenbereinigung (Plausibilisierung, Imputation)
- § Verknüpfung der Querschnittsangaben (incl. manuelle Korrekturen)
- § Hochrechnung (Quer- und Längsschnitt)
 - Hochrechnungsfaktoren und GewichtungsvARIABLEN
 - Berechnung des Stichprobenfehlers
 - Unit- und Item-Nonresponse; Hinweise auf weitere systematische Fehlerquellen
- § Filestruktur
 - Identifikatoren (Quer-/Längsschnitt)
 - Quer- und Längsschnittvariablen
 - Analyseeinheiten (Auswahlbezirk, Wohnung, Haushalt, Person, Familie, Lebensgemeinschaft)
- § Variablen
 - erfragte Merkmale
 - abgeleitete Variablen und Typisierungen
 - Hinweise zur temporalen Vergleichbarkeit bzw. Änderung von Fragen im Zeitverlauf
- § Hinweise zur Auswertungspraxis mit Beispielen
 - Abgrenzung von Subpopulationen bei Quer- und Längsschnittauswertungen (Anwendung der Bevölkerungskonzepte; Umgang mit den Längsschnittidentifikatoren und Kennungen wie z. B. PERKL usw.)
 - Verwendung der Hochrechnungsfaktoren bzw. Gewichte im Quer- und Längsschnitt
 - Statistische Modelle von Übergängen im Längsschnitt und Methoden zur Analyse des evtl. Immobilitätsbias bzw. Verfahren zur Biasreduktion (Gewichtung, Pattern-Mixture Modelle, Hausman-Spezifikationstest usw.)
- § Faktische Anonymisierung, Merkmalsvergrößerungen
- § Zugangs- und Nutzungsbedingungen

Für das Arbeitsfile liegen bereits vielfältige Dokumentationen vor, und zwar sowohl maschinenlesbar (SAS- und SPSS-Files im dokumentierten Format) als auch als Manuskript. Die folgenden Vorschläge beziehen sich auf einzelne zu überarbeitende Punkte und darüber hinaus gehende Dokumentationsmaterialien:

- § Ergänzend zur Datenbereitstellung in Form von Rohdaten sowie SAS- und SPSS-Files im dokumentierten Format sollten auch Stata-Files angeboten werden, da Stata von vielen mit den Scientific Use Files des Mikrozensus arbeitenden Forschern genutzt wird und zu den in den Sozial- und Wirtschaftswissenschaften am häufigsten verwendeten Statistikprogrammen gehört (Lüttinger et al. 2003a: 130; Zühlke und Hetke 2002: 7).
- § Die meisten Variablen der Erhebungszeitpunkte 1996 bis 1999 liegen unverändert vor. In den wenigen Fällen, bei denen sich die Fragen bzw. Antwortkategorien verändert haben, sind diese Änderungen in der Datendokumentation deutlich zu kennzeichnen. Dies betrifft die Variablen Arbeitsstunden (DF145), Arbeitssuche (DF248) und beruflicher Ausbildungsabschluss (DF288-DF289).
- § Die im Panelfile verwendeten Variablennamen entsprechen denen der Querschnittsdaten. Dies erleichtert Forschern, die bereits mit den Querschnittsdaten arbeiten, die Auswertung der Paneldaten. Jedoch führt die Orientierung an den Variablennamen der Querschnittsdaten dazu, dass die 1999 geänderten Variablen zur beruflichen Bildung im Arbeitsfile eine andere Bezeichnung erhalten haben. Da diese Variablen aber substantziell weitgehend mit den früheren Erhebungen vergleichbar sind, sollten sie zur Erleichterung von Panelauswertungen entsprechend umbenannt werden (Variable DF288 in DF289, DF289 in DF290).
- § Nach Abschluss des Panelprojekts können die im Projekt erstellten Arbeitsberichte, ggf. mit Beispielprogrammen und ausführlicher Beschreibung zur Variablenkonstruktion, die oben genannten Dokumentationen ergänzen und den Einstieg in die Analyse des Mikrozensuspanels erleichtern.
- § Insbesondere in Bezug auf Erstnutzer wäre es hilfreich, Kurzfassungen der ausführlichen Datendokumentationen und nicht technische Beschreibungen anzubieten. In erster Linie kommt hierfür das WWW in Frage.
- § Damit Forscher die Paneldaten sachgerecht auswerten können, muss in der Dokumentation deutlich auf die derzeit noch ungelösten Probleme im File hingewiesen werden. Dies betrifft insbesondere den Sachverhalt, dass im Mikrozensuspanel 1996-1999 keine längsschnittplausiblen Ordnungsnummern für Familien und Lebensgemeinschaften zur Verfügung stehen und die abgeleiteten Querschnittsvariablen und

