

Ergebnisse des Projekts

Q-MED/LFS

Quantifizierung von Methodeneffekten
unterschiedlicher Erhebungsinstrumente
auf die Datenqualität im Labour Force Survey



Ein Gemeinschaftsprojekt der Statistischen Ämter
des Bundes und der Länder zur Qualitätssicherung
Im Mikrozensus/ Labour Force Survey

2010



 **STATISTISCHE ÄMTER**
DES BUNDES UND DER LÄNDER

Dieses Projekt wurde unterstützt
von der Europäischen Kommission

Herausgeber:

Statistische Ämter des Bundes und der Länder

Herstellung und Redaktion:

Statistisches Bundesamt

65180 Wiesbaden

Tel.: +49 (0) 611 / 75 24 05

Fax: +49 (0) 611 / 75 33 30

www.destatis.de/kontakt

Fachliche Informationen**zu dieser Veröffentlichung:**

Statistisches Bundesamt

Bereich „Mikrozensus, Arbeitskräftestichprobe, Haushalte und Familie“

Tel.: +49 (0) 2 28 99 / 643 89 55

Fax: +49 (0) 2 28 99 / 643 89 62

mikrozensus@destatis.de

Erscheinungsfolge: einmalig

Erschienen im April 2010

Fotorechte: © Statistisches Bundesamt

© Statistisches Bundesamt, Wiesbaden 2010
(im Auftrag der Herausbergemeinschaft)

Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.

Vorwort

Der vorliegende Bericht dokumentiert die Ergebnisse des Projekts „Quantifizierung von Methodeneffekten auf die Datenqualität im Mikrozensus / Labour Force Survey“ (Q-MED / LFS), das im Jahr 2009 in einer Kooperation zwischen dem Statistischen Bundesamt und den Statistischen Landesämtern Berlin Brandenburg, Hessen, Mecklenburg-Vorpommern, Niedersachsen, Nordrhein-Westfalen, Rheinland-Pfalz, Sachsen und Thüringen durchgeführt wurde. Damit wurde erstmals in einem umfangreichen Methodenexperiment die Bedeutung der Erhebungswege für die Datenqualität im Mikrozensus / LFS systematisch untersucht. In Deutschland wird der Mikrozensus und die darin integrierte Arbeitskräfteerhebung der EU - wie in vielen anderen europäischen Ländern - als Mixed-Mode-Erhebung durchgeführt. Erkenntnisse darüber, welchen Einfluss der Erhebungsweg auf die Ergebnisqualität hat, gibt es bislang nur wenige.

Der Mikrozensus / LFS ist für Deutschland eine der wichtigsten Datenquellen, wenn es um die Darstellung sozioökonomischer Strukturen der Bevölkerung und des Arbeitsmarktes geht. Umso lohnenswerter ist es, das Antwortverhalten der Befragten methodisch zu untersuchen und aus den Ergebnissen Handlungsoptionen für die Weiterentwicklung der Erhebung im Mikrozensus / LFS abzuleiten. Wie der vorliegende Ergebnisbericht belegt, leisten methodisch angelegte Projekte einen wichtigen Beitrag zur Innovationsfähigkeit und damit auch zur Modernisierung der amtlichen Statistik.

Das Projekt Q-MED / LFS ist Bestandteil einer Qualitätsoffensive der Statistischen Ämter des Bundes und der Länder, in der sich die amtliche Statistik der Gegenwart den Anforderungen einer modernen Informationsgesellschaft stellt und die Herausforderungen und Bedarfe von morgen antizipiert.

Viele verschiedene Personen und Institutionen haben zum Gelingen des Projektes beigetragen. Ein besonderer Dank gilt den Kooperationspartnern und Projektgruppenmitgliedern in den beteiligten Statistischen Landesämtern. Sie haben es ermöglicht, dass ein methodisch ambitioniertes Projekt innerhalb eines engen Zeitfensters durchgeführt werden konnte. Sie haben erfolgreich um die Teilnahme am Projekt Q-MED / LFS geworben, sie haben die Feldarbeit organisiert und durchgeführt und sie haben durch konstruktive Diskussionsbeiträge zum Erfolg des Projekts beigetragen.

Ein weiterer Dank gilt den Mitarbeiterinnen und Mitarbeitern vom GESIS-Leibniz-Institut für Sozialwissenschaften, die während der gesamten Projektzeit begleitend und beratend der Projektleitung zur Verfügung standen. Durch die Kooperation mit GESIS konnte das Projekt Q-MED / LFS als positiven Nebeneffekt die Kommunikation zwischen amtlicher Statistik und Wissenschaft weiter befördern.

Nicht zuletzt gilt unser Dank der Europäischen Kommission, die einen Großteil der Projektkosten finanziert hat. Durch die Anregung und Förderung methodischer Untersuchungen unterstreicht EUROSTAT die internationale Bedeutung qualitativer Untersuchungen im Labour Force Survey und unterstützt Strategien zur Qualitätssicherung der amtlichen Statistik in den Mitgliedsstaaten.

Bonn, den 08. April 2010

Karl Müller, Abteilungspräsident „Gesundheit, Soziales, Private Haushalt“

Schaubildverzeichnis

Schaubild 1: Anteil schriftlich Befragter nach Alter und Bildungsniveau	3
Schaubild 2: Stichprobenumfang Q-MED / LFS	7
Schaubild 3: Rekrutiersunterlage zur Gewinnung der Haushalte Beispiel IT.NRW	8
Schaubild 4: Einleitungstext zum Frageblock „Erwerbstätigkeit“	11
Schaubild 5: Beispiel aus dem Selbstausfüllerbogen	12
Schaubild 6: Online-Seite mit Hinweis auf unvollständige Angaben	13
Schaubild 7: Filteranweisungen im Selbstausfüllerbogen	14
Schaubild 8: Plausible und unplausible Angaben nach Filterfehlern	15
Schaubild 9: Teilnahmezeiten der Online-Befragten	17
Schaubild 10: Proxy-Anteile nach Geschlecht und Erhebungsweg	18
Schaubild 11: Strukturmerkmale zur Umgewichtung	21
Schaubild 12: Stellung im Beruf (gegenwärtige Tätigkeit)	22
Schaubild 13: Grund für weniger geleistete Arbeitsstunden in Wochen mit mindestens einem Feiertag	23
Schaubild 14: Frage zur Art der besuchten Schule im Selbstausfüllerbogen	24
Schaubild 15: Höchster beruflicher Abschluss (Befragte mit mittleren Berufsabschlüssen)	25
Schaubild 16: Frage zur aktiven Arbeitsuche in der Online-Befragung	28
Schaubild 17: Frage zum Bezug von Renten und Pensionen (Selbstausfüller)	29
Schaubild 18: Frage zum Hauptgrund für Mehrarbeit in der Berichtswoche (Selbstausfüller)	34
Schaubild 19: Hauptgrund für Mehrarbeit in der Berichtswoche in Prozent der Befragten	34
Schaubild 20: Normalerweise und tatsächlich geleistete Arbeitsstunden Mikrozensus 2008	38
Schaubild 21: Beispiel: Neue Filterführung im Mikrozensus/ LFS	43
Schaubild 22: Anteil der Personen mit Abschluss einer Lehre an allen mit mittleren Berufsabschlüssen/ Anteil CAPI, CATI-Interviews Mikrozensus 2008	45

Tabellenverzeichnis

Tabelle 1: Strukturmerkmale der Stichproben	9
Tabelle 2: Fragen mit mehr als 5 Antwortkategorien	21
Tabelle 3: Fragen mit Mehrfachantworten Suchbemühungen bei der Arbeitsuche	27
Tabelle 4: Fragen mit Mehrfachantworten Bezug von Renten und Pensionen	30
Tabelle 5: Fragen mit Mehrfachantworten Bezug von öffentlichen Zahlungen	31
Tabelle 6: Fragen mit Restkategorien Sonstige und andere Gründe	33
Tabelle 7: Fragen mit möglichen Effekten zur sozialen Erwünschtheit	36
Tabelle 8: Fragen mit Eintrag von Zahlenwerten (Arbeits- und Überstunden)	37

Inhaltsverzeichnis

1	Anlass der Untersuchung	1
1.1	Mode-Effekte im Labour Force Survey	1
1.2	Mixed Mode im Mikrozensus / LFS	2
2	Mixed Mode: Zum Stand der Forschung	4
2.1	Die interviewer-gestützten Befragung: CAPI und CATI	4
2.2	Ohne Interviewer: Selbstausfüller und Online-Erhebung	5
3	Das Projekt Q-MED / LFS	6
3.1	Allgemeine Zielsetzung	6
3.2	Zufallsstichproben im Feldexperiment	7
3.3	Rekrutierung der Probanden	7
3.4	Durchführung des Experiments	8
3.5	Repräsentativität und Vergleichbarkeit der Stichproben	8
3.5.1	Strukturmerkmale der Stichproben	9
3.5.2	Vergleichbarkeit der Online-Stichprobe mit onlinebereiten Selbstausfüllern	10
4	Die Erhebungsinstrumente im Projekt Q-MED / LFS	10
4.1	Das Face-to-Face-Interview (CAPI)	10
4.2	Das telefonische Interview (CATI)	11
4.3	Der Selbstausfüllerbogen	11
4.4	Der Online-Survey	12
5	Ausfüllverhalten in der schriftlichen Befragung	14
5.1	Das Ausfüllverhalten im Selbstausfüllerbogen	14
5.1.1	Viele Filterfehler bei der schriftlichen Befragung	14
5.1.2	Plausible und unplausible Angaben durch Filterfehler	15
5.1.3	Fast in jedem Bogen: Fehlende Angaben bei Pflichtmerkmalen	16
5.2	Ausfüllverhalten im Online-Survey	16
5.2.1	Rücklaufquote hoch - Abbrüche gering	16
5.2.2	Dauer und Teilnahmezeiten der Befragung	17
6	Unterschiede zwischen den Erhebungsinstrumenten	18
6.1	Der Einfluss des Erhebungsinstruments auf die Proxy-Quote	18
6.2	Einfluss der Erhebungsinstrumente auf das Antwortverhalten (Mode-Effekte)	19
6.2.1	Schätzung von Mode-Effekten im Mikrozensus / LFS	20
6.2.2	Listeneffekte bei Fragen mit zahlreichen Antwortmöglichkeiten	20
6.2.3	Effekte bei Fragen mit Mehrfachantworten	27
6.2.4	Fragen mit Restkategorien - „Sonstige und andere Gründe“	33
6.2.5	Fragen mit möglichen Effekten zur „Sozialen Erwünschtheit“	35
6.2.6	Effekte bei der Angabe von Zahlenwerten (Arbeitsstunden)	37
7	Zusammenfassung	39
7.1	Vor- und Nachteile einzelner Erhebungswege	39
7.1.1	Bewertung: Das CAPI-Interview	39
7.1.2	Bewertung: Das CATI-Interview	41
7.1.3	Bewertung: Der Selbstausfüllerbogen	42
7.1.4	Bewertung: Der Online-Bogen	43
7.2	Die Bedeutung von Mode-Effekten im Mikrozensus / LFS	44
	Literaturverzeichnis	46

1 Anlass der Untersuchung

Der Labour Force Survey (im Folgenden: LFS) wird in vielen EU-Mitgliedsstaaten mit unterschiedlichen Erhebungsinstrumenten durchgeführt. Die nationalen Ergebnisse sind oftmals eine Kombination aus persönlichen, telefonischen und in einigen Ländern auch schriftlichen Befragungsangaben. Da sich die Erhebungswege methodisch teilweise erheblich voneinander unterscheiden - etwa durch die An- bzw. Abwesenheit von Interviewern/-innen oder technische Unterstützungen -, hängt die Vergleichbarkeit der Ergebnisse auch davon ab, inwieweit die Befragungsergebnisse des LFS vom verwendeten Erhebungsweg beeinflusst sind. Internationale Forschungsbefunde weisen in diesem Zusammenhang auf eine Vielzahl möglicher Effekte hin, die durch den Einsatz eines bestimmten Erhebungsinstruments hervorgerufen werden (Mode-Effekte): Empirische Studien zeigen, dass persönliche Befragungen in vielen Fällen zu genaueren Ergebnissen führen, da die Interviewer/innen den Befragten bei Verständnisschwierigkeiten helfen können (Biemer / Groves / Lyberg 1991; De Leeuw 1992, 2008). Zugleich lassen sich aber in der persönlichen Befragung bei sensiblen Fragen Effekte sozialer Erwünschtheit beobachten (Groves / Lyberg 1991; De Leeuw 1992, 2008). Technische Möglichkeiten computergestützter Befragungsformen verringern den Item-Nonresponse, da Befragte keine Fragen offen lassen oder ungültige Angaben machen können (De Leeuw / Nicholls 1996; Biemer / Lyberg / Collins 1997). Ferner wird darauf hingewiesen, dass verschiedene Arten der Präsentation von Antwortmöglichkeiten zur bevorzugten Wahl von erst- oder letztgenannten Kategorien führen können (De Leeuw 1992, 2008; Dillman 2007). Diese wenigen Beispiele zeigen, dass Mode-Effekte die Befragungsergebnisse in nicht unerheblichem Umfang beeinflussen können. In großen Mixed-Mode-Erhebungen wie dem Mikrozensus / LFS sollten Mode-Effekte daher genau identifiziert und mit geeigneten Strategien minimiert werden, um eine Vergleichbarkeit der Daten aus unterschiedlichen Erhebungsformen zu gewährleisten.

1.1 Mode-Effekte im Labour Force Survey

Für den LFS liegen bereits erste Untersuchungen zu möglichen Mode-Effekten vor. Im Rahmen der von EUROSTAT initiierten „Task Force on the Quality of the LFS“ wurde in Frankreich, Griechenland, den Niederlanden und im Vereinigten Königreich über Abweichungen bei verschiedenen Merkmalen in Abhängigkeit von der Befragungswelle berichtet. In den Niederlanden werden in der Erstbefragung, die als computergestütztes persönliches Interview durchgeführt wird (Computer Assisted Personal Interview, im Folgenden CAPI) weniger Erwerbstätige und mehr Erwerbslose erfasst als in den darauf folgenden Wellen, die telefonisch erhoben werden (Computer Assisted Telephone Interview, im Folgenden CATI). In einer vom UK Office for National Statistics durchgeführten Untersuchung (Elliot et al. 2006) wurden Mode-Effekte u. a. bei unbezahlten Überstunden (CAPI: 34%, CATI: 39%) festgestellt.

Als Erklärung für diese Befunde kommen zwar auch verschieden hohe Proxy-Anteile, Panelmortalität, Fragebogen- und Selektivitätseffekte in der ersten Befragungswelle in Betracht, zugleich liefern sie erste Hinweise darauf, dass die Befragungsergebnisse im LFS möglicherweise nicht unerheblich vom verwendeten Erhebungsweg beeinflusst werden. Wie De Leeuw (2008) argumentiert, können solche Mode-Effekte für die Vergleichbarkeit von Subgruppen problematisch sein: „As different groups [...] are measured using different modes, this may cause data comparability problems for groups“ (De Leeuw 2008: 307). Für den LFS sind Mode-Effekte relevant:

- a) da auf internationaler Ebene nur Länderdaten verglichen werden können, die unter Verwendung desselben Erhebungsweges produziert wurden und

- b) auf nationaler Ebene nur Subgruppen verglichen werden können, die mit demselben Erhebungsweg oder derselben Kombination von Erhebungswegen befragt wurden.

In welchem Ausmaß bei Vergleichen Einschränkungen zu erwarten sind, hängt wesentlich von der quantitativen Bedeutung der eingesetzten Erhebungswege und der Stärke der beobachtbaren Mode-Effekte ab.

1.2 Mixed Mode im Mikrozensus / LFS

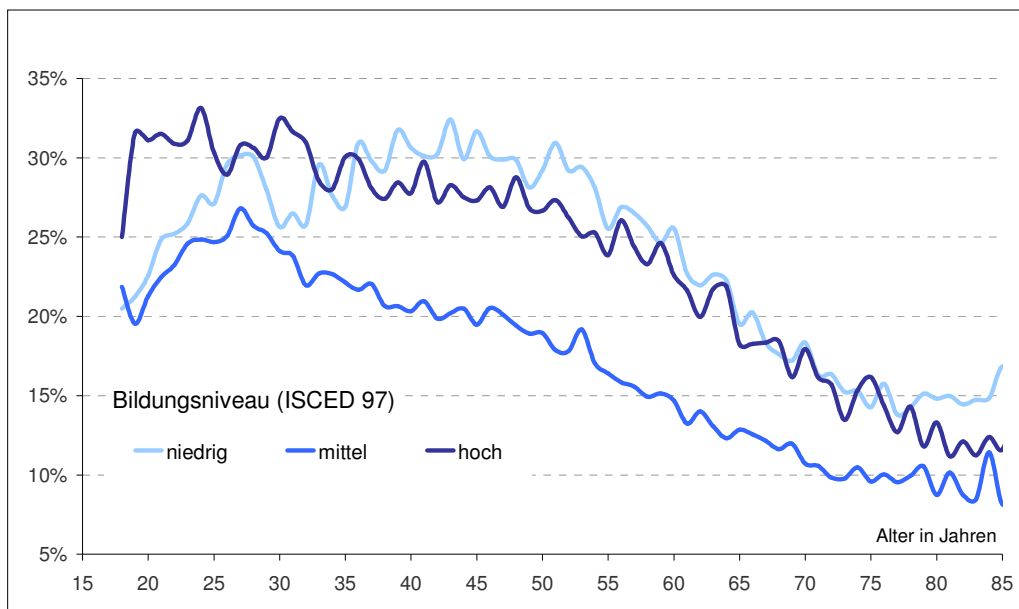
Während in vielen EU-Mitgliedstaaten die erste Befragungswelle als Face-to-Face-Interview und die Folgewellen telefonisch durchgeführt werden, sieht das Erhebungsdesign des deutschen Mikrozensus / LFS den parallelen Einsatz des Face-to-Face-Interviews, der schriftlichen und telefonischen Befragung vor. Der überwiegende Teil der Interviews wird persönlich durchgeführt (2008: 76% der Haushalte; 77% der Befragten). Trotz der bekannten und empirisch belegbaren Nachteile der schriftlichen Befragung werden im deutschen Mikrozensus / LFS auch Selbstausfüllerbogen eingesetzt (2008: 21%). Dies hat unterschiedliche Gründe: Zum einen sind nicht alle Haushalte persönlich anzutreffen. Wenn mehrmalige Kontaktversuche fehlschlagen, erhält der zu befragende Haushalt einen Selbstausfüllerbogen und die Aufforderung, die Auskunft schriftlich zu erteilen. Eine schriftliche Befragung kann aber auch auf Verlangen der Befragten erfolgen. Das Bundesverfassungsgericht hat in seinem Urteil zum Volkszählungsgesetz 1983 (1 BvR 209/83) festgelegt, dass das Recht auf informationelle Selbstbestimmung im überwiegenden Allgemeininteresse eingeschränkt werden kann: "Diese Einschränkungen bedürfen einer verfassungsgemäßen gesetzlichen Grundlage, die dem rechtsstaatlichen Gebot der Normenklarheit entsprechen muss. Der Gesetzgeber hat dabei den Grundsatz der Verhältnismäßigkeit zu beachten und organisatorische und verfahrensrechtliche Vorkehrungen zu treffen, welche der Gefahr einer Verletzung des Persönlichkeitsrechts entgegenwirken." Aus dem Urteil des BVerfG leitet sich das Recht des Auskunftspflichtigen ab, die Auskünfte im Mikrozensus / LFS vollständig anonym zu erteilen. Dies ist derzeit nur in der schriftlichen Form mit Selbstausfüllerbogen möglich. Die telefonische Befragung spielt im Mikrozensus / LFS eine untergeordnete Rolle: In wenigen Fällen und in der Regel nur auf ausdrücklichen Wunsch des Haushalts werden telefonische Befragungen durchgeführt (2008: etwa 3%).

Zwischen den Bundesländern lassen sich im Einsatz der unterschiedlichen Erhebungswege zum Teil erhebliche Unterschiede feststellen: Während die Länder Bayern und Baden-Württemberg mit 17% der befragten Personen die schriftliche Befragung mit Selbstausfüllerbögen am seltensten einsetzen, macht sie in Bremen mit 68% die Mehrheit aller Interviews aus. Insgesamt werden in den alten Bundesländern 19% der Befragten schriftlich befragt, während es in den neuen Bundesländern (einschl. Berlin) 22% der Befragten sind. Dieser Unterschied lässt sich nur zum Teil durch unterschiedliche Verstädterungsgrade in Ost- und Westdeutschland erklären. In dicht besiedelten Agglomerationsräumen füllen sowohl in den alten wie auch in den neuen Bundesländern (einschl. Berlin) etwa 22% der Befragten einen Selbstausfüllerbogen aus. Jedoch sinkt dieser Anteil in den alten Bundesländern mit abnehmender Bevölkerungsdichte und erreicht in ländlichen Räumen mit 15% seinen niedrigsten Wert. In den neuen Bundesländern (einschl. Berlin) zeigt sich ein umgekehrter Zusammenhang: In verstädterten Räumen bleibt der Anteil der schriftlichen Befragung mit 23% ähnlich hoch wie in den Agglomerationsräumen. In ländlichen Regionen steigt er jedoch auf 26%.

Auf Ebene der Haushalte variiert der Anteil der schriftlichen Befragung mit der Haushaltsgröße¹. Der Anteil der schriftlich Befragten in Einpersonenhaushalten ist mit 22% am höchsten und sinkt bei Haushalten mit zwei oder mehr Personen unter 20%. Als mögliche Erklärung für diesen Zusammenhang kommt die häufig schlechtere Erreichbarkeit allein lebender Personen in Betracht. Allgemein kann angenommen werden, dass mit wachsender Haushaltsgröße die Wahrscheinlichkeit steigt, zumindest eine auskunftspflichtige Person anzutreffen und ein persönliches Interview zu führen.

Auf der Personenebene haben Merkmale wie Alter und Bildung Einfluss auf die Wahl des Erhebungsweges. Es fällt auf, dass in allen Bildungsschichten der Anteil der schriftlich Befragten mit dem Lebensalter stark abnimmt und bei Personen ab 75 Jahren nur noch etwa 15% beträgt (vgl. Schaubild 1). Personen mit einem niedrigen oder mittleren Bildungsniveau nehmen dabei etwa gleich häufig schriftlich am Mikrozensus / LFS teil, solange sie unter 30 Jahre. Für Personen zwischen 30 und 75 Jahren ist auffällig, dass niedrig und hoch gebildete Personen deutlich häufiger schriftlich am Mikrozensus / LFS teilnehmen als solche mit einem mittleren Bildungsniveau. Die hoch Gebildeten nehmen bis zu einem Alter von etwa 35 Jahren am häufigsten schriftlich am Mikrozensus / LFS teil, danach sind sie etwas seltener in der schriftlichen Befragung als Personen mit höherem Bildungsniveau. Der insgesamt geringe Anteil mittlerer Bildungsschichten an der schriftlichen Befragung deutet auf eine höhere Bereitschaft zur Teilnahme an der persönlichen Erhebung und möglicherweise auch eine allgemein höhere Befragungsbereitschaft hin. Vergleichende Untersuchungen zwischen dem Mikrozensus / LFS und dem ALLBUS haben bereits in den 1990er Jahren gezeigt, dass mittlere und höhere Bildungsschichten in freiwilligen Erhebungen deutlich besser repräsentiert sind als untere Bildungsschichten (Hartmann / Schimpl-Neimanns 1992).

Schaubild 1: Anteil schriftlich Befragter nach Alter und Bildungsniveau²



Quelle: Mikrozensus 2008

¹ Die schriftliche Befragung kommt mit 20% deutlich häufiger in Privathaushalten zum Einsatz als in Gemeinschaftsunterkünften (15%). Aufgrund ihrer geringen quantitativen Bedeutung werden Personen in Gemeinschaftsunterkünften allerdings nicht gesondert ausgewiesen.

² Die Klassifizierung der Befragten gemäß ISCED 97: Personen mit einem Abschluss der Stufen ISCED 1 oder 2 wurden als niedrig, Personen mit einem Abschluss der Stufen ISCED 3 oder 4 wurden als mittel und Personen mit einem Abschluss gemäß ISCED 5 oder 6 wurden als hoch gebildet eingestuft.

