

methoden daten analysen

ZEITSCHRIFT FÜR EMPIRISCHE SOZIALFORSCHUNG

mda

2010, Jahrgang 4, Heft 2



- Markus Klein und Michael Kühhirt* Sozial erwünschtes Antwortverhalten bezüglich der Teilung häuslicher Arbeit. Die Interaktion von Interviewergeschlecht und Befragtenmerkmalen in Telefoninterviews
- Michael Häder und Mike Kühne* Mobiltelefonerfahrung und Antwortqualität bei Umfragen
- Franziska Kunz* Mahnaktionen in postalischen Befragungen. Empirische Befunde zu Auswirkungen auf den Rücklauf, das Antwortverhalten und die Stichprobenszusammensetzung

Herausgegeben von *Christof Wolf*
Marek Fuchs
Bärbel Knäuper
Petra Stein

Methoden – Daten – Analysen. Zeitschrift für Empirische Sozialforschung

Die Zeitschrift wird herausgegeben von GESIS – Leibniz-Institut für Sozialwissenschaften.

Herausgeber: Christof **Wolf** (Mannheim, geschäftsführend), Marek **Fuchs** (Darmstadt),
Bärbel **Knäuper** (Montreal), Petra **Stein** (Duisburg-Essen)

Wissenschaftlicher

Beirat: Hans-Jürgen **Andreß** (Köln), Andreas **Diekmann** (Zürich), Sabine **Häder** (Mannheim),
Udo **Kelle** (Marburg), Dagmar **Krebs** (Gießen), Frauke **Kreuter** (College Park, Maryland),
Edith **de Leeuw** (Utrecht), Norbert **Schwarz** (Ann Arbor)

Redaktion: Paul **Lüttinger**

GESIS – Leibniz-Institut für Sozialwissenschaften
Postfach 12 21 55
68072 Mannheim
Tel.: 0621 – 1246-268
E-Mail: mda@gesis.org
Internet: www.gesis.org/MDA/

Die MDA deckt alle Fragestellungen aus dem Bereich der Empirischen Sozialforschung ab, insbesondere aus dem Bereich der Umfragemethodik. Im Vordergrund stehen Artikel, die die methodischen und/oder statistischen Kenntnisse der Profession erweitern, sowie Beiträge, die sich mit der Anwendung der Methoden der Empirischen Sozialforschung in der Forschungspraxis beschäftigen, oder solche, in denen ein statistisches Verfahren exemplarisch angewandt wird. Obwohl der Schwerpunkt auf Umfragemethoden liegt, sind Beiträge zu anderen methodischen Bereichen willkommen.

Alle Beiträge, die zur Veröffentlichung in der MDA eingereicht werden, werden von mindestens zwei unabhängigen Gutachtern blind begutachtet.

Der Nachdruck von Beiträgen ist nach Absprache möglich. Die MDA erscheint zweimal im Jahr und steht als Printversion und online zur Verfügung. Die Registrierung für den Bezug der MDA erfolgt über die Web-Seiten von GESIS:

<http://www.gesis.org/forschung-lehre/gesis-publikationen/zeitschriften/mda/bestellung/>

Druck: Concordia-Druckerei König oHG, Mannheim-Sandhofen
Gedruckt auf chlorfrei gebleichtem Papier.

ISSN 1864-6956

4. Jahrgang 2010 © GESIS, Mannheim, Dezember 2010

Inhalt

77 Editorial

FORSCHUNGSBERICHTE

79 Sozial erwünschtes Antwortverhalten bezüglich der
Teilung häuslicher Arbeit. Die Interaktion von Interviewer-
geschlecht und Befragtenmerkmalen in Telefoninterviews
Markus Klein und Michael Kühhirt

105 Mobiltelefonerfahrung und Antwortqualität bei Umfragen
Michael Häder und Mike Kühne

127 Mahnaktionen in postalischen Befragungen. Empirische
Befunde zu Auswirkungen auf den Rücklauf, das
Antwortverhalten und die Stichprobenzusammensetzung
Franziska Kunz

REZENSIONEN

156 Umfrageforschung. Herausforderungen und Grenzen.
Martin Weichbold, Johann Bacher und
Christof Wolf (Hg.), 2009
Johannes Stauder

158 Klassifikationsanalysen in Theorie und Praxis.
Jost Reinecke und Christian Tarnai (Hg.), 2008
Wolfgang Sodeur

162 Statistik. Eine verständliche Einführung.
Udo Kuckartz, Stefan Rädiker, Thomas Ebert und
Julia Schehl, 2010
Sven Stadtmüller

ANKÜNDIGUNGEN

- 166 Ausschreibung: ALLBUS-Preis 2011
- 167 Neue Klassifikation der Berufe 2010 der
Bundesagentur für Arbeit
- 169 Drei neue Studiengänge für sozialwissenschaftliche
Survey-Methoden
- 172 AutorInnen, RezensentInnen, GutachterInnen 2010
- 173 Hinweise für unsere Autorinnen und Autoren

Editorial

Liebe Leserin, lieber Leser,

das vorliegende Heft beschließt den vierten Jahrgang der MDA. Mit insgesamt 22 Einreichungen haben wir in diesem Jahr in etwa das Vorjahresniveau erreicht. Das ist nicht schlecht, wir würden uns jedoch über viele weitere Manuskripte freuen.

Tabelle 1 Manuskripteingang und Verbleib, 2007-2010

	Eingegangen	Angenommen	Abgelehnt	Noch offen	Annahme- quote
2007	28	9	19	-	32 %
2008	14	8	6	-	57 %
2009	25	11	14	-	44 %
2010	22	4	12	6	
Gesamt	89	32	51	6	

Das aktuelle Heft enthält drei Beiträge aus der Forschung und drei Rezensionen. Im ersten Beitrag untersuchen Markus Klein und Michael Kühhirt am Beispiel von Fragen zur Aufteilung der Hausarbeit, wie das Geschlecht des Interviewers mit Merkmalen der Befragten interagiert. Vor dem Hintergrund von Überlegungen aus der Theorie der rationalen Wahl arbeiten die Autoren nicht nur heraus, dass es sich beim Interview um eine soziale Situation handelt, sondern auch, wie das berichtete Verhalten mit dem Geschlecht der befragenden Person variiert. Die Ergebnisse der Studie bestätigen, dass Ausmaß und Richtung von Interviewereffekten nicht bei allen Befragten gleich sind. Vielmehr variiert das Antwortverhalten systematisch zwischen verschiedenen Bevölkerungsgruppen.

Im zweiten Beitrag geht es um die Antwortqualität bei Befragungen am Mobiltelefon. Michael Häder und Mike Kühne untersuchen die Vermutung, dass die Erfahrung mit Mobiltelefonen die Qualität der Antworten bei Befragungen über dieses Medium beeinflusst. Auf Basis der vorgelegten Befunde kann diese Vermutung jedoch nicht bestätigt werden. Für alle, die Befragungen über das Festnetz mit Befragungen über das Mobiltelefon kombinieren wollen, ist dieses Ergebnis eine gute Nachricht.

Im dritten Beitrag fragt Franziska Kunz, welche Auswirkungen Mahnaktionen in postalischen Befragungen auf den Rücklauf, das Antwortverhalten und die Stichprobenszusammensetzung haben. Ausgangspunkt des Beitrags ist die Beobachtung, dass wir zwar relativ sicher wissen, dass Mahnaktionen sich positiv auf den Rücklauf auswirken, dass jedoch deutlich weniger darüber bekannt ist, was daraus für die Zusammensetzung der Stichprobe und die Verteilung der Antworten folgt. Kunz zeigt am Beispiel einer größeren postalischen Erhebung, dass Mahnaktionen zur Beteiligung von zunächst unterrepräsentierten Personen führen: jüngere, im ländlichen Raum lebende sowie einen geringeren sozialen Status und ein geringeres politisches Interesse aufweisende Personen. Wie die Autorin weiter zeigt, weisen diejenigen, die früh antworten, ein deutlich höheres Interesse am Befragungsthema auf als andere. Nachfassaktionen können somit zur Abschwächung anfänglicher Verzerrungen beitragen. Daher empfiehlt Kunz, Nachfassaktionen mit zwei Mahnungen in das Erhebungsdesign postalischer Befragungen einzuplanen.

Neben den drei Beiträgen freue ich mich besonders, dass die MDA wieder drei Rezensionen aktueller Neuerscheinungen enthält. Insgesamt wurden seit dem ersten Jahrgang über 30 Bücher besprochen. Damit trägt die Zeitschrift nicht nur zur Diskussion neuester Forschungsergebnisse, sondern auch zur Rezeption aktueller Debatten und Gesamtdarstellungen bei.

Im Namen der Herausgeber danke ich allen Autoren, Rezensenten und Gutachtern, die an diesem Jahrgang mitgewirkt haben. Außerdem gilt mein Dank der Redaktion der MDA, insbesondere Paul Lüttinger und Christa von Briel, ohne deren tatkräftige Unterstützung kein Heft der MDA erschienen wäre.

Sozial erwünschtes Antwortverhalten bezüglich der Teilung häuslicher Arbeit

Social Desirability and Response Bias in Case of the Division of Household Labour

*Die Interaktion von
Interviewergeschlecht und
Befragtenmerkmalen in
Telefoninterviews*

*The Interaction between
Gender of the Interviewer and
Respondent Characteristics
in Telephone Interviews*

Markus Klein und Michael Kühhirt

Zusammenfassung

Der Beitrag untersucht den Einfluss des Interviewergeschlechts auf das Antwortverhalten von Befragten bezüglich der Teilung häuslicher Arbeit in telefonischen Interviews. Aufgrund einer Diskrepanz zwischen öffentlich vertretenen egalitären Geschlechterrollen und der noch immer geringen männlichen Beteiligung an der Hausarbeit wird auf Basis von Rational-Choice Argumenten erwartet, dass männliche Befragte bei weiblichen Interviewern sozial erwünscht antworten und ihre Beteiligung an der Hausarbeit überschätzen. Dagegen ist anzunehmen, dass Frauen ihren relativen Anteil an der partnerschaftlichen Hausarbeit eher unterschätzen. Interviewereffekte des Geschlechts sollten bei weiblichen Befragten jedoch geringer ausfallen als bei Männern. Da sich vor allem junge bzw. gebildete Befragte egalitärer Rollenvorstellungen bewusst sind, sollten vor allem

Abstract

In this paper we examine the impact of the interviewer's gender on respondents' self-reported share of housework in telephone interviews. Due to a discrepancy between modern, egalitarian gender roles dominating public discussion and men's marginal participation in housework we expect male respondents to answer in a socially desirable way and exaggerate their share of housework vis à vis female interviewers. At the same time, we assume that female respondents underreport their contribution to the couple's housework to female interviewers. The effects of the interviewer's gender should be less strong in case of female respondents than in case of male respondents, though. Additionally, theory suggests that young and educated respondents are particularly susceptible to effects of the interviewer's gender as they are most aware of egalitarian gender roles

diese Gruppen anfällig für derartige Interviewereffekte sein. Für männliche Befragte entsprechen die Ergebnisse weitestgehend den Erwartungen, wobei das Interviewergeschlecht in Abhängigkeit vom Alter des Befragten die stärkeren Effekte zeigt. Für weibliche Befragte zeigen sich über alle Altersgruppen und Bildungsabschlüsse dagegen keine signifikanten Effekte. Die Ergebnisse verdeutlichen, dass sozial erwünschtes Antwortverhalten auch bei Auskünften über konkretes Alltagshandeln auftreten kann. Hinzu kommt eine beträchtliche Heterogenität der Effekte über unterschiedliche Bevölkerungsgruppen.

due to their socialization and environment. For male respondents the results are largely consistent with the expectations. However, the effect of the interviewer's gender varies stronger by age than by education. Regarding female respondents the results do not indicate any significant effects across age groups and educational degrees. Overall, the results show that social desirability may also bias self-reports of everyday behaviour. Moreover, interviewer effects vary considerably between different social groups.

1 Einleitung

Umfragedaten sind für empirisch ausgerichtete Sozialwissenschaftler in überwiegendem Maße die Grundlage statistischer Analysen. Die Reliabilität und Validität der Daten ist dabei einerseits abhängig von der korrekten Operationalisierung und Messung der interessierenden Konstrukte durch den Forscher. Andererseits ist hinlänglich bekannt, dass die Datenqualität bereits im Verlauf des Erhebungsprozesses beträchtlich eingeschränkt werden kann, sei es durch den Ausfall spezifischer Personengruppen oder durch systematische Antwortverzerrungen während des Interviews. So können z. B. Dritte die Teilnahme an einer Befragung behindern oder Einfluss auf das Antwortverhalten der Befragten nehmen (Mohr 1986). Weitaus häufiger jedoch sind es die Interviewer selbst, die den Erfolg der Datenerhebung maßgeblich beeinflussen. Dies gilt sowohl für persönliche Befragungen als auch für telefonische Interviews (Diekmann 2004: 399). So zeigt sich beispielsweise, dass Kontrollüberzeugungen der Interviewer (Weidmann/Schmich/Schiller-Born 2008) oder deren sprachliche Eigenschaften (Steinkopf/Bauer/Best 2010) die Ausschöpfungsquote telefonischer Befragungen signifikant beeinflussen.

In diesem Beitrag gehen wir der Frage nach, inwieweit in einer telefonischen Umfrage die subjektiv berichtete Beteiligung an Hausarbeit durch das Geschlecht der Interviewer verzerrt wird. Im Mittelpunkt steht dabei die mögliche Interaktion des Interviewergeschlechts mit bestimmten Eigenschaften der Befragten. Dadurch sollen für verschiedene Bevölkerungsgruppen unterschiedlich ausfallende Interviewereffekte herausgearbeitet werden.

Wenngleich das Geschlecht der Interviewer schon früh als eine der bedeutendsten Ursachen für systematische Antwortverzerrungen insbesondere in telefonischen Befragungen angesehen wurde (Grimes/Hansen 1984), gibt es auf diesem Gebiet relativ wenige empirische Studien. Aus der bisherigen Forschung stellt sich jedoch vermehrt heraus, dass Befragte zu Themen des Geschlechterverhältnisses gegenüber weiblichen Interviewern tendenziell dazu neigen, häufiger Partei für Frauen zu ergreifen, als gegenüber männlichen Interviewern. Dies wurde vor allem bezüglich der wahrgenommenen Ungleichheit zwischen den Geschlechtern sowie der Diskriminierung von Frauen (Ballou/DelBoca 1980; Kane/Macaulay 1993), Einstellungen zu Geschlechterrollen (Ballou/DelBoca 1980; Grimes/Hansen 1984; Lueptow/Moser/Pendleton 1990; Kane/Macaulay 1993; Huddy et al. 1997) und der Unterstützung von Frauenrechten und politischen Interessenvertretungen zu Gunsten des weiblichen Geschlechts (Kane/Macaulay 1993; Huddy et al. 1997) festgestellt.

Besteht jedoch keine offensichtliche Beziehung zwischen dem Inhalt der Befragung und dem Interviewermerkmal Geschlecht, so ergeben sich uneinheitliche Ergebnisse. Während Befragte bei männlichen Interviewern im Vergleich zu weiblichen Interviewern optimistischere Aussagen zu ihren wirtschaftlichen Aussichten machen, kommen Groves und Fultz (1985) in der gleichen Studie zu dem Schluss, dass Interviewfehler (wie missing values, probing etc.) nicht signifikant zwischen dem Geschlecht des Interviewers variieren.

Noch weitaus weniger ist bislang über die Heterogenität von Interviewereffekten über verschiedene Bevölkerungsgruppen bekannt. Am häufigsten wurde in der Vergangenheit der Effekt des Interviewergeschlechts in Abhängigkeit des Befragten geschlechts untersucht, allerdings mit inkonsistenten Ergebnissen. So zeigen einige Studien, dass sich weibliche Befragte im Vergleich zu männlichen Befragten eher durch das Geschlecht des Interviewers in ihrem Antwortverhalten beeinflussen lassen und der Kontrast der Antworten zwischen weiblichen und männlichen Interviewern bei ihnen höher ausfällt als dies bei männlichen Befragten der Fall ist (Argentino et al. 1977; Grimes/Hansen 1984; Lueptow/Moser/Pendleton 1990). In der Studie von Ballou und DelBoca (1980) ergeben sich zwar bei Fragen zur Thematik der Ungleichheit zwischen den Geschlechtern dieselben Ergebnisse, bei der spezifischen Frage zur Frauenbewegung tritt jedoch genau der umgekehrte Fall ein: Männliche Befragte sind anfälliger für Interviewereffekte des Geschlechts als weibliche Befragte. Anscheinend ist der Grad der Anfälligkeit der Geschlechter für Effekte des Interviewergeschlechts von den angesprochenen Themen in der Umfrage abhängig (Kane/Macaulay 1993: 13). Allerdings kommt die Forschung auch bei spezifischen Themen zu keinem einheitlichen Ergebnis. So stellen Kane und Macaulay (1993) im Gegensatz zu Ballou und DelBoca (1980) fest, dass im Bereich der „gender-related action“-Items, insbesondere dem Aspekt der Frauenbewegung, weibliche Befragte anfälliger für Interviewereffekte des Geschlechts sind.

Die Abhängigkeit von Interviewereffekten von anderen Merkmalen der Befragten untersuchten bisher nur Huddy et al. (1997). In dieser Studie wurden vor allem jüngere und weniger gebildete Befragte durch das Geschlecht des Interviewers in ihrem Antwortverhalten bezüglich verschiedener Items zur Frauenbewegung und Geschlechterungleichheit beeinflusst. Zur Erklärung des Effektes der Bildung argumentieren Huddy et al. (1997: 200), dass Befragte mit geringer Bildung den vermeintlichen Einstellungen des aus ihrer Sicht höher gebildeten Interviewers nicht widersprechen möchten, um eine Stigmatisierung als weniger intelligent zu vermeiden.

Darüber hinaus liegt der Fokus der bisherigen Forschung weitestgehend auf der Betrachtung von Interviewereffekten auf Einstellungsvariablen. Ob Merkmale des Interviewers auch Antwortverzerrungen in Bezug auf konkretes Verhalten hervorrufen, ist dagegen kaum erforscht. Im Unterschied zu Einstellungen, die nicht immer bewusst vorliegen, kann tatsächliches Verhalten erinnert werden, so dass der „wahre Wert“ für den Befragten leichter zu bestimmen ist. Bei Einstellungsfragen besteht zudem das Problem, dass bei manchen Befragten keine Meinung zu einem spezifischen Thema vorliegt. In diesem Fall ist es für Befragte wesentlich einfacher, sozial erwünscht zu antworten, da sie keine Abweichung zu einem „wahren Wert“ zeigen müssen. Im Vergleich zu Einstellungsfragen sollte bei Fragen zu konkreten Tätigkeiten die Hürde damit nochmals höher liegen, den „wahren Wert“ in der Interviewsituation an soziale Erwartungen anzupassen.

Die Befragung zur Arbeitsteilung in Partnerschaften erfasst zum einen konkrete Tätigkeiten. Zum anderen betrifft sie das Geschlechterverhältnis, das mit einem zentralen öffentlichen Merkmal des Interviewers, dem Geschlecht, in unmittelbarer Verbindung steht. Da somit eine Stereotypisierung der Einstellungen vor allem weiblicher Interviewer durch die Befragten mit hoher Wahrscheinlichkeit auftritt, ist dieses inhaltliche Feld sehr gut geeignet, systematische Antwortverzerrungen bezüglich konkretem, alltäglichem Handeln zu untersuchen.

Die Methodenforschung im Bereich der Teilung häuslicher Arbeit hat sich bisher vor allem auf Probleme der Erhebung der Zeitverwendung konzentriert. Im Speziellen konzentriert sich die Forschung meist auf die Frage, ob eher Tagebuchangaben oder Selbstauskünfte valide Daten liefern (Juster/Ono/Stafford 2003) bzw. inwieweit das Geschlecht der Befragten die Angaben verzerrt (Lee/Waite 2005). Darüber hinaus zeigen Press und Townsley (1998), dass Merkmale wie Klasse, Bildung und Erwerbsstatus ebenfalls Antwortverzerrungen hervorrufen können und dass diese Effekte für Männer und Frauen unterschiedlich ausfallen.

Zum Thema Arbeitsteilung in Partnerschaften in Zusammenhang mit Interviewereffekten des Geschlechts gibt es bisher wenig empirische Forschung. In der Studie von Kane und Macaulay (1993: 13) ergeben sich zu Items der Arbeitsteilung

im Haushalt sowohl bei Männern als auch bei Frauen keine signifikanten Effekte. Allerdings geben männliche Befragte im Gespräch mit weiblichen Interviewern häufiger an, sich die Verantwortung für die Kinderbetreuung mit dem Partner zu teilen, als gegenüber männlichen Interviewern (ebd.: 11). In einer älteren Studie von Mohr (1986), deren Untersuchungsschwerpunkt auf dem Einfluss der Anwesenheit Dritter lag, zeigen die Ergebnisse in eine der allgemeinen Forschung zu Interviewereffekten bei „gender attitudes“ widersprechende Richtung: Männliche Befragte nennen, unabhängig davon ob der Partner anwesend ist, gegenüber weiblichen Interviewern im Durchschnitt eine geringere Beteiligung an der Hausarbeit als gegenüber männlichen Interviewern (ebd.: 67). Die Ergebnisse stammen jedoch ausschließlich aus bivariaten Analysen ohne Kontrolle der statistischen Signifikanz.

Insgesamt zeigen die oben genannten Studien vor allem, dass Interviewereffekte besonders dann zu erwarten sind, wenn die Thematik der Befragung mit Eigenschaften der Interviewer in eindeutiger Beziehung steht (Reinecke 1991: 28f.; Johnson et al. 2000: 78). Zudem deutet einiges auf die Heterogenität von Interviewereffekten über verschiedene Bevölkerungsgruppen hin. Beides spricht für eine stärkere theoretische Herleitung, sowie eine differenziertere empirische Untersuchung von Interviewereffekten.

Im Folgenden sollen nun zunächst theoretische Überlegungen dargelegt werden, von denen der zu erwartende Effekt des Interviewergeschlechts auf die berichtete Beteiligung an Hausarbeit in Abhängigkeit bestimmter Merkmale der Befragten abgeleitet wird. Anschließend erläutern wir die Datengrundlage unserer Analysen sowie unsere methodische Vorgehensweise. Zum Abschluss präsentieren wir die empirischen Ergebnisse und betrachten diese im Hinblick auf die theoretischen Erwartungen.