Typisierungen für Haushalte, Familien und Lebensgemeinschaften nicht ohne Weiteres für längsschnittliche Auswertungen verwendet werden können.

Ein weiterer Punkt bezieht sich nicht nur auf das Mikrozensuspanel, sondern betrifft den Mikrozensus als Ganzes. Für einige Merkmale (z. B. Arbeitsstunden und Wirtschaftszweig) werden fehlende Angaben im Zuge der Datenbereinigung ersetzt, jedoch werden diese Datenänderungen im Mikrozensus bislang nicht durch entsprechende Variablen (flags) gekennzeichnet. Für Qualitätseinschätzungen können diese Informationen allerdings sehr wichtig sein. Vor dem Hintergrund des eingangs genannten Transparenzgebots wird den statistischen Ämtern deshalb vorgeschlagen, diese Kennzeichnung künftig vorzunehmen und sich damit der beispielsweise beim Europäischen Haushaltspanel (ECHP) und dem Survey of Income and Program Participation (SIPP) verwendeten Praxis sowie entsprechenden internationalen Empfehlungen (Lynn 2001: 4; Office of Management and Budget 2001: 7-10) anzuschließen.

Zusammenfassend lässt sich sagen, dass verschiedene bereits vorliegende Manuskripte nach Überarbeitung hinsichtlich der für das Scientific Use File des Mikrozensuspanels vorgenommenen Änderungen einen guten Grundstock der Datendokumentation bilden.⁹ Auf dieser Basis und aufbauend auf den nach Projektabschluss vorliegenden Arbeitsberichten kann die Datendokumentation sukzessive weiter ausgebaut werden, so dass schon nach kurzer Zeit ein umfangreiches Angebot von Metadaten vergleichbar zu den Mikrozensus-Grundfiles¹⁰ zur Verfügung stehen wird. Hinsichtlich der spezifischen Dokumentationsanforderungen bei Paneldaten können in Anlehnung an die Dokumentationen zum British Household Panel Survey (BHPS)¹¹, der Panel Study of Income Dynamics (PSID)¹², dem Sozio-oekonomischen Panel (Haisken-DeNew und Frick 2005) und dem Survey of Income and Program Participation (SIPP)¹³ weitere Optimierungen durch edv-gestützte Datenbanksysteme wie das Mikrodaten Informationssystem (MISSY, im Aufbau; siehe Lüttinger et al. 2003b: 161) erreicht werden. Bei all diesen Arbeiten sollten möglichst die internationalen Dokumentationsstandards für sozialwissenschaftliche Daten der Data Documentation Initiative (DDI), an der unter anderem auch die statistischen Ämter in Kanada und in den USA beteiligt sind, umgesetzt werden.¹⁴

⁹ Es handelt sich um: Schlüsselverzeichnis, Systematiken (Wirtschaftszweige, Berufe (KldB, ISCO-COM) und Hauptfachrichtung) und weitere Beschreibungen zu den Hochrechnungsfaktoren.

¹⁰ Siehe dazu die vom Statistischen Bundesamt und ZUMA bereitgestellten Informationen unter www.gesis.org/Dauerbeobachtung/GML/Daten/MZ/index.htm.

¹¹ Siehe www.iser.essex.ac.uk/ulsc/bhps/doc/.