Die höhere Bereitschaft der mittleren Bildungsschichten an Befragungen teilzunehmen, zeigt sich hier bei der Wahl des Erhebungsinstruments: Wenn allerdings (1) die Wahl des Erhebungswegs nicht zufällig ist und strukturellen Selektionsprozessen unterliegt und (2) bedeutsame Mode-Effekte nachgewiesen werden können, bedeutet dies eine Einschränkung der Vergleichbarkeit. Ergebnisunterschiede zwischen Gruppen (oder Ländern) können möglicherweise zum erheblichen Teil auf Mode-Effekte zurückgeführt werden. Umso wichtiger scheint es, genauere Kenntnisse über mögliche Mode-Effekte zu generieren, um die Erhebungsinstrumente weiter zu entwickeln und Mode-Effekte zu minimieren. Nur wenn ausreichend transparent wird, mit welchem Erhebungsinstrument welche Datenqualität erzeugt werden kann, können reale Differenzen zwischen Bevölkerungsgruppen oder regionalen Einheiten von statistischen Artefakten unterschieden werden.

2 Mixed Mode: Zum Stand der Forschung

2.1 Die interviewer-gestützte Befragung: CAPI und CATI

Das persönliche Interview gilt zwar nicht mehr uneingeschränkt als der „Königsweg“ unter den Erhebungstechniken, dennoch konnte es - trotz hoher Kosten und vergleichsweise niedriger Ausschöpfungsquoten - „seine dominierende Position in der Forschungspraxis bewahren“ (Kromrey 2006: 336). Im Gegensatz zu den schriftlichen Erhebungstechniken erfordert die persönliche Befragung die Anwesenheit eines Interviewers oder einer Interviewerin, der oder die die Antworten der Befragten entweder auf einem Papierfragebogen einträgt (PAPI - Paper And Pencil Interview) oder computergestützt erfasst (CAPI und CATI). Der Einsatz computergestützter Fragebogen bietet gegenüber anderen Erhebungsinstrumenten eine Vielzahl an technischen Vorteilen. Die Filterführung erfolgt automatisch, einzelne Felder können nicht übersprungen oder leer gelassen werden (Baker et al. 1995: 416f.) und die Antworten der Befragten werden bereits während des Interviews plausibilisiert. Das persönliche Interview bietet darüber hinaus den Vorteil, Verständnisschwierigkeiten der Befragten beispielsweise durch zusätzliche Informationen, die der Interviewer oder die Interviewerin in Form von Hilfstexten abrufen kann, zu minimieren. Diese Informationen können gegenüber schriftlichen Verfahren sehr viel umfangreicher gestaltet sein, da sie in der persönlichen Befragung nur dann durch den/die Interviewer/in abgerufen werden, wenn sie benötigt werden. Zudem kann der/die Interviewer/in Hilfsmittel vorlegen, wie beispielsweise Listenhefte, die der optischen Unterstützung im Interview dienen. Der/die Interviewer/in kann bei schwer zu beantwortenden Fragen aber auch generelle Hilfestellung geben. In einer 2006 vom UK Office for National Statistics (ONS) durchgeführten qualitativen Studie konnte gezeigt werden, dass Interviewer im Face-to-Face-Interview Probleme der Befragten besser erkannten und vor Ort Einsicht in relevante Dokumente nahmen.

Durch die Gegenwart eines Interviewers oder einer Interviewerin in Face-to-Face-Befragungen können aber auch negative Effekte entstehen. Bereits seine/ihre Anwesenheit löst beim Befragten ein gewisses Bild aus, das bewusst oder unbewusst zu Anpassungsmechanismen führen kann: „Sometimes, the presence of the interviewer in itself is sufficient to create bias“ (Loosveldt 2008: 215). Selbst wenn der Interviewer oder die Interviewerin sich so neutral wie möglich verhält, kann er oder sie nicht verhindern, dass sich die Befragungsperson Vorstellungen über ihn oder sie macht (Atteslander 2003: 152) oder unbewusst auf nonverbale Kommunikation reagiert und damit direkten Einfluss auf den Gesprächsverlauf nimmt (de Leeuw 1992: 17). Wie Dillman zeigt, sind in interviewer-gestützten Befragungsformen auch Effekte sozialer Erwünschtheit stärker ausgeprägt als in schriftlichen Befragungen (Dillman 2007: 226f.). Face-to-Face-Befragungen gelten daher zur Erfassung sensibler Sachverhalte -

wie etwa dem Einkommen - allgemein als weniger geeignet als anonymere Techniken der Befragung (vgl. Gfroerer / Hughes 1991, 1992). De Leeuw weist in diesem Zusammenhang darauf hin, dass die „weiß nicht“- und „unentschieden“- Kategorien in der schriftlichen Befragung zwar im allgemeinen häufiger gewählt werden als im CAPI oder CATI, bei sensiblen Fragen erzielen dagegen schriftliche Befragungen mehr auswertbare Informationen (de Leeuw 1992: 31f.; Bradburn 1983: 294).

Ausschöpfungsraten und Item-Nonresponse spielen im Mikrozensus zwar keine größere Rolle, da die Befragten einer Auskunftspflicht unterliegen und nur wenige Merkmale freiwillige Angaben sind. Dennoch gibt die Auswertung des Item-Nonresponse im Mikrozensus Hinweise darauf, dass die Datenqualität im persönlichen Interview deutlich höher ist als in der schriftlichen Befragung. So lag beispielsweise 2007 die Item-Nonresponse-Quote³ der freiwilligen Fragen (ohne Ad-Hoc-Modul) bei Face-to-Face-Interviews bei nur 3%, bei Telefoninterviews bei 5% und in der schriftlichen Befragung bei 9%.

Gegenüber dem Face-to-Face-Interview können Interviewer in der telefonischen Befragung schlechter auf die Befragten eingehen, da nonverbale Reaktionen im CATI nicht erkennbar sind (vgl. Lyberg / Kasprzyk 1991: 240). Da umgekehrt auch die Befragten keine (nonverbale) Rückmeldung von den Interviewern bekommen, fühlen sie sich oft einem höheren Zeitdruck bei der Beantwortung der Fragen ausgesetzt. Dieser Zeitdruck kann nachteilige Folgen auf Beurteilungs- und Erinnerungseffekte haben (vgl. Dillman 2007). Aus demselben Grund sind in telefonischen Interviews auch die Antworten auf offene Fragen tendenziell kürzer. Sollen detaillierte Informationen erfragt werden, empfehlen Biemer und Lyberg deshalb die Anwendung von CAPI (Biemer / Lyberg 2003: 149). Schließlich wird bei Telefoninterviews häufig der so genannte Recency-Effekt beobachtet. Wenn eine Frage viele Antwortkategorien vorsieht und diese zudem vorgelesen werden, gibt es eine Tendenz zur Wahl der letzten Antwortkategorie (Porst 2000: 53). Dillman (2007) begründet diesen Effekt mit der beim Vorlesen schnellen Abfolge von Antwortkategorien, was dazu führt, dass die zuletzt gehörten Antwortkategorien besser erinnert werden. Diesem Effekt steht der Primacy-Effekt in schriftlichen Befragungen entgegen. Befragte neigen besonders bei optischer Präsentation von Antwortkategorien dazu, die erste als passend empfundene Antwortkategorie zu wählen (Dillman 2007: 228f.).

2.2 Ohne Interviewer: Selbstausfüller und Online-Erhebung

Die schriftliche Befragung mit Papier-Fragebogen (Selbstausfüller) hat in der Sozialforschung mit dem verbreiteten Einsatz von CAPI- und CATI- Befragungen an Bedeutung verloren. Vor allem der hohe Aufwand in der Auf- und Nachbereitung der Daten, der vergleichsweise hohe Item-Nonresponse, die geringere Rücklaufquote, die Fehleranfälligkeit bei komplexen Filtersprüngen sowie eine Verzögerungen des Rücklaufs durch verspäteten Rücksendungen werden in der Fachliteratur als Nachteile gesehen (de Leeuw / Hox / Dillman 2008; de Leeuw 1992; Biemer / Groves / Lyberg 1991). Bei komplexen Befragungen wie dem Mikrozensus / LFS ist vor allem nachteilig, dass während der schriftlichen Befragung keine Plausibilierungen vorgenommen werden können. Implausibilitäten und die Inkonsistenz der Angaben fallen erst bei der Dateneingabe auf. Die Angaben der schriftlichen Befragung müssen dann aufwändig geprüft bzw. nachrecherchiert werden (Lyberg / Biemer / Collins 1997). Die vordergründig niedrigeren Kosten der schriftlichen Befragung (de Leeuw / Hox / Dillman 2008), die den kostenintensiveren Einsatz von Interviewern spart, kann - zumindest bei dem Anspruch an eine hohe Vollständigkeit der Daten - durch nachbereitende Arbeiten schnell die Kosten der Face-to-Face-Befragung übersteigen. Ein gewichtiger Nachteil der schriftlichen Befragung ist zudem, dass bei kom-

³ Die Item-Non-Response-Quote ergibt sich aus dem Verhältnis der nicht beantworteten Fragen zur Gesamtzahl der freiwilligen Fragen, die einem Befragten gestellt wurden.

plizierten Sachverhalten keine Rückfragen möglich sind und der Befragte allein auf Hilfe- und Erläuterungstexte im Fragebogen angewiesen ist. Im Vergleich zur mündlichen Befragung ist daher eine geringere Validität der Antworten zu erwarten, wenn Unklarheiten nicht ausgeräumt werden können. Im Zweifelsfall bleiben Fragen unbeantwortet oder es wird die Antwortkategorie gewählt, die zwar nicht vollständig zutreffend, aber naheliegend erscheint. Hinzu kommt, dass in der Auswertung von schriftlichen Befragungsergebnissen nicht nachvollzogen werden kann, wer die Angaben im Papierbogen oder auch im Online-Bogen tatsächlich gemacht hat.

Neben den genannten Nachteilen werden in der Fachliteratur auch Vorteile der schriftlichen Befragung benannt: So genannte Interviewer-Effekte (vgl. de Leeuw 1992) sind in der schriftlichen Befragung ausgeschlossen. Als weiterer Vorteil der schriftlichen Befragung gilt die Flexibilität der Befragten während des Interviews. Die Befragung kann jederzeit unterbrochen werden, um z. B. bei nichtanwesenden Haushaltsmitgliedern Angaben zu erfragen oder zusätzliche Informationen zu recherchieren (z. B. Abschlussjahr des Studiums o. ä.). De Leeuw et al. konnten nachweisen, dass es hier in der schriftlichen Befragung zu präziseren Antworten kommt (de Leeuw / Hox / Dillman 2008).

Angesichts einer zunehmenden Nutzung des Internets durch private Haushalte - 79% der Haushalte in Deutschland verfügten 2008 über einen Internetzugang (Statistisches Bundesamt 2009: 11) - werden Online-Erhebungen auch in der amtlichen Statistik in Zukunft an Bedeutung gewinnen, sofern die Anforderungen an die Erhebung, die Akzeptanz bei den Befragten und die Qualität der Daten dies ermöglichen. Das Online-Verfahren ist vergleichsweise kostengünstig und bietet wie CAPI und CATI die Möglichkeit technischer Unterstützungen und Filterführungen während des Interviews. Weitere Vorteile bestehen in der schnellen Datenverfügbarkeit, der Möglichkeit zielgruppenspezifischer Formulierungen und der variablen Darstellung von Fragen (Pötschke / Simonson 2001: 12f.). Wie bei der postalischen Befragung ist die Auskunftserteilung unabhängig von Zeit und Ort und ermöglicht den Ausschluss interviewerbedingter Fehler. Nachteile ergeben sich wie auch bei anderen schriftlichen Verfahren dadurch, dass keine persönliche Unterstützung durch einen Interviewer erfolgen kann, der beispielsweise Hilfstexte vorlesen oder zusätzliche Erläuterungen geben kann. Bei Verständnisproblemen, die über die Hilfstexte hinaus bestehen, muss sich der Befragte in der schriftlichen Befragung auf eigene Interpretationen der Fragen verlassen.

3 Das Projekt Q-MED / LFS

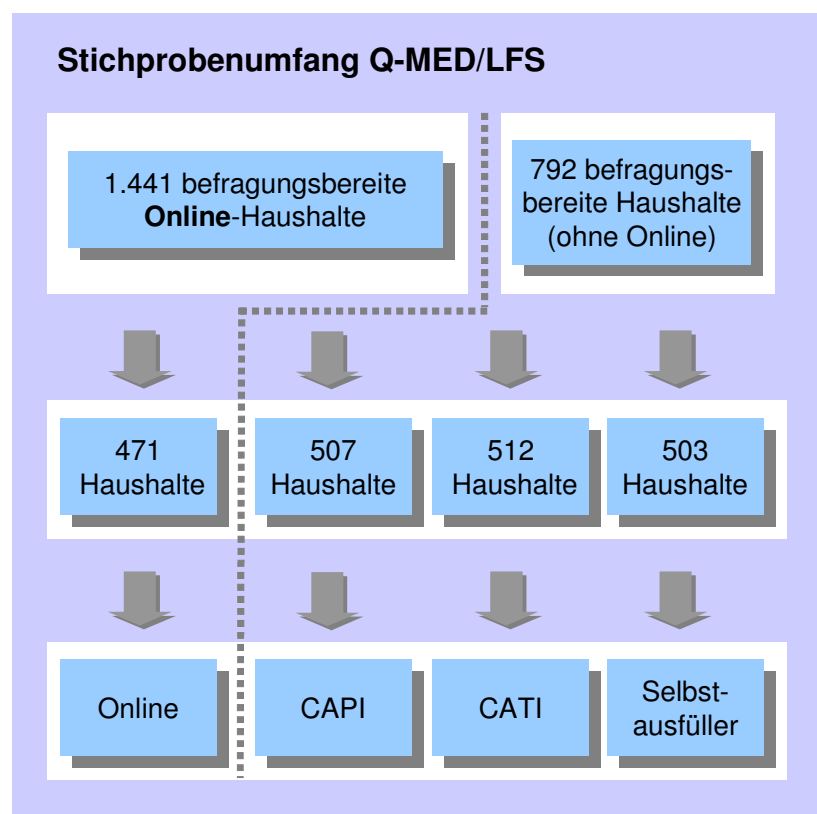
3.1 Allgemeine Zielsetzung

Ziel des Projektes „Quantifizierung der Methodeneffekte unterschiedlicher Erhebungsinstrumente auf die Datenqualität im Labour Force Survey“ (im Folgenden Q-MED / LFS) war es, die Erhebungsinstrumente CAPI, CATI, schriftliche Befragung mit Selbstausfüller sowie den Mikrozensus / LFS als Online-Survey in einem Feldexperiment zu testen und die Unterschiede in der Datenqualität dieser Erhebungswege zu quantifizieren. Die Analysen zur Datenqualität fokussieren dabei auf die Erfassung der erwerbstätigen und erwerbslosen Personen, eröffnen aber ebenso die Möglichkeit der Prüfung anderer Merkmale. Erwartetes Ergebnis des Methodenexperiments ist eine vergleichende und empirisch gesicherte Aussage über die erzielte Datenqualität bei Verwendung unterschiedlicher Erhebungsinstrumente sowie deren Vergleichbarkeit. Zusätzlich werden genauere Erkenntnisse über die Akzeptanz verschiedener Erhebungsformen erwartet.

3.2 Zufallsstichproben im Feldexperiment

Um Effekte der Selbstselektivität auszuschließen, wurde den bundesweit angeworbenen Probanden nach dem Zufallsprinzip einer von vier Erhebungswegen zugewiesen: Befragung durch einen Interviewer, Telefonische Befragung, Schriftliche Befragung (Selbstaussfüller) und Online-Survey. Für die Online-Befragung wurden nur Haushalte ausgewählt, die zuvor ihre Bereitschaft für den Online-Survey mitgeteilt hatten. Aus den online-bereiten Haushalten wurde zunächst die Stichprobe für den Online-Survey gezogen (473 Haushalte aus 1.443). Die nicht gezogenen, onlinebereiten Haushalte wurden gemeinsam mit den übrigen Haushalten auf die Erhebungswege CAPI, CATI und Selbstaussfüller verteilt. Insgesamt konnten 4.118 Interviews in 1.993 Haushalten realisiert werden.

Schaubild 2: Stichprobenumfang Q-MED / LFS



3.3 Rekrutierung der Probanden

Von den teilnehmenden Statistischen Landesämtern Berlin Brandenburg, Hessen, Mecklenburg-Vorpommern, Niedersachsen, Nordrhein-Westfalen, Rheinland-Pfalz, Sachsen und Thüringen wurden Probanden aus den vorhandenen Adressbeständen der generell befragungsbereiten Haushalte aus der Einkommens- und Verbrauchsstichprobe ausgewählt und schriftlich angeworben. Jeder Haushalt musste sich bereit erklären, an einer schriftlichen, einer Face-to-Face-Befragung oder einem telefonischen Interview teilzunehmen, ohne entsprechende Präferenzen äußern zu können. Zusätzlich wurde erfragt, ob der Haushalt bereit sei, die Befragung online durchzuführen. Um eine hohe Anwerbequote zu realisieren, wurde den Haushalten eine Mitwirkungsprämie von 10 Euro pro Haushalt und weiteren 5 Euro pro Haushaltsmitglied und realisiertem Interview angeboten.

Schaubild 3: Rekrutiersunterlage zur Gewinnung der Haushalte
Beispiel IT.NRW

Befragungsformen im Vergleich

Mitmachen ist einfach!

Wenn Sie am Projekt „Befragungsformen im Vergleich“ teilnehmen möchten, füllen Sie bitte die Antwortkarte aus und senden Sie diese im Briefumschlag an uns zurück. Die Mitteilung, in welcher Form wir die Befragung mit Ihnen durchführen werden, erhalten Sie kurze Zeit später.

Haben Sie noch Fragen an uns?

Wenn Sie weitere Fragen zu unserem Projekt oder der Durchführung der Befragung haben, beantworten wir diese gerne unter:

Information und Technik NRW
– Geschäftsbereich Statistik –
Andrea Janssen
Willy-Brandt-Platz 3
46045 Oberhausen
Telefon: +49 (0) 211/94 49-42 14
E-Mail: andrea.janssen@it.nrw.de

Informationen und Hinweise zur Methoden-Studie

Das Projekt wird gefördert von der Europäischen Kommission.

Erstheraus im März 2009
Fotorechte: ©Statistische Ämter des Bundes und der Länder
©Statistisches Bundesamt, Wiesbaden 2009, (im Auftrag der herausgebenden Gesellschaft).
Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.
Statistische Ämter des Bundes und der Länder; Befragungsformen, 2009

Bitte bis spätestens 20.4.2009 zurücksenden!

Information und Technik NRW
– Geschäftsbereich Statistik –
Andrea Janssen
Willy-Brandt-Platz 3
46045 Oberhausen

Name, Vorname (ausdrücklich Anredeformulierung im Hauptteil)
Straße, Hausnummer
Postleitzahl, Wohnort
Telefon (unter Angabe der Telefonnummer mit Vorwahl) (ausdrücklich)
E-Mail-Adresse (optional)

Druck: Befragungsformen_103-0709.indd 1

23.03.2009 10:30:42

3.4 Durchführung des Experiments

Die Probanden wurden zwischen Mai und Juli 2009 mit dem jeweils zufällig zugewiesenen Erhebungsinstrument befragt. Der schriftliche Erhebungsbogen bzw. die für CAPI und CATI eingesetzten Frageprogramme entsprachen dem Original-Frageprogramm des Mikrozensus / LFS 2010 (ohne Ad-Hoc-Modul und Zusatzprogramm⁴). Zusätzlich wurde ein Online-Fragebogen entwickelt, der ebenfalls den vollen Frageumfang des Grundprogramms umfasste, aber nicht als Haushaltsbogen, sondern als personenbezogener Fragebogen entwickelt wurde. Nach der Haupt-Befragung wurde eine schriftliche Nachbefragung der Haushalte durchgeführt. Gegenstände der Nachbefragung waren evtl. auftretende Schwierigkeiten bei der Beantwortung zentraler Fragen und die Präferenz der Befragten für die verschiedenen Erhebungswege. Die Ergebnisse dieser Befragung liefern wichtige Erkenntnisse zur Anwendbarkeit der eingesetzten Instrumente aus Sicht der Befragten.

Die Rekrutierung der Haushalte und die Durchführung der Face-to-Face-Interviews, der telefonischen Befragung sowie die Verschickung der Selbstausfüllerbogen bzw. Zugangsdaten zum Online-Survey erfolgten in den beteiligten Statistischen Landesämtern. Die zufällige Verteilung der Haushalte auf die einzelnen Erhebungsinstrumente erfolgte für die Gesamtstichprobe anonymisiert im Statistischen Bundesamt. Der Online-Survey wurde während der Feldphase zentral vom Statistischen Bundesamt begleitet.

3.5 Repräsentativität und Vergleichbarkeit der Stichproben

Auswahlgrundlage waren Haushalte aus der Einkommens- und Verbrauchsstichprobe (im Folgenden EVS), die ihre Bereitschaft zur Teilnahme am Experiment schriftlich mitgeteilt hatten. Da bereits die EVS als Freiwilligenstichprobe einen Freiwilligen-Bias aufweist und dieser sich durch die Ziehung ei-

⁴ Der Fragebogen des Mikrozensus besteht aus mehreren Teilen. Merkmale des Grundprogramms werden jährlich erhoben. Das Zusatzprogramm rotiert im Vier-Jahres-Zyklus, so dass jeder Haushalt jedes Zusatzprogramm einmal beantworten muss. Zusätzlich gibt es das Ad-hoc-Modul, das nur einer Unterstichprobe von 0,1% gestellt wird und das jährlich wechselnde Themen beinhaltet.

ner weiteren, freiwilligen Unterstichprobe eher vergrößern als ausgleichen dürfte, ist nicht von einer für die Bevölkerung der Bundesrepublik Deutschland repräsentativen Stichprobe auszugehen. Weitere Selektionen ergeben sich aus einer regionalen Beschränkung der Q-MED / LFS-Haushalte auf die teilnehmenden Bundesländer.

3.5.1 Strukturmerkmale der Stichproben

Insgesamt sind die Probanden in Q-MED / LFS mit durchschnittlich 44,3 Jahren etwas älter als der bundesdeutsche Durchschnitt (42,6 Jahren). Die Teilnehmer an der Online-Befragung stellen mit durchschnittlich 39,5 Jahren die jüngste Gruppe. In Q-MED / LFS wie auch in der Subgruppe der Online-Befragten sind verheiratete Personen mit 51% bzw. 52% überrepräsentiert (Mikrozensus 2008 42%). Wie bei vielen anderen Freiwilligen-Befragungen ist die ausländische Bevölkerung nur in geringem Maß vertreten: Lediglich 1% der Q-MED / LFS-Teilnehmer hatten keine deutsche Staatsangehörigkeit. Ebenfalls typisch für freiwillige Befragungen ist die Unterrepräsentanz von Bevölkerungsschichten mit geringem Bildungsniveau und entsprechend einer größeren Teilnahmebereitschaft der höheren Bildungsschichten. 51% der Online-Befragten und 47% der Gesamtstichprobe verfügten über einen hohen Bildungsabschluss, im Mikrozensus sind es 26%⁵. Zudem sind in der Q-MED / LFS-Stichprobe überzufällig viele Beamte, Beamtinnen und Angestellte vertreten (Mikrozensus / LFS: 17%, Q-MED / LFS insgesamt: 37%, Online-Stichprobe: 41,5%).

Tabelle 1: Strukturmerkmale der Stichproben

Strukturmerkmale		Mikrozensus / LFS 2008	Q-MED / LFS Gesamt	Online-Teilnehmer	Kontrollgruppe Selbstausfüller online-bereit
Alter	Mittelwert	42,6	44,3	39,5	39,8
Geschlecht	Männlich	49%	46%	47%	48%
	Weiblich	51%	54%	53%	52%
Familienstand	Ledig	40%	36%	40%	42%
	Verheiratet	47%	51%	52%	47%
	Geschieden	6%	9%	7%	10%
	Verwitwet	7%	3%	1%	1%
Nationalität	Deutsch	91%	99%	99%	99%
	Nicht-Deutsch	9%	1%	1%	1%
Bildungsniveau (ISCED97)	Niedrig	25%	14%	18%	14%
	Mittel	50%	38%	30%	38%
	Hoch	26%	49%	52%	49%
Tätigkeit	Beamte	3%	6%	8%	8*%
	Angestellte	14%	31%	33%	62*%
Haushaltsgröße	Mittelwert	2,05	2,06	2,30	2,18

* Die Frage nach dem Erwerbstatus wurde von 38 Personen (7,8%) nicht beantwortet.