2 Theoretischer Hintergrund

Um das Auftreten von Interviewereffekten zu verstehen, muss das Interview als eine „soziale Situation“ betrachtet werden (Esser 1991: 1), in deren Rahmen sich Interviewer und Befragte wechselseitig aneinander orientieren und aufeinander Bezug nehmen. Der Interviewer bietet dem Befragten Hilfestellung in der spezifischen Interviewsituation, so dass dieser sich in der Rolle des Befragten zurechtfindet und den Anforderungen der Rolle gerecht wird. Es ist explizit Aufgabe des Interviewers, den Frage-Antwort-Prozess zu managen und gegebenenfalls sogar Rückfragen zu stellen (Groves et al. 2004: 269). Dies hat aber zur Folge, dass der Interviewer bewusst oder unbewusst Einfluss auf das Antwortverhalten des Befragten nehmen kann.

Das Resultat der Datenerhebung ist somit Folge eines menschlichen Interaktionsprozesses, in dessen Kontext beim Befragten bestimmte Kognitionen und Assoziationen hervorgerufen werden.

„Der Befragte nutzt alle ihm in der Situation zur Verfügung stehenden Signale, interpretiert die vermuteten Absichten des Interviewers und richtet sein Verhalten danach und nach den eigenen Zielsetzungen aus.“ (Reinecke 1991: 38)

Befragte geben somit in der spezifischen Situation des Interviews nicht immer automatisch ihren „wahren Wert“ zur jeweiligen Frage an, sondern zeigen mitunter reaktives Antwortverhalten.

2.1 Sozial erwünschtes Antwortverhalten als rationale Wahl

Im Rahmen der Theorie der rationalen Wahl kann das Befragtenverhalten als „Spezialfall einer allgemeinen Theorie des situationsorientierten Handelns“ betrachtet werden (Esser 1986, 1991). Es wird argumentiert, dass das Antwortverhalten der Befragten nicht durch fixe Orientierungen und Einstellungen im Vorhinein festgelegt ist, sondern sich erst innerhalb der Situation der Befragung in wechselseitiger Beeinflussung mit dem Interviewer durch den Vergleich verschiedener, alternativ möglicher Antwortoptionen ergibt.

Ausgangspunkt der Überlegungen ist das universelle Streben individueller Akteure nach sozialer Wertschätzung. In der Situation des Interviews kann der Befragte diese erlangen, indem er die Anerkennung seines Gesprächspartners, des Interviewers, erhält bzw. Missbilligung durch den Interviewer vermeidet. Dies geschieht, indem die jeweilige Antwort den Erwartungen entspricht, die der Befragte dem Interviewer auf Basis von bestimmten persönlichen Merkmalen unterstellt. Der „wahre Wert“ des Befragten weicht jedoch unter Umständen von diesen Erwartungen ab. Gegebenenfalls steht der Befragte während des Interviews vor der Entscheidung, den „wahren Wert“ preiszugeben und damit möglicherweise die Missbilligung des Interviewers auf sich zu ziehen oder die Antwort den unterstellten Idealen des Interviewers anzupassen.

Antwortverzerrungen durch soziale Erwünschtheit sind jedoch keineswegs zwingend. Zunächst muss eine Diskrepanz zwischen „wahrem Wert“ und „sozial erwünschtem Wert“ vorliegen. Neben dieser trivialen Voraussetzung sollte der Nutzen einer Antwortanpassung den Nutzen einer Angabe des „wahren Wertes“ übersteigen. Letzterer kann z. B. darin liegen, sich die Mühe zu ersparen, eine unwahre Antwort zu erfinden. Jedoch sollte der dazu nötige kognitive Aufwand bei den meisten sozialwissenschaftlichen Fragestellungen eher gering sein. Folglich kann man davon ausgehen, dass der Nutzen einer sozial erwünschten Antwort diese Kosten in der

Regel übersteigt. Auch die Aufrechterhaltung der eigenen Identität bzw. die Vermeidung kognitiver Dissonanz kann als Nutzen aus der Angabe des „wahren Wertes“ gezogen werden. Ob dieser den Nutzen sozial erwünschten Antwortverhaltens übertrifft, kann in den meisten Fällen wohl nur empirisch überprüft werden. Allerdings ist es plausibel anzunehmen, dass der Anreiz, sozial erwünscht zu antworten noch weiter ansteigt, wenn der „wahre Wert“ neben den vermeintlichen Ansichten des Interviewers auch den eigenen Idealen widerspricht.

Das Ausmaß der Konformität mit den sozialen Erwünschtheitswahrnehmungen ist zudem von zwei Faktoren abhängig: vom Ausmaß des Bedürfnisses nach sozialer Anerkennung, welches durch Social Desirability (SD)-Skalen gemessen werden kann, und von der wahrgenommenen Befürchtung, dass die Offenbarung des „wahren Wertes“ negative Konsequenzen etwa durch Missbilligung des Interviewers nach sich zieht.

Zunächst müssen jedoch drei Vorbedingungen erfüllt sein, die den Einfluss von Interviewermerkmalen auf das Antwortverhalten erst ermöglichen. Erstens muss die Befragung einer *Öffentlichkeit* unterliegen. Ist bei der Beantwortung keine andere Person anwesend, hat der Befragte auch bei einem sehr hohen Bedürfnis nach sozialer Wertschätzung kaum Anlass, von seinem „wahren Wert“ abzuweichen. Zweitens muss die Bedingung der *Sichtbarkeit* erfüllt sein. Damit ein Merkmal des Interviewers das Antwortverhalten des Befragten systematisch verzerren kann, muss die relevante Interviewereigenschaft für den Befragten transparent sein. Drittens müssen sich mit den identifizierten Merkmalen des Interviewers „klar typisierte, unterschiedliche Konsequenzerwartungen verbinden: man muss davon ausgehen können, dass z. B. männliche und weibliche Interviewer evtl. Antworten zu einem bestimmten Thema mit deutlich unterschiedlichen Bewertungen versehen“ (Esser 1986: 326). Dies wird als *Stereotypisierung* bezeichnet.

Der folgende Abschnitt zeigt auf, dass die genannten Bedingungen für unsere Forschungsfrage erfüllt sind und welche Effekte des Interviewergeschlechts auf die berichtete Beteiligung an der Hausarbeit auf Grundlage der oben dargestellten allgemeinen Überlegungen zu erwarten sind.

2.2 Interviewergeschlecht, Befragtenmerkmale und häusliche Arbeitsteilung

Vor dem Hintergrund des Antwortverhaltens in Befragungen als rationaler Wahl sind eindeutige Effekte des Interviewergeschlechts auf die berichtete Beteiligung an häuslicher Arbeit zu erwarten. Aufgrund der telefonischen Befragung sind die Bedingungen der Öffentlichkeit und Sichtbarkeit im hier behandelten Fall zwar

stark eingeschränkt, aber dennoch erfüllt. Obwohl der Befragte dem Interviewer nicht persönlich gegenüber sitzt, so muss er doch seine Antworten dem Interviewer direkt mitteilen und kann an verschiedenen Äußerungen des Interviewers Wertschätzung oder Ablehnung ablesen. In Bezug auf die Sichtbarkeit ist es dem Befragten möglich, das Geschlecht des Interviewers an Hand der Stimme zu identifizieren. So kamen Schejbal, Sachs und Lavrakas (1993) zu dem Ergebnis, dass 97 % der Teilnehmer einer telefonischen Befragung das Geschlecht ihres Interviewers eindeutig erkennen können. Gerade in Telefoninterviews sollte das Geschlecht des Interviewers besonders markant hervortreten, da den Befragten Informationen über andere Merkmale des Interviewers wie sozioökonomischer Status oder Aussehen fehlen (Groves/Fultz 1985: 31f.).

Die inhaltlichen Vorhersagen der zu erwartenden Effekte ergeben sich nun sämtlich aus dem Vorliegen der Stereotypisierung. In der heutigen Gesellschaft haben sich, befördert durch Medien und Werbung (Sullivan/O'Connor 1988; Brosius/Mundorf/Staab 1991), in letzter Zeit aber auch durch die Politik (Henninger/Wimbauer/Dombrowski 2008: 295f.), moderne Geschlechterbilder weitgehend in der öffentlichen Diskussion etabliert und werden zum Teil offensiv eingefordert. Im Mittelpunkt steht dabei vor allem das Idealbild einer selbstbewussten, im Beruf erfolgreichen Frau, aber auch die Vorstellung einer egalitär zu gestaltenden Arbeitsteilung in Partnerschaften (Kane/Macaulay 1993: 4). Dabei werden diese Ideale keineswegs von allen Mitgliedern der Gesellschaft geteilt. Es sind vor allem junge und gebildete Personen, die egalitäre Geschlechterrollen vertreten (van Snippenburg/Ter Voert/Janssen 1990; Wilcox 1991; Braun/Scott 2009). Wenig überraschend ist auch, dass Frauen diesen Ideen aufgeschlossener gegenüberstehen als Männer (Scott/Alwin/Braun 1996).

Sowohl auf Grund der Berichterstattung in den Medien als auch in Folge tatsächlicher lebensweltlicher Umwälzungen, vor allem was das Erwerbsverhalten von Frauen anbelangt, ist davon auszugehen, dass zumindest ein Teil der Befragten gerade weiblichen Interviewern in jungen Jahren und mit hohem Bildungsstand egalitäre Rollenbilder unterstellt und somit annimmt, dass diese verschiedene Antwortmöglichkeiten bezüglich der Geschlechterbeziehungen unterschiedlich positiv bewerten. Im Fall der häuslichen Arbeitsteilung bedeutet dies, dass Befragte besonders von jungen, gebildeten Frauen Zuspruch für eine egalitäre Aufgabenteilung erwarten. Da die Interviews vollständig von Studierenden durchgeführt wurden und den Teilnehmern vor Beginn der Befragung mitgeteilt wurde, dass die Erhebung im Rahmen einer Lehrveranstaltung an der Universität Mannheim stattfindet, sind solche Stereotypisierungsprozesse sehr wahrscheinlich.

Wie kann es nun aber bei weitreichender Verbreitung egalitärer Geschlechterbilder zu Antwortverzerrungen bezüglich der Teilung häuslicher Arbeit kommen? Tatsächlich spiegeln sich die genannten Ideale bisher bei weitem nicht in individuellem Alltagshandeln wider, vor allem was den privaten Bereich und damit auch die Arbeitsteilung betrifft. In der Realität übernehmen noch immer überwiegend Frauen den Großteil der routinemäßigen Hausarbeit, auch bei jüngeren und gebildeten Paaren (Sayer 2005; van der Lippe et al. 2010). Das Beharren auf traditionellen Handlungsmustern im Familienalltag zeigt sich auch deutlich in der anhaltend geringen Nutzung der Elternzeit durch Väter (Henninger/Wimbauer/Dombrowski 2008). Somit besteht bei vielen Befragten eine Diskrepanz zwischen tatsächlichem Verhalten und den Interviewern unterstellten Werten und Erwartungen. Vor dem Hintergrund der theoretischen Überlegungen entsteht dadurch ein Anreiz, das Antwortverhalten an die dem Interviewer unterstellten Erwartungen anzupassen, um Missbilligung zu vermeiden und soziale Wertschätzung zu garantieren. Dies gilt sowohl für Männer, die sich nicht ausreichend an Hausarbeit beteiligen, als auch für Frauen, die nach den normativen Vorstellungen zu viel Hausarbeit verrichten. Daraus ergibt sich die erste Hypothese.

H1: *Männer geben im Gespräch mit weiblichen Interviewern eine höhere Beteiligung an Hausarbeit an als gegenüber männlichen. Frauen hingegen, berichten weiblichen Interviewern einen geringeren Anteil an der Hausarbeit als männlichen.*

Da die Richtung des Effekts des Interviewergeschlechts bei männlichen und weiblichen Befragten entgegengesetzt wirken sollte, wird aus der Formulierung der Hypothesen bereits ersichtlich, dass wir von einem *Interaktionseffekt* zwischen Interviewer- und Befragtengeschlecht ausgehen.

Zusätzlich nehmen wir an, dass männliche Befragte bei weiblichen Interviewern durch Nennung des „wahren Wertes“ größere negative Konsequenzen erwarten als dies bei weiblichen Befragten der Fall ist. In der heutigen Gesellschaft ist der soziale Druck auf Männer, ihren Anteil an der Hausarbeit zu erhöhen, wesentlich stärker als derjenige auf Frauen, weniger Hausarbeit zu übernehmen. Männer haben somit weitaus größere Missbilligung für ihr tatsächliches Verhalten zu befürchten als Frauen. Als Konsequenz daraus sollten männliche Befragte häufiger als weibliche Befragte sozial erwünschtes Antwortverhalten zeigen (Huddy et al. 1997: 200).

H2: *Männer lassen sich in ihrem Antwortverhalten bezüglich der Beteiligung an häuslicher Arbeit stärker vom Geschlecht des Interviewers beeinflussen als Frauen.*

Diese Ausgangshypothesen lassen sich jedoch noch weiter ausdifferenzieren, da sich verschiedene Bevölkerungsgruppen in unterschiedlichem Ausmaß moderner Rollenbilder bewusst sind bzw. diese selbst vertreten (Hochschild/Machung 1989; van Snippenburg/Ter Voert/Janssen 1990; Wilcox 1991; Braun/Scott 2009) und somit auch die Stereotypisierung unterschiedlich stark ausfällt. So wurden ältere Befragte in einer Zeit sozialisiert, in der die traditionelle Arbeitsteilung zwischen Mann und Frau die Regel war. Diese in der Sozialisation erlernten Normen sind so stark internalisiert, dass es für ältere Befragte weniger Anlass gibt, den „wahren Wert“ des Anteils an der Arbeitsteilung in der Partnerschaft zu verschleiern. In gleicher Weise sind egalitäre Geschlechterrollen unter geringer Gebildeten noch immer weniger verbreitet. Somit sollte diese Gruppe weiblichen Befragten auch in geringerem Maße egalitäre Ansichten zuschreiben. Umgekehrt vertreten junge und gebildete Befragte im Schnitt durchaus egalitäre Geschlechterbilder, setzen diese in der Alltagspraxis jedoch nicht hinreichend um. Somit besteht für diese Gruppen nicht nur eine Diskrepanz zwischen Verhalten und den erwarteten Werten des Interviewers, sondern auch zwischen Verhalten und den eigenen Idealen. Weiterhin ist davon auszugehen, dass sich die Befragten dessen durchaus bewusst sind. Dies sollte das Auftreten sozial erwünschten Antwortverhaltens für junge und gebildete Befragte im Gespräch mit Interviewern mit ähnlichen Merkmalen besonders wahrscheinlich machen.

H3: *Je jünger die Befragten, desto stärker wirken die angenommenen Interviewereffekte aus Hypothese 1.*

H4: *Je höher die Bildung des Befragten, desto stärker wirken die angenommenen Interviewereffekte aus Hypothese 1.*

Die aufgeführten Argumente beziehen sich zum größten Teil auf die Furcht vor Missbilligung durch den Interviewer. Sicherlich sind auch Hypothesen über den Einfluss des Bedürfnisses nach sozialer Wertschätzung denkbar. Diese können jedoch mit den hier verwendeten Daten nicht getestet werden, da keine SD-Skalen abgefragt wurden.

3 Methodik

3.1 Die Mannheimer Familienstudie 2005

Die theoretischen Vorhersagen sollen mit den Daten der Mannheimer Familienstudie 2005 (MaFAM2005) empirisch getestet werden. Diese wurden im Rahmen einer Lehrveranstaltung am Lehrstuhl für Methoden der empirischen Sozialforschung an der Universität Mannheim im März 2005 von studentischen Interviewern mittels einer telefonischen Befragung erhoben. Ziel dieses Projektes war es, Einstellungen und Verhalten bezüglich Familie und Partnerschaft in Mannheim, Ludwigshafen und Umgebung zu untersuchen. Befragt wurde die erwachsene, deutschsprachige Wohnbevölkerung in Privathaushalten mit Festnetzanschlüssen.

Die Stichprobenziehung erfolgte durch ein zweistufiges Auswahlverfahren: Zunächst wurde von GESIS ein Pool von Ziffernfolgen möglicher Telefonanschlüsse in der Sampleregion generiert. Durch das Gabler-Häder-Verfahren (Häder/Gabler 1998; Gabler/Häder 2002) wurden nur solche Nummern ausgewählt, die auch tatsächlich als „100er-Block“ existieren. Mit diesem Verfahren ist es möglich, auch Personen zu erreichen, die keine im Telefonverzeichnis eingetragene Nummer haben. Aus dem resultierenden Nummernpool wurden den Interviewern in der Folge per Zufallsauswahl Anschlüsse zugewiesen. Im zweiten Schritt erfolgte die Auswahl der Zielpersonen nach der „Last-Birthday-Methode“. Es wurde nur das Haushaltsmitglied befragt, welches mindestens 18 Jahre alt war und zuletzt Geburtstag hatte. Anhand dieses Verfahrens ist ein „interpenetrated sample assignment“ gegeben, welches Interviewern durch Randomisierung Befragte mit im Durchschnitt äquivalenten Merkmalen zuteilt (Groves et al. 2004: 274). Systematische Unterschiede in den Merkmalen der Befragten zwischen den Befragtengruppen, die weiblichen und männlichen Interviewern zugewiesen werden, sind damit unwahrscheinlich.

Für die Analyse werden lediglich Befragte in Partnerschaften berücksichtigt, die mit dem Partner zusammen leben und Informationen für alle Analysevariablen liefern. 24 Interviews mit homosexuellen Befragten werden ausgeschlossen, da die Hypothesen eine Interaktion zwischen Interviewergeschlecht und Befragtengeschlecht postulieren und die für die Hypothesengenerierung relevante Attribution der Interviewereinstellungen sich auf Traditionalität vs. Egalität der partnerschaftlichen Arbeitsteilung zwischen Mann und Frau bezieht. Insgesamt verbleiben im Analysesample 445 Fälle. Tabelle 1 zeigt, dass für diese Teilstichprobe keine systematischen Unterschiede in den Befragtenmerkmalen zwischen männlichen und weiblichen Interviewern bestehen. Somit ist auszuschließen, dass Unterschiede im Antwortverhalten in den beiden Interviewergruppen auf unterschiedliche Stich-

probenkompositionen zurückzuführen sind. Von den 145 Interviewern, deren Interviews in die Analyse eingehen, sind 53 männlich und 92 weiblich.

Tabelle 1 Verteilung der Befragtenmerkmale nach Interviewergeschlecht

Befragtenmerkmal	männliche Interviewer	weibliche Interviewer	
weiblich	54,60	53,21	p>0,1
<i>Bildung:</i>			p>0,1
kein Abschluss/HS/RS ohne Lehre	12,27	9,29	
HS/RS mit Lehre	42,33	49,29	
Abitur	17,79	12,86	
FH/Uni	27,61	28,57	
Alter	46,90	45,52	p>0,1
verheiratet	82,82	80,00	p>0,1
<i>Kinder im Haushalt:</i>			p>0,1
kein Kind	51,53	48,93	
ein Kind	22,09	24,64	
zwei und mehr Kinder	26,38	26,43	
N Interviewer	53	92	
N Befragte	163	280	

Einträge sind Prozentsätze mit Ausnahme des Alters, für das der Mittelwert berichtet wird. Die Wahrscheinlichkeiten basieren auf χ^2 -Tests aus Kontingenztabellen, in denen das Geschlecht des Interviewers mit dem Geschlecht, der Bildung sowie der Kinderzahl und dem Familienstand der Befragten gekreuzt wurde. Die Wahrscheinlichkeit für Alter der Befragten basiert auf t-Tests der Mittelwertunterschiede.

Die Interviewer sind allesamt Studierende, die über wenig bis gar keine Erfahrung als Interviewer verfügen.¹ Aus diesem Grund fand vor der Befragung eine ausführliche Einweisung der Interviewer einschließlich Probeinterviews statt. Im Schnitt wurden 6,1 Interviews pro Interviewer realisiert, wobei 11 Interviews das Maximum darstellen. Die Befragung erstreckte sich über einen Zeitraum von vier Wochen im März 2005, in denen die Anrufe ausgenommen sonntags täglich durchgeführt wurden. Mit 36,5 % liegt die Ausschöpfungsquote für eine telefonische Befragung relativ hoch. Dies liegt möglicherweise am lokalen Bezug der Studie, der durch Beiträge in regionalen Tageszeitungen weiter befördert wurde.

1 Von den Interviewern im Analysesample haben lediglich 37 bereits Erfahrung. Getrennt nach Geschlecht ergeben sich keine signifikanten Unterschiede in der Interviewererfahrung (Ergebnis nicht ausgegeben).

3.2 Operationalisierungen

Aufgrund des Fokus auf Familie und Partnerschaft beinhaltet die MaFAM2005 zahlreiche Fragen zur Aufgabenteilung in Partnerschaften, darunter auch folgende Items bezüglich der Teilung häuslicher Arbeit: Waschen, Wohnung bzw. Haus putzen, Kochen, Abspülen, kleine Reparaturen im Haus bzw. in der Wohnung, Lebensmittel einkaufen, kranke Familienmitglieder betreuen und Finanzen verwalten. Auf die Frage „Wer macht die folgenden Dinge in ihrem Haushalt?“ standen den Teilnehmern die folgenden Antwortkategorien zur Verfügung: „Immer Sie“, „Meistens Sie“, „Immer Ihr (Ehe-)Partner/Ihre (Ehe-)Partnerin“, „Meistens Ihr (Ehe-)Partner/Ihre (Ehe-)Partnerin“, „Jeder zur Hälfte bzw. gemeinsam“ oder „wird von einer anderen Person gemacht“. Zur besseren Identifikation der Beteiligung des Befragten an der Hausarbeit werden die Kategorien in eine ordinale Rangfolge umkodiert.² Dieses Maß wird auch in zahlreichen Studien zur häuslichen Arbeitsteilung verwendet, vor allem denjenigen, welche auf Daten des International Social Survey Programme (ISSP) basieren (Batalova/Cohen 2002; Fuwa 2004; Breen/Cooke 2005; Geist 2005; Lewin-Epstein/Stier/Braun 2006), aber auch zwei Studien mit ausschließlich deutschen Daten (Schulz/Blossfeld 2006; Grunow/Schulz/Blossfeld 2007).