¹² Siehe <http://psidonline.isr.umich.edu/Data/>

¹³ Siehe www.sipp.census.gov/sipp/

¹⁴ Siehe www.icpsr.umich.edu/DDI/ sowie die Vorschläge zum Vorgehen bei komplexen Datenstrukturen unter www.icpsr.umich.edu/DDI/org/complex-files.pdf.

4 Überlegungen zur künftigen Bereitstellung von Mikrozensuspanels als Scientific Use File

Mit der Bereitstellung des Mikrozensuspanels 1996-1999 können erstmals substanzwissenschaftlich orientierte Längsschnittanalysen des Mikrozensus durchgeführt werden. Des Weiteren bietet das Panel neue methodische Auswertungen. Aufgrund von noch offenen Fragen¹⁵ werden diese Auswertungen in mancher Hinsicht noch einen experimentellen Status haben. Auf Grundlage der im Projekt gesammelten Erfahrungen lassen sich auch schon für die Mikrozensus bis 2004 einige Problempunkte bearbeiten. Ab dem Mikrozensus 2005 können eine Reihe der bisherigen Schwierigkeiten bei der Zusammenführung der Rotationsviertel gelöst werden. Außerdem bieten sich ab 2005 durch die Einführung der unterjährigen Erhebung und anderen Umstellungen weitere Vorteile (siehe Afentakis und Bihler 2005; Lotze und Breiholz 2002a,b).¹⁶

Vor diesem Hintergrund und ausgehend von einem Bedarf der Forschung an weiteren Panels des Mikrozensus sowie dem Ziel, die in Bezug auf Längsschnittdaten in Deutschland im internationalen Vergleich festgestellten Defizite (KVI 2001: 15-49; BMA 2001: 37) zu verringern, stellt sich deshalb die Frage, wie künftig regelmäßig und möglichst zeitnah Mikrozensuspanels konstruiert und von den statistischen Ämtern bereit gestellt werden können. Weitere Zielsetzungen bei diesen Überlegungen sind eine möglichst hohe Flexibilität der Datennutzung durch die Forschung einerseits und eine effektivere und mit geringerem Aufwand verbundene Datenproduktion durch die statistischen Ämter andererseits. Vorbilder für solche Standardlösungen sind die (Rotations-) Paneldaten des U.S.-amerikanischen Current Population Surveys (CPS)¹⁷ oder des UK Labour Force Surveys (UK LFS)¹⁸, bei denen Forscher je nach Aufgabenstellung selbstständig verschiedene Längsschnitte zusammenspielen können (CPS) bzw. Verknüpfungsmöglichkeiten zwischen Querschnitts- und Längsschnittdaten (CPS, UK LFS) gegeben sind. Hierfür müssen allerdings drei zentrale Voraussetzungen erfüllt sein.

1. Für die Verknüpfung der Erhebungs- bzw. Auswertungseinheiten einzelner Rotationsviertel ist eine einheitliche Substichprobenziehung bzw. eine gemeinsame Substichprobenba-

¹⁵ Diese Fragen hängen vor allem mit der nicht vollständigen Längsschnittverknüpfung und mit den oben angesprochenen Problemen bei Auswertungen auf der Ebene von Haushalten, Familien und Lebensgemeinschaften zusammen. Außerdem ist zu beachten, dass die im Verbundprojekt entwickelten Korrekturgewichte nur für ausgewählte Fragestellungen überprüft werden konnten.

¹⁶ Das Rotationsprinzip und der einjährige Abstand der vier Befragungen eines Auswahlbezirks bleiben aber erhalten.

¹⁷ Siehe www.unicon.com/info/cpsfaqs.html#q7 bis (...)#q12.

¹⁸ Siehe www.data-archive.ac.uk/doc/5213%5Cmrdoc%5Cpdf%5Clongitudinal.pdf.

sis unerlässlich. Ansonsten würde der aus der Zusammenführung einzelner Teilstichproben resultierende Auswahlatz als Produkt der jeweiligen Ziehungswahrscheinlichkeiten geringer als der Auswahlatz der einzelnen Teilstichproben sein.