Einige wenige Strukturvergleiche zeigen, dass die Q-MED / LFS-Stichprobe vor allem mittlere soziale Schichten der Bevölkerung repräsentiert. Bei der Interpretation der Ergebnisse und Bewertung von Mode-Effekten muss daher berücksichtigt werden, dass diese vorerst nur bei einer bestimmten Bevölkerungsgruppe nachgewiesen wurden: Bei Personen, die überdurchschnittlich hoch gebildet sind und

⁵ Die Klassifizierung der Befragten erfolgte gemäß ISCED 97, siehe Fußnote 2.

die aufgrund ihrer Berufstätigkeit als Angestellte bzw. Beamte vermutlich gewohnt sind, mit Schriftsprache und Formularen im beruflichen Alltag umzugehen. Allerdings kann man auch davon ausgehen, dass Verständnisprobleme in der Befragung, die sich in Q-MED / LFS empirisch belegen lassen, mindestens genauso häufig in bildungsfernen Bildungsschichten vorkommen dürften.

3.5.2 Vergleichbarkeit der Online-Stichprobe mit onlinebereiten Selbstausfüllern

Eine Vergleichbarkeit der Gruppen CAPI, CATI und Selbstausfüller ist durch die zufällige Aufteilung der Stichprobe auf die Erhebungsformen gegeben. Dies gilt allerdings nicht für die Teilnehmer der Online-Befragung. Durch die Vorselektion der online-bereiten Haushalte, die sich signifikant von den nicht-onlinebereiten Haushalten unterscheiden, kann die Online-Stichprobe nicht ohne weiteres mit den anderen Gruppen verglichen werden. Um dennoch Hinweise zu erhalten, ob und an welchen Stellen sich das Online-Verfahren von dem anderen schriftlichen, nicht interviewer-gestützten Verfahren (Selbstausfüller) unterscheidet, wurde aus der Gruppe der Selbstausfüller-Haushalte eine Kontrollgruppe konstruiert. Da für alle Haushalte die Information vorlag, ob die Haushalte bereit gewesen wären, an einer Online-Erhebung teilzunehmen, konnten die online-bereiten Haushalte aus der Selbstausfüller-Stichprobe zu Vergleichszwecken selektiert werden. Insgesamt 1.085 Befragte beantworteten den Q-MED / LFS-Fragebogen online, die mit den 503 Personen der onlinebereiten Selbstausfüller-Stichprobe verglichen werden können. Betrachtet man hier Versuchs- und Kontrollgruppe als eine Grundgesamtheit, so hatte jeder onlinebereite Haushalt die gleiche Ziehungswahrscheinlichkeit für die Online-Befragung und genauso die gleiche Wahrscheinlichkeit, einen Selbstausfüllerbogen in Papierform zu erhalten. Signifikante Unterschiede zwischen den Onlineteilnehmern und der Kontrollgruppe der online-bereiten Selbstausfüller sind aufgrund der Zufallszuweisung zu den Stichproben daher nicht feststellbar. Einschränkungen in der Interpretation der Ergebnisse gibt es eher aufgrund der Größe der Kontrollgruppe. Bei Teilpopulationen ergeben sich rasch Probleme mit geringen Fallzahlen und großen Stichprobenfehlern. An entsprechender Stelle wird im Ergebnisteil darauf verwiesen.

4 Die Erhebungsinstrumente im Projekt Q-MED / LFS

4.1 Das Face-to-Face-Interview (CAPI)

Die für das persönliche Interview ausgewählten Haushalte wurden vom zuständigen Statistischen Landesamt angeschrieben und es erfolgte eine Terminvereinbarung, um das Face-to-Face-Interview in der Wohnung des Haushalts durchzuführen. Wie auch im regulären Mikrozensus / LFS wird das Face-to-Face-Interview als CAPI-Interview durchgeführt, bei dem die Antworten des Haushalts direkt in den Laptop eingegeben werden. Der CAPI-Fragebogen wurde mit der Software „Blaise“ erstellt und hat den Vorteil der automatischen Filterführung und der Plausibilisierung von Antworten während des Interviews. Zudem unterstützen Signierhilfen und der Zugriff auf Klassifikationen die Zuordnung von offenen Antworten im Interviewverlauf. Die Plausibilitätsprüfungen (im Folgenden PL) erzeugen bei Eingabe von unplausiblen Werten Fehlermeldungen, die den Interviewer oder die Interviewerin auffordert, einen plausiblen Wert einzugeben oder sich beim Befragten rückzuversichern, ob diese Antwort wirklich stimmt. Hier wird zwischen so genannten harten (CHECK) und weichen PLs (SIGNAL) unterschieden. Weiche PLs lassen es zu, dass die Fehlermeldung unterdrückt und das Interview fortgesetzt werden kann, bei harten PLs kann eine Fortführung des Interviews ohne Eingabe eines plausiblen Wertes nicht erfolgen.

Für die Durchführung der Face-to-Face-Interviews im Projekt Q-MED / LFS wurde das Blaise-Frageprogramm des Mikrozensus / LFS vor allem in den Frageformulierungen optimiert. Außerdem wurde zu jedem neuen Frageblock einen kurzer Überleitungstext formuliert, der den Befragten bereits auf wichtige Einzelheiten hinweisen kann. Des Weiteren wurde ein umfangreiches Listenheft zur visuellen Unterstützung von Fragen mit zahlreichen Antwortmöglichkeiten zur Verfügung gestellt.

Schaubild 4: Einleitungstext zum Frageblock „Erwerbstätigkeit“

Erfassung/Bearbeitung aller Interviews - Mikrozensus Projekt QMED

Datensätze Antwort Navigieren Freigabe Haupteingabe

- Person 1 - unbekannt - 37 Jahre - männlich -

Frage 19

Im nächsten Frageblock geht es um Beschäftigung.
 Uns interessieren auch kleinere bezahlte Tätigkeiten, wie z.B. von Rentnern, Hausfrauen, Schülern oder Arbeitslosen.
 Hauptsache es wurde mindestens 1 Stunde pro Woche gegen Bezahlung gearbeitet.
 Denken Sie jetzt bitte an alle Tage der letzten Woche, von Montag bis Sonntag.
 Haben Sie in der vergangenen Woche gegen Bezahlung gearbeitet oder waren Sie als Selbständiger bzw. Freiberufler tätig?

Personen in Ausbildung, Grundwehr-, Zivildienst oder im sozialen Jahr gelten als erwerbstätig.

1. Ja
 8. Nein

Person	Info1	Erwtaet	LandBetr	GeringET	SonstErw	GruNiArb	DauUnt	Lohnzahl	HinzuV	SchulJET	ETStat	FruehErw
unbekannt		8	8	1	1						1	

4.2 Das telefonische Interview (CATI)

Telefonische Befragungen werden bisher im Mikrozensus / LFS nur auf ausdrücklichen Wunsch des Haushalts durchgeführt (Mikrozensus / LFS 2008 Telefoninterview: 3%). Sie werden nur teilweise als computerunterstütztes Interview und ohne entsprechende Infrastruktur durchgeführt. Der bei CATI häufig genannte Vorteil der Interviewerkontrolle durch Supervisoren in einem Call-Center entfällt somit im Mikrozensus / LFS, da die telefonischen Interviews entweder in den Statistischen Landesämtern oder im Ausnahmefall von den Erhebungsbeauftragten von zu Hause aus durchgeführt werden. Da die telefonischen Interviews, sofern sie computergestützt erfolgen, mit dem gleichen Erfassungsprogramm durchgeführt werden wie die Face-to-Face-Interviews, können unplausible Angaben der Befragten schon im Interview durch Rückfragen korrigiert werden. Auch die automatisierte Filterführung und die Signierhilfen erhöhen hier die Datenqualität im Vergleich zur schriftlichen Befragung.

Die telefonischen Interviews im Projekt Q-MED / LFS wurden ausschließlich computer-gestützt durchgeführt. Hierfür wurde wie im Face-to-Face-Interview die Software „Blaise“ eingesetzt, wobei die Intervieweranweisungen und Sprechtexte der Interviewer/innen an die Situation einer telefonischen Befragung angepasst wurden. Lange Listen von Antwortkategorien sind in CATI schwerer zu vermitteln. Um die Befragung zu vereinfachen, wurden teilweise Vorfilter-Fragen verwendet, die es erlaubten, eine Frage mit vielen Antwortmöglichkeiten in zwei oder mehrere kürzere Fragen umzuwandeln.

4.3 Der Selbstausfüllerbogen

Der schriftliche Fragebogen in Q-MED / LFS war wie im regulären Mikrozensus / LFS als Haushaltsfragebogen aufgebaut. Zu jeder Frage gibt es pro Haushaltsmitglied ein Antwortkästchen bzw. ein

Textfeld. Filteranweisungen, die das Überspringen von Fragen anzeigen, sind farbig hinterlegt. Am rechten Rand findet sich bei Sprunganweisungen die Nummer der Frage, die als nächstes zu beantworten ist. Insgesamt gliedern sich die 160 Fragen des Fragebogens in 21 Unterabschnitte von unterschiedlicher Länge.

Schaubild 5: Beispiel aus dem Selbstausfüllerbogen

Fragen zur früheren Erwerbstätigkeit

26 **Haben Sie früher schon einmal gegen Bezahlung gearbeitet?**
Reine Gelegenheitstätigkeiten wie z. B. Ferienjobs zählen nicht dazu.

Ja 1 1 1 1 1

Nein, noch nie 8 8 8 8 8 66

27 **Wann haben Sie Ihre letzte Erwerbstätigkeit beendet?**

	1. Person	2. Person	3. Person	4. Person	5. Person
Monat	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Jahr	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

28 **Warum haben Sie Ihre letzte Erwerbstätigkeit beendet?**
Nennen Sie den Hauptgrund. (Bitte nur eine Antwort)

Entlassung 01 01 01 01 01

Befristeter Arbeitsvertrag 02 02 02 02 02

Eigene Kündigung 03 03 03 03 03

4.4 Der Online-Survey

Neben den klassischen Erhebungsinstrumenten wurde im Rahmen des Projektes Q-MED / LFS auch die Befragung als Online-Survey getestet. Zu diesem Zweck wurde mit Hilfe des Onlinetools EFS-Survey der Firma Globalpark eine Web-basierte Version des Q-MED / LFS-Fragebogens programmiert, der sich im Aufbau an dem des Selbstausfüllers orientierte, aber als Personenfragebogen angelegt wurde. Mit dem EFS-Survey-Onlinetool konnten Passwörter generiert werden, die den Haushalten in der Online-Befragung schriftlich mitgeteilt wurden. Dabei erhielt jedes Haushaltsmitglied ein eigenes Passwort für einen personenbezogenen Login. Proxy-Angaben waren möglich: Hierzu musste die auskunftgebende Person das Passwort des Haushaltsmitglieds angeben, für das die Auskunft erteilt werden sollte und dann in dessen Rolle „schlüpfen“. Die Haushaltsfragen wurden nur von einer Person beantwortet. Diese Person wurde als Ansprechpartner/in geführt und erhielt ein Passwort, das sie als erste Person identifizierte.

Der Versand der Zugangsdaten (URL und Passwort) erfolgte auf dem Postweg durch die Landesämter. Zwischen dem 18.05.2009 und dem 30.06.2009 konnten die Befragten mit ihrem persönlichen Passwort auf die Onlineversion des Q-MED / LFS-Fragebogens zugreifen. Die Landesämter erhielten wöchentlich eine aktualisierte Liste der Haushalte, die den Fragebogen bereits komplett beantwortet hatten. Auf Grundlage dieser Informationen konnten an die übrigen Haushalte gezielt Erinnerungsschreiben versendet werden. Um die Abbrecherquote so gering wie möglich zu halten, war es möglich, die Befragung jederzeit zu unterbrechen und zu einem späteren Zeitpunkt wieder aufzunehmen.

Im Onlinesurvey wurden die Probanden automatisch durch den Fragebogen geleitet. Die Befragten mussten keine Filtersprünge beachten, sondern sahen jeweils nur die Fragen, die auch beantwortet werden sollten. Es war jedoch jederzeit möglich, zu einer bereits beantworteten Frage zurückzukehren, um diese zu korrigieren und somit in einen anderen Zweig der Befragung zu wechseln. Da eine zielgerichtete Filterführung ohne vollständige Angaben der Befragten nicht möglich ist, waren alle Fragen im Onlinesurvey Pflichtfragen: Bei unbeantworteten Fragen konnte die Befragung nicht fortgesetzt werden, der Befragte wurde auf fehlende Antworten hingewiesen.

Schaubild 6: Online-Seite mit Hinweis auf unvollständige Angaben

**STATISTISCHE ÄMTER
DES BUNDES UND DER LÄNDER**

(37) Fragen zur gegenwärtigen (Haupt-)Erwerbstätigkeit

 **Eine oder mehrere Fragen sind für den weiteren Verlauf des Fragebogens wichtig.**

Wie viele Stunden arbeiten Sie normalerweise pro Woche?
Einschließlich regelmäßig geleisteter Überstunden.

Anzahl der vollen Stunden


Und wie viele Stunden haben Sie tatsächlich in der letzten Woche gearbeitet?
Einschließlich aller geleisteter Überstunden.

Anzahl der vollen Stunden

In der letzten Woche nicht gearbeitet

Bei technischen Problemen: Wenden Sie sich bitte an andreas.lingnau@destatis.de oder an unsere Hotlinenummer 0228 99 643 8821

Neben der automatischen Filterführung wurden zahlreiche Plausibilitätsprüfungen hinterlegt, die die Antworten bereits während der Eingabe auf Plausibilität und Konsistenz kontrollierten. Wurde beispielsweise als Antwort auf die Frage, in welchem Jahr die derzeitige Tätigkeit aufgenommen wurde, ein Jahr angegeben, in dem der Befragte noch keine 15 Jahre alt war, erschien eine Fehlermeldung, die den Befragten detailliert über die Art des Fehlers informierte. Der Befragte muss in diesem Fall seine Eingabe abändern um die Befragung fortsetzen zu können.

Der Online-Fragebogen bietet gegenüber dem Papierfragebogen die Möglichkeit, umfangreiche Hilfetexte zu hinterlegen. Dies wurde über die Bereitstellung von Hilfsbuttons  realisiert, die direkt zu den Fragen zur Verfügung standen. Außerdem wurde auf jeder Seite der Online-Befragung die E-Mail-Adresse eines Projektmitarbeiters als Ansprechpartner genannt sowie eine Hotline angeboten, an die sich die Befragten bei technischen Problemen wenden konnten.

5 Ausfüllverhalten in der schriftlichen Befragung

Bevor die Ergebnisse der unterschiedlichen Erhebungsinstrumente vergleichend analysiert werden, erschien es erkenntnisreich, zunächst die Besonderheiten zum Ausfüllverhalten und der Datenqualität in den nicht-interviewergestützten Befragungsformen Selbstausfüller und Online-Survey gesondert zu betrachten.

5.1 Das Ausfüllverhalten im Selbstausfüllerbogen

Im regulären Mikrozensus / LFS werden die ausgefüllten Selbstausfüllerbogen an die Statistischen Landesämter zurückgeschickt und dort zunächst auf Vollständigkeit und Plausibilitäten geprüft. Sind die Angaben weitgehend vollständig, erfolgt die Eingabe direkt in das Programm Blaise. Unnötige Angaben werden nicht eingegeben; Unplausibilitäten müssen, sofern dies ohne Rückfrage beim Haushalt möglich ist, bereits bei der Eingabe korrigiert werden. Um im Rahmen des Q-MED / LFS-Projekts zu prüfen, welche Datenqualität der schriftliche Bogen ohne Nachbereitung liefert, wurden die Fragebogen ohne Plausibilitätskontrollen und Filteranweisungen erfasst. Die unplausibilisierten Daten liefern interessante Hinweise, wie die Probanden mit den typischen Problemen der schriftlichen Befragung umgehen. Wie aus der Forschung bekannt ist, ergeben sich in der schriftlichen Befragung u. a. auch Probleme durch die Komplexität der Filterführungen. Besonders schwierig ist die Filterführung nachvollziehbar, wenn es sich wie beim Mikrozensus-Fragebogen um einen Haushaltsbogen handelt, der für mehrere Personen hintereinander - in der Praxis wohl eher parallel und nebeneinander - auszufüllen ist.

5.1.1 Viele Filterfehler bei der schriftlichen Befragung

Das Befolgen der Filteranweisungen kostet Mühen des Lesens und eine hohe Konzentration. Wird der Fragebogen für mehrere Personen gleichzeitig ausgefüllt, müssen durch Zurück- und Vor-Blättern jeweils individuelle Befragungsverläufe verfolgt werden - eine kaum lösbare Aufgabe. Daher überrascht es wenig, dass es trotz plakativer Hinweise zur Filterführung insgesamt nur sehr wenigen Probanden in der schriftlichen Befragung gelang, keine Filterfehler zu machen. In 93% der Fälle wurde mindestens ein Mal ein Filter übersehen und eine oder mehrere Frage unnötigerweise beantwortet.

Schaubild 7: Filteranweisungen im Selbstausfüllerbogen

Das Diagramm zeigt einen Ausschnitt aus einem Fragebogen mit folgenden Elementen:

- Frage 9: **Bitte ordnen Sie sich zu ...**
 - Verheiratete Person: Weiter mit Nr. 10
 - Ledige Person: Weiter mit Nr. 11
 - Sonstige Person: Weiter mit Nr. 13
- Frage 10: **Lebt Ihr/-e Ehepartner/-in in diesem Haushalt?**
 - Ja: fünf Auswahlkästchen mit der Ziffer 1.
 - Personennummer Ehepartner/-in (siehe Beiblatt): fünf Auswahlkästchen mit der Ziffer 8.
 - Nein: fünf Auswahlkästchen mit der Ziffer 8.

Am rechten Rand des Fragebogens befinden sich Sprunganweisungen (Pfeile) mit den Nummern 14 und 15. Ein Hinweisfeld erklärt:

Hinweis zu den kurzen Pfeilen:
Wenn Sie eine Antwort angekreuzt bzw. eingetragen haben, auf die am rechten Rand ein Pfeil folgt, geht es für Sie mit der Fragennummer weiter, die im Pfeil angegeben ist.

Im Durchschnitt machen die Probanden im Verlauf der Befüllung des Bogens 6,9 Filterfehler und beachteten Sprunganweisungen nicht. Am häufigsten passierten Fehler bei Haushaltsmitgliedern in

Mehr-Personen-Haushalten, die nicht als erste Person eingetragen wurden (Mittelwert: 7,3 Filterfehler). Intellektuelle Fähigkeiten helfen hier nur begrenzt: Die Probanden mit Hochschulabschluss machten durchschnittlich 6,6 Filterfehler; die Fehlerrate bei nicht-akademischen Versuchspersonen lag mit durchschnittlich 7,3 Filterfehlern aber nur geringfügig höher.

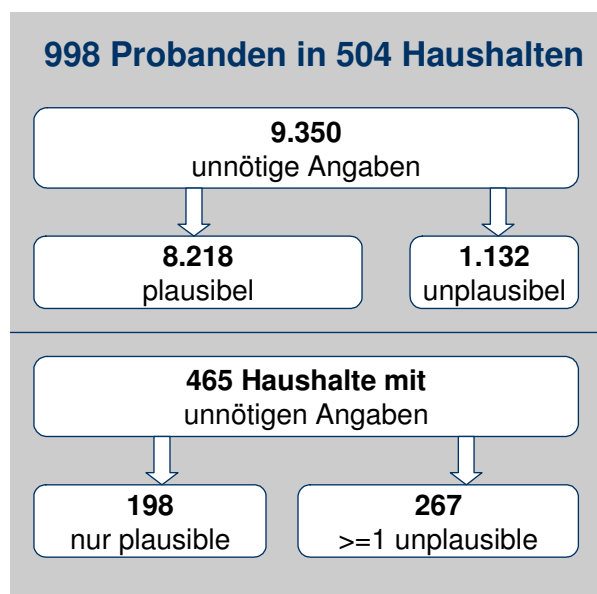
Besonders häufig kommt es zu Filterfehlern, wenn die nächste Frage plausibel bzw. beantwortbar erscheint. So übersehen zum Beispiel 53% aller Verheirateten in der Stichprobe, dass die Frage nach der Mutter (Frage 11) oder dem Vater (Frage 13) im Haushalt nur von Ledigen zu beantworten ist. Auch der Filtersprung, wenn die Frage nach einer bezahlten Tätigkeit in der Berichtswoche bejaht wurde, wird von 66% der in der Berichtswoche Erwerbstätigen ignoriert und alle weiteren Fragen zum ILO-Erwerbsstatus beantwortet.

Die Filterfehler können zu plausiblen und nicht-plausiblen Angaben führen: Wenn ein Proband die Frage nach einem weiteren Wohnsitz verneint und die Wohnung am Befragungsort in der Folgefrage als Hauptwohnung bezeichnet, kann die Angabe ignoriert werden, ohne dass die Datenqualität hierdurch beeinträchtigt würde. Auch die Angabe der Verheirateten, dass die eigene Mutter im Haushalt lebt (oder auch nicht), kann ignoriert werden. Problematisch sind Filterfehler erst dann, wenn sie zu unplausibeln oder inkonsistenten Angaben führen. Wer angibt, in der Berichtswoche erwerbstätig gewesen zu sein und den Filter übersieht und dann in einer weiteren Frage ankreuzt, normalerweise zu arbeiten, aber gerade in der Berichtswoche nicht, stellt die Erfassungskraft vor ein Problem: Welcher Aussage soll man trauen? In der Regel sind solche Unplausibilitäten ohne Nachfrage beim Haushalt nicht auflösbar.⁶

5.1.2 Plausible und unplausible Angaben durch Filterfehler

Insgesamt haben knapp 1.000 Probanden in der schriftlichen Befragung über 9.000 Angaben zu viel gemacht. Allerdings konnten über 8.000 Angaben als plausibel eingestuft werden und nur ein kleiner Teil der unnötigen Angaben beinhaltet unplausible bzw. widersprüchliche Informationen.

Schaubild 8: Plausible und unplausible Angaben nach Filterfehlern



⁶ Aus Zeit- und personellen Kapazitätsgründen wurde auf Nachfragen bei Q-MED / LFS-Haushalten verzichtet. Eine Aufklärung von Implausibilitäten war daher nicht möglich.

D. h. das Ausmaß der Filterfehler auf die Datenqualität ist als gering zu betrachten. Ärgerlich sind Filterfehler zunächst nur für den Haushalt, der zu viele Fragen beantwortet, dadurch seine Auskunftslast selbst erhöht und sich zum Teil über sich wiederholende Fragen wundert. Bricht man die Differenzierung in plausible und unplausible Angaben nach Filterfehlern auf die Ebene der Haushalte runter, so findet sich in der Mehrzahl der Haushalte mindestens eine unklare Angabe, die nur durch eine Rückfrage plausibilisiert werden könnte.

5.1.3 Fast in jedem Bogen: Fehlende Angaben bei Pflichtmerkmalen

Im Vergleich zu unnötigen Kreuzen stellen fehlende Angaben das größere Problem dar. Bei jedem Proband blieben im Durchschnitt 5,5 Fragen unbeantwortet. Umgerechnet fehlten in der Gesamtstichprobe etwa 5.400 Angaben, von denen allerdings nur ein kleiner Teil (10% der fehlenden Angaben) bei den als freiwillig gekennzeichneten Fragen fehlte. Der Großteil des Item-Non-Response bezog sich auf Fragen, die im regulären Mikrozensus / LFS der Auskunftspflicht unterliegen. Schaut man auch an dieser Stelle darauf, in welchem Ausmaß die Haushalte insgesamt fehlende Angaben aufweisen, zeigt sich, dass bei 82% aller Haushalte in der Selbstaussfüller-Stichprobe mindestens eine Angabe zu einem Pflichtmerkmal fehlte. Kann die fehlende Angabe durch Rückfragen nicht in Erfahrung gebracht werden, bleibt der Item-Nonresponse im Datensatz bestehen. Ein Item-Nonresponse der durch den Erhebungsweg hervorgerufen wird, ist zwangsläufig in einer regulären Befragung selbstselektiv und stellt ein Problem für die Datenqualität dar.