In einer Faktorenanalyse ergibt sich, dass die drei Items Waschen, Putzen und Kochen von allen acht Items am höchsten auf dem Faktor laden, der die traditionellen Tätigkeiten der Frau im Haushalt repräsentiert.³ Da die Hypothesen implizit das Geschlechterrollenverhältnis betreffen und die stereotype Zuweisung der Interviewereinstellungen in traditionelle vs. egalitäre Ansichten zur Arbeitsteilung zentral ist, werden nur diese traditionell weiblichen Tätigkeiten in den Analysen berücksichtigt.

Für die zentralen Analysen fassen wir die drei Items zu einem additiven Index zusammen. Dieser umfasst die Werte 0 bis 12. Niedrige Indexwerte zeigen dabei eine geringe Beteiligung an der traditionellen Hausarbeit an. Wir verwenden den Index anstatt des errechneten Faktors, da neben den drei Items kein anderes Item eine annähernd hohe Faktorladung aufweist. Zudem sind die Ergebnisse auf Basis des Index inhaltlich anschaulicher interpretierbar. Neben Interviewer- und Befragtingeschlecht werden für die Konstruktion der theoretisch abgeleiteten Dreifach-Interaktionen Alter und Bildung des Befragten verwendet. Die Bildungsvaria-

2 Die Antwortkategorien wurden folgendermaßen umbenannt und in den Analysen verwendet: „Immer Ihr (Ehe-)Partner/Ihre (Ehe-)Partnerin“ = keine Beteiligung (0); „Meistens Ihr (Ehe-)Partner/Ihre (Ehe-)Partnerin“ = niedrige Beteiligung (1); „Jeder zur Hälfte“ = egalitäre Arbeitsteilung (2); „Meistens Sie“ = hohe Beteiligung (3); „Immer Sie“ = sehr hohe Beteiligung (4). Die Kategorie „wird von einer anderen Person gemacht“ setzen wir auf „missing“, da diese keinem Rangplatz zugeordnet werden kann.

3 Faktorladungen: Waschen: 0,8384, Putzen: 0,8607, Kochen: 0,8394.

ble besteht aus 4 Gruppen: kein Abschluss bzw. Haupt- oder Realschule ohne Lehre, Haupt- oder Realschule mit Lehre, Abitur, sowie Fachhochschul- bzw. Hochschulabschluss. Wir verwenden hier die objektiven Indikatoren Alter und Bildung anstatt der ebenso im Datensatz enthaltenen direkten Angaben zu Einstellungen bezüglich Familie und Geschlechterrollen, da davon auszugehen ist, dass letztere potentiell ebenso durch Effekte sozialer Erwünschtheit verzerrt sind.

Um tatsächlich sicherzustellen, dass der Effekt des Interviewergeschlechts nicht auf die Komposition der jeweiligen Stichproben zurückzuführen ist, kontrollieren wir in allen multivariaten Analysen für Bildung, Alter, Familienstand und Kinderzahl im Haushalt.

3.3 Analysestrategie

Für die Analyse greifen wir auf ein lineares Regressionsmodell mit dem Index der häuslichen Arbeitsteilung als abhängiger Variable zurück.⁴ Im ersten Schritt schätzen wir den allgemeinen Interaktionseffekt zwischen Interviewer- und Befragten-geschlecht, um Unterschiede zwischen Männern und Frauen hinsichtlich der Anfälligkeit für sozial erwünschtes Antwortverhalten zu untersuchen. Anschließend soll die Frage beantwortet werden, ob Befragte, die sich moderneren Rollenbildern besonders bewusst sind, häufiger dazu neigen, ihre Angaben bezüglich der Beteiligung an häuslicher Arbeit diesen Idealvorstellungen anzupassen. Dazu modellieren wir eine Dreifachinteraktion zwischen Interviewergeschlecht, Befragten-geschlecht und Alter bzw. Bildung des Befragten.

Wie Dijkstra (1983) und Groves (1989) betonen, werden in der Methoden-forschung zu Interviewereffekten Interviews häufig als unabhängige Beobachtungen betrachtet. Dabei wird vernachlässigt, dass jeder Interviewer ein Cluster an Interviews generiert. In Analysen kann die Nicht-Berücksichtigung dieser Clusterung auf Grund von Autokorrelation der Fehlerterme zu verzerrten Schätzungen der Standardfehler führen. Daher erweist sich für einen Test von Interviewereffekten die hohe Zahl an Interviewern und geringe Zahl an Interviews pro Interviewer in der gewählten Datengrundlage als vorteilhaft (Groves/Fultz 1985: 34). Zwar ist die Anzahl der Interviewer wie in jeder Erhebung kleiner als die Anzahl der Befragten, jedoch ist es durch die geringe Cluster-Bildung sehr unwahrscheinlich, dass einzelne Interviewer desselben Geschlechts durch ihre Art der Interviewführung den in

4 Unsere Ergebnisse sind durch eine Mehrebenenanalyse (Random Intercept) vollständig replizierbar. Da wir nicht an einer Varianzerklärung auf Interviewer- oder Befragtenebene interessiert sind und zudem lediglich ein Merkmal auf der Interviewerebene betrachten können, haben wir uns für die Darstellung des einfacheren Regressionsmodells entschieden.

dieser Studie relevanten Interaktionseffekt zwischen Interviewer- und Befragten-geschlecht verantworten. Für die korrekte Schätzung der Standardfehler werden in den Regressionsanalysen zudem die Befragtencluster der einzelnen Interviewer berücksichtigt.

4 Ergebnisse

4.1 Deskriptive Analyse

Tabelle 2 zeigt konditionale Kreuztabellierungen zwischen den drei relevanten Items der Hausarbeit und dem Interviewergeschlecht getrennt für männliche und weibliche Befragte. Bereits diese bivariate Analyse deutet darauf hin, dass Männer gegenüber weiblichen Interviewern eine höhere Beteiligung an den drei häuslichen Tätigkeiten angeben als gegenüber männlichen. Ein besonders deutlicher Effekt des Interviewergeschlechts zeigt sich jedoch lediglich bei der Kategorie „keine Beteiligung“ an Hausarbeit. So berichten 54 % der Männer männlichen Interviewern keine Beteiligung am Waschen, während dies gegenüber weiblichen Interviewern lediglich bei 38 % der Fall ist. Ein noch größerer Unterschied zeigt sich beim Kochen, bei dem 41 % der Männer gegenüber männlichen Interviewern und nur 23 % gegenüber weiblichen Interviewern keine Beteiligung angeben. Beim Putzen ist der Unterschied mit 26 % zu 17 % geringer, aber dennoch auffällig. Zudem berichten Männer weiblichen Interviewern auch häufiger eine sehr hohe Beteiligung. Diese Unterschiede sind jedoch weitaus geringer.

Für Frauen zeigen sich generell geringere Unterschiede zwischen den Angaben gegenüber männlichen und weiblichen Interviewern. Anders als bei Männern sind diese Unterschiede eher am oberen Ende der Skala zu finden. Vor allem bezüglich der Tätigkeiten Kochen und Waschen geben Frauen bei weiblichen Interviewern seltener eine sehr hohe Beteiligung an als bei männlichen.

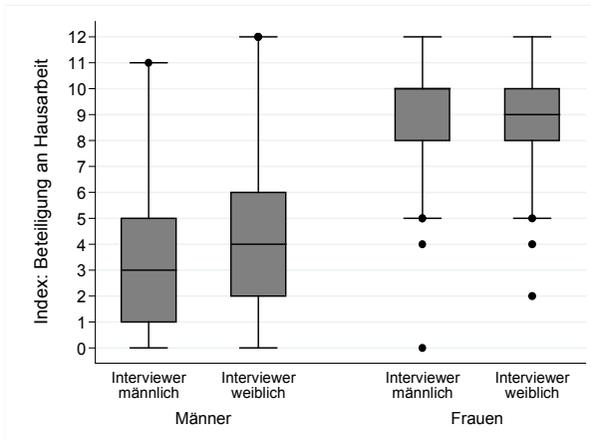
Insgesamt sprechen diese Ergebnisse für die in Hypothese 1 formulierte Vermutung, dass Männer gegenüber Frauen tendenziell ihre Beteiligung an Hausarbeit überschätzen, während Frauen diese eher unterschätzen, wenn sie von Frauen befragt werden. Die generell größeren Effekte für Männer sprechen weiterhin dafür, dass der normative Druck zu einer egalitären Hausarbeit für Männer als größer einzuschätzen ist, wie in Hypothese 2 postuliert.

Tabelle 2 Antwortverhalten zur Beteiligung an der Hausarbeit von Männern und Frauen nach Interviewergeschlecht, in Prozent

Beteiligung am	Männer		Frauen	
	Interviewergeschlecht männlich	Interviewergeschlecht weiblich	Interviewergeschlecht männlich	Interviewergeschlecht weiblich
<i>Waschen</i>				
keine	54,05	38,17	2,25	0,67
niedrig	17,57	21,37	1,12	2,01
egalitär	14,86	25,19	16,85	13,42
hoch	8,11	5,34	17,98	28,70
sehr hoch	5,41	9,92	61,80	55,70
<i>Cramers V=0,19, $Ch^2=7,14$ ($p=0,13$)</i>			<i>Cramers V=0,14, $Ch^2=4,55$ ($p=0,34$)</i>	
<i>Putzen</i>				
keine	25,68	16,79	2,25	0,67
niedrig	25,68	28,24	3,37	2,01
egalitär	45,95	47,33	43,82	40,27
hoch	1,35	3,82	25,84	30,87
sehr hoch	1,35	3,82	24,72	26,17
<i>Cramers V=0,14, $Ch^2=3,96$ ($p=0,41$)</i>			<i>Cramers V=0,10, $Ch^2=0,10$ ($p=0,70$)</i>	
<i>Kochen</i>				
keine	40,54	22,90	4,49	1,34
niedrig	28,38	37,40	2,25	6,71
egalitär	20,27	24,43	23,60	26,85
hoch	6,76	10,69	24,72	28,86
sehr hoch	4,05	4,58	44,94	36,24
<i>Cramers V=0,19, $Ch^2=7,33$ ($p=0,12$)</i>			<i>Cramers V=0,16, $Ch^2=6,05$ ($p=0,20$)</i>	
N Befragte	74	131	89	149

In Abbildung 1 sind nun die entsprechenden Ergebnisse für den additiven Index der Beteiligung an Hausarbeit an Hand von Boxplots dargestellt. Es zeigt sich das erwartete Muster. Männer berichten generell weniger Beteiligung an der Hausarbeit als Frauen, weisen jedoch auch eine höhere Streuung auf. Bezüglich des Einflusses des Interviewergeschlechts zeichnet sich ein den Ergebnissen für die einzelnen Items ähnliches Bild ab. Gegenüber Frauen geben Männer häufiger eine höhere Beteiligung an als gegenüber Männern. Sowohl Median als auch die 25%- und 75%-Quartile liegen bei weiblichen Interviewern jeweils einen Indexpunkt höher als bei männlichen Interviewern. Frauen wiederum äußern gegenüber weiblichen Interviewern eine geringere Beteiligung. Hier unterscheidet sich jedoch lediglich der Median um einen Indexpunkt.

Abbildung 1 Verteilung des Index der Beteiligung an der Hausarbeit nach Interviewer- und Befragtengeschlecht



4.2 Multivariate Analyse

Im Folgenden soll der Effekt des Interviewergeschlechts auf die Angaben von Männern und Frauen bezüglich der Beteiligung an Hausarbeit weiter herausgearbeitet werden. Tabelle 3 zeigt zunächst die Interaktion von Befragten- und Interviewergeschlecht unter Kontrolle von Alter, Bildung, Familienstand und Haushaltskomposition. Damit wird der Effekt um unterschiedliche Zusammensetzungen der Stichproben von männlichen und weiblichen Interviewern bereinigt. Der auf dem 0,1%-Niveau signifikant positive Effekt weiblicher Interviewer signalisiert, dass Männer gegenüber Frauen eine höhere Beteiligung an Hausarbeit berichten. Obwohl der Interaktionseffekt zwischen weiblichen Interviewern und weiblichen Befragten nicht signifikant ist, hat der Koeffizient substanzielle Implikationen. So zeigen Interaktionseffekt und Haupteffekt der weiblichen Interviewer zusammen genommen tendenziell einen negativen Effekt weiblicher Interviewer auf die berichtete Beteiligung an Hausarbeit. Der Effekt unterscheidet sich jedoch nicht signifikant von null.

Die im Modell enthaltenen Kontrollvariablen haben weitestgehend die erwartbaren Effekte. So verrichten Frauen im Schnitt mehr Hausarbeit als Männer. Weiterhin geben verheiratete Männer eine geringere Beteiligung an der Hausarbeit an als unverheiratete. Das umgekehrte Bild zeigt sich für Frauen.

Im Ganzen repliziert die erste Regressionsanalyse weitestgehend den bivariaten Befund. Vor allem Männer sind vom Geschlecht der Interviewer beeinflussbar

und berichten gegenüber Frauen, einen größeren Teil der Hausarbeit zu verrichten als gegenüber Männern. Frauen hingegen zeigen sich nicht signifikant vom Geschlecht der Interviewer beeinflussbar. Damit liefern die bisherigen Ergebnisse weitere Evidenz für Hypothese 1 und 2.

Tabelle 3 OLS-Regression zur Erklärung der Beteiligung an der Hausarbeit

Variablen	β	SE
weibliche Interviewerin (Ref.: männlicher Interviewer)	0,73*	(0,39)
weibliche Befragte (Ref.: männlicher Befragter)	4,15***	(0,57)
weibl. Int. x weibl. Befr.	-0,80	(0,49)
Alter	0,00	(0,01)
HS/RS m. Lehre (Ref.: kein Abschluss/HS/RS oder Lehre)	0,13	(0,35)
Abitur	-0,11	(0,45)
FH/Uni	-0,11	(0,38)
verheiratet (Ref.: Kohabitation)	-1,52***	(0,54)
verheiratet x weibl. Befr.	1,72***	(0,59)
ein Kind im HH (Ref.: kein Kind)	0,22	(0,52)
zwei und mehr Kinder im HH	-0,03	(0,47)
ein Kind im HH x weibl. Befr.	0,44	(0,61)
zwei und mehr Kinder im HH x weibl. Befr.	0,41	(0,57)
Konstante	4,45***	(0,67)
R ²	0,58	
N Befragte	443	
N Interviewer	145	

* $p < 0,1$; *** $p < 0,01$

Im nächsten Schritt wird das bisherige Modell durch mehrere Zweifach- und Dreifachinteraktionen aus Befragtengeschlecht, Interviewergeschlecht und Alter bzw. Bildung des Befragten erweitert (vgl. Tabelle A1 im Appendix). Damit soll getestet werden, ob die Effekte des Interviewergeschlechts – wie theoretisch vorhergesagt – besonders ausgeprägt sind für Personen, die sich egalitären Rollenbildern bewusst sind. In Abbildung 2 ist zunächst der aus diesen Modellen geschätzte marginale Effekt des Interviewergeschlechts nach Befragtenalter getrennt für Männer und Frauen zusammen mit dem 95 %-Konfidenzintervall abgetragen. Es zeigt sich eindeutig, dass jüngere Männer weiblichen Interviewern eine signifikant höhere Beteiligung an der Hausarbeit berichten als männlichen. Der Effekt des Interviewergeschlechts liegt für 20-jährige Männer bei ca. 2 Indexpunkten und nimmt mit steigendem Alter ab. Für Männer ab 50 Jahren bestehen keine signifikanten Unterschiede im Antwortverhalten gegenüber weiblichen und männlichen Interviewern. Zwar deuten die Ergebnisse für weibliche Befragte in jungem Alter

ebenfalls auf den theoretisch zu erwartenden negativen Effekt einer geringeren Angabe der Beteiligung an der Hausarbeit bei weiblichen Interviewern hin. Allerdings unterscheiden sich die Angaben bei männlichen und weiblichen Interviewern für keine Altersgruppe signifikant.

Abbildung 2 Marginaler Effekt weiblicher Interviewer nach Alter der Befragten

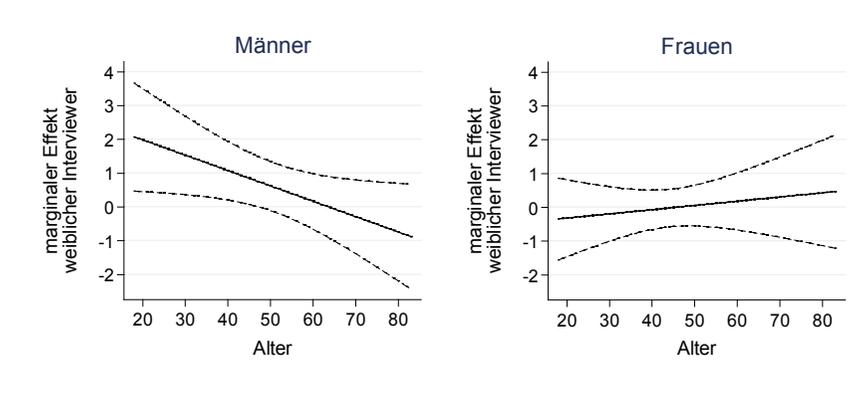
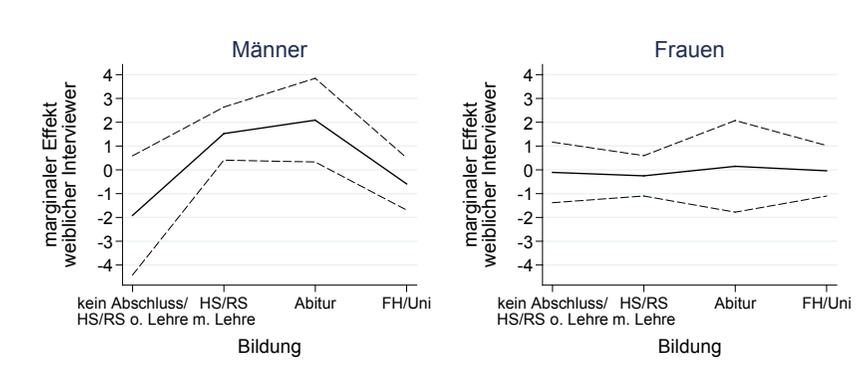


Abbildung 3 Marginaler Effekt weiblicher Interviewer nach Bildung der Befragten



Der Einfluss des Interviewergeschlechts nach der Bildung des Befragten, dargestellt in Abbildung 3, ist auch für Männer weniger deutlich. Für die drei ersten Kategorien zeigt sich zunächst das erwartete Muster eines mit der Bildung ansteigenden positiven Effekts weiblicher Interviewer auf die berichtete Beteiligung an

Hausarbeit. Sowohl Männer, die einem Hauptschul- bzw. Realschulabschluss eine Berufsausbildung angeschlossen haben, als auch Abiturienten berichten weiblichen Interviewern einen signifikant größeren Anteil an der Hausarbeit als männlichen Interviewern, während für Männer mit der geringsten Bildung kein signifikanter Effekt des Interviewergeschlechts besteht. Mit steigender Bildung scheint es somit zu einer höheren Wahrscheinlichkeit von Interviewereffekten zu kommen, da sich diese Männer der gesellschaftlichen Vorstellung einer egalitären Arbeitsteilung bewusst sind, diese aber meist im Haushalt nicht in konkretes Verhalten ummünzen. Dem widerspricht jedoch der Befund für Männer mit Hochschulabschluss, da für diese kein signifikanter Effekt des Interviewergeschlechts auftritt. Eine Erklärung könnte darin bestehen, dass Befragte mit Hochschulabschluss generell eine höhere Kooperationsbereitschaft gegenüber wissenschaftlichen Studien aufweisen. Wie Stocké (2004) zeigen konnte, wirkt eine solche kooperative Rahmung des Interviews Antwortverzerrungen entgegen. Dadurch kann möglicherweise die grundlegend vorhandene Tendenz zu einer Überschätzung der Beteiligung an der Hausarbeit gegenüber Frauen ausgeglichen werden. Diese Hypothese kann hier allerdings nicht getestet werden. Eine weitere Erklärungsmöglichkeit besteht darin, dass Männer mit Hochschulabschluss tatsächlich einen größeren Anteil der Hausarbeit verrichten als Männer der anderen Bildungsgruppen und somit weniger Anreiz haben, ihren Anteil zu überschätzen. Tatsächlich ergeben unsere Analysen, dass Männer mit Hochschulabschluss auch im Gespräch mit Männern eine höhere Beteiligung an Hausarbeit angeben als Männer in den anderen drei Bildungsgruppen (siehe Tabelle A1). Die Unterschiede zu den mittleren beiden Bildungsgruppen sind auf dem 5 %-Niveau signifikant. Da uns kein objektives Maß der Arbeitsteilung zur Verfügung steht, muss jedoch auch dies eine Vermutung bleiben.

Für Frauen zeigt sich abermals kein signifikanter Effekt des Interviewergeschlechts. Auch in der Tendenz unterscheiden sich die Bildungsgruppen nicht bezüglich des Effekts.

Insgesamt bestätigen sich damit die Hypothesen weitestgehend nur für Männer. Diese geben gegenüber weiblichen Interviewern an, mehr Hausarbeit zu verrichten, als gegenüber männlichen Interviewern. Insbesondere jüngere und höher gebildete Männer, von denen wir annehmen, modernere Geschlechterrollenbilder zu vertreten, zeigen dieses Verhalten. Auf die Aussagen von Frauen bezüglich ihrer Beteiligung an Hausarbeit hat das Geschlecht des Interviewers dagegen kaum Auswirkungen. Wie in Hypothese 2 erwartet, scheint der soziale Druck auf Männer größer zu sein, ihre Angaben an die Erwartungen des weiblichen Interviewers anzupassen, als dies für Frauen der Fall ist.

5 Diskussion

An Hand von Daten der Mannheimer Familienstudie 2005, die mit Hilfe von 145 studentischen Interviewern durchgeführt wurde, hat dieser Beitrag den Effekt des Interviewergeschlechts auf die berichtete Beteiligung an häuslicher Arbeit untersucht. Im Mittelpunkt stand dabei insbesondere die Interaktion des Interviewergeschlechts mit dem Geschlecht des Befragten sowie zusätzlich mit den Befragtenmerkmalen Alter und Bildung, welche als Proxy für individuelles Bewusstsein egalitärer Geschlechterrollen in unserer Studie fungieren.