Wie in Abschnitt 3.1 ausgeführt, kommt als gemeinsame Substichprobenbasis für die Verknüpfung der Einheiten der Rotationsviertel zu Längsschnittdaten in erster Linie eine Substichprobe auf der Basis von Wohnungen in Frage. Die Umstellung von Haushalts- auf Wohnungssubstichproben bei den Querschnittsdaten der Mikrozensus Scientific Use Files dürfte unproblematisch sein, zumal schon gegenwärtig für die Jahre, in denen die Fragen des Zusatzprogramms zur Wohnsituation gestellt werden, Wohnungssubstichproben gezogen werden.

2. Damit diese Einheiten zu einem Längsschnittdatensatz zusammengeführt werden können, müssen identische und im Zeitverlauf stabile Identifikatoren bzw. Ordnungsnummern für die Einheiten vorliegen.

Dies betrifft insbesondere die Wohnungen. Die Längsschnittplausibilität der Ordnungsnummern für Wohnungen ist bis zum Mikrozensus 2004 eingeschränkt und lässt sich voraussichtlich erst ab dem Mikrozensus 2005 in ausreichender Güte herstellen. Daher ist bis einschließlich Mikrozensus 2004 für die Längsschnittfiles eine andere Substichprobenziehung als für Querschnittsdaten notwendig.

Aus Datenschutzgründen können im Scientific Use File nur systemfreie Ordnungsnummern bereitgestellt werden. Für Längsschnittfiles sind daher Übergangsschlüssel zwischen den Ordnungsnummern des vollen Mikrozensus und zeitstabilen Identifikatoren des Scientific Use Files notwendig.

3. Die Verknüpfung von Querschnittsangaben muss datenschutzrechtlich unbedenklich sein. Infolge des bei den Unterstichproben regional variierenden Auswahlatzes können Unterstichprobenmerkmale für die vierte Befragung des Mikrozensuspanels 1996-1999 aus Datenschutzgründen nicht weitergegeben werden (siehe Abschnitte 3.1 und 3.4). Dies betrifft alle Erhebungen bis einschließlich 2004. Da ab dem Mikrozensus 2005 auch für die Unterstichprobenmerkmale der einheitliche Auswahlatz von einem Prozent gilt, ist eine Weitergabe ab 2005 unproblematisch.

Aus diesen Gründen ist für ein Konzept zur Bereitstellung aktueller Längsschnittdaten als Nachfolger des Mikrozensuspanels 1996-1999 zwischen den Erhebungsjahren bis einschließ-

lich 2004 und ab 2005 zu unterscheiden. Dieses mittelfristige Konzept sowie das mittel- bis längerfristige Konzept werden im Folgenden skizziert.

Mittelfristiges Konzept (Mikrozensus bis 2004)

In Bezug auf die bis einschließlich 2004 möglichen Längsschnitte kann das für das Mikrozensuspanel 1996-1999 dargestellte Konzept übernommen werden. Zu klären sind die Erhebungszeitpunkte und der Grad der Überlappung. Mehrere sich zeitlich überlappende Längsschnitte¹⁹ bieten zwar umfassende Analysemöglichkeiten, jedoch ist der dafür notwendige Konstruktionsaufwand als sehr hoch einzuschätzen. Außerdem können diese Überlappungen bei einer einheitlichen Stichprobenbasis für Quer- und Längsschnittdaten im mittel- bis längerfristigen Konzept ab 2005 effektiver erreicht werden.

Die Forschung ist voraussichtlich stärker an zeitnahen Paneldaten als einer Reihe älterer Daten interessiert. Daher wäre mittelfristig die Zusammenführung des dritten Rotationsviertels der fünften Mikrozensusstichprobe als nächstes, aktuelleres Mikrozensuspanel anzustreben. Dieses Längsschnittfile deckt die Zeitpunkte 2001 bis 2004 ab. Neben dem Vorteil aktueller Daten bieten sich mit der Weitergabe der Angaben zu Selbst- und Fremdauskünften (Proxy-Interviews) interessante methodische Analysen. Erfreulicherweise plant die Mikrozensusgruppe des Statistischen Bundesamtes die Erstellung des Panelfiles 2001-2004. Nach Möglichkeit sollen auch Personen in Gemeinschaftsunterkünften einbezogen werden, so dass, falls dies gelingt, das File zusätzliche Analysepotenziale zu einer in anderen Umfragen i. d. R. überhaupt nicht vertretenen Population eröffnen wird.