5.2 Ausfüllverhalten im Online-Survey

Durch die technische Ausstattung des Online-Surveys konnten in der web-basierten Befragung Filterfehler und fehlende Werte ausgeschlossen werden. Dennoch gibt es auch hier, interessante Befunde zum Ausfüllverhalten der Onlineteilnehmer. Das EFS-Onlinetool von Globalpark bietet die Möglichkeit, Metadaten wie Zeitstempel, Reihenfolge der aufgerufenen Seiten, Länge des Gesamtinterviews, Abbruch und Wiederaufnahme der Befragung u. ä. abzurufen. Durch die Auswertung dieser Metainformationen können wichtige Erkenntnisse über die Durchführung und mögliche technische Schwierigkeiten der Probanden generiert werden.

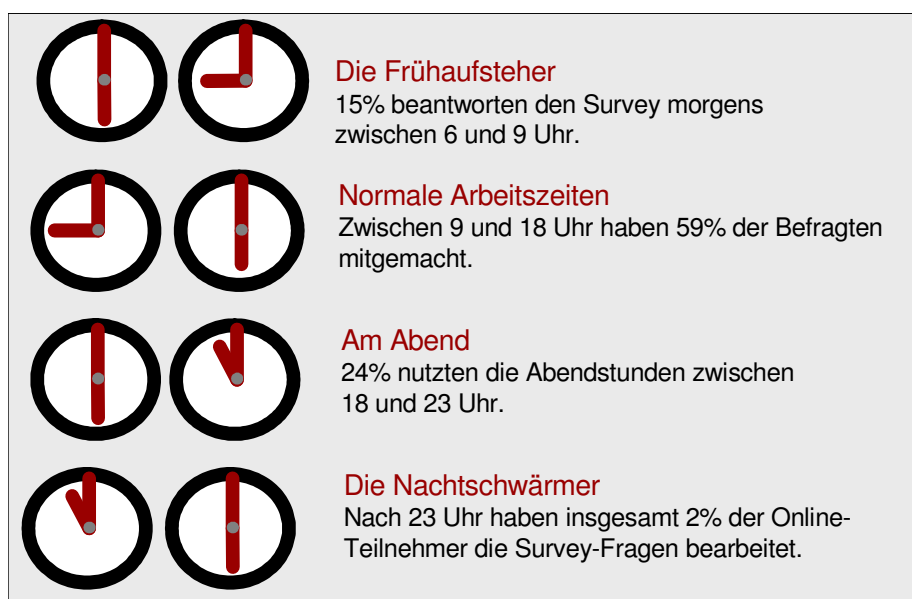
5.2.1 Rücklaufquote hoch - Abbrüche gering

Insgesamt wurden 1.108 Personen in die Q-MED / LFS-Stichprobe gezogen, von denen 1.086 den auszufüllenden Online-Fragebogen aufgerufen haben. Dies entspricht einer Rücklaufquote in der Feldphase von 98%. Von den 1.086 Personen, die den Fragebogen aufgerufen haben, brach lediglich eine Person das Interview frühzeitig ab. Die Beendigungsquote liegt somit über 99,9%. Dies ist insbesondere vor dem Hintergrund bemerkenswert, dass ein Großteil der zu beantwortenden Fragen nicht übersprungen werden konnte und die Bearbeitungsdauer häufig deutlich über der aus der Fachliteratur bekannten Empfehlung von 10 bis 15 Minuten lag (Czaja et al. 2005: 43). Die sehr hohe Rücklauf- und Beendigungsquote ist hierbei auf mehrere Faktoren zurück zu führen: (1) Erstens hatten die Teilnehmer die Möglichkeit, die Befragung jederzeit zu unterbrechen und zu einem späteren Zeitpunkt wieder aufzunehmen. Von dieser Möglichkeit haben etwa 6% der Probanden Gebrauch gemacht. (2) Zweitens ist es wahrscheinlich, dass sich die Mitwirkungsprämie auf das vollständige Ausfüllen des Fragebogens positiv ausgewirkt hat (vgl. hierzu auch Stadtmüller 2009: 182).

5.2.2 Dauer und Teilnahmezeiten der Befragung

Die durchschnittliche Bearbeitungszeit des Online-Surveys betrug 16 Minuten pro Einzelinterview. Das schnellste Interview wurde nach etwa 1,5 Minuten (!) beendet. Die längste Bearbeitungsdauer betrug 80 Minuten. In der Altersgruppe der unter 15-Jährigen⁷ lag die Spannweite zwischen 1,5 und 22 Minuten, im Durchschnitt beantworteten die jungen Teilnehmer den Survey nach 4 Minuten. Da die Zahl der zu beantwortenden Fragen auch vom Alter abhängt, ist wenig erstaunlich, dass die Älteren länger brauchten, um den Survey zu bearbeiten: Durchschnittlich 17 Minuten brauchten die 15- bis 65-Jährigen zur vollständigen Beantwortung der Fragen. Spitzenwert in dieser Altersgruppe: 63 Minuten. Die längste Bearbeitungsdauer hatte mit durchschnittlich fast 19 Minuten die Gruppe der über 65-Jährigen.

Schaubild 9: Teilnahmezeiten der Online-Befragten



Ein weiterer Vorteil von Online-Erhebungen ist die zeitliche Ungebundenheit. Die Möglichkeit, den Survey zu jeder beliebigen Tages- und Nachtzeit zu bearbeiten, wurde von den Teilnehmern auch genutzt. So entfallen zwar gut 59% der geführten Onlineinterviews auf die Zeit zwischen 9 und 18 Uhr. Fast 15% der Q-MED / LFS-Teilnehmer nutzten aber auch die frühen Morgenstunden zwischen 6 und 9 Uhr, um die Fragen zu beantworten, während ein Viertel der Befragten die Teilnahme in die Abendstunden verlegten. Nur wenige waren in der Nacht online unterwegs, um sich am Projekt Q-MED / LFS zu beteiligen.

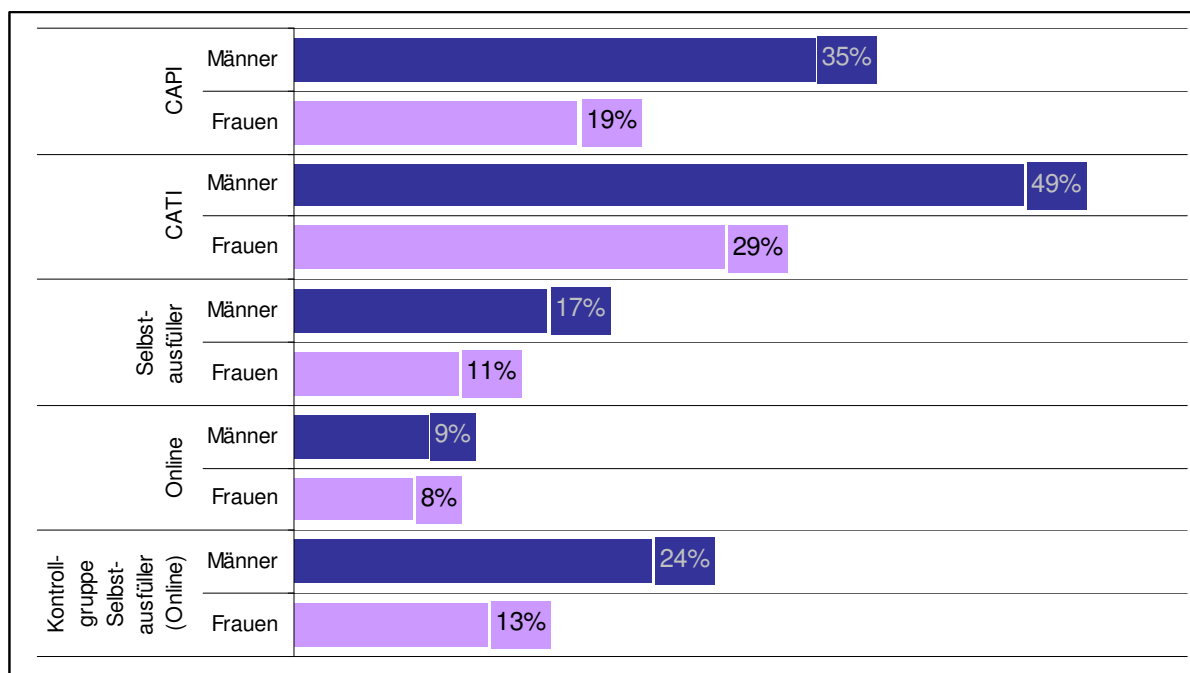
⁷ Berücksichtigt wurden hier nur Personen mit Selbstauskünften, da bei Proxy-Interviews keine Zuordnung von Alter und Ausfülldauer möglich ist.

6 Unterschiede zwischen den Erhebungsinstrumenten

6.1 Der Einfluss des Erhebungsinstruments auf die Proxy-Quote

Proxy-Interviews werden stellvertretend für Personen geführt, die nicht anwesend sind oder aus anderen Gründen das Interview nicht selbst führen können oder möchten. Aus Untersuchungen des Mikrozensus ist bekannt, dass die Proxy-Quote bei Kindern und Jugendlichen sowie bei alten Menschen sehr hoch ist. Aber auch in den mittleren Altersgruppen sind nicht immer alle Haushaltsmitglieder anwesend, so dass das Proxy-Interview zur gängigen Praxis im Mikrozensus gehört. Aus der Perspektive einer möglichst hohen Datenqualität ist der Anteil der Proxy-Interviews möglichst gering zu halten. Zwar konnte in der Nachbefragung im Mikrozensus - einem Projekt der Statistischen Ämter des Bundes und der Länder - kein direkter Zusammenhang zwischen Proxy-Angaben und der Qualität der Ergebnisse im Mikrozensus / LFS nachgewiesen werden (Köhne-Finster / Lingnau 2009: 1045), dennoch ist davon auszugehen, dass Eigenauskünfte verlässlicher als Schätzungen durch dritte Personen sind. Aus diesem Grund kann der Anteil der Proxy-Interviews als ein Indikator, der die Datenqualität unterschiedlicher Erhebungsinstrumente kennzeichnet, gewertet werden.

Schaubild 10: Proxy-Anteile nach Geschlecht und Erhebungsweg



Im Folgenden werden nur Proxy-Quoten für Personen über 14 Jahren betrachtet. Die niedrigste Proxy-Quote konnte im Rahmen des Projektes Q-MED / LFS für die Befragten in nicht interviewergestützten Interviews ermittelt werden. Die Proxy-Quoten lagen mit 14% (Selbstaussfüller) bzw. 9% (Online) deutlich unter den Quoten der anderen Befragungsformen. Die höchste Proxy-Quote wurde bei CATI-Interviews mit 38% der Probanden erreicht; bei CAPI waren es 25%. Zudem zeigen sich Unterschiede zwischen männlichen und weiblichen Befragten: deutlich häufiger wird in den persönlichen Interviewformen CAPI und CATI über Männer Fremdauskunft gegeben; möglicherweise weil Frauen tagsüber im Haushalt besser erreicht werden können. In der Onlinebefragung gehen die geschlechtsspezifischen Unterschiede deutlich zurück: Durch die zeitliche Flexibilität des Instruments ist nicht nur

die Proxy-Quote besonders niedrig; das Instrument wird auch von Frauen und Männern gleichermaßen für Selbstauskünfte genutzt.

Bei der Betrachtung des Anteils der Proxy-Interviews in Abhängigkeit vom Alter fällt auf, dass gerade junge Mitglieder der Q-MED / LFS-Stichprobe telefonisch eher selten selber Auskünfte erteilt haben. Insbesondere Personen unter 25 Jahren, die der CATI- oder CAPI-Stichprobe zugeteilt wurden, waren während des Experiments kaum direkt zu erreichen. Die Proxy-Quote lag bei den jungen Selbstausfüllern und Onlinebefragten (20- bis 24-Jährige) mit unter 30% deutlich niedriger als in den interviewer-gestützten Befragungsformen. Jugendliche und junge Erwachsenen unter 20 Jahre haben nur etwa zur Hälfte bei der Online-Befragung selbst mitgemacht. Insgesamt erreicht die Onlinebefragung in fast allen Altersgruppen die geringste Proxy-Quote, lediglich bei älteren Personen über 65 Jahre und insbesondere bei den über 75-Jährigen sind die Proxy-Quoten für CATI und CAPI etwa gleich hoch und bei den Selbstausfüllern sogar etwas niedriger.

Die Projektergebnisse belegen die eingangs formulierte These, dass schriftliche Verfahren eine höhere zeitliche Flexibilität aufweisen und hierdurch niedrigere Proxy-Quoten aufweisen. Kritisch zu sehen ist die sehr hohe Proxy-Quote bei telefonischen Interviews. Während im CAPI-Interview oftmals mehrere Haushaltsmitglieder anwesend sind, die sich möglicherweise gegenseitig bei der Beantwortung der Fragen unterstützten, findet die Kommunikation am Telefon ausschließlich zwischen Interviewer und einem Befragten statt. Denkbar sind hier Einschränkungen in der Datenqualität. Ein weiterer positiver Befund ist die sehr gute persönliche Erreichbarkeit durch das Onlineverfahren und die entsprechend niedrige Proxy-Quote. Insbesondere junge Erwachsene, die möglicherweise in einem teilnahmebereiten Haushalt lebten, dessen Teilnahme durch die Mutter oder den Vater initiiert wurde, konnten im Q-MED-Projekt mit dem Online-Verfahren besser erreicht werden.

6.2 Einfluss der Erhebungsinstrumente auf das Antwortverhalten (Mode-Effekte)

Es wurde aufgezeigt, dass die Erhebungsinstrumente einen Einfluss auf die Art der Beantwortung der Fragen (Ausfüllverhalten), auf die zeitliche Flexibilität und die Proxy-Quoten haben. Weiteres Ziel der Untersuchung ist es, zu analysieren, welchen direkten Einfluss der Erhebungsweg auf das Antwortverhalten und damit auf die Ergebnisse hat. Hierzu wurden die Daten aus der schriftlichen Befragung mit Selbstausfüllern nachträglich plausibilisiert und Filtersprünge so nachgebildet, dass unnötige Angaben zur Feststellung signifikanter Unterschiede als „missings“ unberücksichtigt blieben⁸. Durch das experimentelle Design der Untersuchung und die zufällige Verteilung der befragungsbereiten Haushalte auf die Erhebungsinstrumente weisen statistisch signifikante Ergebnisunterschiede zwischen den Gruppen auf Mode-Effekte hin.

Um statistisch signifikante von nicht-signifikanten Effekten zu unterscheiden, wurden im Folgenden für alle kategorialen Variablen ein Chi-Quadrat-Test und für alle metrischen Merkmale einfaktorielle Varianzanalysen durchgeführt. Als statistisch signifikanter Mode-Effekt gilt im Folgenden, wenn die Wahrscheinlichkeit für Unterschiede in der Grundgesamtheit mindestens bei 90% liegt bzw. die Irrtumswahrscheinlichkeit $p \leq 0,10$ beträgt. Die durch dieses Verfahren festgestellten Mode-Effekte werden im Folgenden zusammenfassend beschrieben. Sofern der Mode-Effekt sich auf Unterschiede zwi-

⁸ Sofern im Selbstausfüllerbogen fehlende Werte vorkamen, konnten diese nicht plausibilisiert oder ersetzt werden. Der Item-Nonresponse kann im Einzelfall die Verteilung der gültigen Angaben beeinflussen und zu Ergebnisabweichungen führen. An den Stellen, an denen Mode-Effekte sichtbar werden, wurde geprüft, ob ein hoher Item-Nonresponse ggf. die Unterschiedlichkeit der Ergebnisse beeinflussen könnte. In der Regel ist der Item-Nonresponse bei einzelnen Merkmalen allerdings gering.

schen CAPI und Selbstausfüllern bezieht, kann anhand der Stichproben im Mikrozensus / LFS 2008 geprüft werden, ob und in welchem Umfang dieser Effekt für die Ergebnisse aus der regulären Erhebung relevant ist. Da die Erhebungswege im regulären Mikrozensus / LFS selektiv sind, besteht bei der Überprüfung im Mikrozensus das Problem, dass nicht genau zwischen Mode-Effekten und Effekten der Selbstselektivität der Stichprobe unterschieden werden kann. Um dennoch näherungsweise Effekte im regulären Mikrozensus / LFS abschätzen zu können, wurde eine Umgewichtung vorgenommen, um eine strukturelle Annäherung der nach Erhebungsweg zuzuordnenden Gruppen zu erhalten. Dieses Verfahren wird im Folgenden skizziert.

6.2.1 Schätzung von Mode-Effekten im Mikrozensus / LFS

Wie in Kapitel 1.2 dargestellt, ergeben sich aus der ungleichen, nicht zufälligen Verteilung der schriftlichen Befragung im regulären Mikrozensus Effekte der Selbstselektion. Die ungleiche Stichprobenstruktur bei Selbstausfüllern und persönlich Befragten führt zu Ergebnisunterschieden, die nicht oder nur bedingt auf den Einfluss der Erhebungsmodi zurückzuführen sind. Eine Möglichkeit, die Effekte der Selbstselektivität zu minimieren, ist eine nachträgliche Umgewichtung der Original-Stichproben im Mikrozensus. Ziel des Verfahrens ist es, Strukturunterschiede zwischen Selbstausfüllern und persönlich Interviewten durch Umgewichtung weitgehend zu reduzieren. Die Wahl des Erhebungswegs ist im regulären Mikrozensus / LFS von unterschiedlichen Faktoren abhängig: Es gibt deutliche Unterschiede zwischen den Bundesländern, zwischen Altergruppen, zwischen Bevölkerungsgruppen mit unterschiedlichem Bildungsniveau und zwischen Großstädten und ländlichen Gebieten. Auch die Größe des Haushalts hat Einfluss auf die Erreichbarkeit und damit auf den gewählten Weg der Auskunftserteilung.

Schaubild 11: Strukturmerkmale zur Umgewichtung

Alters- klasse	unter 45 Jahren	45 - 74 Jahre	75 Jahre oder älter
Bildungs- niveau	ISCED 1 - 2	ISCED 3 - 4	ISCED 5 - 6
Haushalts- größe	1- Personen- Haushalt	Mehr- Personen- Haushalt	Umgewichtung pro Bundesland
Gemeinde- größe	unter 500 Tsd.	500 Tsd. oder mehr	

Ausgehend von den genannten Faktoren ergeben sich Strukturunterschiede zwischen den Stichproben, die sich durch eine nachträgliche Gewichtung zwar nicht vollständig ausgleichen, zumindest aber nivellieren lassen. Hierzu wird für jede Teilgruppe, die sich aus einer bestimmten Merkmalskombination ergibt (siehe Schaubild 11), ein spezifischer Eckwert ermittelt. Dieser Eckwert dient der Justierung der Selbstausfüller- bzw. CAPI-/CATI-Stichprobe im Mikrozensus / LFS. Beide Stichproben werden anhand der aus dem Mikrozensus / LFS ermittelten (hochgerechneten) Eckwerte neu gewichtet, d. h. in beiden Stichproben wird durch eine Umgewichtung eine vergleichbare Struktur erzeugt. Diese Struktur entspricht im Hinblick auf die berücksichtigten Merkmale der Grundstruktur des Mikrozensus / LFS bzw. der Grundstruktur der Bevölkerung in der Bundesrepublik Deutschland. Die nach diesem Verfahren erzeugte Projektdatei dient der Überprüfung, ob die in Q-MED / LFS festgestellten Mode-Effekte auch im Mikrozensus / LFS erkennbar sind, wenn Strukturunterschiede ausgeglichen wären.

6.2.2 Listeneffekte bei Fragen mit zahlreichen Antwortmöglichkeiten

Aus der Forschungsliteratur ist bekannt, dass insbesondere Fragen mit sehr vielen Antwortkategorien je nach Erhebungsinstrument und Darstellungsform der Kategorien (z. B. Listen, Vorlesen, Zuordnen

von frei formulierten Antworten) zu unterschiedlichen Ergebnissen führen können. Auch im Mikrozensus / LFS finden sich viele Merkmale, die sehr differenziert erhoben werden und die entsprechenden Fragen zahlreiche Antwortkategorien haben. Während der Proband in der schriftlichen Befragung alle Antwortkategorien durchlesen kann, bevor er die Antwort gibt, müssen im persönlichen Interview entweder alle Kategorien genannt (CATI), Listen⁹ verwendet (CAPI) oder freie Antworten zugeordnet werden (CAPI und CATI). Wie die Ergebnisse zeigen, führt die unterschiedliche Wahrnehmung, ob Antwortmöglichkeiten gesehen, gehört oder frei formulierte Antworten durch den Interviewer bzw. die Interviewerin zugeordnet werden, zum Teil zu deutlich unterschiedlichen Ergebnissen. Insbesondere bei Fragen mit sehr vielen Antwortkategorien sind die Antwortmöglichkeiten unübersichtlich und in der schriftlichen Befragung muss nicht unerhebliche Lesezeit darauf verwendet werden, die zutreffende Kategorie zu finden. In der Gesamtbetrachtung der Fragen mit vielen Antwortmöglichkeiten zeigt sich, dass die meisten Unterschiede zwischen CAPI-Interviews und der schriftlichen Befragung bestehen. Weniger Differenzen lassen sich zwischen CAPI- und CATI-Interviews feststellen und am geringsten sind die Unterschiede zwischen Selbstausfüllern und Online-Befragten.