Vor dem Hintergrund einer wachsenden Akzeptanz moderner Geschlechterrollenbilder in der heutigen Gesellschaft bei gleichzeitiger Diskrepanz zwischen erwünschter und tatsächlicher Arbeitsteilung in Partnerschaften wurde auf Basis der Theorie der rationalen Wahl erwartet, dass Männer gegenüber weiblichen Interviewern ihre Beteiligung an der Hausarbeit übertreiben, um den sozial erwünschten Idealen zu entsprechen. Frauen hingegen sollten aus den gleichen theoretischen Gründen bei weiblichen Interviewern ihre Beteiligung an der Hausarbeit unterschätzen. Da moderne Geschlechterrollen vor allem von Männern eine Anpassung ihres Verhaltens im Haushalt fordern, ist der soziale Druck in der Interviewsituation für männliche Befragte größer als für weibliche. Folglich sollten Männer im Vergleich zu Frauen anfälliger für Interviewereffekte des Geschlechts sein. Weiterhin wurden derartige Interviewereffekte vor allem bei solchen Personen erwartet, von denen man annehmen kann, dass sie sich moderner Geschlechterrollen bewusst sind und diese möglicherweise selbst vertreten. Aufgrund ihrer Sozialisation sollte dies vor allem auf jüngere sowie gebildete Befragte zutreffen. Gerade für diese ist der Druck, sozial erwünscht zu antworten, besonders hoch, da in ihrem unmittelbaren Umfeld das Ideal egalitärer Arbeitsteilung am weitesten verbreitet ist, sie dieses in der Realität aber meist noch immer nicht erreichen.

Unsere empirischen Ergebnisse lassen nun darauf schließen, dass diese theoretischen Überlegungen weitestgehend zutreffend sind. Allerdings zeigen sich Frauen in Bezug auf die Frage der Arbeitsteilung in Partnerschaften nicht signifikant vom Geschlecht der Interviewer beeinflussbar. Wie erwartet, ergeben sich besonders für Männer in jüngeren Jahren sowie für höher gebildete Männer ausgeprägte Effekte des Interviewergeschlechts. Einzig Hochschulabsolventen weichen von diesem Muster ab. Möglicherweise ist dies mit einer erhöhten Kooperationsbereitschaft gegenüber wissenschaftlichen Befragungen dieser Personengruppe bzw. mit deren tatsächlich erhöhter Beteiligung an Hausarbeit zu erklären.

Mit den vorliegenden Daten besteht jedoch keine Möglichkeit zu testen, ob die Antworten bei weiblichen Interviewern oder jene bei männlichen Interviewern

näher am „wahren Wert“ liegen. Ohne objektive Messung der Arbeitsteilung ist beispielsweise die Behauptung durchaus zulässig, dass Männer nicht Frauen gegenüber ihre Beteiligung an Hausarbeit überschätzen, sondern diese bei Männern untertreiben. Theoretisch erscheint uns der hier ausführlich beschriebene Mechanismus jedoch plausibler.

Ungeachtet dieser theoretischen Frage zeigt der Beitrag jedoch eindeutig, dass Auskünfte über konkretes Alltagshandeln vom Geschlecht der Interviewer beeinflusst werden. Dies ist erstaunlich, da davon auszugehen ist, dass Befragte ungefähr einschätzen können, in welchem Ausmaß sie sich in Relation zum Partner an der Hausarbeit beteiligen. Somit sollte Befragten durchaus bewusst sein, dass sie nicht den „wahren Wert“ angeben. Interviewereffekte sind darüber hinaus nicht homogen, sondern verschieden stark für spezifische Bevölkerungsgruppen. Dies macht die Identifizierung und noch vielmehr die Korrektur daraus resultierender Antwortverzerrungen weitaus komplizierter, da derartige Interaktionen zuvor stets theoretisch herausgearbeitet werden müssen.

Die Ergebnisse sind auch dahingehend interessant, dass derartig komplexe Interviewereffekte in telefonischen Interviews, in denen die Öffentlichkeit und Sichtbarkeit von Interviewermerkmalen erheblich eingeschränkt sind, auftreten können. Dies liegt womöglich daran, dass das Geschlecht des Interviewers als einziges Merkmal auch in telefonischen Befragungen erkenntlich ist und somit Interviewereffekte des Geschlechts deutlicher hervortreten. Interessant wäre es dennoch, die Ergebnisse zu telefonischen Interviews mit denjenigen aus persönlichen Befragungen zu vergleichen. Hier sind die Verzerrungen möglicherweise noch komplexer, da neben dem Geschlecht noch andere Eigenschaften des Interviewers eine Rolle spielen können. Eventuell sind Interviewereffekte des Geschlechts jedoch schwieriger festzustellen, da sie von Effekten anderer Interviewermerkmale überlagert werden.

Unsere empirische Untersuchung macht darauf aufmerksam, dass bei inhaltlichen Fragestellungen gleich welcher Art stets auch Aspekte des Datenerhebungsprozesses Berücksichtigung finden müssen. Dazu benötigt es aber einerseits eine ausgeprägte Methodenforschung, die etwaige Antwortverzerrungen empirisch herausarbeitet und andererseits eine starke theoretische Herleitung derartiger Verzerrungen. Bei Forschungsfragen zur Arbeitsteilung in Partnerschaften bietet es sich nun vor allem im Rahmen von Robustheitsanalysen an, nach Möglichkeit das Interviewergeschlecht sowie die hier behandelten Interaktionseffekte zu kontrollieren, um Effekte inhaltlicher Variablen von methodischen Artefakten zu trennen.

Nach den vorgestellten Ergebnissen ist mit Blick auf die empirische Forschung im Bereich häuslicher Arbeitsteilung vermehrt vor der Nutzung der hier

verwendeten Indikatoren der Arbeitsteilung zu warnen. Denn selbst berichtete Einschätzungen der relativen Beteiligung an Hausarbeit verstärken möglicherweise das Auftreten von Interviewereffekten. Anders als bei einfachen zeitlichen Auskünften zur eigenen Person ist der Interviewer bei relativen Angaben direkt darüber informiert, ob die Arbeitsteilung egalitär abläuft bzw. welcher Partner den Hauptanteil an der Erledigung der Hausarbeit trägt. Da den Interviewern die Arbeitsteilung in der jeweiligen Partnerschaft unmittelbar offenbart wird, sollte bei relationalen Angaben für Befragte der Anreiz noch größer sein, sozial erwünscht zu antworten. Genauerem Aufschluss zu dieser Frage kann allerdings nur zukünftige Forschung zum Vergleich der Interviewereffekte über die verschiedenen Maße der absoluten und relativen Zeitverwendung bringen.

Literatur

- Argentino, C., M. Aline, C. Kidd und K. Bogart, 1977. The Effects of Experimenter's Sex and Subject's Sex on the Attitudes toward Fraternity, Sorority, and Mixed-dormitory Residents. *Journal of Community Psychology* 5: 186-188.
- Ballou, J. und F. K. DeBoca, 1980. Gender Interaction Effects on Survey Measures in Telephone Interviews. Paper presented at the annual meeting of the American Association of Public Opinion Research, Mason, Ohio.
- Batalova, J. und P. Cohen, 2002. Premarital Cohabitation and Housework: Couples in Cross-National Perspective. *Journal of Marriage and Family* 64: 743-755.
- Braun, M. und J. Scott, 2009. Gender Role Egalitarianism – Is the Trend Reversal Real? *International Journal of Public Opinion Research* 21: 362-367.
- Breen, R. und L. P. Cooke, 2005. The Persistence of the Gendered Division of Domestic Labour. *European Sociological Review* 21: 43-57.
- Brosius, H.-B., N. Mundorf und J. F. Staab, 1991. The Depiction of Sex Roles in American and German Magazine Advertisements. *International Journal of Public Opinion Research* 3: 366-383.
- Diekmann, A., 2004. *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*. 11. Auflage. Reinbek: Rowohlt.
- Dijkstra, W., 1983. How Interviewer Variance Can Bias the Results of Research on Interviewer Effects. *Quality and Quantity* 17: 179-187.
- Esser, H., 1986. Können Befragte lügen? Zum Konzept des „wahren Wertes“ im Rahmen der handlungstheoretischen Erklärung von Situationseinflüssen bei der Befragung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 38: 314-336.
- Esser, H., 1991. Die Erklärung systematischer Fehler in Interviews: Befragtenverhalten als „Rational Choice“. S. 59-78. in: R. Wittenberg (Hg.): *Person – Situation – Institution – Kultur*. Günter Büschges zum 65. Geburtstag. Berlin: Duncker & Humblot.
- Fuwa, M., 2004. Macro-Level Gender Inequality and the Division of Household Labor in 22 Countries. *American Sociological Review* 69: 751-767.
- Gabler, S. und S. Häder (Hg.). 2002. *Telefonstichproben. Methodische Innovationen und Anwendungen in Deutschland*. Münster u. a.: Waxmann-Verlag.
- Grimes, M. D. und G. L. Hansen, 1984. Response Bias in Sex-Role Attitude Measurement. *Sex Roles* 10: 67-72.
- Groves, R. M., 1989. *Survey Errors and Survey Costs*. New York: Wiley.

- Groves, R. M., F. J. Fowler, M. P. Couper, J. M. Lepkowski, E. Singer und R. Tourangeau, 2004. *Survey Methodology*. Hoboken: Wiley.
- Groves, R. M. und N. H. Fultz, 1985. Gender Effects Among Telephone Interviewers in a Survey of Economic Attitudes. *Sociological Methods and Research* 14: 31-52.
- Grunow, D., F. Schulz und H.-P. Blossfeld, 2007. Was erklärt die Traditionalisierungsprozesse häuslicher Arbeitsteilung im Eheverlauf: soziale Normen oder ökonomische Ressourcen? *Zeitschrift für Soziologie* 36: 162-181.
- Häder, S. und S. Gabler, 1998. Ein neues Stichprobendesign für telefonische Umfragen in Deutschland. S. 43-57 in: Gabler, S., S. Häder und J. H. P. Hoffmeyer-Zlotnik (Hg.): *Telefonstichproben in Deutschland*. Opladen: Westdeutscher Verlag.
- Henninger, A., C. Wimbauer und R. Dombrowski, 2008. Demography as a Push toward Gender Equality? Current Reforms of German Family Policy. *Social Politics* 15: 287-314.
- Hochschild, A. und A. Machung, 1989. *The Second Shift: Working Parents and the Revolution at Home*. New York: Viking.
- Huddy, L., J. Billig, J. Bracciodieta, L. Hoeffler, P. J. Moynihan und P. Pugliani, 1997. The Effect of Interviewer Gender on the Survey Response. *Political Behaviour* 19: 197-220.
- Johnson, T. P., M. Fendrich, C. Shaligram, A. Garcy und S. Gillespie, 2000. An Evaluation of the Effects of Interviewer Characteristics in an RDD Telephone Survey of Drug Use. *Journal of Drug Issues* 30: 77-102.
- Juster, F. T., H. Ono und F. P. Stafford, 2003. An Assessment of Alternative Measures of Time Use. *Sociological Methodology* 33: 19-54.
- Kane, E. W. und L. J. Macaulay, 1993. Interviewer Gender and Gender Attitudes. *Public Opinion Quarterly* 57: 1-28.
- Lee, Y.-S. und L. J. Waite, 2005. Husbands' and Wives' Time Spent on Housework. A Comparison of Measures. *Journal of Marriage and Family* 67: 328-336.
- Lewin-Epstein, N., H. Stier und M. Braun, 2006. The Division of Household Labor in Germany and Israel. *Journal of Marriage and Family* 68: 1147-1164.
- Lueptow, L. B., S. L. Moser und B. F. Pendleton, 1990. Gender and Response Effects in Telephone Interviews about Gender Characteristics. *Sex Roles* 22: 29-42.
- Mohr, H.-M., 1986. Dritte beim Interview. Ergebnisse zu Indikatoren aus dem Bereich Ehe und Partnerschaft mit Daten des „Wohlfahrtssurvey 1984“. *ZA-Information* 19: 52-71. http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/za_information/ZA-Info-19.pdf (14.10.2010).
- Press, J. E. und E. Townsley, 1998. Wives' and Husbands' Housework Reporting: Gender, Class, and Social Desirability. *Gender and Society* 12: 188-218.
- Reinecke, J., 1991. Interviewer- und Befragtenverhalten. Theoretische Ansätze und methodische Konzepte. Opladen: Westdeutscher Verlag.
- Sayer, L. C., 2005. Gender, Time and Inequality: Trends in Women's and Men's Paid Work, Unpaid Work and Free Time. *Social Forces* 84: 285-303.
- Schejbal, J. A., H. Sachs und P. J. Lavrakas, 1993. "Hello, do you remember us?": What Respondents Remember about the Interview. Paper presented at the Annual Meeting of the Midwest Association for Public Opinion Research. Chicago.
- Schulz, F. und H.-P. Blossfeld, 2006. Wie verändert sich die häusliche Arbeitsteilung im Eheverlauf? Eine Längsschnittstudie der ersten 14 Ehejahre in Westdeutschland. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 58: 23-49.
- Scott, J., D. F. Alwin und M. Braun, 1996. Generational Changes in Gender-Role Attitudes: Britain in a Cross-National Perspective. *Sociology* 30: 471-492.
- Steinkopf, L., G. Bauer und H. Best, 2010. Nonresponse und Interviewer-Erfolg im Telefoninterview. Empirische Untersuchungen zum Einfluss stimmlicher Eigenschaften der Interviewer. *Methoden – Daten – Analysen*, 4 (1): 3-26. http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/mda/Vol.4_Heft_1/01_Best.pdf (14.10.2010).

- Stocké, V., 2004. Entstehungsbedingungen von Antwortverzerrungen durch soziale Erwünschtheit. Ein Vergleich der Prognosen der Rational-Choice Theorie und des Modells der Frame-Selektion. *Zeitschrift für Soziologie* 33: 303-320.
- Sullivan, G. L., P. J. O'Connor, 1988. Women's Role Portrayals in Magazine Advertising: 1958-1983. *Sex Roles* 18: 181-188.
- van der Lippe, T., J. de Ruijter, E. de Ruijter und W. Raub, 2010. Persistent Inequalities in Time Use between Men and Women. *European Sociological Review: Advance Access*. <http://esr.oxfordjournals.org/content/early/2010/01/19/esr.jcp066.full.pdf+html> (14.10.2010).
- van Snippenburg, L. B., M. Ter Voert und J. Janssen, 1990. Social Positions of Women and Sex-Role Attitudes: Differences between Women in the Netherlands in the Eighties. *European Sociological Review* 6: 275-284.
- Weidmann, C., P. Schmich und S. Schiller-Born, 2008. Der Einfluss von Kontrollüberzeugungen der Interviewer auf die Teilnahme an telefonischen Befragungen. *Methoden – Daten – Analysen* 2 (2): 125-147. http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/mda/Vol.2_Heft_2/04_Schmich.pdf (14.10.2010).
- Wilcox, C., 1991. Support for Gender Equality in West Europe. A Longitudinal Analysis. *European Journal of Political Research* 20: 127-147.

Anschrift der Autoren

Markus Klein
Mannheimer Zentrum für Europäische
Sozialforschung (MZES)
Universität Mannheim
A5, 6
68131 Mannheim
Markus.Klein@mzes.uni-mannheim.de

Michael Kühhirt
Mannheimer Zentrum für Europäische
Sozialforschung (MZES)
Universität Mannheim
Parkring 47
68131 Mannheim
Michael.Kuehhirt@mzes.uni-mannheim.de

Appendix

Tabelle A1 OLS-Regression zur Interaktion zwischen Geschlecht, Interviewergeschlecht und Alter bzw. Bildung

Variablen	Modell Alter	Modell Bildung
weibliche Interviewerin (Ref.: männlicher Interviewer)	2,88** (1,17)	-1,91 (1,27)
weibliche Befragte (Ref.: männlicher Befragter)	2,31* (1,34)	4,63*** (1,16)
weibl. Int. x weibl. Befr.	-3,46** (1,62)	1,81 (1,38)
weibl. Int. x weibl. Befr. x Alter	0,06* (0,03)	
weibl. Int. x Alter	-0,05** (0,02)	
weibl. Befr. x Alter	0,05* (0,03)	
weibl. Int. x weibl. Befr. x HS/RS m. Lehre		-3,58** (1,43)
weibl. Int. x weibl. Befr. x Abitur		-3,75* (1,91)
weibl. Int. x weibl. Befr. x FH/Uni		-1,25 (1,54)
weibl. Int. x HS/RS m. Lehre		3,44** (1,37)
weibl. Int. x Abitur		4,00*** (1,53)
weibl. Int. x FH/Uni		1,32 (1,33)
weibl. Befr. x HS/RS m. Lehre		0,95 (1,02)
weibl. Befr. x Abitur		0,02 (1,47)
weibl. Befr. x FH/Uni		-2,16* (1,22)
Alter	-0,01 (0,02)	0,00 (0,01)
HS/RS m. Lehre (Ref.: kein Abschluss/HS/RS o. Lehre)	0,18 (0,35)	-0,79 (1,07)
Abitur	-0,10 (0,43)	-0,75 (1,24)
FH/Uni	0,06 (0,39)	1,16 (1,07)
verheiratet (Ref.: Kohabitation)	-0,77 (0,60)	-1,33** (0,54)
verheiratet x weibl. Befr.	0,40 (0,69)	1,42** (0,58)
ein Kind im HH (Ref.: kein Kind)	-0,38 (0,51)	0,10 (0,51)
zwei und mehr Kinder im HH	-0,76 (0,52)	0,24 (0,45)
ein Kind im HH x weibl. Befr.	1,58** (0,62)	0,59 (0,62)
zwei und mehr Kinder im HH x weibl. Befr.	1,68** (0,66)	0,18 (0,56)
Konstante	4,69*** (1,11)	4,41*** (1,15)
N Befragte	443	443
N Interviewer	145	145
R ²	0,61	0,60

* $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$; Standardfehler in Klammern

Mobiltelefon- erfahrung und Antwortqualität bei Umfragen

Experience with Mobile Phones and Survey Response Quality

Michael Häder und Mike Kühne

Zusammenfassung

Telefonische Befragungen über das Mobilfunknetz sind inzwischen ein Bestandteil der sozialwissenschaftlichen Datengewinnung. Da es sich beim Mobilfunk jedoch noch um eine relativ junge Befragungstechnologie handelt, stellt sich die Frage, inwieweit die Routine im Umgang mit dem Handy als eine Determinante der Antwortqualität wirkt. Dieser Frage wurde im Rahmen des Projektes CELLA nachgegangen (Telefonbefragungen in der Allgemeinbevölkerung über das Mobilfunknetz 2009, GESIS Köln, Deutschland, DOI: 10.4232/1.4875). CELLA wurde durch die Deutsche Forschungsgemeinschaft gefördert und steht für telefonische Befragungen via Mobilfunk (*CELL* phone) und via Festnetz (*LAnd*line phone). Zunächst werden die Operationalisierung und die Bildung eines Index beschrieben, anhand dessen die Routine im Umgang mit dem Mobilfunk gemessen werden kann. Auf der Basis von sechs Kriterien der Antwortqualität wird dann nach den entsprechenden Zusammenhängen gesucht. Dabei stellt sich heraus, dass sich – entgegen unseren ursprünglichen Erwartungen – solche Zusammenhänge jedoch kaum nachweisen lassen. Die Befunde sind für den Einsatz von telefonischen Befragungen über das Mobilfunknetz insgesamt positiv zu bewerten. Sie

Abstract

Telephone surveys via mobile phones have become increasingly important for the social sciences. The use of mobile phones in interviews is still a novelty. Thus the question arises to what extent the respondents' experience with mobile phones influences the response quality. The German Research Foundation funded a CELLA project to investigate how to conduct mobile phone surveys. CELLA stands for *CELL* phone and *LAnd*line phone surveys. Technische Universität Dresden, Germany, and GESIS – Leibniz Institute for Social Sciences, Germany, cooperated for that purpose. Data presented in the following section were collected within the frame of CELLA (Telefonbefragungen in der Allgemeinbevölkerung über das Mobilfunknetz 2009, GESIS Köln, DOI: 10.4232/1.4875).

Firstly, an index for the experience in handling mobile phones is presented. This is followed by an examination of correlations between the index and six aspects of response quality. A significant influence of the experience was – against our expectations – not found. The results support the conduct of telephone surveys via mobile phones. Nevertheless, there are numerous determinants with stronger influence, which suggests that further research needs to be done.

verdienen es aber auch, bei der Diskussion um die Determinanten der Antwortqualität bei Umfragen allgemein berücksichtigt zu werden. So zeigt sich vor allem, dass von einer ganzen Reihe an Determinanten ein stärkerer Einfluss auf die Antwortqualität ausgeht als von der Routine.

1 Einleitung¹

Telefonische Befragungen in der Allgemeinbevölkerung, die auf Kontakte über das Mobilfunknetz verzichten, liefern inzwischen aufgrund von Coverage-Problemen nur noch verzerrte Ergebnisse. Verantwortlich dafür ist vor allem die Gruppe der Mobile-Onlys (Kuusela/Callegaro/Vehovar 2008; Blumberg/Luke 2008). Dabei handelt es sich um Personen, die nur noch über ein Mobiltelefon erreichbar sind, die also nicht mehr über einen Festnetzanschluss verfügen. Dies betrifft gegenwärtig etwa 11 % der Bevölkerung der Bundesrepublik. Vermutlich wird dieser Anteil in der nächsten Zeit noch ansteigen. Insbesondere ist zu beachten, dass sich dieser Personenkreis durch ganz spezielle sozio-demografische Charakteristika auszeichnet und nicht etwa einen Querschnitt der Gesellschaft darstellt (vgl. Graeske/Kunz 2009: 57ff.).

Die damit notwendig werdenden Befragungen über das Mobilfunknetz bringen jedoch für die Umfrageforschung auch methodische Probleme mit sich. Spätestens seit den Arbeiten von Dillman (vgl. 2000) ist bekannt, dass die Routine im Umgang mit dem Medium, über welches eine Befragung vermittelt wird, für die Qualität der Antworten mit verantwortlich ist. Entsprechende Effekte sind beispielsweise von ihm und seinen Kollegen bei Internetbefragungen nachgewiesen worden, die sie dann den Ergebnissen von postalischen und persönlich-mündlichen Befragungen gegenüber gestellt haben (vgl. Christian/Dillman/Smyth 2008: 257).