Mittel- bis längerfristiges Konzept (Mikrozensus ab 2005)

Abhängig vom Erreichen der oben genannten Voraussetzungen einer einheitlichen Substichprobe für Quer- und Längsschnittdaten, konsistenten Längsschnittidentifikatoren für Wohnungen und Kennungen für Wohnungstrennungen und –zusammenführungen können ab 2005 bei der Bereitstellung des Mikrozensuspanels wesentliche Verbesserungen erreicht werden. Die Verfügbarkeit einzelner Substichproben von Rotationsgruppen bzw. die Weitergabe von entsprechenden Kennungen in den Querschnittfiles ermöglicht es dann Forschern je nach Fragestellung und benötigtem Stichprobenumfang eigenständig Panels unterschiedlicher Beobachtungszeiträume zu konstruieren. Damit entfällt auch die aufwändige Erstellung je-

¹⁹ Betrachtet man nur Längsschnitte über vier Erhebungszeitpunkte, sind Zusammenführungen von Rotationsvierteln denkbar, die eine maximale zeitliche Überlappung von drei Jahren aufweisen; z. B.: 1997-2000, 1998-2001, 1999-2002, 2000-2003, 2001-2004. Wählt man jede zweite Rotationsgruppe für die Zusammenstellung von vierjährigen Längsschnitten aus, ergeben sich mit einer Überlappung von zwei Jahren die Panels 1996-1999, 1998-2001, 2000-2003, 2002-2005 usw. Siehe hierzu Abb. 1 und 2.

weils spezieller Files durch die statistischen Ämter. Zu diesen Effizienz- und Flexibilitätsge-
winnen kommen ab 2005 infolge der Umstellung des Mikrozensus auf eine kontinuierliche
Befragung und der Erhebung der Unterstichprobenmerkmale mit vollem Auswahlatz weitere
Analysepotenziale hinzu; beispielsweise die Untersuchung saisonaler Einflüsse auf dem
Arbeitsmarkt oder Informationen über die Dauer der Arbeitssuche. Sobald detaillierte Ergeb-
nisse zum Mikrozensus 2005 und 2006 vorliegen, kann die Konzeption von Längsschnittfiles
konkretisiert werden.

Es geht aber nicht nur um die Zusammenführung von Rotationsgruppen zu einem Längs-
schnittfile. Damit Nutzer in der Lage sind, mit dem bei Flächenstichproben infolge von Fort-
und Zuzügen immanenten Problem selektiver Ausfälle sachgerecht umzugehen, benötigen sie
auch Korrekturgewichte für die jeweiligen Längsschnitte. Diese Gewichte wurden für das
Arbeitsfile Mikrozensuspanel 1996-1999 unter Verwendung der Ziehungswahrscheinlichkei-
ten, teilweise auch mit geschätzten Immobilitätswahrscheinlichkeiten sowie von Eckwerten
aus verschiedenen Bevölkerungsdaten, d. h. für die Gesamtpopulation, nachträglich ermittelt
und dem File zugespielt (Rohloff 2005). Für das Mikrozensuspanel 1996-1999 wurden Ge-
wichte für drei Längsschnitte (1996-1997, 1996-1998 und 1996-1999) konstruiert. Nach dem
mittel- bis längerfristigen Konzept ab 2005 wird die Konstruktion von 2-, 3- oder 4-Jahres-
Längsschnitten mit unterschiedlichen Beobachtungszeiträumen möglich. Um auch hinsicht-
lich der Korrekturfaktoren für selektive Ausfälle eine flexible Analyse zu gewährleisten, sind
somit Gewichte für mehrere Kombinationen von Rotationsgruppen zu erstellen.