Tabelle 2: Übersicht zu signifikanten Unterschieden bei Fragen mit mehr als 5 Antwortkategorien

Fragen mit mehr als 5 Antwortkategorien	Signifikanzen im Vergleich der Instrumente		
	CAPI/ Selbst- ausfüller	CAPI/ CATI	Selbst- ausfüller/ Online
EF85 Grund für Nichtarbeit in der letzten Woche			
EF92 Wichtigster Grund für die Beendigung der letzten Tätigkeit		●	
EF93 Stellung im Beruf in der letzten Tätigkeit			
EF117 Stellung im Beruf (gegenwärtige Tätigkeit)	●		
EF130 Grund für Teilzeittätigkeit		●	
EF147 Grund für weniger geleistete Arbeitszeit in der Berichtswoche ohne Feiertag mit Feiertag	●●●	●	
EF224 Grund der Arbeitsuche als Erwerbstätiger			
EF233 Grund für Nichtarbeitsuche	●		
EF243 Grund der Arbeitsuche als Nichterwerbstätiger			
EF289 Art der besuchten Schule	●		
EF310 Höchster allgemeiner Schulabschluss	●●		
EF312 Höchster beruflicher Ausbildungs-, (Fach-)Hochschulabschluss	●●●	●●●	
EF316 Welchen höchsten weiteren beruflichen Abschluss?			
EF401 Überwiegender Lebensunterhalt	●		
EF443 Erwerbsbeteiligung vor 12 Monaten			

● p < 0,1
 ●● p < 0,05
 ●●● p < 0,01

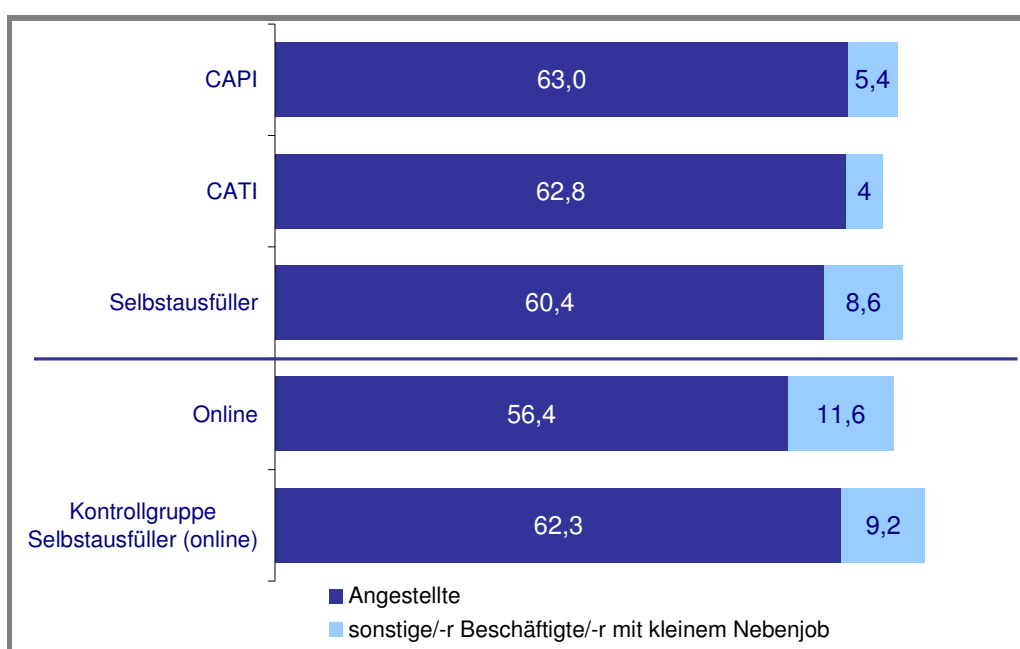
Listeneffekte bei Fragen im Themenbereich „Erwerbstätigkeit / Arbeitsuche“

Betrachtet man exemplarisch einzelne Fragen, zeigen sich unterschiedliche Muster: Die signifikanten Unterschiede zwischen CAPI-Befragten und Selbstausfüllern bei der Frage nach dem derzeit ausgeübten Beruf lassen sich auf die unterschiedlichen Anteile in der Restkategorie „sonstige/-r Beschäftig-

⁹ Für das Face-to-face-Interview erhielten die Interviewer ein Listenheft. Im Blaiseprogramm wurde an den entsprechenden Stellen auf den Einsatz des Listenheftes hingewiesen.

te/-r mit bezahltem Nebenjob“ zurückführen. Wie Schaubild 12 verdeutlicht, wird diese Kategorie bei der Frage nach der Stellung im Beruf der letzten Tätigkeit in den schriftlichen Befragungsformen häufiger gewählt. Diese letzte Kategorie wurde 2007 in den Mikrozensus eingeführt, um Personen mit kleinen Nebenjobs, denen häufig eine Zuordnung zu den Standardkategorien schwer fällt, eine Antwortmöglichkeit anzubieten (z. B. Babysitter, Nachhilfe). Sie werden normalerweise in der Datenaufbereitung umgesetzt und für alle weiteren Analysen als Angestellte betrachtet. Addiert man die Anteile dieser Restkategorie zu den Anteilen der Befragten, die sich selbst als Angestellte klassifizieren (oder von dem Interviewer bzw. der Interviewerin als Angestellte eingeordnet werden), zeigen sich keine signifikanten Unterschiede zwischen den Erhebungswegen.

Schaubild 12: Stellung im Beruf (gegenwärtige Tätigkeit)
in Prozent der gültigen Angaben



Ähnliche Phänomene lassen sich auch im regulären Mikrozensus 2008 feststellen: Die Restkategorie wird in der schriftlichen Befragung deutlich häufiger gewählt. Nach Umsetzung, d. h. Zusammenlegung mit der Kategorie „Angestellte“ jedoch bestehen keine Unterschiede zwischen den Erhebungswegen mehr.

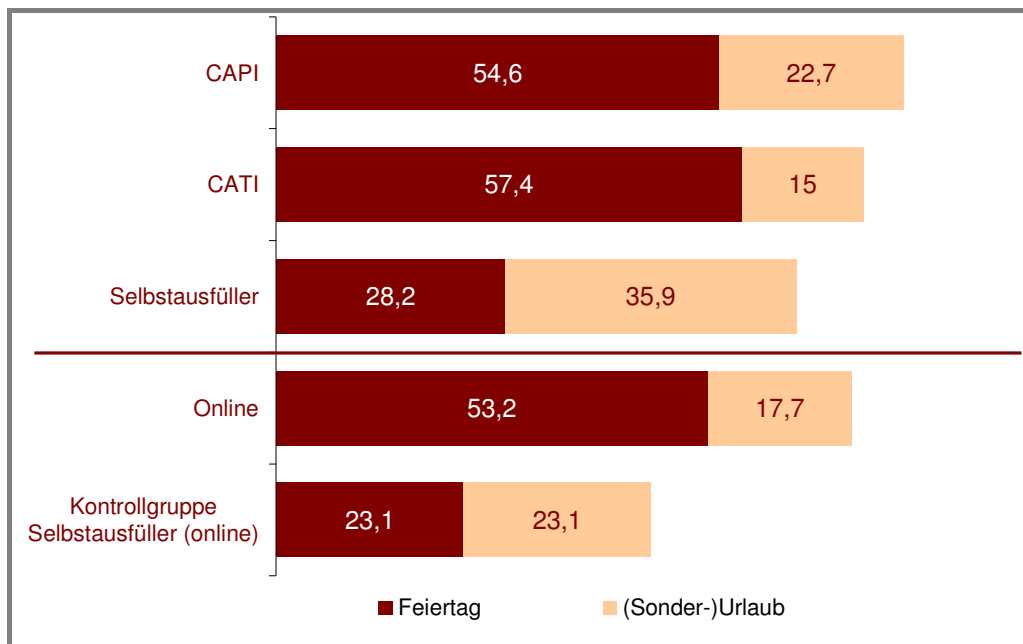
Deutlicher sind die Unterschiede zwischen CAPI-Befragten und Selbstausfüllern bei der Frage nach dem Hauptgrund für weniger geleistete Arbeitsstunden in der Berichtswoche. Alle Erwerbstätigen, die in der Berichtswoche weniger Arbeitsstunden als normalerweise geleistet haben, müssen den Hauptgrund für die weniger geleisteten Arbeitsstunden benennen - hierzu gehören u. a. auch die gesetzlichen Feiertage. Interessanterweise zeigen sich signifikante Unterschiede nur in den Berichtswochen, in denen ein Feiertag lag. Viele Arbeitnehmer/innen nehmen in Wochen mit Feiertagen auch Urlaubstage, so dass im Einzelfall nicht genau angegeben werden kann, ob der Urlaub oder ein Feiertag der Hauptgrund für die weniger geleisteten Stunden war. Es fällt jedoch auf, dass der Feiertag in der schriftlichen Befragung deutlich seltener genannt wird. Hingegen ist der Anteil derjenigen, die den Urlaub als Grund benennen, deutlich höher. Möglicherweise wird diese Kategorie in der schriftlichen Be-

fragung zuerst gesehen und als zutreffend empfunden, auch wenn der Feiertag der Hauptgrund für die weniger geleisteten Arbeitsstunden war.

Eine andere Erklärungsmöglichkeit ist, dass der Interviewer oder die Interviewerin in Feiertagswochen die Antwortkategorie aktiv als möglichen Grund anbietet und auf die Nennung aller Kategorien bzw. die Vorlage einer Liste verzichtet. Dieses Interviewerverhalten würde auch die Unterschiede zwischen CATI und CAPI erklären. In den telefonischen Interviews ist für Feiertagswochen der Anteil derjenigen, die aus „Feiertags-Gründen“ weniger gearbeitet haben, noch höher (57%) als im CAPI-Interview - entsprechend niedriger ist der Wert für den Urlaub als Hauptgrund (15%). Signifikante Unterschiede bestehen zudem zwischen Selbstausfüllern (onlinebereite Kontrollgruppe) und Onlineteilnehmern. Die Onlineteilnehmer gaben signifikant häufiger den Feiertag als Grund für weniger geleistete Stunden an (53%).

Schaubild 13: Grund für weniger geleistete Arbeitsstunden
in Wochen mit mindestens einem Feiertag
in Prozent der gültigen Angaben

(Erwerbstätige mit mindestens 1 Stunde Arbeit in der Berichtswoche)



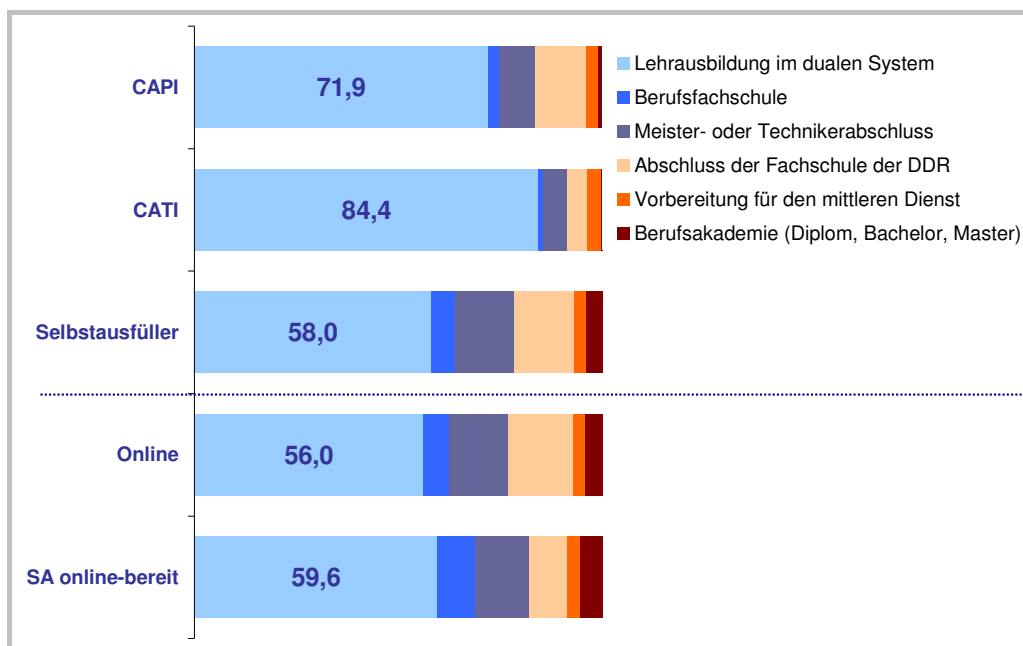
Auch im Mikrozensus 2008 zeigen sich Unterschiede: In der schriftlichen Befragung mit einem Feiertag in der Berichtswoche nannten 13% der Befragten den Feiertag als Hauptgrund für die weniger geleisteten Arbeitsstunden. In der persönlichen Befragung waren es mit etwa 50% fast viermal so viele (Nach Strukturausgleich: Selbstausfüller: 12% CAPI: 49%). Eine Quantifizierung der Unterschiede ist nicht möglich, da in der schriftlichen Befragung die Woche vor postalischem Eingang als Berichtswoche gilt. Verzögert sich die Rücksendung der ausgefüllten Bögen kann es Unterschiede zwischen der zugeordneten und tatsächlichen Berichtswoche geben.

Signifikante Unterschiede beim Grund, warum keine Erwerbstätigkeit gesucht wird (bei Nicht-Erwerbstätigen) ergeben sich aus den unterschiedlich hohen Anteilen bei der Kategorie „Dauerhaft verminderte Erwerbsfähigkeit oder Behinderung“: Im Selbstausfüller nannten 7% der nichtarbeitenden Nicht-Erwerbstätigen dies als Grund; bei den persönlich Befragten waren es bei CAPI und CATI jeweils 2%. Diese Unterschiede lassen sich allerdings in der regulären Mikrozensus-Erhebung nicht erkennen.

Die Ergebnisunterschiede bei der Art der besuchten Schule lassen sich im Mikrozensus / LFS 2008 nicht beobachten. Hier werden Förder- und Sonderschulen, Gesamt- und Waldorfschulen ebenso häufig genannt wie in der schriftlichen Befragung.

Unterschiedliche Ergebnisse in Q-MED / LFS lassen sich auch bei der Frage nach dem höchsten Schulabschluss beobachten: Hierbei fällt auf, dass die „Polytechnische Oberschule der DDR mit Abschluss der Klasse 10“ in der schriftlichen Befragung mit einem Anteil von 16% signifikant häufiger als bei CAPI (10%) oder CATI (11%) genannt wurde.¹⁰

Schaubild 15: Höchster beruflicher Abschluss
in Prozent der gültigen Angaben
(Befragte mit mittleren Berufabschlüssen)



Am deutlichsten treten die Unterschiede zwischen den Erhebungswegen bei den beruflichen Abschlüssen hervor: Bei den mittleren beruflichen Abschlüssen ist in der mündlichen Befragung signifikant häufiger die Lehrausbildung im dualen System anzutreffen. Im CAPI-Interview beträgt der Anteil 72% aller mittleren Abschlüsse; bei CATI sind es sogar 84%. Die Angaben in der schriftlichen Befragung zeigen eine größere Differenzierung: Hier sind es „nur“ 58% mit einer Lehrausbildung im dualen System. Deutlich häufiger nennen die Selbstausfüller zum Beispiel einen Abschluss an einer Berufsfachschule (schulische Ausbildung), Abschluss einer Meister- oder Technikerschule oder den Abschluss an einer Fachschule der DDR. Ähnlich differenziert wie beim Selbstausfüller sind auch die Angaben zum höchsten beruflichen Abschluss aus der Online-Befragung.

Im Bereich der akademischen Berufe lassen sich ähnliche Tendenzen zur Hauptkategorie im persönlichen Interview nachweisen: Höhere Anteile in der mündlichen Befragung bei „normalen“ Universitätsabschlüssen (CAPI: 52%, CATI: 53%, Selbstausfüller: 45% an allen akademischen Abschlüssen), ge-

¹⁰ Die Differenzierung des Abschlusses einer Polytechnischen Oberschule in Abschlüsse nach der 8. Klasse und Abschlüsse nach der 10. Klasse wurde erst mit der Erhebung 2010 eingeführt. Daher ist es möglich, dass die routinierten Interviewer/innen die neue Kategorie noch nicht bewusst wahrgenommen haben.

ringere Anteil bei Promovierten (CAPI: 4%, CATI: 1%, Selbstaufüller: 9% an allen akademischen Abschlüssen).

Auch im regulären Mikrozensus / LFS lassen Unterschiede zwischen den Erhebungswegen auf starke Mode-Effekte bei den beruflichen Abschlüssen schließen: In der schriftlichen Befragung wird doppelt so oft ein berufsqualifizierender Abschluss an einer Berufsfachschule genannt (Mikrozensus / LFS 2008: CAPI, CATI: 2%, Selbstaufüller: 4% an allen mittleren Abschlüssen) und auch Meister- und Technikerabschlüsse (CAPI, CATI: 10%, Selbstaufüller: 13%) oder Abschlüsse einer Fachschule der DDR (CAPI, CATI: 2%, Selbstaufüller: 3%) kommen bei schriftlich Befragten häufiger vor. Weitere Mode-Effekte sind bei den akademischen Abschlüssen erkennbar: in der schriftlichen Befragung lag der Anteil der promovierten Akademiker/innen 2008 bei 11%; im mündlichen Interview waren es nur 7%. Auch nach einer Strukturgewichtung bleiben die Unterschiede bestehen. Unterstellt man den Wert der schriftlichen Befragung als näherungsweise „wahren“ Wert, würden beispielsweise in der mündlichen Befragung etwa 250.000 Doktorinnen und Doktoren untererfasst.

Listeneffekte bei Fragen im Themenbereich „Einkommen und Lebensunterhalt“

Im Themenbereich Einkommen und Unterhalt wird bei der Frage nach dem überwiegenden Lebensunterhalt eine Liste mit sieben möglichen Antwortkategorien angeboten. Signifikante Unterschiede zeigen sich an dieser Stelle nur zwischen CAPI-Interviews und Selbstaufüllern. Die Befragten in der schriftlichen Befragung gaben etwas häufiger an, überwiegend von eigener Erwerbstätigkeit (CAPI: 45%, Selbstaufüller: 43%) oder von Rente, Pension zu leben (CAPI: 29%, Selbstaufüller: 27%), hingegen wird etwas seltener der Unterhalt, zum Beispiel durch Familienangehörige, genannt (CAPI: 23%, Selbstaufüller: 19%). Möglicherweise handelt es sich hier um eine Art Primacy-Effekt, bei dem die Frage nur flüchtig zur Kenntnis genommen und die zuerst plausibel erscheinende Antwort gewählt wird.

Der Mikrozensus / LFS zeigt ähnliche Unterschiede: In der schriftlichen Befragung ist der Anteil der Personen, die überwiegend von Erwerbstätigkeit leben (CAPI, CATI: 41%, Selbstaufüller: 49%), höher - geringer ist der Anteil bei denjenigen, die hauptsächlich von privaten Unterstützungen leben (CAPI, CATI: 28%, Selbstaufüller: 26%). Nach einer Strukturumgewichtung beider Stichproben verringern sich die Abstände zwar; das Grundmuster bleibt jedoch bestehen. Unterstellt man, dass in der schriftlichen Befragung die Erwerbstätigkeit als Haupteinkommensquelle überschätzt wird und legt man den Wert aus der persönlichen Befragung als plausibleren Wert zugrunde, wäre die Erwerbstätigkeit als Haupteinnahmequelle etwa mit 1 Mio. Personen überschätzt.

Zwischenfazit: Listeneffekte bei Fragen mit zahlreichen Antwortkategorien

Die Analyse der Listeneffekte hat zwei unterschiedliche und teilweise gegenteilige Befunde als Ergebnis geliefert: Bei den Fragen, bei denen aus Sicht der Befragten mehrere Antworten plausibel erscheinen, kann der Interviewer oder die Interviewerin Hilfestellung geben und ggf. nochmals nachfragen. In der schriftlichen Befragung wird im Zweifelsfall tendenziell eher die zuerst plausibel erscheinende Antwort gewählt - zum Beispiel Urlaub, Sonderurlaub (6. Position im Fragebogen) statt Feiertag (10. Position) bei der Frage nach dem Grund für weniger Arbeitsstunden in der Berichtswoche.

Bei Fragen mit relativ eindeutigen Antwortkategorien ist dieser Primacy-Effekt nicht erkennbar. Deutliche Ergebnisunterschiede sind hierbei nicht auf die Reihenfolge der Antwortkategorien zurückzuführen.

ren. Es ist zu vermuten, dass Ergebnisunterschiede sowohl in Q-MED / LFS als auch im regulären Mikrozensus / LFS darauf zurückzuführen sind, dass der Interviewer bzw. die Interviewerin die Listen nicht konsequent verwendet hat, sondern freie Antworten der Probanden zuordnete. Antwortet der oder die Befragte zum Beispiel auf die Frage nach dem beruflichen Abschluss mit „Ausbildung“, wählt der Interviewer oder die Interviewerin vermutlich den zuerst plausibel erscheinenden bzw. den häufigste Wert, in diesem Fall die „Lehrausbildung im dualen System“. Dies führt in der Konsequenz zu einer unzulässigen Standardisierung der Ergebnisse und einer deutlich eingeschränkten Datenqualität bei der Abbildung differenzierter Entwicklungen und Verteilungen.

6.2.3 Effekte bei Fragen mit Mehrfachantworten

Bei insgesamt sieben Fragen werden mehrere Items in einer Frage mit Mehrfachantwortmöglichkeit zusammengefasst. In der schriftlichen Befragung sind die entsprechenden Fragen durch den Hinweis „Kreuzen Sie bitte alle zutreffenden ... an“ gekennzeichnet. Im persönlichen Interview sollte der Interviewer eine Liste vorlegen (CAPI) oder alle Antwortkategorien vorlesen (CATI).

Bemühungen bei der Arbeitsuche

Arbeitsuchende können sich auf unterschiedliche Art und Weise um Stelle bemühen. In der Befragung konnten die Probanden aus einem Fragekatalog von neun Einzelposten zutreffende Suchbemühungen auswählen.

Tabelle 3: Übersicht zu signifikanten Unterschieden bei Fragen mit Mehrfachantworten
Suchbemühungen bei der Arbeitsuche

Fragen mit Mehrfachantworten Suchbemühungen bei der Arbeitsuche	Signifikanz im Vergleich der Instrumente		
	CAPI/ Selbst- ausfüller	CAPI/ CATI	Selbst- ausfüller/ Online
Filterfrage: EF247 Aktive Arbeitsuche			●
EF254 Suche über die Agentur für Arbeit			
EF255 Suche über private Vermittler			
EF256 Aufgabe von Inseraten		● ●	●
EF257 Bewerbung auf Inserate			
EF258 Bewerbung auf eine nicht ausgeschriebene Stelle			
EF259 Suche über persönliche Verbindung	● ● ●		
EF260 Durchsehen von Inseraten		● ● ●	
EF261 Tests, Vorstellungsgespräche, Prüfungen			
EF262 Sonstige Bemühungen		● ●	

Signifikante Unterschiede zwischen CAPI-Befragten und Selbstausfüllern gibt es lediglich bei dem Item „Suche über persönliche Verbindungen“, das von Befragten in der schriftlichen Befragung deutlich häufiger gewählt wird (CAPI: 38%; Selbstausfüller 46%). Möglicherweise wurde die Antwortmöglichkeit im persönlichen Interview nicht angeboten und nur dann gewählt, wenn sie - wie im Selbstausfüller - selbst gelesen wurde. Bei insgesamt drei Items ließen sich Unterschiede zwischen CAPI und CATI feststellen: deutlich häufiger gaben CATI-Befragte an, selbst Stellenanzeigen aufzugeben (CATI: 21%, CAPI 7%) und seltener wurden Anzeigen durchgesehen (CATI: 59%, CAPI: 80%). Möglicherweise wurden im Telefoninterview die Antwortkategorien versehentlich vertauscht. Ein weiterer Unter-

schied besteht bei den „Sonstigen Bemühungen“. Auch diese Kategorie wird im CATI-Interview signifikant seltener gewählt (CATI: 13%, CAPI 29%). Auch hier ist zu vermuten, dass bei Nennung mindestens einer der anderen Kategorien „Sonstige Bemühungen“ am Telefon nicht explizit nachgefragt wurde.

Bei der Auswertung der Online-Ergebnisse fällt auf, dass es bereits bei der Filterfrage, ob überhaupt aktiv nach Arbeit gesucht wird, signifikante Unterschiede bestehen: Die arbeitssuchenden Onlineteilnehmer antworten zu 23%, dass sie nicht aktiv suchen, während es in der Kontrollgruppe nur 9% sind.

Schaubild 16: Frage zur aktiven Arbeitsuche in der Online-Befragung

Haben Sie in den letzten 4 Wochen eine andere oder zusätzliche Tätigkeit gesucht?
Jede Art der Suche ist gemeint.
Zum Beispiel auch:

- Suche über Freunde, Bekannte, Verwandte
- Durchsehen von Stellenanzeigen

Ja

Nein

Vermutlich bietet die Einsicht in die Vielzahl der Suchbemühungsmöglichkeiten, wie sie der Selbstausfüller gewährleistet, einen anderen Stimulus als lediglich der Hinweis, dass auch „niedrig schwellige“ Suchbemühungen wie die Suche über Freunde oder das Durchsehen von Stellenanzeigen gemeint sind. Wenn die Filterfrage bejaht wurde, kamen Onlineteilnehmer und Probanden der Kontrollgruppe bei der Auswahl von Suchbemühungen zu vergleichbaren Ergebnissen. Lediglich die „Aufgabe von Stellenanzeigen“ wurde von Online-Teilnehmern etwas häufiger genannt. Möglicherweise ist dieser Unterschied darauf zurückzuführen, dass die Antwortkategorien gerade beim Online-Verfahren nur überflogen werden und die erste Kategorie, die den Begriff „Stellenanzeige“ enthält, gewählt wird. Zählt man insgesamt die Zahl der genannten Suchbemühungen ergeben sich folgende Werte: Durchschnittlich werden zwischen 2,77 bei arbeitssuchenden CAPI-Befragten und 3,35 Suchbemühungen bei Onlineteilnehmern genannt. Die Unterschiede sind gering und nicht signifikant.

Die in Q-MED / LFS festgestellten Unterschiede zwischen CAPI-Befragten und Selbstausfüllern bei der „Suche über private Kontakte“ lassen sich in der regulären Erhebung Mikrozensus / LFS 2008 nicht nachweisen.