Auch bei der Kommunikation mittels Mobilfunk handelt es sich um eine noch relativ junge Technologie. Trotz der hohen Penetrationsrate – statistisch gesehen verfügt jeder Haushalt in der Bundesrepublik bereits über mehr als eine Mobilfunknummer (vgl. Häder/Gabler/Heckel 2009) – kann noch nicht davon ausgegangen werden, dass der Umgang mit diesen Geräten, etwa im Vergleich mit Festnetztelefonen, bereits bei allen Nutzern zur Routine geworden ist.

1 Die Autoren bedanken sich bei den beiden anonymen Gutachtern für die zahlreichen Hinweise zu ihrem Manuskript. Sie waren uns eine große Hilfe bei der Fertigstellung des vorliegenden Beitrages.

Der vorliegende Text wendet sich anhand empirischer Daten deshalb der Frage zu, ob ein und – falls gegeben – welcher Einfluss von der Routine im Umgang mit den Mobilfunkgeräten auf die Qualität der Antworten bei einer sozialwissenschaftlichen Befragung via Mobiltelefon ausgeht.

Bei der für diese Analysen benutzten Datenbasis handelte es sich um die Studie CELLA 1² (*Cell-Phone* und *Land-Line*). Es wurden hier insgesamt 1.009 zufällig ausgewählte Personen über das Festnetz und 1.162 Personen über den Mobilfunk mit den gleichen Instrumenten befragt. Grundlage der Stichprobenziehung für die Befragung über das Festnetz war das Gabler-Häder-Design (vgl. Gabler/Häder 2009). Für die Auswahl der Teilnehmerinnen und Teilnehmer für die Mobilfunkbefragung wurde ein modifiziertes RDD-Verfahren verwendet (vgl. Häder/Gabler/Heckel 2009: 21ff.).

Im folgenden Abschnitt wird, nachdem einige theoretische Überlegungen zur Erklärung des Antwortverhaltens dargestellt wurden, die Operationalisierung der Routine im Umgang mit dem Mobiltelefon und der drauf aufbauende Index präsentiert. Dieser drückt aus, inwieweit die Zielpersonen im Umgang mit ihrem Handy vertraut sind. Im Anschluss werden dann anhand des genannten Datensatzes (vgl. Häder/Häder 2009: 269ff.) die verschiedenen Kriterien für die Antwortqualität mit dem zuvor gebildeten Routineindex in Beziehung gesetzt. Abschließend wird versucht, die empirischen Befunde auch theoretisch einzuordnen.

2 Kognitionspsychologische Modelle der Antwortgenerierung: Implikationen für die Antwortqualität

Kognitive Modelle des Antwortprozesses (vgl. z. B. Sudman/Bradburn/Schwarz 1996; Tourangeau/Rips/Rasinski 2000) stellen eine hilfreiche Basis zur Erklärung des Antwortverhaltens bereit. Danach sind erstens das Verständnis der Frage, zweitens der Abruf von Informationen aus dem Gedächtnis, drittens die Bildung einer Antwort sowie viertens das Formulieren der Antwort die wesentlichen Schritte, die eine Person bei einer Befragung zu absolvieren hat. Insbesondere die Theorien paralleler

2 Der Datensatz der CELLA 1 Untersuchung ist beim Datenarchiv der GESIS – Leibniz Institut für Sozialwissenschaften – erhältlich (DOI: 10.4232/1.4875). Eine ausführliche Darstellung der gesamten Untersuchung einschließlich des Fragebogens findet sich bei Häder/Häder (2009). Die in diesem Beitrag verwendeten Variablenbezeichnungen sind mit denen im Datensatz identisch. Die Studie CELLA 2 wird 2010 durchgeführt. Die Untersuchung CELLA 1 zu Telefonbefragungen über das Mobilfunknetz wurde von der DFG gefördert. Die Förderung für CELLA 2 erfolgt ebenfalls von dort.

Verarbeitungswege haben der empirischen Umfrageforschung für die Darstellung zahlreicher empirischer Phänomene den theoretischen Rahmen liefern können.

Aufgrund der Komplexität des Antwortprozesses erscheint es jedoch fraglich, inwieweit die Befragten tatsächlich, wie in den hier ausgearbeiteten kognitionspsychologischen Pfadmodellen beschrieben (vgl. Strack 1994: 53; Sudman/Bradburn/Schwarz 1996: 56; Tourangeau/Rips/Rasinski 2000: 8; Jobe/Herrmann 1996), den kompletten Prozess der Antwortgenerierung durchlaufen (vgl. Tourangeau/Rips/Rasinski 2000: 354). Es ist vielmehr davon auszugehen, dass es bei ungünstigen Bedingungen während des Interviews – beispielsweise bei kognitiv besonders anspruchsvollen Sachverhalten, bei einer niedrigen Salienz des in der Frage aufgegriffenen Gegenstandes oder, wie an dieser Stelle angenommen, bei mangelnder Routine im Umgang mit dem Kommunikationsmedium, über welches die Befragung geführt wird – zum Einsatz von Heuristiken bei der Antwortgenerierung kommt (vgl. Strack/Martin 1988). Diesem Modell weiter folgend liegt der Antwort dann keine kognitiv anspruchsvolle Urteilsbildung zugrunde, welche auf den individuellen bisherigen Wissensbeständen zum erfragten Sachverhalt basiert. Vielmehr entscheiden sich die Befragten bei ihrer Antwort für eine vermeintlich einfache kognitive Lösung. Wenn überhaupt eine Antwort übermittelt wird, ist diese relativ spontan. Problematisch ist, dass damit Beeinträchtigungen der Datenqualität einhergehen, da sowohl eine erhöhte Tendenz zu Item-Nonresponse als auch zu einer zufälligen Entscheidung für eine Antwortkategorie feststellbar sind.

Krosnick und Alwin (1987; vgl. auch Strack/Martin 1988) entwickelten ein Modell paralleler Verarbeitungswege, das über den Spezialfall der Meinungsfragen hinaus auf alle Phasen des Antwortprozesses angewendet werden kann. Dabei wird insbesondere die Motivation der Befragten in den Mittelpunkt gestellt; und es wird davon ausgegangen, dass – beispielsweise wiederum aufgrund fehlender Routine im Umgang mit dem Mobiltelefon – weniger motivierte Befragte die kognitiven Verarbeitungsprozesse nur oberflächlich durchführen (Weak Satisficing) und einige diese sogar vollständig überspringen (Strong Satisficing). Sowohl das Strong- als auch das Weak Satisficing führen so zu Reihenfolgeeffekten (vgl. Payne 1949; Hyman/Sheatsley 1950; Dillman et al. 1995; für einen deutschsprachigen Überblick: Faulbaum/Prüfer/Rexroth 2009).

In zahlreichen Studien haben Umfrageforscher eine Vielzahl von weiteren Effekten nachgewiesen, die infolge der genannten Satisficing-Strategien hervorgebracht werden und die damit zu systematischen Verzerrungen des wahren Wertes führen. Dazu zählen beispielsweise auch Skaleneffekte (vgl. Stadtler 1983; Schwarz/Hippler/Deutsch/Strack 1985; Bishop 1987; Böcker 1988) und Kontexteffekte (vgl. Cantril 1944, zitiert nach Strack 1994; Schuman/Presser 1981: 47; für einen Über-

blick: Schuman 1992). Je stärker die hier aufgezählten Effekte auftreten, desto geringer ist die Qualität der Antwort. Daraus lassen sich die folgenden Aspekte als Grundlage der sich anschließenden Analysen ableiten. Basierend auf den Annahmen der kognitionspsychologischen Modelle der Antwortgenerierung lässt sich eine hohe Antwortqualität erwarten, wenn die Befragten alle vier Phasen des Antwortprozesses ungestört durchlaufen. Im Gegensatz dazu sollte eine mangelnde Routine im Umgang mit dem Mobilfunk zu Ablenkung im Antwortprozess und damit zu weniger elaborierten Antworten und damit zu einer geringeren Antwortqualität führen.

3 Operationalisierung der unabhängigen Variablen

Zwei Befragungen, die sich auf die gleiche Grundgesamtheit beziehen und das gleiche Instrument verwenden, müssen nicht zwingend zu den gleichen Befunden führen. Insbesondere, wenn verschiedene Erhebungsmodes verwendet werden, kann es zu Unterschieden im Antwortverhalten kommen (Häder/Kühne 2009a: 165ff.). Ursachen dafür liegen unter anderem in modespezifischen Eigenschaften der Erhebung. Für eine Befragung über den Mobilfunk lassen sich einige Besonderheiten vermuten (Häder/Kühne 2009a: 165). Schwerpunkt der folgenden Analysen wird die Routine im Umgang mit dem Mobiltelefon sein.

Routine im Umgang mit dem Mobiltelefon

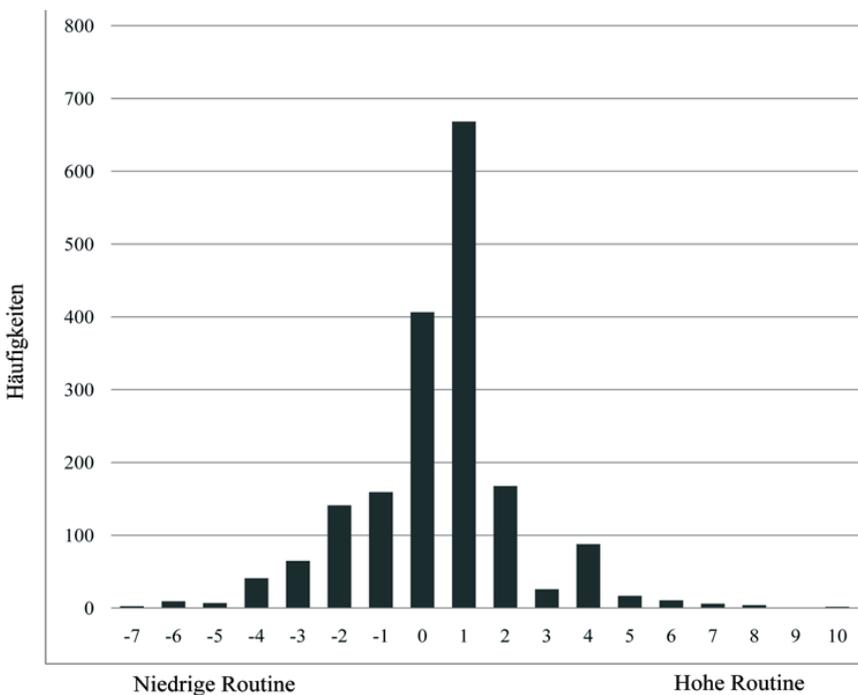
Die Erfahrung im Umgang mit Mobiltelefonen lässt sich nur schwer anhand einer einzelnen Frage messen. Es ist aber davon auszugehen, dass ein Anstieg in der Intensität der Telefonnutzung zu einem Zuwachs an Erfahrung im Umgang führt. Die Intensität der Handynutzung wurde anhand der folgenden Indikatoren³ gemessen:

- Anzahl der Handynummern, über die die Zielperson privat erreichbar ist (v4),
- Häufigkeit, mit der auf dem eigenen Handy Anrufe eingehen (v9),
- Häufigkeit, mit der das Handy eingeschaltet ist (v8_1) und
- Häufigkeit, mit der das Handy auch unterwegs mitgenommen wird (v11).

3 Es wurde auch getestet, ob die Einbeziehung weiterer Variablen in die Berechnung des Index zu einer Verbesserung von dessen Güte führt. Dabei handelte es sich um Fragen danach, ob eingehende Anrufe verpasst werden (v10), um die Anzahl der Rufnummern, die auf dem eigenen Handy gespeichert sind (v16) und um eine Frage danach, ob Anrufe in der Öffentlichkeit als peinlich empfunden werden (v17). Da dies nicht der Fall war, wurde auf die Berücksichtigung dieser Indikatoren verzichtet.

Diese Variablen wurden standardisiert und zu einem additiven Index zusammengefasst, bei dem ansteigende Werte eine zunehmende Routine repräsentieren. Der Index weist eine ausreichende interne Konsistenz (Cronbachs Alpha = 0,77) auf und ist eindimensional. Aus Abbildung 1 ist die Verteilung des Index zu entnehmen. Dabei ist zu erkennen, dass die Routine im Umgang mit dem Mobilfunkgerät in der Stichprobe durchaus unterschiedlich vorhanden ist und nicht etwa nur Personen mit einer ausgeprägten Technikaffinität befragt worden sind.

Abbildung 1 Häufigkeitsverteilung des Routine-Index



Eine erste nähere Betrachtung macht deutlich, dass es – wie auch nicht anders zu erwarten war – einen Zusammenhang des Grades an Routine mit dem Alter der Befragten gibt. Der Korrelationskoeffizient nach Pearson beträgt $r=-0,39$ ($p<0,001$). Jüngere Personen besitzen damit eine höhere Routine im Umgang mit dem Handy. Auch ein klarer Geschlechterunterschied ist anhand der Daten zu verzeichnen: Männer verfügen über eine größere Routine als Frauen ($M_{\text{Männer}} = 3,30 > M_{\text{Frauen}} = 3,13$; $p<0,001$).

Unsere Bemühungen gehen dahin, wahre Effekte, die von der Routine auf das Antwortverhalten ausgehen, zu ermitteln. Damit sind vor allem multivariate Auswertungsverfahren erforderlich, um einen möglichen Einfluss weiterer Drittvariablen kontrollieren zu können. Die folgenden Analysen beziehen sich ausschließlich auf die Befragten ($n=1.162$), die in der CELLA 1 Studie über ein Mobiltelefon kontaktiert und befragt wurden.

Kontrollvariablen

Personen können am Mobilfunk in verschiedenen sozialen Situationen und Umgebungen erreicht werden. Diese Umgebungen können sich in einzelnen Aspekten (z. B. Grad der Privatheit) erheblich unterscheiden. In der Befragung wurde deshalb auch erhoben, wo sich die befragten Personen zum Zeitpunkt des Interviews befanden. Der größte Teil der Befragten (63,2 %) wurde auch über den Mobilfunk zu Hause befragt.⁴

Bei der Auswertung wird der Ort des Interviews dichotomisiert verarbeitet. Der Wert 0 besagt, dass die Befragung zu Hause, und der Wert 1, dass die Befragung an einem anderen Ort stattfand.

Aufgrund der Heterogenität der Orte, an denen die Befragten sich befanden, wurde auch erhoben, ob zum Zeitpunkt des Interviews Dritte anwesend waren. Ist der Befragte zum Zeitpunkt des Interviews nicht allein, vergrößert sich das Publikum in der Interviewsituation. Die Wirkung der Anwesenheit Dritter auf den Antwortprozess ist allerdings nicht eindeutig. Bei günstigen Bedingungen – hierzu zählt beispielsweise auch eine hohe Routine im Umgang mit dem betreffenden Kommunikationsmedium – können damit die Anstrengungen, eine sozial erwünschte Antwort zu geben, intensiver ausfallen als in einem ungünstigen Setting.

Für die Auswertung wird die Anwesenheit dritter Personen während des Interviews ebenfalls dichotomisiert erhoben. Hier besagt der Wert 0, dass keine dritte Person während des Interviews anwesend war. Der Wert 1 bringt zum Ausdruck, dass sich zumindest zeitweise ein Dritter während des Interviews in unmittelbarer Nähe aufhielt.

Um potentielle Mediatoreffekte der Routine auf die Anwesenheit Dritter und auf den Ort des Interviews zu kontrollieren, wurden in den multivariaten Analysen diese Interaktionsterme mit einbezogen. Zusätzlich fanden die soziodemografischen Variablen Alter (in Jahren) Bildung (in Jahren) und das Geschlecht Eingang in die Modelle.

4 Aufgrund der Heterogenität der anderen Antworten wird diese Variable dichotom (zu Hause ja/nein) in den Analysen berücksichtigt.

4 Indikatoren der Antwortqualität

In der Umfrageforschung lassen sich vier zentrale Dimensionen zur Beurteilung der Qualität von Umfragen aufzeigen. Dabei handelt es sich – unter Verwendung der inzwischen gängigen englischen Begriffe – um: Sampling, Coverage, Response und Measurement (de Leeuw et al. 2008: 2f.). In diesem Aufsatz stehen das Antwortverhalten und damit insbesondere der Bereich der Messfehler im Vordergrund.

Zwei Gruppen von Indikatoren wurden für die Ermittlung der Qualität des Antwortverhaltens bzw. zur Aufdeckung von Messfehlern verwendet. Erstens globale Indikatoren, die die Qualität verschiedener Schritte der Antwortfindung (vgl. Abschnitt 2) zugleich abbilden, und zweitens Indikatoren, die die Qualität eines bestimmten Schrittes der Antwortfindung wiedergeben.⁵ Zur ersten Gruppe von Indikatoren zählen:

- **Pseudo-Opinions:** Eine Reaktion auf fiktive Fragegegenstände könnte sowohl auf ein mangelndes Verständnis der Frage als auch auf einen mangelhaften Informationsabruf oder auf sozial erwünschtes Antwortverhalten zurückgeführt werden (Kühne/Böhme 2006). Als eine Folge geringer Routine im Umgang mit der Kommunikationstechnik lässt sich deshalb ein stärkeres Auftreten von Pseudo-Opinions vermuten.
- **Item-Nonresponses:** Ursachen für die Nichtbeantwortung einzelner Fragen können in allen vier Phasen des kognitionspsychologischen Antwortmodells vermutet werden. Mangelnde Routine sollte insbesondere zu einer höheren Wahrscheinlichkeit von Satisficing-Strategien führen, die wiederum Item-Nonresponse zur Folge haben können (Krosnick 2000). Substanzielle Antworten werden im umgekehrten Fall aufgrund eines höheren kognitiven Einsatzes gegeben, sodass seltener Satisficing-Strategien genutzt werden (Shoemaker/Eichholz/Skewes 2002: 196).
- **Antwortstabilität:** Die Ursache für instabile individuelle Einstellungen werden in der Umfrageforschung kontrovers diskutiert (u. a. Converse 1964, 1970; Kriesi 2004: 242; Zaller 1992; Alvarez/Brehm 1995, 2002). Allerdings lässt sich vermuten, dass instabile Einstellungen auf Ablenkungen im Antwortprozess zurückführbar sind (Kühne/Böhme 2006: 49f.). Insofern sollte eine geringere Routine zu einem Anstieg instabiler Antworten führen.

Zur zweiten Gruppe von Indikatoren zählen zunächst solche, die ganz konkret auf den Informationsabruf für die Urteilsbildung zielen:

5 In CELLA 1 wurden zahlreiche weitere Indikatoren untersucht, die an dieser Stelle nicht präsentiert werden können. Unter anderem wurden der Zeitpunkt des Abbruchs und die Schwierigkeit der Beantwortung der Fragestellung analysiert (Häder/Häder 2009).

- **Erinnerungsleistung:** Die Erinnerungsleistung kann als Indiz dafür gelten, wann die Informationssuche – der Schritt zwei – für die Antwortfindung abgebrochen wurde bzw. mit welcher Intensität diese überhaupt geführt wurde (Häder/Kühne 2009b: 225ff.). Fehlende Routine im Umgang mit dem Kommunikationskanal kann bei den Befragten zu geringeren Erinnerungsleistungen führen. Es ist bekannt, dass für die Findung einer Antwort nicht alle zur Verfügung stehenden Informationen aus dem Gedächtnis abgerufen werden, sondern dass dieser Prozess abgebrochen wird, sobald die Zielperson meint, genügend erinnert zu haben. Dieser Abbruch erfolgt vermutlich um so eher, je geringer der Grad an Routine ist.
- **Reihenfolgeeffekte:** Treten in einer Befragung Reihenfolgeeffekte auf, kann das unter anderem auf Unterschiede bei dem Abruf von Informationen für die Antwortgenerierung hindeuten (Häder/Kühne 2009c: 207ff.). Bei fehlender Routine sollten systematische Effekte durch die Änderung der Fragereihenfolge stärker auftreten. Bei einem höheren Grad an Routine können sich die Befragten verstärkt auf ihre Antwort konzentrieren und sind gegenüber den genannten Einflüssen resistenter.

Die CELLA 1 Studie bietet damit zahlreiche Möglichkeiten, um den Einfluss der Routine auf die Qualität des Antwortverhaltens analog zu den formulierten Vermutungen empirisch zu überprüfen. Dies macht einen besonderen Wert der Studie aus.

5 Ergebnisse

5.1 Globale Indikatoren der Antwortqualität

Pseudo-Opinions

Pseudo-Opinions stellen einen beliebten Indikator der Antwortqualität dar (Bishop/Tuchfarber/Oldendick 1986; Reuband 2000; Mayerl/Urban 2008; Krosnick et al. 2002; Kühne/Böhme 2006). CELLA 1 enthält eine Frage zu einem Sachverhalt (konkret: die Bewertung eines Vorschlags des Europäischen Parlaments zum Ausbau eines gemeinsamen, europaweiten Mobilfunknetzes, tp3), den es nicht gegeben hat. Eine positive Antwort auf eine solche Frage deutet auf eine geringere Antwortqualität hin.

Überraschend ist zunächst, dass die Befragten, die angaben, diesen Vorschlag zu kennen ($n=143$), über eine etwas höhere Routine verfügen als jene ($n=772$), die dies verneinen ($M_{\text{bekannt}} = 0,35 > M_{\text{nicht bekannt}} = -0,04$; $p < 0,05$; gewichteter Datensatz).

Diejenigen Befragten, die auf das Pseudo-Opinion reagieren, haben also eine etwas größere Routine im Umgang mit dem Handy. Weiteren Aufschluss kann das Regressionsmodell liefern (vgl. Tabelle 1). In dieses Modell werden auch zwei Interaktionsterme aufgenommen. Der eine betrifft die Routine und die Anwesenheit Dritter (*Routine x Anwesenheit Dritter*), der andere die Routine und den Ort (*Routine x Ort des Interviews*), an dem das Interview stattfindet. Damit soll ein moderierender Einfluss der Routine kontrolliert werden. Es ist zu erkennen, dass auch unter Kontrolle der einbezogenen Größen der Einfluss der Routine auf das Antwortverhalten erhalten bleibt, während von den Interaktionstermen kein signifikanter Effekt ausgeht.