Mithilfe von Anpassungsdaten bzw. einer Dokumentation der Eckwerte aus den Bevölke-
rungsdaten und (Spezial-) Programmen könnten Nutzer die Gewichte auch eigenständig be-
rechnen und somit zur Entlastung bei der Datenproduktion beitragen. Jedoch dürfte die
Summe des dafür von einzelnen Nutzern benötigten Aufwandes die Kosten einer einmaligen
Konstruktion und zentralen Bereitstellung bei weitem übersteigen. Berücksichtigt man zu-
sätzlich, dass einige Eckwerte für Regionaleinheiten verwendet werden, die aus Datenschutz-
gründen nicht im Scientific Use File ausgewiesen werden, ist eine selbstständige Konstruktion
von Gewichten durch Forscher nur eingeschränkt möglich. Die Zusammenstellung entspre-
chender Eckwerte und die Konstruktion von Gewichtungsvariablen werden also auch ab dem
Zeitpunkt 2005 mit einem nicht vermeidbaren Aufwand verbunden sein. Es ist aber davon
auszugehen, dass bei der Konstruktion der speziell für das Scientific Use File bereitzustellen-
den Gewichtungsvariablen weitestgehend auf die bereits für die Original-Paneldaten im Sta-
tistischen Bundesamt gebildeten Variablen zurückgegriffen werden kann, so dass für das

Scientific Use File kein wesentlicher Zusatzaufwand anfällt. Insgesamt steht dem Aufwand für die Bildung des Mikrozensuspanels im Vergleich zu den derzeitigen Querschnittauswertungen ein erheblich höheres Analysepotenzial für Längsschnittauswertungen gegenüber.

Abschließend ist festzuhalten, dass das mit dem mittel- bis längerfristigen Filekonzept verfolgte Ziel, regelmäßig zeitnahe und flexibel auswertbare Längsschnittdaten des Mikrozensus bereitzustellen, ganz entscheidend von verbesserten Aufbereitungsprozessen in den statistischen Ämtern abhängig ist. Nur wenn es gelingt, künftig mit zeitlich konsistenten Ordnungsnummern der zentralen Erhebungs- und Auswertungseinheiten Person, Haushalt, Wohnung, Gebäude und Auswahlbezirk zu arbeiten, können effektivere und mit geringerem Aufwand verbundene Wohnungssubstichproben einzelner Rotationsviertel gezogen und der Forschung als Scientific Use File zur Verfügung gestellt werden.