Bezug von Renten und Pensionen

Im Themenblock „Unterhalt und Einkommen“ werden Mehrfachantwort-Fragen dazu verwendet, um unterschiedliche Einkommensarten abzufragen. Bei den Hinterbliebenen-Renten gibt es keine signifikanten Unterschiede bei den Anteilen der Personen, die mindestens eine Hinterbliebenen-Rente beziehen. Differenzen zeigen sich jedoch bei den Angaben, welche Rentenart bezogen wurde. Bei CAPI-Befragten mit Hinterbliebenenrente wird fast ausschließlich die erste Kategorie - „Rente aus der Deutschen RV Bund“ - gewählt (98%). Bei den schriftlich Befragten sind es nur 83%. Bei der Kategorie „Sonstige öffentliche Rente“ ist es entsprechend umgekehrt: Diese Restkategorie wird von 12% der schriftlich Befragten mit Hinterbliebenenrente gewählt und von keinem in der mündlichen Befragung. Die CATI-Befragten weisen ebenfalls bei der Kategorie „Rente aus der Deutschen RV Bund“ etwas niedrigere Anteile (87%) als die CAPI-Befragten aus. Dafür gaben signifikant mehr telefonisch Befragte an, eine Hinterbliebenenrente aus einer Unfallversicherung zu beziehen. Auch im Onlineverfahren wurde die Hinterbliebenenrente etwa doppelt so häufig genannt wie in der schriftlichen Befragung (Online: 5%, Vergleichsgruppe: 2%). Insgesamt ist die Teilgruppe der Probanden mit Hinterbliebenenrente allerdings so gering, dass sich weitere Unterschiede nicht sinnvoll interpretieren lassen.

Schaubild 17: Frage zum Bezug von Renten und Pensionen (Selbstaussfüller)

<p>106 Beziehen Sie eine oder mehrere eigene (Versicherten-) Rente/-n, Pension/-en? Betriebsrenten sind hier nicht gemeint. Kreuzen Sie bitte alles Zutreffende an.</p> <p>Ja, und zwar ...</p> <p>aus der Deutschen Rentenversicherung Bund (früher BfA, LVA) oder der Deutschen Rentenversicherung Knappschaft-Bahn-See</p> <p>eine öffentliche Pension</p> <p>eine Kriegsopferrente</p> <p>aus der Unfallversicherung</p> <p>Rente aus dem Ausland</p> <p>eine sonstige öffentliche Rente</p> <p>Nein</p>	<input type="checkbox"/> 01	<input type="checkbox"/> 01	<input type="checkbox"/> 01	<input type="checkbox"/> 01	<input type="checkbox"/> 01
	<input type="checkbox"/> 02	<input type="checkbox"/> 02	<input type="checkbox"/> 02	<input type="checkbox"/> 02	<input type="checkbox"/> 02
	<input type="checkbox"/> 03	<input type="checkbox"/> 03	<input type="checkbox"/> 03	<input type="checkbox"/> 03	<input type="checkbox"/> 03
	<input type="checkbox"/> 04	<input type="checkbox"/> 04	<input type="checkbox"/> 04	<input type="checkbox"/> 04	<input type="checkbox"/> 04
	<input type="checkbox"/> 05	<input type="checkbox"/> 05	<input type="checkbox"/> 05	<input type="checkbox"/> 05	<input type="checkbox"/> 05
	<input type="checkbox"/> 06	<input type="checkbox"/> 06	<input type="checkbox"/> 06	<input type="checkbox"/> 06	<input type="checkbox"/> 06
	<input type="checkbox"/> 88	<input type="checkbox"/> 88	<input type="checkbox"/> 88	<input type="checkbox"/> 88	<input type="checkbox"/> 88

Deutlich größer ist die Gruppe in allen Vergleichsgruppen, die eine eigene Rente beziehen. Dabei gibt es den auffälligsten Unterschied zwischen CAPI-Befragten und Selbstaussfüllern bei der Anzahl der Personen, die eine öffentliche Rente beziehen.

Tabelle 4: Übersicht zu signifikanten Unterschieden bei Fragen mit Mehrfachantworten
Bezug von Renten und Pensionen

Fragen mit Mehrfachantworten Bezug von Renten/ Pensionen	Signifikanzen im Vergleich der Instrumente		
	CAPI/ Selbst- ausfüller	CAPI/ CATI	Selbst- ausfüller/ Online
Hinterbliebenenrente oder -pension (mindestens eine Nennung)			
EF403 Bezug aus der Deutschen Rentenversicherung Bund			
EF406 Bezug von Öffentlicher Pension			
EF407 Bezug von Kriegsopferrente			
EF408 Bezug von Rente aus Unfallversicherung			
EF409 Bezug von Rente aus dem Ausland			
EF410 Bezug von sonstiger öffentlicher Rente			
Eigene Rente oder Pension (mindestens eine Nennung)			
EF412 Bezug aus der Deutschen Rentenversicherung Bund			
EF415 Bezug von Öffentlicher Pension			
EF416 Bezug von Kriegsopferrente			
EF417 Bezug von Rente aus Unfallversicherung			
EF418 Bezug von Rente aus dem Ausland			
EF419 Bezug von sonstiger öffentlicher Rente			

p < 0,1
 p < 0,05
 p < 0,01

Bei der CAPI-Befragung gaben 29% der Probanden mindestens eine eigene Rente, Pension an. In der schriftlichen Befragung ist der Anteil mit knapp 15% nur etwa halb so hoch. Unterschiede zwischen der Art der Bezüge gibt es zwischen CAPI-Befragten und Selbstausfüllern jedoch nicht. Es dominiert der Bezug von Renten der Deutschen Rentenversicherung oder bei einem geringeren Teil auch der Bezug einer öffentlichen Pension. Bei den online-bereiten Haushalten ist der Anteil der Rentner/-innen mit eigenen Versorgungsansprüchen mit 9% deutlich geringer als in den anderen Gruppen. Unterschiede zwischen Online-Teilnehmern und der Kontrollgruppe zeigen sich jedoch nur bei wenigen Kategorien.

Der wichtigste Befund in diesem Themenblock ist, dass bei Selbstausfüllern deutlich weniger Rentner/-innen mit eigenen Rentenbezügen nachgewiesen werden konnten. Möglicherweise entsteht durch umfangreiche Texte in der Frage, der Erläuterung und der ersten Antwortkategorie eine Unübersichtlichkeit, die zur Folge hat, dass sich einige Rentner/-innen nicht wiederfinden und daher „Nein“ ankreuzen.

Auch im regulären Mikrozensus / LFS ist der Anteil der Personen mit eigenem Rentenbezug mit 16% deutlich niedriger als in der persönlichen Befragung (Selbstausfüller: 16%; CAPI, CATI: 25%). Berücksichtigt man Strukturunterschiede und gewichtet Alter, Bildungsniveau und Haushaltgröße nach, bleibt ein Abstand von 5 Prozentpunkten bestehen (Mikrozensus / LFS 2008 nach Ungewichtung: Selbstausfüller: 14%; CAPI, CATI: 19%). Unterstellt man, dass es in der schriftlichen Befragung zu einer Unterschätzung der Rentner/-innen kommt und der Wert der mündlichen Befragung nach Strukturausgleich näherungsweise dem „wahren“ Wert entspricht, würden etwa 500.000 Personen als Rentner/-innen mit eigenen Versorgungsansprüchen unterschätzt.

Bezug von öffentlichen Zahlungen

Auch der Bezug von öffentlichen Zahlungen wird durch eine Frage mit Mehrfachantwortmöglichkeiten erfragt. Hierbei gibt es insgesamt im Vergleich der Erhebungsinstrumente nur wenige Unterschiede. Etwa 22% der Probanden in der CAPI- und CATI-Stichprobe sowie bei den Selbstausfüllern erhielten mindestens eine öffentliche Zahlung. Bei den onlinebereiten Haushalten war es mit 28% aufgrund des höheren Anteils von Studierenden etwas mehr. Unterschiede bei der Angabe, welche öffentlichen Transferleistungen in Anspruch genommen werden, zeigen sich zwischen CAPI-Befragten und Selbstausfüllern nur beim Bezug von Leistungen nach Hartz IV (CAPI: 22%, Selbstausfüller: 14%). Hierbei betreffen die Unterschiede nur Haushalte mit mehreren Personen, bei denen in der schriftlichen Befragung häufig der Hartz IV-Bezug nur der ersten Person zugeordnet wird. In den mündlichen Interviews werden hingegen signifikant häufiger Hartz IV-Bezüge auch den anderen Personen des Haushalts zugeordnet.

Bei allen anderen Bezugsarten zeigen sich kaum Unterschiede: Lediglich die Gruppe der Onlineteilnehmer gab etwas häufiger an, sonstige öffentliche Leistungen wie zum Beispiel Kindergeld oder BAföG zu beziehen (85% aller Online-Probanden mit mindestens einer öffentlichen Transferleistung). In der Vergleichsgruppe der onlinebereiten Selbstausfüller waren es zum Vergleich nur 78%.

Auch an dieser Stelle zeigt sich, dass Abweichungen weniger aus der Art der Mehrfach-Frage resultieren, sondern daraus, dass der Interviewer oder die Interviewerin über spezifisches Wissen verfügt und bei der Beantwortung der Fragen Hilfestellungen geben kann.

Tabelle 5: Übersicht zu signifikanten Unterschieden bei Fragen mit Mehrfachantworten
Bezug von öffentlichen Zahlungen

Fragen mit Mehrfachantworten Bezug von öffentlichen Zahlungen	Signifikanz im Vergleich der Instrumente		
	CAPI/ Selbst- ausfüller	CAPI/ CATI	Selbst- ausfüller/ Online
Bezug von öffentlichen Zahlungen (mindestens eine Nennung)			
EF423 Arbeitslosengeld I			
EF424 Leistungen nach Hartz IV (ALG II, Sozialgeld)	●●		
EF422 Laufende Hilfe zum Lebensunterhalt			
EF421 Wohngeld			
EF439 Elterngeld			
EF426 Sonstige öffentliche Zahlungen, z.B. Kindergeld, BAföG u.ä.			●●
EF425 Leistungen nach Pflegestufe 1 - 3/ Härtefall			

● p < 0,1
●● p < 0,05
●●● p < 0,01

Im regulären Mikrozensus / LFS 2008 ist diese Hilfestellung durch den Interviewer bzw. die Interviewerin in Hartz IV-Haushalten nicht erkennbar. Korrekterweise müssten alle Personen in einer Bedarfsgemeinschaft Leistungen nach Hartz IV zugeordnet werden, wenn dies bei einer Person der Fall ist. In der schriftlichen Befragung gelten 66% der Probanden als Hartz IV-Empfänger, wenn mindestens ein weiteres Haushaltsmitglied Leistungen nach Hartz IV bezieht. In der mündlichen Befragung sind es nur 58%. Diese Differenzen bleiben auch nach einer Strukturangleichung der Stichproben bestehen.

Bezug von anderen Einkommen

Der Bezug von anderen Einkommen wird signifikant häufiger in der schriftlichen Befragung (29%) als bei CAPI (23%) angegeben. Dies geht vor allem auf Unterschiede bei der Angabe von Einkommen aus eigenem Vermögen, Zinsen (CAPI: 48%, Selbstausfüller: 58% aller Probanden mit anderen Einkommen) und Leistungen aus Lebensversicherungen, privaten Rentenversicherungen zurück (CAPI: 0%, Selbstausfüller: 7%). Auch in der telefonischen Befragung werden Versicherungsleistungen als Einkommensquelle häufiger genannt (CATI: 3%). Die Onlineteilnehmer geben im Vergleich zur Kontrollgruppe etwas seltener andere Einkommen an (Online: 24%, Kontrollgruppe: 28%). Auffällig ist hier, dass im Online-Survey besonders häufig eine Betriebsrente angegeben wurde (Online: 20%, Kontrollgruppe: 7%). Geringer ist hingegen der Anteil der Probanden, die Einkommen aus Vermietung und Verpachtung (Online: 25%, Kontrollgruppe: 34%) beziehen oder eine private Unterstützung, Unterhalt (Online: 16%, Kontrollgruppe: 24%) erhalten.

Andere Einkommen werden im regulären Mikrozensus / LFS - im Gegensatz zu den Q-MED / LFS-Ergebnissen - häufiger von Personen angegeben, die mündlich befragt wurden (CAPI, CATI: 4%, Selbstausfüller: 2%). Allerdings werden Einkommen aus Vermögen und Zinsen nur von 3% aller mündlich Befragten, die über zusätzliche Einkommen verfügen, angegeben. In der schriftlichen Befragung sind es mit 13% etwa viermal so viele.

Unterstellt man den Anteilswert aus der schriftlichen Befragung als näherungsweise „wahren“ Wert würden etwa 140.000 Angaben zu Einkommen aus Vermögen und Zinsen unterschätzt. Auch Leistungen aus Lebens-, privaten Rentenversicherungen werden im Papierbogen im Vergleich zum mündlichen Interview mit 3% doppelt so oft genannt.

Zwischenfazit: Fragen mit Mehrfachantworten

Die Gesamtanalyse der Fragen mit Mehrfachantworten zeigt, dass die Abweichungen insgesamt ein geringeres Ausmaß haben, als dies bei Fragen mit zahlreichen Antwortkategorien, aus denen nur eine ausgewählt werden kann (Listeneffekt) der Fall ist. Dennoch gibt es interessante Hinweise auf Modeeffekte durch Unterschiede bei einzelnen Items: Antwortkategorien, die eher selten sind - wie zum Beispiel die „Arbeitsuche über private Kontakte“, - werden in der schriftlichen Befragung deutlich häufiger genannt. Es ist zu vermuten, dass nicht allen Probanden entsprechende Listen vorgelegt oder alle Antwortkategorien vorgelesen wurden und daher Antworten, die nicht häufig auftauchen, auch im mündlichen Interview seltener gegeben werden, da der Proband diese Antwort nicht von sich aus anbietet. Aus ähnlichen Gründen können möglicherweise auch bestimmte Einkommensarten wie z. B. Einkommen aus Vermögen und Zinsen beim Selbstausfüller häufiger auftauchen, wenn diese Kategorie durch die schriftliche Form dem oder der Befragten „in den Sinn“ kommt. Voraussetzung für die schnelle Wahrnehmung aller Antwortkategorien in der schriftlichen Befragung ist eine übersichtliche Darstellung. Sind einzelne Antwortkategorien eher „textlastig“ (wie zum Beispiel bei der Frage nach eigenen oder Hinterbliebenenrenten), kann es dazu kommen, dass in der schriftlichen Befragung keine Antwort oder eine Nein-Antwort gegeben wird, da die zutreffende(n) Kategorie(n) nicht mit einem Blick erfasst werden können.

Um mit unterschiedlichen Erhebungswegen vergleichbare Ergebnisse zu erzielen, sollte daher bei Listenfragen und Fragen mit Mehrfachantworten in der schriftlichen Befragung besonderes Augenmerk auf eine übersichtliche Gestaltung der Frage gelegt werden. Unterschiede zwischen Selbstausfüllern und CAPI, CATI können nur vermieden werden, wenn konsequent Listen zum Einsatz kommen oder

alle Antwortkategorien einzeln vorgelesen werden. Nur so kann gewährleistet werden, dass der Proband versteht, welche Antwortmöglichkeiten gegeben werden können.

6.2.4 Fragen mit Restkategorien - „Sonstige und andere Gründe“

Bei einer Reihe von Fragen im Mikrozensus / LFS wird nach Ursachen für bestimmte Verhaltensweisen („Aus welchem Grund suchen Sie eine Arbeit?“) oder Situationen („Warum haben Sie in der vergangenen Woche nicht gearbeitet?“) gefragt. So vielfältig und unterschiedlich die Verhaltensweisen und Situationen sein können, so divers sind auch die Ursachen und Gründe dafür. Zur Vereinfachung werden die Probanden an vielen Stellen gebeten, nicht alle Gründe zu nennen, sondern sich auf einen Hauptgrund festzulegen. Angeboten werden als Antwortkategorien die Gründe, die besonders häufig zutreffen. Da aber immer auch individuelle Ursachen für bestimmte Verhaltensweisen oder das Zustandekommen von Situationen eine Rolle spielen können, finden sich am Ende vieler Antwort-Liste so genannte Restkategorien, in denen sich alle einordnen können, die die übrigen Kategorien als nicht zutreffend empfinden.

Tabelle 6: Übersicht zu signifikanten Unterschieden bei Fragen mit Restkategorien
Sonstige und andere Gründe

Fragen mit Restkategorien Sonstige und andere Gründe	Signifikanz im Vergleich der Instrumente		
	CAPI/ Selbst- ausfüller	CAPI/ CATI	Selbst- ausfüller/ Online
EF85 Grund für Nichtarbeit in der letzten Woche			
EF92 Wichtigster Grund für die Beendigung der letzten Tätigkeit			
EF126 Grund der Befristung der Tätigkeit			
EF130 Grund für Teilzeittätigkeit		●●●	
EF148 Grund für Mehrarbeit in der Berichtswoche	●●●	●●●	
EF147 Grund für weniger geleistete Arbeitszeit in der Berichtswoche			
EF139 Hauptgrund, warum man innerhalb von 2 Wochen nicht mehr arbeiten kann			
EF224 Grund der Arbeitsuche als Erwerbstätiger			
EF233 Grund für Nichtarbeitsuche			
EF243 Grund der Arbeitsuche als Nichterwerbstätiger			
EF277 Grund der Nichtverfügbarkeit innerhalb von 2 Wochen			

● p < 0,1
 ●● p < 0,05
 ●●● p < 0,01

Wie die Übersicht zeigt, gibt es erstaunlich wenig Unterschiede zwischen den Erhebungsmodi. Lediglich bei der Frage, warum eine Person in der Berichtswoche mehr als normalerweise gearbeitet hat, greifen die Befragten in der schriftlichen Befragung und am Telefon doppelt so häufig (CAPI: 23%, Selbstausfüller 44%; CATI 48%) auf die Antwortkategorie „Sonstige Gründe“ zurück.

Schaubild 18: Frage zum Hauptgrund für Mehrarbeit in der Berichtswoche (Selbstaussfüller)

47a **Warum haben Sie in der vergangenen Woche mehr Stunden als normalerweise gearbeitet?**
 Nennen Sie den Hauptgrund. (Bitte nur eine Antwort)
 Erläuterung [6] „Überstunden“ auf Seite 45.

Ausgleich für zu wenig geleistete Arbeitsstunden an anderen Arbeitstagen, -wochen (z. B. gleitende Arbeitszeit) 1

Bezahlte Überstunden 2

Unbezahlte Überstunden 3

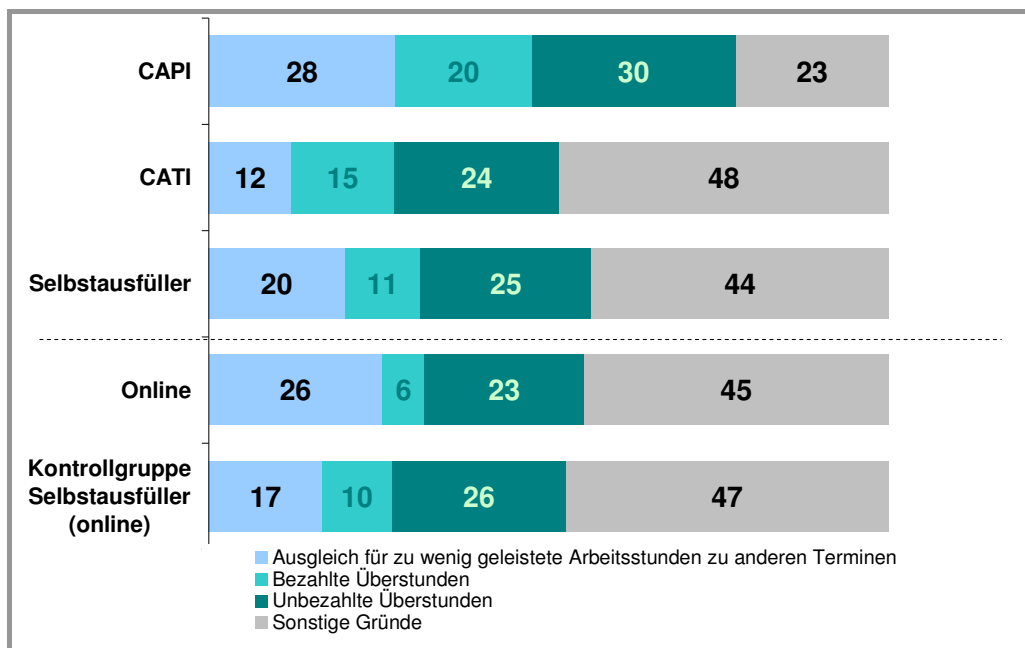
Sonstige Gründe 4

mit Nr. ...

48

Hier sind die Antwortkategorien ohne Interviewerhilfestellung schwer verständlich. Zudem werden möglicherweise viele Probanden den Grund für die Mehrarbeit in der Arbeit selbst sehen (z. B. „weil viel Arbeit zu erledigen war“) und weniger in der gefragten Verrechnungs- und Überstundenlogik und daher auf die Restkategorie „Sonstige Gründe“ zurückgreifen.

Schaubild 19: Hauptgrund für Mehrarbeit in der Berichtswoche in Prozent der Befragten



Auch im telefonischen Interview wird bei einzelnen Fragen häufiger auf die Restkategorie „Sonstige oder andere Gründe“ zurückgegriffen. Bei der Frage nach dem Hauptgrund für die Ausübung einer Teilzeittätigkeit finden sich zum Beispiel bei der telefonischen Befragung mehr Probanden in den Kategorie „Sonstige familiäre oder persönliche Gründe“ (CAPI: 11%, CATI: 17%) oder bei „Vollzeittätigkeit aus anderen Gründen nicht gewünscht“ (CAPI: 28%, CATI: 38%). Fast genauso häufig wird bei der Frage nach dem Grund für Mehrarbeit in der Berichtswoche im CATI-Interview „Sonstige Gründe“ angegeben. In der Online-Befragung zeigen sich keine signifikanten Unterschiede zum schriftlichen Verfahren mit Papierfragebogen. Die Restkategorie „Sonstige oder andere Gründe“ wird in der Online-Befragung genauso häufig oder selten gewählt wie bei den Selbstausfüllern.

Der Mode-Effekte bei dem Hauptgrund für mehr geleistete Stunden in der Berichtswoche ist auch im regulären Mikrozensus / LFS deutlich erkennbar. In der schriftlichen Befragung lag der Anteil der Kategorie „Sonstige Gründe“ bei 60%, während im mündlichen Interview nur 20% die Restkategorie gewählt hatten. Entsprechend unterschiedlich sind die übrigen Verteilungen: Den Ausgleich von zu wenig geleisteten Stunden an anderen Terminen nannten beim Selbstausfüller nur wenige als Hauptgrund (CAPI, CATI: 34%, Selbstausfüller: 8%). Und auch bei den bezahlten Überstunden (CAPI, CATI: 23%, Selbstausfüller: 14%) genauso wie bei den unbezahlten Überstunden (CAPI, CATI: 24%, Selbstausfüller: 16%) sind deutliche Mode-Effekte erkennbar.

Auch nach einer Strukturanpassung zwischen den Stichproben bleiben die Abstände zwischen den Erhebungswegen bestehen. Geht man davon aus, dass im Selbstausfüller die Kategorie „Sonstige Gründe“ stark überbesetzt ist, weil die Begriffe „Mehrarbeit“ und „Überstunden“ erläuterungsbedürftig sind und geht man weiter davon aus, dass im persönlichen Interview die „wahre“ Verteilung näherungsweise richtig gemessen wird, würden sich hochgerechnet in der schriftlichen Befragung etwa 1 Mio. Beschäftigte zu viel in der sonstigen Kategorie finden und entsprechend zu wenige Beschäftigte bei den anderen Kategorien.

Zwischenfazit: Fragen mit Restkategorien

In der Gesamtanalyse der Fragen mit so genannten Restkategorien - hier „Sonstige und andere Gründe“ - zeigt sich, dass Restkategorien kein prinzipielles Methodenproblem darstellen und in den meisten Fällen nicht zu mode-spezifischen Effekten führen. Dies gilt allerdings nur unter der Bedingung, dass sich ein Großteil der Befragten in den übrigen Kategorien wiederfindet und die Restkategorie tatsächlich nur dazu dient, einzelnen Befragten, die sich den Hauptkategorien nicht zuordnen können, eine Antwortmöglichkeit zu geben.