Tabelle 1 Ergebnisse einer binär logistischen Regression zur Erklärung des Antwortverhaltens bei einer Pseudo-Opinionfrage, Effektkoeffizienten

Konstante	8,50***
Routine	0,79**
Anwesenheit Dritter (Ref. Kat. = niemand anwesend)	1,14
Ort des Interviews (Ref. Kat. = zu Hause)	1,06
Alter	0,99
Geschlecht (Ref. Kat. = männlich)	1,48
Bildung	0,95
Routine x Anwesenheit Dritter	1,10
Routine x Ort des Interviews	1,21

$n = 949$, $Pseudo-R^2 = 0,02$, * = 0,05, ** = 0,01, *** = 0,001;

Abhängige Variable: Pseudo-Opinion, Ref. Kat. = Reaktion auf Pseudo-Opinion

Item-Nonresponse

Die Stärke des Item-Nonresponse kann ebenfalls als Qualitätskriterium für das Antwortverhalten angesehen werden. Zu vermuten ist, dass bei größerer Routine für die Zielpersonen die Antwortfindung weniger kognitiv anspruchsvoll ist und ihnen der Abruf von Informationen aus dem Gedächtnis dadurch erleichtert wird. Weiter ist naheliegend, dass bei der Antwortfindung weniger Heuristiken eingesetzt werden und dass es damit zu weniger Item-Nonresponse kommt.

Für die Ermittlung des Ausmaßes an Item-Nonresponse sind insgesamt 50 Variablen der Studie CELLA 1 herangezogen worden, darunter beispielsweise auch

die Einkommensfrage.⁶ Bereits die bivariaten Analysen zwischen der Häufigkeit des Nonresponse und dem Routineindex verweisen auf einen fehlenden Zusammenhang. Der Korrelationskoeffizient nach Pearson ist minimal und nicht signifikant ($r=-0,013$; $p=0,67$; $n=1.123$, gewichteter Datensatz). Damit tritt der erwartete Zusammenhang nicht auf.

Eine detaillierte Aussage soll dabei wieder ein Regressionsmodell liefern (vgl. Tabelle 2). Ganz offenbar stützen die multivariaten Analysen die bivariaten Ergebnisse. Von den einfachen Kontrollvariablen hat einzig das Geschlecht einen Einfluss auf den Item-Nonresponse. Männer verweigern häufiger bei einzelnen Fragen die Antwort.

Tabelle 2 Ergebnis einer linearen Regression zur Erklärung der Häufigkeit des Item-Nonresponse, unstandardisierte und [standardisierte] Regressionskoeffizienten

Konstante	1,57***
Routine	-0,06 [-0,09]
Anwesenheit Dritter (Ref. Kat. = niemand anwesend)	0,03
Ort des Interviews (Ref. Kat. = zu Hause)	-0,08
Alter	0,01 [0,07]
Geschlecht (Ref. Kat. = männlich)	-0,40**
Bildung	0,02 [0,05]
Routine x Anwesenheit Dritter	0,10 [0,10]
Routine x Ort des Interviews	-0,13 [-0,12]

$n = 368$, $R^2 = 0,06$, * $=0,05$, ** $=0,01$, *** $=0,001$;

Abhängige Variable: Anzahl des Item-Nonresponse von 0 bis 14

Antwortstabilität

Wird innerhalb einer Befragung die Bewertung eines Sachverhalts wiederholt, so sollten dabei möglichst identische Befunde ermittelt werden. Ist dies nicht der Fall, so liegt der Schluss nahe, dass von den Zielpersonen lediglich Heuristiken bei der Antwortfindung benutzt werden und diese damit instabil sind. Die Antwort ist dann mehr oder weniger zufällig zustande gekommen. Dies liegt nahe, da bei einer hohen Qualität des Antwortverhaltens ein einmal gebildetes Urteil wiederholt

6 Es handelt sich konkret um die Folgenden: v2, v13, v8, v9, v10, v11, v13, f2a, f2b, tp1, tp2, tp3, tp4, v15_1, v15_2, v15_3, v15_4, v15_5, v17, pt5, v17a, v18_1, v18_2, v18_3, v18_4, v18_5, v18_6, v18_7, v18_8, v18_9, v18-1014, v19a, s1, s1_1, f1, f3_4, f3_5, f3_6, f3_7, f3_8, f3_9, s3, t4, s4, s4_1, s5, s5_2, s7, v22 und v28.

werden sollte. Werden die entsprechenden Fragen jedoch relativ willkürlich beantwortet, so liegt eine nur geringe Qualität des Antwortverhaltens vor.

In der CELLA 1 Studie waren zwei Fragen enthalten, die auf die Bewertung eines nahezu identischen Sachverhalts zielen. Relativ zu Beginn wurde gefragt (tp1), ob der Ausbau der Mobilfunknetze die Lebensqualität eher positiv oder eher negativ beeinflussen wird. Gegen Ende der Befragung wurde Auskunft darüber eingeholt (tp5), ob nach Meinung der Befragten beim Ausbau der Mobilfunknetze die Vorteile oder die Nachteile überwiegen (vgl. dazu auch Kühne/Häder 2009: 217ff.). Damit kann nun untersucht werden, ob eine höhere Routine im Umgang mit dem Handy zu stabileren Antworten führt.⁷ Das Ergebnis kann aus Tabelle 3 entnommen werden. Die Routine im Umgang mit dem Handy sowie auch die Interaktionsterme haben keinen Einfluss auf diesen Aspekt der Qualität des Antwortverhaltens. Die Kontrollvariablen Geschlecht und Bildung besitzen einen Einfluss. Frauen und Personen mit niedriger Bildung neigen demnach eher zu instabilem Antwortverhalten.

Tabelle 3 Ergebnisse einer binär logistischen Regression zur Erklärung der Antwortstabilität, Effektkoeffizienten

Konstante	3,78**
Routine	0,98
Anwesenheit Dritter (Ref. Kat. = niemand anwesend)	0,86
Ort des Interviews (Ref. Kat. = zu Hause)	1,08
Alter	1,00
Geschlecht (Ref. Kat. = männlich)	0,67*
Bildung	0,95*
Routine x Anwesenheit Dritter	1,03
Routine x Ort des Interviews	1,10

n = 891, *Pseudo-R*² = 0,02, * = 0,05, ** = 0,01, *** = 0,001;

Abhängige Variable: Antwortstabilität, Ref. Kat. = keine Antwortstabilität

7 Untersuchungen zur Stabilität von Einstellungen beruhen zum überwiegenden Teil auf Längsschnittdaten (Converse 1964; Schuman/Presser 1981; Craig/Kane/Martinez 2000). Es konnte aber bereits gezeigt werden, dass Antwortstabilität auch in Querschnittsdesigns gemessen werden kann (Slaby 1998; Kühne/Böhme 2006).

5.2 Indikatoren zum Informationsabruf und zum Editieren der Antworten in Richtung der sozialen Erwünschtheit

Erinnerungsleistung

Um die Erinnerungsleistungen der Zielpersonen zu ermitteln, wurde ein an die Methode des Paraphrasing (vgl. Driscoll 2007) angelehntes Vorgehen gewählt. Das Frageprogramm der CELLA 1 Studie enthält einen entsprechenden Indikator (v18_10). Nach der Abfrage einer Reihe von Sachverhalten – insgesamt wurden neun Einzelfragen gestellt – bestand die Aufgabe für die Zielperson darin, möglichst viele der zuvor von ihr bewerteten Dinge aus dem Gedächtnis zu wiederholen. Die Anzahl der erinnerten Sachverhalte gibt eine Auskunft über die Intensität der Informationssuche und erfüllt damit eine Indikatorfunktion für die Antwortqualität bei den übrigen Fragen.

Die Korrelation der Anzahl der erinnerten Gegenstände mit dem Routineindex zeigt keinen Zusammenhang ($r=-0,01$, $p=0,72$; $n=1.132$, gewichteter Datensatz). Das Ergebnis der Schätzung einer linearen Regression zur Erklärung der Anzahl von erinnerten Gegenständen aus den vorangegangenen neun Fragen zeigt Tabelle 4. Die Bildung und das Alter stellen sich, wie zu erwarten war, als Bestimmungsgrößen für die Erinnerungsleistung heraus. Auch vom Geschlecht geht ein signifikanter Effekt aus. Die Routine im Umgang mit dem Mobiltelefon sowie die beiden Interaktionsterme haben dagegen ein weiteres Mal keinen Einfluss.

Tabelle 4 Ergebnis einer linearen Regression zur Erklärung der Erinnerungsleistung, unstandardisierte und [standardisierte] Regressionskoeffizienten

Konstante	3,54***
Routine	-0,02 [-0,02]
Anwesenheit Dritter (Ref. Kat. = niemand anwesend)	0,04
Ort des Interviews (Ref. Kat. = zu Hause)	-0,20
Alter	-0,02 [-0,19]***
Geschlecht (Ref. Kat. = männlich)	0,33**
Bildung	0,09 [0,16]***
Routine x Anwesenheit Dritter	-0,06 [-0,04]
Routine x Ort des Interviews	0,02 [0,01]

$n = 1.084$, $R^2 = 0,06$, * = 0,05, ** = 0,01, *** = 0,001;

Abhängige Variable: Anzahl des erinnerten Items von 0 bis 9

Fragereihenfolge

Weitere deutliche Hinweise auf die Qualität des Antwortverhaltens liefert die Analyse von Fragereihenfolgeeffekten. Dabei liegt es nahe, dass mit steigender Größe dieses Einflusses die Qualität der Antworten sinkt – oder mit anderen Worten: Je mehr sich die Befragten bei ihren Antworten von der Reihenfolge der Fragen leiten lassen, desto weniger werden eigene Informationen aus dem Gedächtnis für die Antwortfindung aktiviert, desto weiter entfernt sind diese Angaben vom wahren Wert usw.

In unserem Zusammenhang ist zu analysieren, ob Menschen mit einer größeren Routine im Umgang mit dem Mobiltelefon resistenter gegenüber solchen Fragereihenfolgeeffekten sind. Schließlich wird angenommen, dass sich jene Befragten aufgrund ihrer Sicherheit im Umgang mit dem Medium Handy kognitiv stärker auf den Informationsabruf konzentrieren können als andere Personen.

In der Studie CELLA 1 war ein Split ballot enthalten: Einem zufällig rekrutierten Teil der Befragten wurde eine Frage zum allgemeinen Lebensglück (v17)⁸ vor und dem anderen Teil nach einer speziellen Indikatorenbattery zur Verzichtbarkeit bzw. Unverzichtbarkeit von materiellen Dingen (v18)⁹ gestellt. Mithilfe dieses Designs ist es gelungen, einen Reihenfolgeeffekt zu erzeugen. So kann im Weiteren festgestellt werden, ob die Stärke dieses Effekts im Zusammenhang steht mit der Routine, die die Zielpersonen im Umgang mit dem Mobilfunkgerät besitzen. Tabelle 5 zeigt das Ergebnis der Schätzung eines entsprechenden linearen Regressionsmodells.

Der Befund ist sehr aufschlussreich: So kann ein deutlicher Reihenfolgeeffekt ausgemacht werden (vgl. die Variable *Split Reihenfolge* in Tabelle 5). Auch die anderen zur Kontrolle in das Modell aufgenommenen Größen besitzen – wie zu erwarten war – teilweise eine starke erklärende Wirkung. Ohne Einfluss bleibt dagegen der Index zur Abbildung der Routine im Umgang mit dem Handy. Beachtung verdienen auch die drei in dieses Modell aufgenommenen Interaktionsvariablen. Neben den bereits bekannten wurde zusätzlich ein Term integriert (Variable *Routine x Split Reihenfolge*), der die Wirkung von Routine sowie den Split berücksichtigt. Der an dieser Stelle wiederum fehlende Effekt belegt, dass die Routine ein weiteres Mal ohne Einfluss auf die Antwortqualität bleibt.

8 "Wenn Sie einmal Ihr Leben betrachten, was würden Sie sagen, wie glücklich oder unglücklich sind Sie alles in allem?"

9 "Nun einmal zu einem anderen Thema. Es gibt verschiedene Meinungen darüber, welche Dinge man in Deutschland zum Leben braucht. Ich lese Ihnen im Folgenden einige Dinge vor und Sie sagen mir jeweils, was für Sie verzichtbar und was unverzichtbar ist. Markenartikel kaufen zu können, halten Sie für verzichtbar oder unverzichtbar. Ein neues Auto kaufen zu können, halten Sie für verzichtbar oder unverzichtbar. Über ein neues Handy zu verfügen, halten Sie für verzichtbar oder unverzichtbar", usw.

Tabelle 5 Ergebnis einer linearen Regression zur Erklärung des Antwortverhaltens bei der Frage nach dem Lebensglück, unstandardisierte und [standardisierte] Regressionskoeffizienten

Konstante	1,95***
Routine	-0,04 [-0,10]
Anwesenheit Dritter (Ref. Kat. = niemand anwesend)	-0,13 [-0,08]*
Ort des Interviews (Ref. Kat. = zu Hause)	-0,04
Alter	0,01 [0,10]**
Geschlecht (Ref. Kat. = männlich)	-0,01
Bildung	-0,02 [-0,07]*
Routine x Anwesenheit Dritter	0,01 [0,01]
Routine x Ort des Interviews	0,02 [0,02]
Split Reihenfolge	0,10 [0,06]*
Routine x Split Reihenfolge	0,01 [0,04]

$n = 1.051$, $R^2 = 0,03$, * $=0,05$, ** $=0,01$, *** $=0,001$;

Abhängige Variable: Einschätzung des Lebensglücks 1 = voll und ganz glücklich bis 5 = überhaupt nicht glücklich

Frageformulierung

Einen guten Hinweis auf die Qualität des Antwortverhaltens können auch Frageformulierungseffekte liefern. Sicher ist dabei: Je stärker von unterschiedlichen Frageformulierungen die Effekte auf das Antwortverhalten – bei identischem Inhalt der Frage – ausfallen, desto geringer wird die Antwortqualität. In der CELLA 1 Studie wurde ein in der Umfrageforschung seit längerem bekannter Test (vgl. Rugg 1941) modifiziert. Reuband (2003: 87ff.) verweist bereits auf eine ganze Reihe an Replikationen dieses Experiments (vgl. Schuman/Presser 1981; Hippler/Schwarz 1986; Waterplas et al. 1988; Glendall/Hoek 1990; Loosveldt 1997 und Holleman 2000). Die Zielpersonen sollten bei CELLA 1 angeben (f2), ob Handy-Sendemasten auf Schulgebäuden „erlaubt“ oder „verboten“ (Splitvariante a) bzw. „erlaubt“ oder „nicht erlaubt“ (Splitvariante b) sein sollten.

Wichtig ist zu hinterfragen, ob der Grad an Routine nun Effekte, die sich aus der Formulierung der Antwortvorgaben ergeben, überdeckt, oder ob diese unabhängig davon auftreten. Um dazu eine Aussage liefern zu können, wurde eine logistische Regression berechnet (vgl. Tabelle 6).

Der Befund ist abermals durchaus bemerkenswert. So geht von der den Zielpersonen angebotenen Antwortvorgabe („verboten“ bzw. „nicht erlaubt“) tatsächlich ein deutlicher Effekt auf deren Antwortverhalten aus (vgl. die Variable *Split Formulierung* in Tabelle 6). Das Alter sowie die absolvierten Bildungsjahre besitzen hier einen Einfluss. Ohne klaren Effekt bleibt jedoch wiederum die Routine, die die Zielpersonen im Umgang mit ihrem Mobilfunkgerät haben.

Auch hier wurden wieder die bekannten Interaktionsterme in das Modell aufgenommen. Alle drei bleiben allerdings ohne Einfluss. Der Einfluss sowie die Erklärungskraft des Modells, wie bereits von allen vorangegangenen, sind äußerst gering.

Tabelle 6 Ergebnisse einer binär logistischen Regression zur Erklärung des Antwortverhaltens bei Unterschieden in der Frageformulierung, Effektkoeffizienten

Konstante	3,78**
Routine	1,15
Anwesenheit Dritter (Ref. Kat. = niemand anwesend)	0,98
Ort des Interviews (Ref. Kat. = zu Hause)	1,02
Alter	1,03***
Geschlecht (Ref. Kat. = männlich)	1,29
Bildung	0,94*
Routine x Anwesenheit Dritter	1,05
Routine x Ort des Interviews	1,15
Split Frageformulierung	1,79***
Routine x Split Frageformulierung	0,85

n = 774, Pseudo- R^2 = 0,05, * = 0,05, ** = 0,01, *** = 0,001;

Abhängige Variable: Frageformulierung, Ref. Kat. = Zustimmung

6 Zusammenfassung

Nachdem insgesamt sechs verschiedenen Kriterien zur Antwortqualität nachgegangen wurde, kann ein recht differenziertes Urteil zum Einfluss der Routine im Handygebrauch auf die Antwortqualität bei einer Befragung über das Mobilfunknetz gefällt werden. Dazu ist zu sagen:

Die Botschaft für die Mobilfunkbefragungen ist zunächst eine sehr positive. So dürfte die unterschiedliche Routine, die im Umgang mit dieser noch relativ neuen Technik in der Bevölkerung besteht, nicht zu deutlich verzerrten Umfrageergebnissen führen. Offenbar stellt die Mobilfunktechnik nicht so starke technische Ansprüche an den Benutzer, dass dadurch dessen Telefonverhalten bei Umfragen negativ beeinflusst würde. Dies steht im Kontrast zu über das Internet vermittelten Umfragestudien, wo solche Effekte aufgetreten sind. Das Internet ist (noch) schwieriger handhabbar als das Mobiltelefon. Für diesen Befund sprechen auch die insgesamt erfreulich geringen Werte für die erklärte Varianz (R^2 bzw. Nagelkerkes R^2). Dies bedeutet, dass vor allem modellexterne Variablen für das Antwortverhalten verantwortlich gemacht werden können.

Ein weiterer, in diesem Zusammenhang zukünftig stärker zu beachtender Aspekt dürfte die Qualität des Fragebogens sein. Der Erhebungsstandard der CELLA 1 Untersuchung enthält durchschnittlich komplizierte Fragestellungen. Der Mittelwert bei einer Frage zur Einschätzung der von den Zielpersonen empfundenen Schwierigkeit (v28) liegt auf einer fünfstufigen Skala bei 1,6 (alle 2.171 Befragten). Das heißt, die Erhebung wird zwischen „überhaupt nicht“ (Skalenwert 1) und „kaum schwierig“ (Skalenwert 2) eingeschätzt. Wie sich die Situation dagegen bei schwieriger zu beantwortenden Fragen oder bei längeren Umfragen verhält, muss an dieser Stelle offen bleiben.

Alle telefonischen Befragungen – zumal solche über den Mobilfunk – stellen einen Eingriff in die Privatsphäre dar. Nicht ausgeschlossen werden konnte deshalb, dass es gerade im Mobilfunkmodus, beispielsweise aufgrund von diversen Umgebungseinflüssen, zu einer starken Ablenkung der Zielpersonen kommen kann. Dies könnte zu stärkeren kognitiven Anstrengungen bei einer Befragung führen und sich damit negativ auf die Qualität des Antwortverhaltens auswirken. Die Befunde stützen eine solche Vermutung jedoch nicht. Bedenkt man, dass die Zielpersonen bei der Befragung über den Mobilfunkmodus zumeist auch zu Hause (vgl. die Ergebnisse von Frage v21 sowie Häder et al. 2009: 77ff.) erreicht worden sind, so ist dieser Befund durchaus sehr plausibel.

Die für die Suche nach möglichen Zusammenhängen zwischen der Routine und dem Antwortverhalten bei CELLA 1 eingesetzten vielfältigen methodischen Instrumente haben sich als ausreichend sensibel erwiesen. So zeigen insbesondere die Splitvarianten zu den Fragereihenfolgeeffekten sowie zum Effekt der Frageformulierung genau die erwarteten Muster. Auch die Fragestellungen zur Erinnerungsleistung und zum Pseudo-Opinion haben sich als besonders aussagestarke Indikatoren für die Antwortqualität erwiesen (vgl. die entsprechenden Tabellen). Der in der Regel nicht erfolgte Nachweis eines Effekts der Routine auf die Qualität des Antwortverhaltens deutet damit höchst wahrscheinlich nicht auf einen methodischen Fehler bzw. auf einen Zufall hin, sondern spiegelt die Realität richtig wider.

Ein weiteres Argument zur Erklärung der fehlenden Effekte könnte sein, dass in die Stichprobe vor allem Personen mit einer relativ hohen Routine gelangt sind, beispielsweise, weil andere ihre Teilnahme verweigert haben. Dass dies jedoch nicht so ist, konnte in Abbildung 1 gezeigt werden.

Welche Schlussfolgerungen können nun aus den Befunden von CELLA 1 für die Diskussion um die Qualität der Antwortfindung bei Umfragen allgemein gezogen werden? Erstens soll nochmals unterstrichen werden, dass bei nahezu allen Tests keine Effekte der Routine auf das Antwortverhalten ermittelt werden konnten.

Für eine ganze Reihe Annahmen gab es keine empirische Unterstützung anhand der CELLA 1 Daten. Damit muss davon ausgegangen werden, dass die Routine im Umgang mit dem Kommunikationskanal, über den die Befragung geführt wird, insgesamt nur einen relativ schwachen Einfluss auf die Antwortqualität hat. So spielen zum Beispiel der Bildungsstand der Zielperson, deren Alter, die Reihenfolge, in der die Fragen gestellt werden, gegebenenfalls das Geschlecht der Zielperson sowie die konkrete Formulierung der Antwortvorgaben eine stärkere Rolle als die Routine im Umgang mit dem Handy. Dies könnte Anlass dazu geben, insgesamt die Rolle der Routine im Umgang mit dem Vermittlungskanal einer Befragung – hier mit der Mobilfunktechnik – unter den Determinanten der Antwortqualität zu relativieren.

Die empirischen Ergebnisse geben Anlass zu der Vermutung, dass die Routine im Umgang mit dem Mobiltelefon nur dann Bedeutung für die Qualität des Antwortverhaltens erlangt, wenn technische Probleme mit dem Gerät auftreten. Da dies in der Regel offenbar nicht der Fall ist, laufen bei einer Mobilfunkbefragung ähnliche Prozesse ab, wie bei einer Erhebung über das Festnetz. Weder die Motivation, sich kognitiv mit der jeweiligen Fragestellung auseinander zu setzen, noch die Ablenkung infolge von Umgebungseinflüssen während der Befragung unterscheiden sich in Abhängigkeit von der Routine.

Jedoch sind Replikationsstudien erforderlich, um beispielsweise auch den Einfluss verschiedener Befragungsthemen als Bestimmungsgrößen für die Antwortqualität zu ermitteln. So ist nicht auszuschließen, dass bei kognitiv anspruchsvolleren Fragestellungen auch die Routine als Bestimmungsgröße für das Antwortverhalten wieder etwas an Einfluss gewinnt.