Literatur

- Afentakis, Anja, und Wolf Bihler, 2005: Das Hochrechnungsverfahren beim unterjährigen Mikrozensus ab 2005. *Wirtschaft und Statistik*, 10: 1039-1048.
- Bihler, Wolf, 2005: Verfahren der Unterstichprobenziehung für das faktisch anonymisierte Längsschnittfile 1996-1999. Wiesbaden: Statistisches Bundesamt, Gruppe II A (unveröffentlichtes Ms.).
- BMA [Bundesministerium für Arbeit und Sozialordnung], 2001: Lebenslagen in Deutschland. Daten und Fakten. Materialband zum ersten Armuts- und Reichtumsbericht der Bundesregierung. Bonn: BMA. URL: www.bma.bund.de/de/sicherung/armutsbericht/ARBDatenFakten.pdf.
- Data Documentation Initiative (DDI) (WWW-Site). URL: www.icpsr.umich.edu/DDI/.
- ESRC United Kingdom Longitudinal Studies Centre: The BHPS User Documentation and Questionnaires (WWW-Site). URL: iser.essex.ac.uk/ulsc/bhps/doc.
- GESIS, German Microdata Lab: Voraussetzungen für die Datenweitergabe nach § 16 Abs. 6 BStatG 1987 (WWW-Site). URL: www.gesis.org/Dauerbeobachtung/GML/Daten/MZ/MZ-Info/Datenweitergabe.htm
- GESIS, German Microdata Lab: Leitfaden zur faktischen Anonymisierung statistischer Einzelangaben des Mikrozensus (WWW-Site). URL: www.gesis.org/Dauerbeobachtung/GML/Daten/MZ/MZ-Info/Leitfaden.htm.
- Haisken-DeNew, John P., und Joachim R. Frick (Hrsg.), 2005: Desktop Companion to the German Socio-Economic Panel (SOEP). Berlin: SOEP. URL: www.diw.de/deutsch/sop/service/dtc/dtc.pdf.
- Heidenreich, Hans-Joachim, 2002: Längsschnittdaten aus dem Mikrozensus: Basis für neue Analysemöglichkeiten. *Allgemeines Statistisches Archiv* 86(2): 213-231.
- Heidenreich, Hans-Joachim, und Michaela Nöthen, 2002: Der Wandel der Lebensformen im Spiegel des Mikrozensus. *Wirtschaft und Statistik*, 1: 26-38.
- Herter-Eschweiler, Robert, 2003a: Längsschnittdaten aus dem Mikrozensus. Basis für neue Analysemöglichkeiten. Bonn: Statistisches Bundesamt (unveröffentlichtes Manuskript).
- Herter-Eschweiler, Robert, 2003b: (SPSS-) Programmablauf zur Erstellung des faktisch anonymisierten Mikrozensus-Längsschnitts 1996-1999. Bonn: Statistisches Bundesamt (unveröffentlichtes Ms.).
- ICPSR [Inter-university Consortium for Political and Social Research], 2005: Guide to Social Science Data Preparation and Archiving. Best Practice Throughout the Data Life Cycle. Ann Arbor, MI: ICPSR. 3. Auflage. URL: www.icpsr.umich.edu/access/dataprep.pdf
- KVI [Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik] (Hrsg.), 2001: Wege zu einer besseren informationellen Infrastruktur. Gutachten der vom Bundesministerium für Bildung und Forschung eingesetzten Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik. Baden-Baden: Nomos.
- Lotze, Sabine, und Holger Breiholz, 2002a: Zum neuen Erhebungsdesign des Mikrozensus. Teil 1. *Wirtschaft und Statistik*, 5: 359-366.
- Lotze, Sabine, und Holger Breiholz, 2002b: Zum neuen Erhebungsdesign des Mikrozensus. Teil 2. *Wirtschaft und Statistik*, 6: 454-459.

- Lüttinger, Paul, Holger Breiholz und Yvonne Lechert, 2003a: Ergebnisse der zweiten Befragung von Nutzern der Mikrozensus-Scientific Use Files. ZUMA-Nachrichten Nr. 53: 128-135.
- Lüttinger, Paul, Bernhard Schimpl-Neimanns, Heike Wirth und Georg Papastefanou, 2003b: Mikrodaten (Geman Microdata Lab): Das Servicezentrum für amtliche Mikrodaten bei ZUMA. ZUMA-Nachrichten 52: 153-172.
- Lynn, Peter, 2001: Standard Quality Profiles for Longitudinal Studies (Draft). Essex: ESRC United Kingdom Longitudinal Studies Centre. URL: www.iser.essex.ac.uk/ulsc/methods/standards/qualprof/template/2001-09-28.pdf.
- Office of Management and Budget, 2001: Measuring and Reporting Sources of Error in Surveys. Statistical Policy Working Paper 31. URL: www.fcs.gov/01papers/SPWP31_final.pdf.
- Panel Study of Income Dynamics: Data and Documentation (WWW-Site). URL: psidonline.isr.umich.edu/Data/.
- Rohloff, Sandra, 2005: Das Hochrechnungsverfahren für Längsschnittauswertungen aus dem Mikrozensus. Methodenverbund „Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“, Arbeitspapier Nr. 6. Wiesbaden: Statistisches Bundesamt. URL: www.destatis.de/download/d/mv/arbeitspapier6.pdf.
- Särndal, Carl-Erik, Bengt Swensson und Jan Wretman, 1997: Model Assisted Survey Sampling. New York: Springer.
- Singer, Judith D., und Willett, John, B., 2003: Applied longitudinal data analysis: modeling change and event occurrence. New York: Oxford University Press.
- Taylor, Marcia Freed, 2000: Dissemination issues for panel studies. Metadata and documentation. S. 146-162 in Rose, David (Hrsg.): Researching Social and Economic Change. London: Routledge.
- Taylor, Marcia Freed, 2000: Dissemination issues for panel studies. Metadata and documentation. S. 146-162 in: Rose, David (Hrsg.): Researching Social and Economic Change. London: Routledge.
- UK Data Archive: Labour Force Survey datasets (WWW-Site). URL: www.data-archive.ac.uk.
- Unicon Research: Frequently asked questions about the CPS and its Supplements (WWW-Site). URL: www.unicon.com.
- U.S. Census Bureau: Survey of Income and Program Participation (WWW-Site). URL: www.sipp.census.gov/sipp/.
- Verordnung (EG) Nr. 577/98 des Rates vom 9.3.1998 zur Durchführung einer Stichprobenerhebung über Arbeitskräfte in der Gemeinschaft. URL: www.bundesstatistik.de/download/d/stat_ges/erwerb/085.pdf.
- Vermunt, Jeroen, 1997: Log-linear models for event histories. Advanced Quantitative Techniques in the Social Sciences; 8. Thousand Oaks: Sage.
- Wolter, Kirk M., 1985: Introduction to Variance Estimation. New York: Springer.
- Zühlke, Sylvia, und Uwe Hetke, 2002: Datenbedarf der Wissenschaft – Ein Bericht des Forschungsdatenzentrums der statistischen Landesämter über die erste Nutzerbefragung. Statistische Analysen und Studien Nordrhein-Westfalen, Band 6: 3-10.