Wie das Beispiel „Grund für mehr geleistete Stunden in der Berichtswoche“ zeigt, führen unklare oder schwer verständliche Begriffe dazu, dass in der schriftlichen Befragung die „Sonstigen Kategorie“ überdurchschnittlich häufig genutzt wird. Hier sind zum Teil starke Mode-Effekte erkennbar und die Vergleichbarkeit der Stichproben ist deutlich eingeschränkt. Eine bedingte Vergleichbarkeit ist auch zu erwarten, wenn beim telefonischen Interview die Hauptkategorien nicht ausreichend erklärt werden (können) und daher die zuletzt genannte Kategorie überproportional stark besetzt ist.

6.2.5 Fragen mit möglichen Effekten zur „Sozialen Erwünschtheit“

Bei interviewergestützten Befragungsformen besteht immer die Möglichkeit, dass Phänomene „Sozialer Erwünschtheit“ das Antwortverhalten des Probanden beeinflussen und statt der „wahren“ Antwort eine Antwort gegeben wird, von der der Befragte glaubt, dass er durch diese Antwort den Erwartungen des Interviewers oder der Interviewerin entspricht. Dieses Phänomen gilt als eines der wichtigsten Interviewer-Effekte und ist besonders bei heiklen Fragen eine mögliche Fehlerquelle. Im Gegensatz zu Einstellungsfragen handelt es sich bei den Sachverhalten im Mikrozensus / LFS überwiegend um Faktenfragen, die zwar auch Effekten sozialer Erwünschtheit unterliegen können, insgesamt aber sicherlich robuster gegenüber derartigen Einflüssen sind. Bei einigen wenigen Fragen kann jedoch durchaus vermutet werden, dass der Proband „gefällig“ antwortet und die Antwort gibt, von der er glaubt, dass sie erwünscht ist.

Tabelle 7: Übersicht zu signifikanten Unterschieden bei Fragen mit möglichen Effekten zur sozialen Erwünschtheit

Fragen mit möglichen Effekten durch "Soziale Erwünschtheit"	Signifikanzen im Vergleich der Instrumente		
	CAPI/ Selbst- ausfüller	CAPI/ CATI	Selbst- ausfüller/ Online
EF231 Berufliche oder bezahlte Tätigkeit gesucht - Nichterwerbstätige			
EF235 Wunsch nach Erwerbstätigkeit bei Nichtarbeitsuche - Nichterwerbstätige	●●●		●●
EF247 Aktive Suche innerhalb der letzten vier Wochen			●
EF276 Verfügbarkeit für eine neue Arbeitsstelle innerhalb von 2 Wochen			

Exemplarisch werden im Folgenden vier Fragen untersucht, bei denen es um die Arbeitsuche geht. Bei keiner der Fragen lassen sich Phänomene sozialer Erwünschtheit nachweisen. Signifikante Unterschiede bestehen sogar in umgekehrter Erwartung: In der schriftlichen Befragung wurde die Frage „Auch wenn Sie keine Erwerbs- oder Berufstätigkeit suchen, würden Sie trotzdem gerne arbeiten?“ doppelt so häufig mit „Ja“ beantwortet als im persönlichen Interview (CAPI: 12%, Selbstaufüller: 24%). Bei allen anderen Fragen zeigen sich keine signifikanten Unterschiede im Antwortverhalten. Weder bei der Frage, ob Nichterwerbstätige eine Arbeit suchen, noch dabei, ob es aktive Suchbemühungen gibt oder die befragte Person dem Arbeitsmarkt innerhalb von zwei Wochen zur Verfügung steht, hat der Interviewer bzw. die Interviewerin einen sichtbaren Einfluss auf die Ergebnisse.

[Ergebnisunterschiede gibt es hier allerdings zwischen Online-Befragten und Selbstaufüllern. Online-Teilnehmer geben signifikant häufiger an, dass sie gerne arbeiten würden, auch wenn sie keine Arbeit suchen (Online: 42%, Kontrollgruppe: 27%). Hingegen stehen sie etwas seltener dem Arbeitsmarkt innerhalb der nächsten zwei Wochen zur Verfügung (Online: 77%, Kontrollgruppe: 90%). Diese Unterschiede sind nach gegenwärtigem Analysestand nicht aus der unterschiedlichen Gestaltung der Erhebungsinstrumente erklärbar.]

Auch der reguläre Mikrozensus / LFS weist bei der Frage, ob eine Arbeit gewünscht wird, auch wenn keine aktive Arbeitsuche angezeigt wurde, Mode-Effekte auf: in der schriftlichen Befragung gaben knapp 8% aller Befragten, die nicht erwerbstätig sind und keine Arbeit suchen, an, dass sie trotzdem gerne arbeiten würden. Bei den persönlich Befragten waren es nur knapp 5%. Auch nach Strukturumgewichtung bleibt der Abstand zwischen den Stichproben bestehen.

Unterstellt man, dass hier ein spezifischer Effekt des persönlichen Interviews vorliegt, und eher der Wert der schriftlichen Befragung den genaueren Umfang der „Stillen Reserve“ anzeigt, wären die nichtarbeitsuchenden Nichterwerbstätigen, die jedoch trotzdem gerne arbeiten würden, um etwa 750.000 Personen unterschätzt.

Zwischenfazit: Soziale Erwünschtheit

Die Gesamtbewertung von Effekten der sozialen Erwünschtheit muss dahingehend relativiert werden, dass das durchgeführte Forschungsdesign nicht primär darauf ausgerichtet war, Interviewer-Effekte - insbesondere Effekte sozialer Erwünschtheit - aufzuspüren. Allerdings geben die Ergebnisse keinen Hinweis darauf, dass dieser Effekt eine wesentliche Rolle spielt. Zumindest wären bei einzelnen Fragen, bei denen soziale Erwünschtheit eine Rolle spielen könnte, Ergebnisunterschiede zwischen Interviewer-gestützten Erhebungsformen und schriftlichen Verfahren zu erwarten. Anhand der Q-MED /

LFS-Daten lassen sich diese Effekte jedoch nicht aufzeigen. Dass es keine signifikanten Unterschiede zwischen den Stichproben gibt, die auf Effekte sozialer Erwünschtheit schließen lassen, bedeutet jedoch nicht, dass es diese Effekte nicht doch geben könnte. In amtlichen Befragungen, also bei Auskünften die der Staat einholt, werden ggf. die Effekte auch in der schriftlichen Befragung wirksam (z. B. Angabe aktiver Arbeitsuche von Hartz IV-Empfängern), die aus der Befürchtung resultieren, die Daten würden weitergegeben oder mit anderen Informationen der öffentlichen Verwaltung abgeglichen.

6.2.6 Effekte bei der Angabe von Zahlenwerten (Arbeitsstunden)

An einigen Stellen müssen die erwerbstätigen Befragten Angaben zu Arbeitsstunden machen: Zu normalerweise geleisteten Arbeitsstunden, zu geleisteten Stunden in der Berichtswoche, zu geleisteten Überstunden sowie zu der Gesamtzahl der gewünschten Wochenarbeitsstunden, falls die aktuelle Wochenarbeitszeit ausgeweitet oder verringert werden möchte.

Die stärksten Mode-Effekte gibt es zwischen CAPI-Befragten und Selbstausfüllern: Bei allen betrachteten Fragen geben die Befragten in der schriftlichen Befragung einen höheren Wert an. Die normale Arbeitszeit wird von Selbstausfüllern mit einem durchschnittlichen Wert von 35,3 Stunden angegeben. Bei CAPI und CATI sind es mit 33,4 bzw. 34,0 Stunden signifikant weniger. Auch bei der Angabe der tatsächlichen Arbeitszeit in der Berichtswoche sind die Werte in der schriftlichen Befragung im Durchschnitt höher: in den Wochen ohne Feiertage gaben die schriftlich Befragten an, falls sie in der Woche nicht frei hatten, durchschnittlich 35,9 Stunden gearbeitet zu haben. Bei den CAPI- und CATI-Befragten sind es mit jeweils 33,3 Stunden deutlich weniger. Betrachtet man die Wochen, in denen ein Feiertag lag und berücksichtigt nur Erwerbstätige, die in der Woche auch tatsächlich gearbeitet haben, so reduziert sich die durchschnittliche Anzahl der tatsächlich geleisteten Stunden bei CAPI-Befragten auf 29,4 und bei CATI-Befragten auf 30,9 Stunden. Nur in der Gruppe der Selbstausfüller bleibt der Durchschnittswert konstant bei 36 Stunden.

Tabelle 8: Übersicht zu signifikanten Unterschieden bei Fragen mit Eintrag von Zahlenwerten (Arbeits- und Überstunden)

Fragen mit Eintrag von Zahlenwerten Arbeits- und Überstunden	Signifikanz im Vergleich der Instrumente		
	CAPI/ Selbst- ausfüller	CAPI/ CATI	Selbst- ausfüller/ Online
EF131 Normalerweise geleistete Stundenzahl pro Woche	●●		●
EF134 Tatsächlich geleistete Stundenzahl in der Berichtswoche	●●		●
			●
			●
EF162 Anzahl der Überstunden	●●●		
EF133 Stundenzahl bei gewünschter Ausweitung der Wochenarbeitszeit	●		●
EF140 Stundenzahl bei gewünschter Reduzierung der Wochenarbeitszeit		●	●●

● p < 0,1
 ●● p < 0,05
 ●●● p < 0,01

Deutliche Unterschiede gibt es auch bei Frage nach den Überstunden. Die Befragten in der schriftlichen Befragung gaben im Durchschnitt 2,77 Überstunden an. Bei den persönlich Befragten waren es im CAPI-Interview durchschnittlich 1.0 und im telefonischen Interview 1,2 Stunden. Abweichungen zwischen Online-Teilnehmern und der Kontrollgruppe der onlinebereiten Selbstausfüller zeigen, dass

bei den normalerweise und tatsächlich geleisteten Arbeitsstunden die Durchschnittswerte aus der Onlinebefragung niedriger sind und eher dem Niveau der Ergebnisse aus der CAPI- und CATI-Befragung entsprechen. Nur bei der Frage nach den geleisteten Überstunden haben die Onlinebefragten ebenso hohe Werte (Mittelwert der Überstunden im Onlinesurvey 2,57 Std.) wie die Kontrollgruppe.

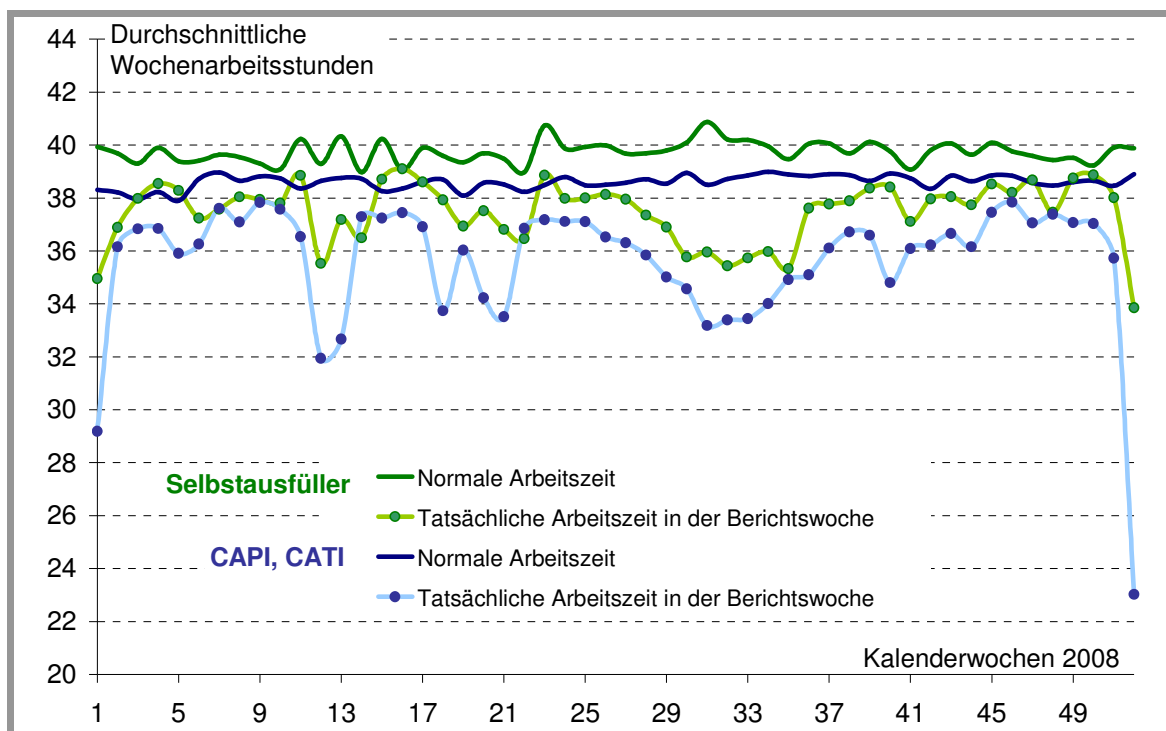
Die Fragen zu den Arbeitszeiten und die abweichenden Ergebnisse verweisen möglicherweise darauf, dass hier Mode-Effekte entstehen, weil die Begriffe „tatsächliche“, „normalerweise“ und „Überstunden“ durch den Interviewer oder die Interviewerin erklärt werden. In der schriftlichen Befragung wird die Interpretation der Begriffe dem Befragten selbst überlassen. Dies führt an dieser Stelle zu offenkundig unplausiblen Angaben. Lag in der Berichtswoche ein Feiertag reduzierte sich die Zahl der tatsächlichen Stunden gegenüber den normalerweise geleisteten Stunden bei CAPI- und CATI-Befragten durchschnittlich um etwa 3,8 Stunden; bei den Selbstausfüllern waren es weniger als 0,3 Stunden.

Auch der Mikrozensus / LFS 2008 weist entsprechende Unterschiede bei der Zahl der Stunden auf. Insgesamt ist erkennbar, dass in der regulären Erhebung im Selbstausfüllerbogen deutlich mehr wöchentliche Arbeitsstunden angegeben werden, sowohl bei den normalerweise geleisteten Stunden als auch bei den tatsächlichen Stunden in der Berichtswoche. Warum allerdings der Papierbogen zu einer Angabe von höheren Stundenzahlen „verleitet“, kann aus keinen anderen Untersuchungen abgeleitet werden und muss an dieser Stelle vorerst unerklärt bleiben.

Es ist aber davon auszugehen, dass bei der Ermittlung der tatsächlichen Stunden der Interviewer oder die Interviewerin durch Erläuterungen den Unterschied zwischen normalerweise geleisteten Stunden und den tatsächlich in der Berichtswoche geleistete Stunden besser vermitteln kann.

Wie Schaubild 20 verdeutlicht, haben Feiertage und Ferienzeiten beim mündlichen Interview einen größeren Einfluss auf die Differenz zwischen normalerweise und tatsächlich geleisteten Arbeitszeiten. Die tatsächliche Arbeitszeit geht sehr viel stärker zurück als dies bei den Befragten in der schriftlichen Befragung der Fall ist. Am deutlichsten ist dieser Effekt am Ende des Jahres zur Weihnachtszeit zu erkennen.

Schaubild 20: Normalerweise und tatsächlich geleistete Arbeitsstunden
Mikrozensus 2008



Zwischenfazit: Zahlenwerte / Arbeitstunden

Abweichungen bei der Angabe von Zahlenwerten zu den Arbeitstunden sind hier weniger auf die Schwierigkeit der Abschätzung von Werten oder gegebenenfalls Rundungsfehler zu erklären. Der Mode-Effekt wird hier durch einen positiven Interviewer-Effekt ausgelöst. Was sind normalerweise geleistete Stunden, was tatsächliche und wann ist eine mehr geleistete Stunde eine Überstunde? Bei diesen Angaben ist es hilfreich, Verständnisprobleme während der Befragung zu (er-)klären. Nur so kann begründet werden, warum die durchschnittlichen Stundenzahlen zwischen schriftlicher und mündlicher Befragung sowohl in der Q-MED / LFS-Stichprobe als auch im regulären Mikrozensus / LFS gleichermaßen voneinander abweichen.

Der nicht unerhebliche Mode-Effekt führt allerdings dazu, dass unterjährige Schwankungen bei den tatsächlichen Arbeitsstunden wahrscheinlich nicht ausschließlich konjunkturell bedingt sind, sondern auch davon abhängen, wie hoch der Anteil von Selbstausfüllern in der Stichprobe ist. Auch der Vergleich zwischen einzelnen Bundesländern ist eingeschränkt, wenn die Zusammensetzung des Mixed Mode in den länderspezifischen Erhebungen so unterschiedlich ausfällt.

7 Zusammenfassung

In der folgenden Zusammenfassung geht es um eine systematische Zusammenstellung der zentralen Befunde. Dabei stehen folgende Fragen im Vordergrund: (1) Was sind die instrumentenspezifischen Vor- und Nachteile? Mit welchem Erhebungsinstrument kann welche Datenqualität erlangt werden? (2) Welche Bedeutung haben Mode-Effekte für die Ergebnisse im Mikrozensus / LFS?

7.1 Vor- und Nachteile einzelner Erhebungswege

Auf Basis der vorhandenen Forschungsliteratur und anhand der vorgestellten empirischen Analysen zu Mode-Effekten in Q-MED / LFS konnten einzelne Vor- und Nachteile der Erhebungsinstrumente aufgezeigt werden. Dabei muss beachtet werden, dass nur ein Ausschnitt der Realität und damit auch nur ein Ausschnitt der Mode-Effekte betrachtet wurde. Aus den Ergebnissen des Experiments Q-MED / LFS ist nicht ableitbar, welche Vor- und Nachteile einzelne Erhebungsinstrumente haben, wenn sie unter nicht freiwilligen Bedingungen eingesetzt werden. Es ist davon auszugehen, dass die Q-MED-Haushalte in hohem Maße motiviert und auskunftsfreudig waren. Von dieser Ausgangslage ist nicht bei allen Haushalten im regulären Mikrozensus / LFS auszugehen. Daher bleibt eine nicht abzuschätzende Größe, welche Effekte die Erhebungsinstrumente haben, wenn keine Freiwilligkeit vorliegt und der Haushalt eine - aus seiner Sicht - „lästige Pflicht“ erfüllen muss.

Trotz dieser Einschränkung hat das Experiment interessante Ergebnisse zu Tage gefördert, die nicht nur einen wichtigen Beitrag zur Bewertung der unterschiedlichen Erhebungswege leistet, sondern auch Impulse liefert, den Mikrozensus / LFS zur Minimierung von Mode-Effekten weiterzuentwickeln.

7.1.1 Bewertung: Das CAPI-Interview

Als „Königsweg“ gilt das persönliche Interview, bei dem geschulte Interviewer/innen standardisierte Befragungen durchführen. Die Vorteile liegen auf der Hand: Bei schwierigen Sachverhalten kann der Interviewer oder die Interviewerin zusätzliche Erklärungen geben, Nachfragen sind möglich - von beiden Seiten. Wenn keine sensiblen Themen angesprochen werden - und dies ist im Mikrozensus / LFS

weitgehend der Fall -, sind Interviewereffekte kaum vorhanden oder in ihrem Ausmaß vertretbar, denn es überwiegen weitere Vorteile: Die Befragungsdaten sind schnell verfügbar, Plausibilisierungen erfolgen während des computer-gestützten Interviews, es werden durch automatische Filterführungen nur die Fragen angezeigt, die zu stellen sind usw.

Die aus Theorie und Praxis bekannten Vorteile haben dazu geführt, dass das Face-to-Face-Interview seit Anbeginn des deutschen Mikrozensus / LFS das am häufigsten eingesetzte Instrument ist. Die Ergebnisse des Q-MED-Experiments liefern zahlreiche Belege für die These, dass die durch die persönliche Befragung erhobenen Daten präziser und genauer sind als die Ergebnisse einer schriftlichen Befragung in Papierform (Selbstaussfüller). Es konnte zum Beispiel nachgewiesen werden, dass seltener „Sonstige Gründe“ gewählt werden, wenn ein Interviewer oder eine Interviewerin Frage und Antwortmöglichkeiten genau erklären. Auch das Beispiel der normalerweise und tatsächlich geleisteten Arbeitsstunden zeigt, dass der Interviewer bzw. die Interviewerin durch die Möglichkeit, bei Verständnisproblemen Erklärungen zu liefern, einen wichtigen Beitrag zur Datenqualität leistet. Angesichts der Komplexität der behandelten Themen und der herausragenden Bedeutung des Mikrozensus / LFS für nationale Analysen, aber auch für internationale Vergleiche, bleibt es bei dem Befund: Das persönliche Interview ist und bleibt für weite Teile der Bevölkerung der Königsweg, um eine hohe Datenqualität zu erlangen.

Diese hohe Datenqualität (und die Rechtfertigung der damit verbundenen hohen Ausgaben) kann allerdings nur erreicht werden, wenn die Interviews professionell und standardisiert durchgeführt werden. Hier zeigen sich Mängel, die jedoch weniger dem Interviewerverhalten anzulasten wären, als vielmehr Resultat eines Zielkonflikts sind: Datenproduzenten und Datennutzer haben hohe Anforderungen an die Qualität und Differenziertheit der Daten, der auskunftspflichtige Haushalt ist vor allem daran interessiert, die Befragung schnell und unkompliziert „zu erledigen“. Der Interviewer oder die Interviewerin hat dabei die Aufgabe zwischen beiden Zielen abzuwägen. (Aus einem eigenen persönlichen Interesse kann er oder sie sich möglicherweise eher mit der Interessenlage des Haushalts identifizieren.)

Obwohl die im Projekt Q-MED / LFS eingesetzten Erhebungskräfte die Anweisung erhielten, alle Fragen und Erläuterungstexte im Wortlaut vorzulesen und entsprechend vorbereitete Listen zu verwenden, zeigen die Ergebnisse, dass dieser Anweisung nicht immer Folge geleistet wurde. Nicht anders ist zu erklären, dass bei Merkmalen mit vielen Ausprägungen - wie zum Beispiel bei den beruflichen Abschlüssen - in der schriftlichen Befragung (Selbstaussfüller oder Online-Survey) eine weitaus höhere Differenzierung erreicht wird; sowohl im Experiment Q-MED / LFS als auch im Mikrozensus / LFS. Dieses und weitere Beispiele (Art der besuchten Schule, höchster Schulabschluss) verweisen damit auf eine grundlegende Problematik oder auch Widersprüchlichkeit des Mikrozensus / LFS selbst. Mit der Diversifizierung und Pluralisierung von Lebenslagen und Beschäftigungsverhältnissen wird es immer wichtiger, differenzierte Daten über diese gesellschaftlichen und wirtschaftlichen Entwicklungen zu erheben. Komplexitätsfortschritte der Gesellschaft stehen dabei in linearem Zusammenhang zur Anforderung an die Komplexität (und Differenziertheit) der Daten, die diese gesellschaftliche Entwicklung adäquat nachzeichnen sollen.

Dies steht aber im Widerspruch zu den begrenzten Möglichkeiten einer Haushaltsbefragung wie dem Mikrozensus / LFS. Würden alle Fragen tatsächlich im Wortlaut vorgelesen und zu allen Fragen mit mehr als drei Antwortkategorien Listen vorgelegt, würde die Befragung ein völlig anderes Ausmaß annehmen. Statt der geschätzten 10 bis 20 Minuten pro Haushalt müsste für ein Interview eine wesentliche längere Zeit veranschlagt werden. Es ist fraglich, ob (1) eine zeitliche Ausweitung der Interviews bei der Mehrzahl der Haushalte auf Akzeptanz stoßen würde und (2) für die bislang verausgabten Finanzmittel eine vollständig standardisierte Form der Befragung finanzierbar wäre und (3) ausreichend

Erhebungsbeauftragte gefunden würden, die eine stark standardisierte und zeitlich ausgeweitete Befragung der Haushalte durchführen könnten.