Literatur

- Alvarez, R. M. und J. Brehm, 1995: American Ambivalence Towards Abortion Policy: Development of a Heteroskedastic Probit Model of Competing Values. *American Journal of Political Science* 39: 1055-1082.
- Alvarez, R. M. und J. Brehm, 2002: *Hard Choices, Easy Answers: Values, Information, and American Public Opinion*. Princeton: Princeton University Press.
- Bishop, G. F., A. J. Tuchfarber und R. W. Oldendick, 1986: Opinions on Fictitious Issues: The Pressure to Answer Survey Questions. *Public Opinion Quarterly* 50: 240.
- Bishop, G. F., 1987: Experiments with the Middle Response Alternative in Survey Questions. *Public Opinion Quarterly* 51 (2): 220-232.
- Blumberg, S. J., und J. V. Luke, 2008: *Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, January - June 2008*. Atlanta: CDC. <http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless200812.pdf> (04.07.2010).
- Böcker, F., 1988: Scale Forms and their Impact on Rating's Reliability and Validity. *Journal of Business Research* 17 (1): 15-26.

- Christian, L. M., D. A. Dillman und J. D. Smyth, 2008: The Effects of Mode and Format on Answering to Scalar Questions in Telephone and Web Surveys. S. 250-274 in: J. M. Lepkowski, C. Tucker, J. M. Brick, E. de Leeuw, L. Japac, P. J. Lavrakas, M. W. Link und R. L. Sangster (Hg.): *Advances in Telephone Survey Methodology*. New York: Wiley.
- Converse, P. E., 1964: The Nature of Belief Systems. S. 206-261 in: D. E. Apter (Hg.): *Ideology and Discontent*. New York: Free Press.
- Converse, P. E., 1970: Attitudes and Non-attitudes: Continuation of a Dialogue. S. 186-189 in: E. R. Tuftte (Hg.): *The Quantitative Analysis of Social Problems*. Reading: Addison-Wesley.
- Craig, S. C., J. G. Kane und M. D. Martinez, 2000: Ambivalence, Attitude Strength, and Response Instability: A Two-Wave Panel Study of Abortion Attitudes in Florida. Washington, DC: American Political Science Association Meetings.
- Dillman, D. A., 2000: *Mail and Internet Surveys. The Tailored Design Method*. New York: Wiley.
- Dillman, D. A., T. L. Brown, J. Carlson, E.H. Carpenter, F. O. Lorenz, M. Robert, S. John und R. L. Sangster, 1995: Effects of Category Order on Answers to Mail and Telephone Surveys. *Rural Sociology* 60 (4): 674-687.
- Driscoll, D. L.; 2007: *Paraphrase: Write it in Your Own Words*. Purdue University Online Writing Lab.
- Faulbaum, F., P. Prüfer und M. Rexroth, 2009: *Was ist eine gute Frage? Die systematische Evaluation der Fragenqualität*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gabler, S. und S. Häder, 2009: Gewichtung für die CELLA Studie. S. 51-55 in: M. Häder und S. Häder (Hg.): *Telefonbefragungen über das Mobilfunknetz. Konzept, Design und Umsetzung einer Strategie zur Datenerhebung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Glendall, Ph. und J. Hoek, 1990: A Question of Wording. *Marketing Bulletin* 1: 25-36.
- Graeske, J. und T. Kunz, 2009: Stichprobenqualität der CELLA-Studie unter besonderer Berücksichtigung der Mobile-onlys. S. 57-70 in: M. Häder und S. Häder (Hg.): *Telefonbefragungen über das Mobilfunknetz. Konzept, Design und Umsetzung einer Strategie zur Datenerhebung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Häder, M. und S. Häder (Hg.), 2009: *Telefonbefragungen über das Mobilfunknetz. Konzept, Design und Umsetzung einer Strategie zur Datenerhebung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Häder, S., S. Gabler und C. Heckel, 2009: Stichprobenziehung für die CELLA Studie. S. 21-49 in: M. Häder und S. Häder (Hg.): *Telefonbefragungen über das Mobilfunknetz. Konzept, Design und Umsetzung einer Strategie zur Datenerhebung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Häder, S., M. Häder, J. Graeske, T. Kunz und G. Schneiderat, 2009: Realisierung der Stichprobe. S. 71-82 in: M. Häder und S. Häder (Hg.): *Telefonbefragungen über das Mobilfunknetz. Konzept, Design und Umsetzung einer Strategie zur Datenerhebung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Häder, M. und M. Kühne, 2009a: Theoretischer Rahmen und Untersuchungsdesign. S. 165-174 in: M. Häder und S. Häder (Hg.): *Telefonumfragen über das Mobilfunknetz. Konzept, Design und Umsetzung einer Strategie zur Datenerhebung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Häder, M. und M. Kühne, 2009b: Erinnerungsleistung der Befragten in beiden Modes. S. 225-228 in: M. Häder und S. Häder (Hg.): *Telefonbefragungen über das Mobilfunknetz. Konzept, Design und Umsetzung einer Strategie zur Datenerhebung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Häder, M. und M. Kühne, 2009c: Fragereihenfolgeeffekte. S. 207-216 in: M. Häder und S. Häder (Hg.): *Telefonbefragungen über das Mobilfunknetz. Konzept, Design und Umsetzung einer Strategie zur Datenerhebung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hippler, H. J. und N. Schwarz, 1986: Not Forbidding isn't Allowing: The Cognitive Basis of the Forbid – Allow Asymmetry. *Public Opinion Quarterly* 50 (1): 87-96.

- Holleman, B., 2000: The Forbid/Allow Asymmetrie. On the Cognitive Mechanism Underlying Wording Effects in Surveys. Amsterdam Atlanta: Radopi.
- Hyman, H. H. und P. B. Sheatsley, 1950: The Current Status of American Public Opinion. S. 11-34 in: Payne, J. C. (Hg.): The Teaching of Contemporary Affairs. Twenty-first Yearbook of the National Council of Social Studies. New York: National Education Association.
- Jobe, J. B. und D. J. Herrmann, 1996: Implications of Models of Survey Cognition for Memory Theory. S. 193-205 in: D. J. Herrmann, M. Johnson, Ch. Herzog und P. Hertel (Hg.): Basic and Applied Memory Research. Bd. 2. Practical Applications. Hillsdale, NJ: Erlbaum.
- Kriesi, H., 2004: Stability and Change of Opinion: The Case of Swiss Policy Against Pollution Caused by Cars. S. 242-267 in: W. E. Saris und P. M. Sniderman (Hg.); Studies in Public Opinion: Gauging Attitudes, Nonattitudes, Measurement Error and Change. Princeton: Princeton University Press.
- Krosnick, J. A. und D. F. Alwin, 1987: An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement. *Public Opinion Quarterly* 51 (2): 201-219.
- Krosnick, J. A., 2000: The Threat of Satisficing in Surveys: The Shortcuts Respondents Take in Answering Questions. *Survey Methods Newsletter* 20 (1): 4-8.
- Krosnick, J. A., A. L. Holbrook, M. K. Berent, R. T. Carson, W. M. Hanemann, R. J. Kopp, R. C. Mitchell, S. Presser, P. A. Ruud, V. K. Smith, W. R. Moody, M. C. Green und M. Conaway, 2002: The Impact of No Opinion Response Options on Data Quality: Non-Attitude Reduction Or an Invitation to Satisfice? *Public Opinion Quarterly* 66: 371-403.
- Kuusela, V., M. Callegaro und V. Vehovar, 2008: The Influence of Mobile Telephones on Telephone Surveys. S. 87-112 in: J. M. Lepkowski, C. Tucker, J. M. Brick, E. de Leeuw, L. Japac, P. J. Lavrakas, M. W. Link und L. Sangster (Hg.): *Advances in Telephone Survey Methodology*, New Jersey: Wiley.
- Kühne, M. und R. Böhme, 2006: Effekte von Informationsstand, Wissen und Einstellungsstärke von Befragten auf die Antwortstabilität. *ZUMA-Nachrichten* 59: 42-71. http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten/zn_59.pdf (22.07.2010).
- Kühne, M. und M. Häder, 2009: Konsistenzen und Inkonsistenzen im Antwortverhalten. S. 217-224 in: M. Häder und S. Häder (Hg.): *Telefonbefragungen über das Mobilfunknetz. Konzept, Design und Umsetzung einer Strategie zur Datenerhebung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- de Leeuw, E. D. und J. H. Jopp, 2008: The Cornerstones of Survey Research. S. 1-17 in: E. D. de Leeuw, J. H. Joop und D. A. Dillman (Hg.): *International Handbook of Survey Methodology*. New York: Lawrence Erlbaum Associates.
- Loosveldt, G., 1997: Interaction Characteristics in Some Question Wording Experiments. In: *Bulletin de Méthodologie Sociologique (BMS)* 56: 20-31.
- Mayerl, J. und D. Urban, 2008: *Antwortreaktionszeiten in Survey-Analysen*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Payne, S. L., 1949: Case Study in Question Complexity. *Public Opinion Quarterly* 13 (4): 653-658.
- Reuband, K. H., 2000: „Pseudo Opinions“. *ZA-Information* 46: 26-38. http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/za_information/ZA-Info-46.pdf (22.07.2010).
- Reuband, K. H., 2003: Variationen der Permissivität: Wie Frageformulierungen unterschiedliche Antwortverteilungen erbringen, wenn von 'Erlauben' oder 'Verbieten' die Rede ist. *ZA-Information* 53: 86-96. http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/za_information/ZA-Info-53.pdf (22.07.2010).
- Rugg, D. (1941): Experiments in Wording Questions. *Public Opinion Quarterly* 5 (1):91-92.
- Schuman, H., 1992: Context Effects: State of the Past / State of the Art. S. 5-20 in: N. Schwarz und S. Sudman (Hg.): *Context Effects in Social And Psychological Research*. New York: Springer.

- Schuman, H. und S. Presser, 1981: Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context. New York: Academic.
- Schwarz, N., H.-J. Hippler, B. Deutsch und F. Strack, 1985: Response Scales: Effects of Category Range on Reported Behaviour and Comparative Judgements. *Public Opinion Quarterly* 49 (3): 388-395.
- Shoemaker, P. J., M. Eichholz und E. A. Skewes, 2002: Item Nonresponse: Distinguishing Between Don't Know and Refuse. *International Journal of Public Opinion Research* 14 (4): 193-201.
- Slaby, M., 1998: Eine Untersuchung zur Konstanz des Meinungsurteils von Befragten im Interviewverlauf dargestellt am Fallbeispiel „Bewertung der Gentechnik“. Stuttgart: IfS.
- Sudman, S., N. M. Bradburn und N. Schwarz, 1996: Thinking about Answers. The Application of Cognitive Processes to Survey Methodology. San Francisco: Jossey-Bass.
- Stadler, K., 1983: Die Skalierung in der empirischen Forschung. München: Infratest.
- Strack, F., 1994: Zur Psychologie der standardisierten Befragung. Kognitive und kommunikative Prozesse. Berlin: Springer.
- Strack, F., L. L. Martin, 1988: Thinking, Judging, and Communicating: A Process Account of Context Effects in Attitude Surveys. S. 123-148 in: H.-J. Hippler (Hg.): *Social Information Processing and Survey Methodology*. 2. Auflage. New York: Springer.
- Tourangeau, R., L. J. Rips und K. A. Rasinski, 2000: *The Psychology of Survey Response*. Cambridge: University Press.
- Waterplas, L., J. Billet und G. Loosveldt, 1988: De verbieden versus niet toelaten asymmetrie. Een stabiele formuleringseffect in survey-onderzoek? *Mens en Maatschappij* 63: 399-417.
- Zaller, J. R., 1992: *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge UP.

Anschrift der Autoren

Prof. Dr. Michael Häder
Technische Universität Dresden
Philosophische Fakultät, Institut für Soziologie
Lehrstuhl für Methoden der empirischen Sozialforschung
01062 Dresden
michael.haeder@tu-dresden.de

Dr. Mike Kühne
Technische Universität Dresden
Philosophische Fakultät, Institut für Soziologie
Lehrstuhl für Methoden der empirischen Sozialforschung
01062 Dresden
mike.kuehne@tu-dresden.de

Mahnaktionen in postalischen Befragungen

*Empirische Befunde zu
Auswirkungen auf den
Rücklauf, das Antwort-
verhalten und die Stich-
probenzusammensetzung*

Follow-Ups in Mail Surveys

*Empirical Findings on the
Effects on the Response
Rate, the Responses and
the Sample Composition*

Franziska Kunz

Zusammenfassung

Empirische Befunde zur rücklaufsteigernden Wirkung von Remindern sind konsistent und zahlreich. Nur wenige Studien beschäftigen sich mit dem Einfluss von Nachfassaktionen auf das Antwortverhalten der Respondenten und die soziale Zusammensetzung der Befragtenstichprobe. Vorliegender Beitrag untersucht diese Aspekte anhand der Daten einer im Jahr 2009 im Raum Südbaden durchgeführten postalischen Befragung (N=3.555) unter 49- bis 81-Jährigen zum Thema „Kriminalität im höheren Lebensalter“. Die Analysen ergeben, dass das Ausmaß fehlender Antworten (Item-Nonresponse) mit den Mahnaktionen zunimmt, wobei der Zusammenhang zwischen der Zahl der Nachfassaktionen und der Zahl fehlender Antworten kurvilinear ist. Mahnaktionen rekrutieren v. a. Gruppierungen in die Befragungsstichprobe, die nach einmaligem Anschreiben üblicherweise unterrepräsentiert sind – nämlich Personen, die jünger sind, im ländlichen Raum wohnen sowie einen schwächeren Sozialstatus und geringeres politisches Interesse aufweisen. Mahnaktio-

Abstract

The fact that reminders increase the response rates of mail surveys is empirically documented. However, only a few studies have dealt with the effect of follow-up mailings on both response behaviour and sample composition. This article investigates these aspects using data from a regionally representative mail survey (N=3.555) on crime at advanced ages among 49 to 81 years-olds that was conducted in Südbaden/Baden-Württemberg in 2009. The analyses showed that the use of follow-ups increased the number of missing values (item-nonresponse), with a curvilinear relation between the number of follow-ups and the extent of item-nonresponse. Follow-ups primarily involve people that are usually underrepresented in the first wave, i. e., people who are younger, living in rural areas, having a lower social status and weaker political interest. Using reminders therefore contributes to an enhanced social composition of the net sample. Nevertheless, the respondents of the first mailing do not differ fundamentally from those of the

nen tragen damit zu einer Verbesserung der sozialen Zusammensetzung der Stichprobe bei. Hinsichtlich sozio-demographischer Merkmale sowie psychosozialer Befindlichkeiten unterscheiden sich die Befragten des Erstversands allerdings nicht gravierend von denen der gesamten Befragungsstichprobe. Eine Ausnahme stellt das kriminelle Handeln der Befragten über die Lebensspanne dar: mit zunehmenden Kontaktierungen werden deutlich weniger delinquente Personen in die Befragung einbezogen. Die Antworten der Respondenten zum Befragungsthema stehen demnach in deutlichem Zusammenhang zum Erhebungszeitpunkt. Nachfassaktionen können somit auch hier zur Abschwächung anfänglicher Verzerrungen beitragen. Aus methodischer Sicht sollten Mahnaktionen deshalb prinzipiell in das Erhebungsdesign postalischer Befragungen eingeplant werden. Insgesamt drei Erhebungszeitpunkte erscheinen dabei optimal.

complete survey sample in terms of their social variables as well as their psycho-social characteristics. However, one exception is the respondents' criminal behaviour over the life-course: with increasing numbers of mailings the respondents' delinquency strongly decreases. Obviously, responses on the survey topic are thus correlated with the timing of data collection. This again shows how reminders can help diminishing biases. From a methodological point of view it is therefore recommended to always include follow-up mailings in postal surveys. Three mailings seem to be ideal.

1 Einleitung¹

Im Vergleich zu anderen Erhebungsmodi herrscht gegenüber der postalischen Befragung als wissenschaftlicher Methode der Erhebung empirischer Daten unter Lehrenden und Sozialwissenschaftlern nach wie vor Skepsis.² Neben anderen Kritikpunkten ist ein weit verbreiteter und für die meisten wohl der schwerwiegendste Einwand gegen ihre Verwendung der, dass die erreichbaren Rücklaufquoten zu gering seien, Repräsentativität der Ergebnisse aufgrund massiver Selbstrekrutierung entsprechend nicht sichergestellt werden könne (u. a. Diekmann 1998: 333f., 439ff.; Häder 2006: 237; Atteslander 2008: 147; Schnell/Hill/Esser 2008: 359).

1 Den Herausgebern und zwei anonymen Gutachtern danke ich für wertvolle Kommentare und Anregungen.

2 So finden sich in weithin rezipierten Lehrbüchern der Methoden empirischer Sozialforschung eher abwertende bzw. einseitig kritische Darstellungen der postalischen Befragung (z. B. Friedrichs 1990; Atteslander 2008; Schnell/Hill/Esser 2008). Im Gegensatz dazu kann man bei Häder (2006) sowie Diekmann (1998; 2007) positivere bzw. ausgewogenere Abhandlungen nachlesen. Auch die Anwendung dieser Methode in der Forschungspraxis ist angesichts ihrer vielfältigen Möglichkeiten und den neueren Problemen anderer Erhebungsmodi (etwa sinkende Ausschöpfung in Face-to-Face Befragungen) bislang eher gering: „Den Jahresberichten der ADM-Institute folgend [...] werden von den genannten Instituten in Deutschland gegenwärtig etwa 20 % der quantitativen Interviews postalisch durchgeführt“ (Häder 2006: 234).

Diese Argumentation spiegelt nach Ansicht der Autorin in zweifacher Hinsicht veraltete bzw. uninformierte Auffassungen wider: *Erstens* stellt die Ausschöpfungsquote von Befragungen lediglich *einen* – und nicht einmal einen besonders aussagekräftigen – Indikator für die Güte von Erhebungen dar (Groves 1989: 133, 147; Platek/Särndal 2001; Biemer/Lyberg 2003)³. Und *zweitens* liegen mittlerweile zahlreiche Forschungsergebnisse vor, die für verschiedene Anwendungsbezüge zeigen, dass postalische Befragungen erfolgreich einsetzbar sind und „zufriedenstellende“ Ausschöpfungsquoten, die teilweise sogar jene von renommierten Face-to-Face-Befragungen (z. B. ALLBUS) übertreffen, erzielt werden können (vgl. z. B. Hippler/Seidel 1985; Niedermayer 1989; De Leeuw 1992; Blasius/Reuband 1996; Arzheimer/Klein 1998; Reuband 1999a, 1999b, 2001; Becker/Imhof/Mehlkop 2007; Kunz 2007; Mehlkop/Becker 2007). Sämtliche Publikationen weisen unisono darauf hin, dass qualitativ hochwertige postalische Befragungen (inklusive hoher Ausschöpfungsquoten) v. a. durch die Umsetzung zentraler Empfehlungen der Total bzw. Tailored Design Method (Dillman 1978, 2000; Dillman/Smyth/Christian 2009) gelingen.

Dass mehrmalige Kontaktierungen der angeschriebenen Personen – sog. Mahn- bzw. Nachfassaktionen – für die Steigerung der Rücklaufquote von zentraler Bedeutung sind⁴, ist vielfach belegt (Toops 1926; Yu/Cooper 1983; Guadagnoli/Cunningham 1989; Sobal/Ferentz 1989; Lamnek/Trepl 1991; Yammarino/Skinner/Childers 1991; Blasius/Reuband 1996 u. v. a.).

Doch welchen Einfluss haben Mahnaktionen auf das Antwortverhalten der Respondenten, d. h. auf das Ausmaß an Item-Nonresponse sowie auf die Antworten? Und inwiefern haben mehrfache Kontaktierungen Auswirkungen auf die soziale Zusammensetzung der Befragtenstichprobe? Welche Bedeutung haben Mahnaktionen damit, jenseits ihrer rücklaufsteigernden Wirkung, für die Güte der Befragung? Anders ausgedrückt: Ist es möglich, dass durch Mahnaktionen zwar die Ausschöpfungsquote erhöht, die Repräsentativität der Stichprobe aber verschlechtert wird? Erbringen geringere Ausschöpfungsquoten also u. U. validere Ergebnisse? Nur vergleichsweise wenige Studien beschäftigten sich bislang mit diesen Fragestellungen. Vorliegender Artikel untersucht die benannten Aspekte anhand der Daten einer im Jahr 2009 im Raum Südbaden durchgeführten postalischen Befragung (N=3.555) mit insgesamt drei Erhebungszeitpunkten zum Thema „Kriminalität im höheren Lebensalter“.

3 Einige Autoren konnten anhand empirischer Analysen zeigen, dass höhere Ausschöpfungsquoten nicht zwangsläufig mit verbesserter Stichprobenqualität, d. h. mit geringerem Ausmaß systematischer Antwortausfälle, einhergehen (Groves 1989; Guadagnoli/Cunningham 1989; Sobal/Ferentz 1989; Koch 1998; Schneekloth/Leven 2003; Stoop 2005: 286ff.). Entsprechend umstritten ist auch, welche Ausschöpfungsquoten als „gut“ bzw. „schlecht“ zu bewerten sind (für einen Überblick zu verschiedenen Positionen vgl. Stoop 2005: 23).

4 Daneben kann u. a. auch die Verwendung von Anreizen deutlich zur Rücklaufsteigerung beitragen (vgl. dazu kurz Abschnitt 4.2).

Der folgende Abschnitt führt zunächst einige theoretische Überlegungen zur allgemeinen Wirkungsweise von Nachfassaktionen in postalischen Befragungen an. Anschließend werden der Forschungsstand und die analytische Vorgehensweise der Studie skizziert (Abschnitt 3). Im vierten Abschnitt wird die Befragung, welche die Datenbasis liefert, kurz vorgestellt. Die Darstellung der Studienergebnisse erfolgt in Abschnitt 5. In einem letzten Schritt werden die Befunde zusammengefasst und kritisch reflektiert sowie methodologische Implikationen für die Durchführung bzw. Bewertung postalischer Befragungen benannt (Abschnitt 6).

2 Theoretische Überlegungen zur Wirkungsweise von Nachfassaktionen

Mehrmalige Kontaktierungen der Befragungspopulation werden durchgeführt, um auch Personen, die nicht unmittelbar nach dem Erstversand des Fragebogens antworten, zu einer Survey-Teilnahme zu bewegen. Die Wirkungsweise von Erinnerungsschreiben beruht dabei auf verschiedenen Mechanismen.