Die erschienenen Arbeitsberichte im Überblick

Nr.	Autor(en): Titel
1	Edin Basic, Ulrich Rendtel: The use of the German Microcensus as a tool for longitudinal data analysis: Methods for the control the effects of the non-coverage of residential mobility.
2	Michael Konold: Ausmaß und Folgen unbeobachteter Übergänge im Rahmen des Mikrozensus-Panels: Ergebnisse empirischer Analysen.
3	Edin Basic, Ivo Marek, Ulrich Rendtel: The German Microcensus as a tool for longitudinal data analysis: An evaluation using SOEP data.
4	Robert Herter-Eschweiler: Der Mikrozensus als Panel: Längsschnittverknüpfung und Selektivitätsanalysen im Bereich der Art der Erwerbsbeteiligung und familialen Lebensformen.
5	Edin Basic, Ivo Marek, Ulrich Rendtel: The German Microcensus as a tool for longitudinal data analysis: An evaluation using SOEP data. Erschienen in: Schmoller's Jahrbuch - Journal of Applied Social Science Studies, Vol. 125, Number 1, 2005.
6	Sandra Rohloff: Das Hochrechnungsverfahren für Längsschnittauswertungen aus dem Mikrozensus.
7	Ulrich Rendtel: Wie geeignet ist der Mikrozensus für Längsschnittanalysen.
8	Edin Basic, Ulrich Rendtel: Estimation strategies in the presence of non-coverage in the German Microcensus-Panel: An evaluation using SOEP data.
9	Edin Basic: Stabilität von Ergebnissen bei unterschiedlichen Arbeitsmarktabgrenzungen.
10	Ivo Marek: Weighting adjustments in the presence of non-coverage due to residential mobility in the German Microcensus-Panel.
11	Heike Wirth: Anonymisierung des Mikrozensuspanels im Kontext der Bereitstellung als Scientific-Use-File
12	Bernhard Schimpl-Neimanns: Filekonzept zum Mikrozensus

Statistisches Bundesamt 65180 Wiesbaden Zweigstelle Bonn Graurheindorfer Str. 198 52117 Bonn	Freie Universität Berlin Fachbereich Wirtschaftswissenschaften Garystr. 21 14195 Berlin	Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen Postfach 101105 40002 Düsseldorf	Zentrum für Umfragen, Methoden und Analysen Postfach 122155 68072 Mannheim
--	--	---	--