Mit Blick auf die Datenqualität und die Notwendigkeit, über Differenzierungsprozesse der Gesellschaft auch differenzierte Daten zu erheben, ist die Standardisierung des CAPI-Interviews jedoch alternativlos. Mode-Effekte des persönlichen Interviews resultieren den Ergebnissen des Q-MED-Experiments zufolge vor allem aus der teilweise standardisierten Zuordnung von frei formulierten Antworten. Mit der Konsequenz, dass die im persönlichen Interview erreichte Differenzierung bei einzelnen Fragen mit zahlreichen Antwortmöglichkeiten hinter der Qualität der schriftlichen Befragung zurückbleibt. In der Konsequenz führt dies zu einer deutlichen Einschränkung der Vergleichbarkeit der Daten, die aus schriftlichen und mündlichen Interviews gewonnen werden. Dies gilt insbesondere, wenn die Zusammensetzung des Mixed Mode im Mikrozensus / LFS räumlich und/ oder zeitlich stark variiert.

7.1.2 Bewertung: Das CATI-Interview

Die telefonische Befragung hat im deutschen Mikrozensus / LFS im Vergleich zu anderen Mitgliedsstaaten bislang einen geringen Stellenwert. Telefoninterviews werden derzeit ausschließlich „passiv“, d. h. nur auf ausdrücklichen Wunsch des auskunftspflichtigen Haushalts durchgeführt und sind mit 3% aller Interviews eher von nachrangiger Bedeutung. Das Experiment Q-MED / LFS hat gezeigt, dass das Mikrozensus / LFS-Frageprogramm prinzipiell „telefonfähig“ ist. Weder die befragten Haushalte noch die Mitarbeiter/innen der Statistischen Landesämter, die an der telefonischen Befragung beteiligt waren, berichteten von besonderen Schwierigkeiten während des Interviews. Die Unterschiede in den Ergebnissen der CAPI- und CATI-Interviews sind insgesamt gering und weniger auffällig als die Unterschiede zwischen CAPI-Interviews und Selbstausfüllern. Der Vorteil des CATI-Interviews liegt in der vergleichsweise günstigeren Erhebungsweise, hat aber den Nachteil, dass keine Listen verwendet werden können und der Interviewer oder die Interviewerin noch stärker als im CAPI-Interview auf die Zuordnung frei formulierter Antworten angewiesen ist. Wie die Ergebnisse des Experiments Q-MED / LFS belegen, geht die telefonische Befragung teilweise zu Lasten einer Differenzierung der Antworten: Bei den beruflichen Abschlüssen wurden in der CAPI-Stichprobe 84% der Probanden mit mittleren beruflichen Abschluss einer Lehrausbildung im dualen System zugeordnet. In der CATI-Stichprobe waren es „nur“ 72% und bei den Selbstausfüllern 58%. Werden die Telefoninterviews nicht standardisiert durchgeführt, besteht hier noch stärker als im persönlichen Interview die Gefahr, dass Differenzierungsprozesse nicht ausreichend abgebildet werden können. Als weiterer Nachteil ergibt sich im telefonischen Interview, dass nonverbale Reaktionen des Befragten nicht erfasst werden können. Verständnisschwierigkeiten, die sich zum Beispiel in einem Stirnrunzeln äußern, werden am Telefon nicht wahrgenommen. Der Interviewer kann sich nur auf die verbalen Reaktionen des Befragten verlassen. Die Q-MED / LFS-Daten zeigen, dass es zum Teil zu einer stärkeren Besetzung der letztgenannten Kategorie kommt. Zum Beispiel bei schwierigen Fragen, wie der Frage nach geleisteten Überstunden wird signifikant häufiger die „Sonstige-Kategorie“ in der telefonischen Befragung gewählt. Außerdem fühlen sich die Befragten, dies zeigen andere Untersuchungen, am Telefon eher zeitlich unter Druck gesetzt und geben möglicherweise irgendeine Antwort, auch wenn die Frage nicht ausreichend verstanden wurde. Dadurch können nicht alle Vorteile der interviewer-gestützten Befragung im telefonischen Interview realisiert werden.

Bei einem Telefoninterview, das nach den Regeln des standardisierten Interviews durchgeführt wird, müssten alle Antwortmöglichkeiten zunächst vorgelesen werden, bevor der Befragte eine Antwort gibt. Diese Vorgehensweise ist bei der derzeitigen Länge des Mikrozensus-Fragebogens allerdings dem Haushalt kaum zumutbar. Aber auch hier gilt: Aus Sicht der Datennutzer und der Datenqualität kann

es keine Alternative zu einer starken Standardisierung der CATI-Interviews geben, wenn Telefoninterviews im Mikrozensus / LFS in größerem Umfang eingesetzt werden sollten.

7.1.3 Bewertung: Der Selbstausfüllerbogen

Durch die unplausibilisierte Eingabe der Selbstausfüllerbogen im Projekt Q-MED / LFS war es erstmals möglich, das Ausfüllverhalten der Befragten quantitativ-empirisch zu prüfen. An den unplausibilisierten Daten konnte festgestellt werden, dass es kaum einem Probanden gelang, den Bogen ohne Filterfehler auszufüllen. Insbesondere bei Mehr-Personen-Haushalten sind unterschiedliche Befragungsverläufe und unterschiedliche Filtersprünge zu beachten. Dies macht das Ausfüllen des Fragebogens sehr kompliziert und die Ergebnisse des Experiments haben gezeigt, dass viele der Befragten im Zweifelsfall alle Fragen beantworten, die ihnen irgendwie plausibel erscheinen. Zahlreiche unnötige Angaben, die auf diese Weise gemacht werden, können zwar als plausible Angaben ignoriert werden. Insgesamt führt es aber dazu, dass der Befragte sich nicht zurecht findet, die Logik der Filterführung nicht erkennen kann und hierdurch eine unnötige bzw. zusätzliche Befragungslast der Befragten entsteht.

Für die Datenqualität problematischer sind allerdings weniger die Angaben, die unnötigerweise gemacht werden, sondern die fehlenden Angaben. Bei 83% der befragten Haushalte in der schriftlichen Befragung fehlte mindestens ein Wert zu einer als Pflichtmerkmal ausgewiesenen Frage. Möglicherweise bleiben Fragen unbeantwortet, weil (1) Filterführungen nicht korrekt beachtet werden oder (2) die Frage nicht verstanden wird oder (3) keine zutreffende Antwortkategorie gefunden wird. Fehlende Angaben, insbesondere wenn sie häufig nur bei einem Erhebungsinstrument auftauchen und daher selektive Ausfälle sind, beeinträchtigen auf jeden Fall die Datenqualität und auch die Vergleichbarkeit der Erhebungsinstrumente untereinander. Das Ergebnis aus der schriftlichen Befragung in Q-MED / LFS hat gezeigt, dass eine Vervollständigung und Plausibilisierung der Daten einen enormen Aufwand bedeutet.

Aber auch bei den Fragen, die beantwortet wurden, kann aufgrund der Komplexität der Themen keine hohe Datenqualität erwartet werden. Wie aus der Forschungsliteratur bekannt ist, hat die schriftliche Befragung prinzipiell den Nachteil, dass der Befragte keine Rückfragen stellen kann und bei Verständnisschwierigkeiten sich selbst helfen muss. Dies wirkt sich umso nachteiliger aus, je schwieriger die Sachverhalte sind, die in einer Erhebung erfragt werden. Wie die Q-MED / LFS-Daten belegen, kann es zu erheblichen Mode-Effekten kommen, wenn die Befragten aus Unsicherheit eine Restkategorie wählen, z. B. „Sonstige Gründe“ bei der Frage nach dem Grund für Mehrarbeit in der Berichtswoche oder durch flüchtiges Lesen den Unterschied zwischen der Frage nach den normalerweise geleisteten Stunden und den tatsächlichen Arbeitsstunden in der Berichtswoche nicht wahrnehmen und jeweils die gleiche Antwort geben.

Allerdings zeigt die Gesamtbetrachtung der Mode-Effekte, dass nicht in jedem Fall die schriftliche Befragung eine geringere Datenqualität produziert: Bei Fragen mit vielen Antwortmöglichkeiten hat die schriftliche Form den Vorteil, dass zunächst alle Antwortkategorien wahrgenommen werden, bevor die zutreffende Antwort gegeben wird. Gering besetzte Kategorien wie z. B. der Besuch einer Gesamtschule oder Waldorfschule, der Abschluss einer Meister- oder Technikerschule oder eine abgeschlossenen Promotion ebenso wie der Bezug von Einkommen aus Vermögen, Zinsen werden in der schriftlichen Befragung signifikant häufiger genannt. Ein weiterer Vorteil der schriftlichen Befragung ergibt sich aus der zeitlichen Flexibilität beim Ausfüllen des Bogens und einer vergleichsweise geringen Proxy-Quote. Insgesamt gilt für die schriftliche Befragung: Papier ist geduldig... Ob es die Haushalte bei der Befragung immer sind, bleibt fraglich. Wenn Filterführungen nicht verstanden und zahlreiche Fra-

gen gar nicht oder nur ungenau beantwortet werden, und wenn Mode-Effekte darauf hindeuten, dass viele Fragen nur oberflächlich gelesen und deshalb möglicherweise anders als im mündlichen Interview beantwortet werden, muss festgestellt werden, dass durch die schriftliche Befragung in Papierform nur eine sehr begrenzte Datenqualität erreicht werden kann. Dies gilt umso mehr, wenn man berücksichtigt, dass sich in der Q-MED / LFS-Befragung vor allem höhere gebildete Bevölkerungsschichten, die häufig als Angestellte oder Beamte/Beamtinnen beschäftigt sind, beteiligt haben. Im regulären Mikrozensus / LFS sind es aber gerade die bildungsfernen Bevölkerungsgruppen, die besonders häufig die schriftliche Befragung wählen. Es ist anzunehmen, dass die Probleme mit dem Selbstausfüllerbogen bei diesen Gruppen nicht unbedingt geringer ausfallen.

Für die Weiterentwicklung des Selbstausfüllerbogens hat das Projekt Q-MED / LFS zahlreiche Anregungen geliefert. Die Erkenntnis, dass Filter und Sprunganweisungen mehrheitlich nicht beachtet werden, hat bereits für die Erhebung 2011 zu einer grundlegenden Überarbeitung des Layouts geführt.

Schaubild 21: Beispiel: Neue Filterführung im Mikrozensus / LFS

41 Falls Ihr Vertrag/Ihre Tätigkeit befristet ist						
Aus welchem Grund haben Sie eine befristete Tätigkeit?						
Falls mehrere Gründe zutreffen, bitte den Hauptgrund angeben.						
Ausbildung	1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dauerstelle nicht zu finden	2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dauerstelle nicht gewünscht	3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Probezeit-Arbeitsvertrag	4	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Aus anderen Gründen	5	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Verbesserungen am Selbstausfüller erscheinen lohnenswert. Dennoch bleibt festzustellen, dass trotz aller Anstrengungen zur Überarbeitung des Layouts und zur Vereinfachung der Frageformulierungen die eingeschränkten Möglichkeiten des Papierfragebogens bestehen bleiben.

7.1.4 Bewertung: Der Online-Bogen

Eine Alternative zum Selbstausfüller könnte das Angebot des Mikrozensus / LFS als Online-Survey darstellen. Das Medium Internet eröffnet zahlreiche neue Möglichkeiten und Befragungstechniken. Auch hier liegen die Vorteile auf der Hand: Filterführungen können hinterlegt werden, Plausibilisierungen erfolgen während der Eingabe und eine Vollständigkeit der Daten wird gewissermaßen „erzwungen“. Zudem kann der Befragte die Teilnahmezeit frei wählen, die Befragung jederzeit unterbrechen und - so zeigen es die Ergebnisse aus Q-MED / LFS –die Online-Befragung wird eher dafür genutzt, dass jede Person die Fragen für sich selbst beantwortet¹¹. Von allen Erhebungsinstrumenten war die Proxy-Quote in der Online-Befragung am geringsten. Nachteile ergeben sich aber auch hier aus der Tatsache, dass der Befragte mit Verständnisproblemen alleine bleibt. Zwar besteht die Möglichkeit, umfangreiche Hilfstexte und Zusatzerläuterungen bereitzustellen, diese ersetzen aber nicht die Funktion des Interviewers, auf individuelle Situationen zu reagieren und entsprechende Hilfen zu geben.

¹¹ Im Gegensatz zum Selbstausfüller war der Online-Bogen als Personenbogen konzipiert, für den jedes Haushaltsmitglied ein persönliches Passwort erhielt. Vermutlich ist die auffallend geringe Proxy-Quote auch durch den personenspezifischen Zugang zu erklären.

Insgesamt zeigen die Befragungsergebnisse aus der Online-Erhebung, dass es kaum systematische Abweichungen zwischen der schriftlichen Befragung mit Selbstausfüller und der Online-Befragung gibt. So wurde zum Beispiel bei Fragen mit vielen Antwortkategorien (Listen) überhaupt kein signifikanter Unterschied zwischen Selbstausfüllern und Onlineteilnehmern festgestellt. Bei anderen Fragen kam es teilweise zu Abweichungen, die jedoch auf den ersten Blick keinem bestimmten Muster folgen. Spezifische Online-Effekte, die teilweise aus anderen Studien bekannt sind, wie zum Beispiel der Primacy-Effekt (häufige Wahl der ersten Antwortkategorie), konnte im Experiment Q-MED / LFS nicht nachgewiesen werden.

Generell hat der Onlinetest in Q-MED / LFS gezeigt, dass eine web-basierte Darstellung des Mikrozensus / LFS-Fragebogens mit geeigneten Werkzeugen und moderner Software möglich ist, d. h. aus technischer Sicht ist der Fragebogen „onlinefähig“. Die meisten Haushalte in der Online-Befragung berichteten, dass sie keine Probleme mit dem Ausfüllen des Online-Bogens hatten und sich auch bei freier Wahl zwischen persönlicher, telefonischer oder schriftlicher Befragung in Papierform für das Online-Medium entschieden hätten. Betont werden muss allerdings, dass sich der Test in Q-MED / LFS ausschließlich auf die Durchführbarkeit der Befragung und die zu erzielenden Ergebnisse bezog. Eine Aussage über die technischen Schwierigkeiten in der Organisation, Prüfung der Vollständigkeit der Haushalte, Datensicherheit, Datenhaltung etc. wurde bewusst ausgeklammert. Erst wenn entsprechende Organisationstests durchgeführt wurden und diese zu einem positiven Ergebnis kommen, kann die „Onlinefähigkeit“ des Mikrozensus / LFS abschließend bewertet werden.

7.2 Die Bedeutung von Mode-Effekten im Mikrozensus / LFS

Bei der Betrachtung einzelner Mode-Effekte wurden in Kapitel 6 bereits erste Schätzungen vorgenommen, welches Ausmaß und auch welche Relevanz die analysierten Mode-Effekte für die Datenqualität im Mikrozensus / LFS haben. Die zum Teil deutlich voneinander abweichenden Ergebnisse zeigen zunächst, dass es Mode-Effekte im Mikrozensus / LFS gibt. Ob Mode-Effekte als kritischer Störfaktor zu sehen sind, hängt davon ab, (1) ob die Wahl des Erhebungsinstruments stark selektiv ist und (2) wie stark die einzelnen Mode-Effekte sind.

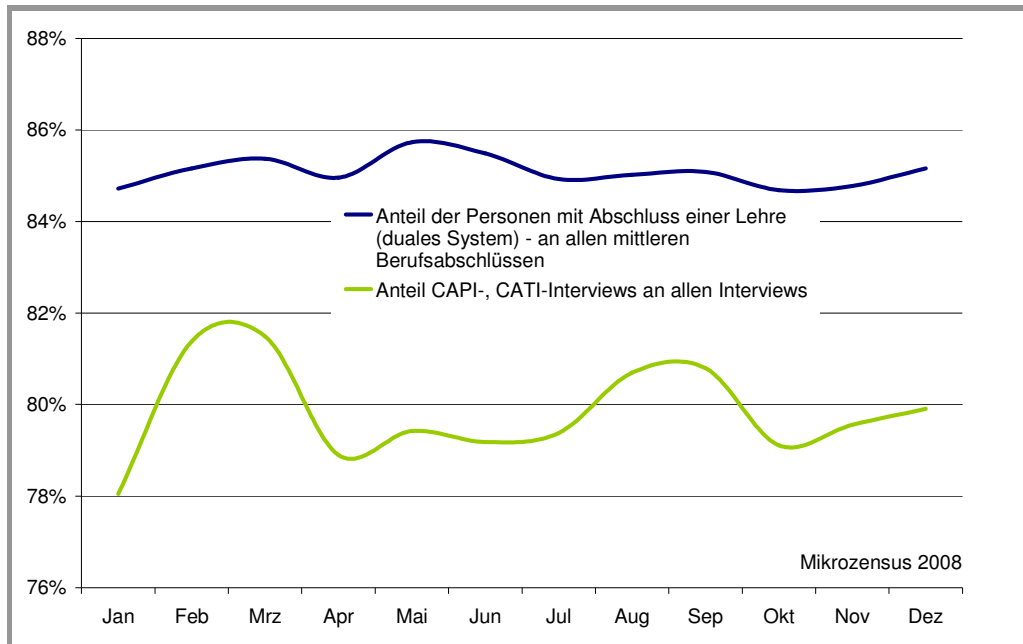
Wie in Kapitel 1 gezeigt werden konnte, ist die Auswahl des Erhebungsinstruments nicht zufällig. Der Anteil von Selbstausfüllern variiert zwischen den Bundesländern, zwischen kleinen und großen Haushalten, zwischen verschiedenen Bildungsschichten, zwischen Stadt und Land, aber auch im Zeitverlauf.

Einer der wichtigsten Mode-Effekte, der im Experiment Q-MED / LFS aufgezeigt werden konnte, ist der Einfluss nicht standardisierter Fragetechniken auf die Ergebnisse im Mikrozensus / LFS. Am deutlichsten tritt dieser Effekt bei langen Listen mit vielen Antwortmöglichkeiten auf. In Schaubild 22 ist dargestellt, welche Auswirkung der unterschiedliche Anteil von persönlichen Interviews (hier zusammengefasst CAPI- und CATI-Interviews) auf ein Merkmal hat, das Mode-Effekten unterliegt. Deutlich erkennbar sind ähnliche Wellenverläufe: Steigt der Anteil der persönlichen Interviews, steigt auch der Anteil der Personen mit einer Lehrausbildung (an allen Personen mit mittleren Abschlüssen). Die sichtbaren „Wellenlinien“ deuten weniger auf unterjährige Verläufe hin als vielmehr auf statistische Artefakte.

Auch an anderen Beispielen kann man zeigen, dass die Ergebnisse nicht durchgehend unabhängig von dem angewendeten Erhebungsweg sind. Schwierige Sachverhalte lassen sich genauer im persönlichen Interview erfragen (zum Beispiel Arbeitszeiten, Überstunden etc.). Die Genauigkeit der Gesamtergebnisse hängt demnach auch von dem Anteil der persönlich befragten Haushalte in der Stich-

probe ab. Entsprechend genau oder ungenau sind Bundes- und Länderergebnisse, je nachdem wie hoch der Anteil der schriftlich oder persönlich befragten Haushalte ist. Um Mode-Effekte und ihren negativen Einfluss auf die Vergleichbarkeit von verschiedenen Zeitpunkten, regionalen Einheiten oder auch verschiedenen Bevölkerungsgruppen zu minimieren, müssen (1) alle Erhebungswege weiter verbessert und weiterentwickelt werden und (2) Strategien entworfen werden, um die Anteile des Mixed Mode (Anteil CAPI, CATI, Selbstaussfüller) möglichst zeitlich und regional konstant zu halten.

Schaubild 22: Anteil der Personen mit Abschluss einer Lehre
an allen mit mittleren Berufsabschlüssen / Anteil CAPI, CATI-Interviews
Mikrozensus 2008



Literaturverzeichnis

- Atteslander**, Peter (2003): Methoden der empirischen Sozialforschung, Berlin.
- Baker**, Reginald P./ **Bradburn**, Norman M./ **Johnson**, Robert A. (1995): Computer-assisted Personal Interviewing. An Experimental Evaluation of Data Quality and Cost; Journal of Official Statistics, Vol. 11, No. 4, S.413-431.
- Biemer**, Paul P./ **Lyberg**, Lars E. (2003): Introduction to Survey Quality, Wiley Series in Survey Methodology, Hoboken, New Jersey.
- Bradburn**, Norman M. (1983): Response Effects; in: Rossi, Peter H./ Wright, James D./ Anderson, Andy B. (Hrsg.): Handbook of Survey Research. Quantitative Studies in Social Relations, Orlando.
- Czaja**, Ronald./ **Blair**, Johnny. (2005): Designing Surveys: A Guide to decisions and procedures. Thousand Oaks.
- Dillman**, Don. (2007): Mail and Internet Surveys - The Tailored Design Method (2007), New Jersey
- Elliot**, Dave/**Rainford**, Laura/ **Eldridge**, Jack (2006): Interview Mode Effects in the UK Local Labour Force Survey. Methods Bulletin 57: 50-60, UK Office for National Statistics.
- Gfroerer**, Joseph C./ **Hughes**, Arthur L. (1991): The Feasibility of Collecting Drug Abuse Data by Telephone; Public Health Reports 106, S. 384-393.
- Gfroerer**, Joseph C./ **Hughes**, Arthur L. (1992): Collecting Data on Illicit Drug Abuse by Phone; in: Turner, Charles/ Lessler, Judith/ Gfroerer, Joseph C. (Hrsg.): Survey Measurement of Drug Abuse, Washington.
- Hartmann**, Peter/ **Schimpl-Neimanns**, Bernhard (1992): Zur Repräsentativität soziodemographischer Merkmale des Allbus - Multivariate Analysen zum Mittelschichtbias der Umfrageforschung, ZUMA Arbeitsbericht 92/01.
- Köhne-Finster**, Sabine/ **Lingnau**, Andreas (2008): Untersuchung der Datenqualität erwerbsstatistischer Angaben im Mikrozensus. Ergebnisse des Projekts „Nachbefragung im Mikrozensus/LFS“; Wirtschaft und Statistik 12/2008, S. 1067-1088.
- De Leeuw**, Edith D. (1992): Data Quality in Mail, Telephone and Face to Face Surveys, Amsterdam.
- De Leeuw**, Edith D./ **Hox**, Joop J./ **Dillmann**, Don A. (2008): Mixed-mode-Surveys: When and Why; In: de Leeuw, Edith D./ Hox, Joop J./ Dillman, Don A. (Hrsg.) International Handbook of Survey Methodology, New York, S. 299-316.
- De Leeuw**, Edith D./**Nicholls**, William (1996): Technological Innovations in Data Collection: Acceptance, Data Quality and Costs; Sociological Research Online, Vol. 1, No. 4., <http://www.socresonline.org.uk/socresonline/1/4/leeuw.html>, abgerufen am 04.01.2010.
- Loosveldt**, Geert (2008): Face-to-face interviews; in: de Leeuw, Edith D./ Hox, Joop J./ Dillman, Don A.: International Handbook of Survey Methodology, New York, S. 201-220.
- Lyberg**, Lars E./ **Kasprzyk**, Daniel (1991): Data Collection Methods and Measurement Error: An Overview; in: Biemer, Paul P./ Groves, Robert M./ Lyberg, Lars E./ Mathiowetz , Nancy A./ Sudman , Seymour (Hrsg.): Measurement Error in Surveys, New York, S. 237-257.
- Pötschke**, Manuela/ **Simonson**, Julia (2001): Online-Erhebungen in der empirischen Sozialforschung. Erfahrungen mit einer Umfrage unter Sozial-, Markt- und Meinungsforschern; ZA-Information 49, S. 6-28.
- Porst**, Rolf (2000): Praxis der Umfrageforschung. Studienskripten zur Soziologie 126, Stuttgart/ Leipzig/ Wiesbaden.
- Stadtmüller**, Sven. (2009): Rücklauf gut, alles gut? Zu erwünschten und unerwünschten Effekten monetärer Anreize in postalischen Befragungen; in: Methoden - Daten - Analysen, 2009 / 2, S. 167 - 185.
- Statistisches Bundesamt** (Hrsg.) (2009): Wirtschaftsrechnungen. Private Haushalte in der Informationsgesellschaft - Nutzung von Informations- und Kommunikationstechnologien (IKT); Fachserie 15 Reihe 4, Wiesbaden.