1. Ein großer Teil des Nonresponse ist keine stabile, sondern eine situationsabhängige Verhaltensweise (Schnell 1997: 191). Viele Personen sind prinzipiell zu einer Teilnahme bereit, vergessen jedoch das Ausfüllen des Fragebogens – etwa weil der Erstkontakt zu einem ungünstigen Zeitpunkt stattfand – oder schieben die Beantwortung der Fragen hinaus. Diese Personengruppe wird durch das Nachfassen schlicht an den Erhalt des Fragebogens erinnert. Hierdurch dürfte die Teilnehmerate steigen.

Aus Sicht der sozialen Austauschtheorie (Homans 1961; Blau 1964), die v. a. von Dillman (1978, 2000; Dillman/Smyth/Christian 2009) im Rahmen seiner Total/Tailored Design Method zur Erklärung des Teilnahmeverhaltens in postalischen Befragungen herangezogen wird, sind Reminder eine (weitere) Maßnahme des Forschers mit dem Ziel, das Nutzenkalkül der angeschriebenen Personen in Bezug auf die Befragungsteilnahme positiv zu beeinflussen. Damit dies gelingt, sollte durch die erneute Kontaktierung der subjektive Nutzen der Respondenten gesteigert und/oder deren Kosten gesenkt und/oder deren Vertrauen gesteigert werden. Es kommen entsprechend zwei weitere Wirkungsmechanismen in Betracht:

2. Zwar bleiben durch Follow-ups die Kosten einer Befragungsteilnahme konstant. Der Nutzen jedoch wird gesteigert, indem deutlich gemacht wird, dass die Meinung und Unterstützung des Angeschriebenen für den Erfolg der Untersuchung von zentraler Bedeutung sind. Zudem unterstreicht die Erinnerungsaktion die Sinnhaftigkeit der Studie und die Ernsthaftigkeit der Forscherabsichten. Dadurch kann das Vertrauen der Befragten gewonnen bzw. stabilisiert werden (Petermann 2005: 62f.).

3. In vorliegender Untersuchung wurde allen Angeschriebenen bereits beim Erstkontakt eine Fünf-Euro-Banknote bedingungslos und gerahmt als „kleines Dankeschön“ beigelegt (Näheres siehe Abschnitt 4.2). In Wechselwirkung mit der Vergabe eines initialen (monetären) Anreizes erscheint nun noch eine weitere Funktionsweise von Remindern plausibel, die im Vergleich zur vorher skizzierten eine stärkere normative Konnotation besitzt. Follow-ups erinnern an den Erhalt des Incentives und verstärken damit das Gefühl der sozialen Verpflichtung (Reziprozitätsnorm). Der Anreiz für eine – wenn auch verspätete – Befragungsteilnahme besteht nun v. a. in der Vermeidung eines schlechten Gewissens.

3 Stand der Forschung und Vorgehensweise der Studie

Zunächst werden bisherige empirische Befunde zu den Auswirkungen von Mahnaktionen auf den Item-Nonresponse, die soziale Zusammensetzung der Stichprobe und die Teilnahmebereitschaft von Menschen mit verschiedenen Einstellungen, Eigenschaften, Erfahrungen und Verhaltensweisen überblicksartig vorgestellt.

Auswirkungen auf das Ausmaß an Item-Nonresponse (INR): Hippler/Seidel (1985) und Reuband (1999b, 2001) konstatieren einen kontinuierlichen Anstieg fehlender Antworten mit jeder Nachfassaktion. So steigerte sich etwa bei Reuband (1999b: 95) der durchschnittliche Anteil fehlender Antworten an allen gestellten Fragen von 4,8 % beim Erstkontakt auf 7,2 % bei der vierten Kontaktierung.

Auswirkungen auf die soziale Zusammensetzung der Stichprobe: Während einige Forscher keine Unterschiede zwischen Respondenten, die früher oder später antworten konstatieren (Koenig/Martin/Seiler 1977; Goudy 1978; Guadagnoli/Cunningham 1989; Sobal/Ferentz 1989), stellen andere Studien fest, dass das Alter der Befragten mit den Nachfassungen variiert. Der konsistente Befund lautet hier, dass Menschen ab einem Alter von 60/65 Jahren überwiegend bereits beim Erstkontakt antworten, während mit weiteren Nachfassaktionen v. a. jüngere Personen in die Stichprobe rekrutiert werden (Hippler/Seidel 1985: 50; Blasius/Reuband 1996; Reuband 1999b: 88, 2001: 320). Scott (1961), Brambilla/McKinlay (1987), Blasius/Reuband (1996), Reuband (1999b; 2001) sowie Petermann (2005) konnten ferner zeigen, dass Angehörige unterer sozialer Schichten, d. h. Personen mit geringer Bildung sowie wenig qualifizierten Berufen, erst mithilfe von Nachfassaktionen stärker repräsentiert werden. Allerdings konnten *keine* Unterschiede im Nettoeinkommen der Befragten verschiedener Erhebungszeitpunkte festgestellt werden (Reuband 1999b; Petermann 2005: 72f.). Die einzige Studie, die Hinweise darauf fand, dass Mahnaktionen die Stichprobenanteile der Geschlechter beeinflussen, stammt von Niedermayer (1989). Hier antworteten Frauen später als Männer.

Aus der Surveyforschung ist bekannt, dass die Thematik einer Befragung bzw. deren Bewertung durch die Respondenten einen Einfluss auf die Teilnahmebereitschaft der Befragten hat (Kanuk/Berenson 1975: 449; Dommeyer 1985; Hippler/Seidel 1985; Schnell 1997: 170ff.), und dass in Populationen, deren Merkmale eng mit dem Befragungsthema zusammenhängen (sog. Spezialpopulationen), zumeist höhere Rücklaufquoten erzielt werden als in allgemeinen Bevölkerungsstichproben (Hippler 1988). Bei genauerer Betrachtung der hier angeführten Studien fällt nun auf, dass die jeweils untersuchten Populationen und Themen ebenfalls einen Einfluss auf die Wirkungsweise von Nachfassaktionen haben. Für Spezialpopulationen im oben genannten Sinne können im Vergleich zu allgemeinen Bevölkerungsstichproben viel seltener Unterschiede zwischen den Respondenten verschiedener Erhebungszeitpunkte festgestellt werden. Nachfassaktionen erhöhen hier lediglich den Rücklauf, haben jedoch kaum Auswirkungen auf die Zusammensetzung der Netto-Stichprobe. Für allgemeine Bevölkerungsstichproben bzw. Populationen, deren Merkmale nicht mit dem Befragungsthema korrelieren, tragen Reminder hingegen nicht nur zur Erhöhung der Ausschöpfungsquote, sondern auch zu einer Veränderung (bzw. Verbesserung) der Stichprobenszusammensetzung bei.⁵

Insgesamt kann festgehalten werden, dass sich die Befragten der einzelnen Mahnaktionen zwar hinsichtlich verschiedener sozio-demografischer Merkmale teilweise deutlich voneinander unterscheiden. Bisherige Untersuchungen deuten aber darauf hin, dass sich die sozialen Zusammensetzungen der gesamten Befragtenstichproben *vor* und *nach* der Durchführung von Mahnaktionen in Bezug auf die Merkmale Alter, Geschlecht und Bildung sehr ähnlich sind. „Die Prozentpunktunterschiede sind nicht allzu groß – und dies gilt selbst in gewissem Maße für die Variablen, für die nennenswerte Effekte der Mahnaktion festgestellt werden können“ (Reuband 2001: 328).

Dennoch: während Reuband (2001: 328) zu der Feststellung gelangt, dass „[...] Mahnaktionen zur Verbesserung der sozialen Zusammensetzung [...] beitragen und deshalb prinzipiell sinnvoll sind“, kommen etwa Koch (1998) sowie Schneekloth/Leven (2003) anhand ihrer Analysen verschiedener ALLBUS-Erhebungen⁶ zu dem Schluss, dass die ohnehin nur geringen Abweichungen sozialer Merkmale der Befragten von denen der Grundgesamtheit durch Nachbearbeitungen weder abgeschwächt noch beseitigt werden. Es erhöht sich vorrangig die Fallzahl, nicht die Qualität der Stichprobe. Petermann (2005) behauptet gar, dass bestehende Ver-

5 Zu den gleichen Ergebnissen gelangt Leslie (1972) in seiner Metastudie.

6 Zwar wird der ALLBUS mittels persönlicher Interviews erhoben und stellt somit keine postalische Befragung dar. Dennoch weisen die Forschungsfragen bzw. Erkenntnisse beider Studien mode-unabhängig einen engen Bezug zu den im vorliegenden Beitrag behandelten Fragestellungen auf.

zerrungen in postalischen Befragungen durch Nachfassungen (zumindest minimal) verstärkt werden und deshalb auf Mahnaktionen sogar verzichtet werden sollte.⁷

Angesichts der einhelligen⁸ Feststellung, dass sich die sozialen Merkmale der Befragten vor und nach Nachfassungen nur geringfügig unterscheiden und zudem kaum von entsprechenden Verteilungen in der Grundgesamtheit abweichen,⁹ schlussfolgern sämtliche benannten Autoren übereinstimmend, dass auch niedrige Ausschöpfungsquoten eine gute Repräsentativität der Stichprobe erbringen können.

Auswirkungen auf die Teilnahmebereitschaft von Menschen mit verschiedenen Einstellungen, Eigenschaften, Erfahrungen und Verhaltensweisen: Zahlreiche weitere Merkmale, die in keinem oder nur sehr schwachem Zusammenhang mit den geprüften Sozio-Demografika stehen, sind für die Charakterisierung der Grundgesamtheit bzw. der Stichprobe von Bedeutung. Deshalb erbringt ein Vergleich verschiedener Merkmale der Befragten einzelner Erhebungswellen Erkenntnisse darüber, ob Mahnaktionen zu Veränderungen des Samples führen.¹⁰

Reuband (1999b, 2001) und Petermann (2005) konnten feststellen, dass mit zunehmenden Mahnaktionen der Anteil politisch Uninteressierter steigt, während thematisch interessierte Personen vorwiegend zu Beginn der Erhebung teilnehmen.

Keine Auswirkungen der Mahnaktionen fand Reuband (1999b: 93ff.) hinsichtlich des Vertrauens der Befragten in ihre Mitmenschen, in Bezug auf ihre Lebenszufriedenheit, die Anzahl möglicher Sorgen, hinsichtlich der Kriminalitätsfurcht, eigener Opfererfahrungen sowie in Bezug auf die politische Orientierung.

Zusammenfassend kommen die Studien zu dem Schluss, dass sich die Befragten nach dem Erstversand auch hinsichtlich „weicher“ Merkmale kaum von denen der Gesamtheit nach allen Mahnaktionen unterscheiden. Die Effekte der Reminder sind schwach bzw. nicht signifikant (Reuband 1999b: 96). Einstellungen und psychosoziale Befindlichkeiten der Respondenten werden durch Mahnaktionen reproduziert.

Vorgehensweise der Studie: Zwar ist die rücklaufsteigernde Wirkung von Nachfassungen bereits umfangreich belegt. Der Vollständigkeit halber soll dies dennoch anhand der Südbadener postalischen Befragung kurz veranschaulicht und kom-

7 Zu ähnlichen Ergebnissen gelangt Stoop (2005: 287) in Bezug auf den Nonresponse im European Social Survey (ESS): auch hier wurde die Stichprobenqualität durch einseitige Konvertierung anfänglicher Verweigerer eher verschlechtert.

8 Ausnahme ist bei Reuband (1999b: 91) das Bildungsniveau; dieses ist in der Stichprobe im Vergleich zur Population (wie häufig in postalischen Befragungen) leicht erhöht.

9 Es sei an dieser Stelle noch einmal darauf hingewiesen, dass derartige „Repräsentanzabgleiche“ lediglich als Indiz, keinesfalls als Beweis für eine gelungene Abbildung der Grundgesamtheit durch die Stichprobe gelten können (vgl. zum „Repräsentanzbegriff“ auch Schnell/Hill/Esser 1995: 287).

10 Da ein Abgleich von Einstellungen, persönlichen Erfahrungen, Verhaltensweisen der Befragten etc. mit offiziellen Daten nicht möglich ist, kann allerdings nicht festgestellt werden, welche Auswirkungen dies auf eventuell bereits beim Erstkontakt bestehende Verzerrungen hat.

mentiert werden. Anschließend werden die Auswirkungen von Mahnaktionen auf a) das Ausmaß an Item-Nonresponse, b) die soziale Zusammensetzung der Stichprobe sowie auf c) die Teilnahmebereitschaft von Menschen mit verschiedenen Einstellungen, Eigenschaften, Erfahrungen und Verhaltensweisen untersucht. Die Ergebnisse werden den bisherigen Forschungsbefunden jeweils gegenübergestellt. Um die Auswirkungen von Mahnaktionen auf die benannten Aspekte analysieren zu können, wurden die eingehenden Fragebögen anhand ihres Eingangsstempels den einzelnen Erhebungswellen zugeordnet.¹¹

Vor Beginn der eigentlichen Analyse in Abschnitt 5 wird jedoch zunächst die verwendete Datenbasis vorgestellt.

4 Die Datengrundlage

Die Fragestellungen des vorliegenden Beitrages werden auf der Grundlage der Daten einer regionalen postalischen Befragung mit dem Titel „Alltag & Gesetze. Erfahrungen und Einstellungen in der zweiten Lebenshälfte“ untersucht. Diese Befragung wurde im Jahr 2009 im Rahmen eines Forschungsprojekts zum Thema „Ältere Menschen und Kriminalität“ am Max-Planck-Institut für ausländisches und internationales Strafrecht in Freiburg durchgeführt.¹²

4.1 Grundgesamtheit, Stichprobe und Fragebogen

Zur Grundgesamtheit der Befragung zählen deutsche Staatsbürger, die zwischen dem 01.01.1929 und dem 01.01.1959 geboren wurden und in Privathaushalten der 302 Gemeinden des Regierungsbezirkes (RB) Freiburg, d. h. in der Region Südbaden, wohnhaft sind.¹³ Die Bruttostichprobe beruht auf einer geschichteten, geklumpten, zweistufigen Zufallsauswahl und enthält 3.555 Personen im Alter zwischen 49 und 81 Jahren, die aus den Adressregistern von Einwohnermeldeämtern zufällig gezogen wurden.

11 Da die Abstände zwischen den einzelnen Erhebungswellen recht groß waren, dürften die Unschärfen, die sich aus dieser Kodierung ergeben, geringfügig sein.

12 Projektbeschreibung: <http://www.mpicc.de/ww/de/pub/forschung/forschungsarbeit/kriminologie/aelteremenschen.htm> (07.06.2010).

13 Menschen, die in Institutionen wie etwa Alters- oder Pflegeheimen leben, sowie Personen mit ausländischer Staatsbürgerschaft sind nicht in der Population enthalten; für diese können im Rahmen der Studie entsprechend keine Aussagen getroffen werden.

Der verwendete Fragebogen umfasst 24 Seiten und beinhaltet 85 Fragen bzw. 240 einzelne Items.¹⁴ Er wurde als DIN A4-Broschüre auf farbiges Papier gedruckt. Der Bogen beinhaltet Fragen zu Einstellungen gegenüber Gesetzen und Gesetzesübertretungen, zu Viktimisierungserfahrungen, zu Kriminalitätsfurcht sowie zu bisheriger polizeilicher Registrierung, kriminellen Verhalten innerhalb verschiedener Lebensabschnitte und zur Delinquenzbereitschaft. Daneben wurden Informationen zu Persönlichkeitseigenschaften, Werteinstellungen, bedeutenden Lebensereignissen sowie zum emotionalen Wohlbefinden, dem Gesundheitszustand und zu Einstellungen gegenüber dem Altern erhoben. Sozio-demografische Angaben wurden am Ende des Bogens abgefragt.¹⁵ Aufgrund der sensitiven Forschungsthematik „selbstberichtete Delinquenz“ wurde die Befragung vollständig anonym, d. h. ohne Aufdruck einer Identifikationsnummer, durchgeführt.¹⁶ Um dennoch effizient nachfassen zu können, wurden adressierte und frankierte Antwortpostkarten beigelegt, die nach erfolgter Beantwortung des Fragebogens – quasi als Benachrichtigung – seitens der Respondenten mit Namen versehen separat an uns zurückgeschickt werden sollten.

4.2 Datenerhebung, Ausschöpfung und weitere methodenbezogene Befunde

Die Erhebungsphase erstreckte sich von Juni bis August 2009, betrug ca. neun Wochen und beinhaltete drei Erhebungszeitpunkte.¹⁷ Beim Erstkontakt wurde an alle Befragten ein DIN C4-Umschlag verschickt, welcher ein doppelseitiges Anschreiben, den Fragebogen, einen adressierten und frankierten Rückumschlag sowie eine adressierte und frankierte Antwortpostkarte enthielt. Als forschungsdurchführende Institutionen wurden das Max-Planck-Institut Freiburg sowie das Institut für Soziologie der Universität Freiburg genannt. Zur Steigerung des Rücklaufs wurde allen Ange-

14 Davon sollten 76 Fragen bzw. 164 Items von allen Befragten beantwortet werden, während weitere 9 Fragen bzw. 76 Items als sog. Filterfragen nur in Abhängigkeit vorheriger Antworten gegebenenfalls zu beantworten waren.

15 Das Erhebungsinstrument wurde vor dem Einsatz mittels „Zwei-Phasen-Pretesting“ (Prüfer/Rexroth 2000), welches einen kognitiven Pretest (n=8) und einen Standardpretest (N=400) beinhaltete, schrittweise geprüft und optimiert.

16 Die Ergebnisse eines im Vorfeld durchgeführten Methodenexperiments zu den Auswirkungen anonymer bzw. nicht-anonymer Befragung auf die Rücklaufquote, das Antwortverhalten und die Stichprobenszusammensetzung trugen ebenfalls zu dieser Entscheidung bei.

17 Die Konzeption der postalischen Befragung orientiert sich im Wesentlichen an der Total/Tailored Design Method (Dillman 1978, 2000; Dillman/Smyth/Christian 2009). Die Anzahl, die Zeitpunkte sowie die Form der einzelnen Mahnaktionen weichen allerdings etwas von den Empfehlungen ab.

schriebenen bereits beim Erstkontakt eine Fünf-Euro-Banknote beigelegt.¹⁸ Diese Gabe wurde im Anschreiben mit folgendem Satz „gerahmt“: „*Als kleine Aufwandsentschädigung und Dankeschön für Ihre Mitwirkung haben wir Ihnen 5 Euro beigelegt. Diese können Sie in jedem Fall behalten.*“ Um die Angeschriebenen darüber hinaus zur Unterstützung des Forschungsvorhabens zu motivieren, wurde dieses in einer regionalen Sonntagszeitung, in den jeweiligen Amts-/Gemeindeblättern der gesampelten Ortschaften sowie im Hörfunk vorgestellt.

Etwa zweieinhalb Wochen nach dem Erstversand wurde ein kurzes Erinnerungsschreiben an diejenigen Personen verschickt, von denen bisher keine Antwortkarte mit dem Hinweis bei uns eingegangen war, dass der Fragebogen ausgefüllt retourniert wurde. Dieses Schreiben enthielt sowohl die erneute Bitte um Teilnahme als auch – für den Fall bereits erfolgter Rücksendung des Fragebogens – Dank für die geleistete Unterstützung. Nach weiteren drei Wochen wurden die Befragten ein letztes Mal kontaktiert, indem noch einmal sämtliche Unterlagen des Erstversands sowie ein letztes Erinnerungsschreiben zugeschickt wurden.

Von den 3.555 angeschriebenen Personen fielen 132 neutral aus,¹⁹ 1.997 Fragebögen konnten in die Auswertung einbezogen werden. Auf der Basis der bereinigten Bruttostichprobe erzielte die Befragung somit eine Ausschöpfungsquote²⁰ in Höhe von 58,4 %.

Insgesamt wurden 9 % aller vergebenen Incentives zurückerhalten. Die beigelegte Fünf-Euro-Banknote wurde dabei sowohl von Personen, die aus verschiedenen Gründen nicht an der Befragung teilnehmen konnten (neutrale Ausfälle), von Personen, die nicht an der Befragung teilnehmen wollten (Verweigerer), als auch von Angeschriebenen, die den Fragebogen beantwortet hatten, zurückgegeben.

Die *separate-Postkarten-Technik* (sPT), die zur Rücklaufkontrolle eingesetzt wurde, funktionierte in Anbetracht der zusätzlichen Komplexität der Aufgabenstellung und des Alters der Befragten recht gut. Insgesamt teilten uns 1.431 Personen mit, dass sie einen Fragebogen retourniert haben. Geht man davon aus, dass

18 Der Einsatz von Anreizen bei der Durchführung postalischer Befragungen wird u. a. von Dillman (2000: 150) und Dillman/Smyth/Christian (2009: 238ff.) im Rahmen der Tailored Design Method empfohlen. Zur rücklaufsteigernden Wirkung monetärer Incentives speziell in postalischen Befragung zu kriminalitätsbezogenen Themen vgl. Becker/Imhof/Mehlkop (2007); Mehlokop/Becker (2007). Dort werden auch zahlreiche weitere Studien zum Einsatz monetärer Anreize in postalischen Befragungen benannt.

19 Als neutrale Ausfälle gelten: unbekannt/verzogen, verstorben, Krankheit/Pflege/Behinderung, kaum deutsche Sprachkenntnisse, Person gehört nicht zur Grundgesamtheit. Die bereinigte Bruttostichprobe umfasst somit 3.423 Personen.

20 Die Ausschöpfungsquote (A) wurde hier wie folgt berechnet (Schnell/Hill/Esner 1995: 289; Diekmann 1998: 359): $A = (\text{Anzahl auswertbarer Fragebögen} / \text{Umfang der bereinigten Bruttostichprobe}) * 100 \%$.

dies lediglich Personen waren, die tatsächlich an der Befragung teilgenommen haben, dann entspricht dies einer beachtlich guten Benachrichtigungsquote²¹ von 71,7 %. Die Ergebnisse zeigen zweierlei: 1) die Anzahl der retournierten Postkarten ist *nicht* – wie im Lehrbuch von Schnell/Hill/Esser (2008: 363) nach wie vor behauptet – höher als die der retournierten Fragebögen (vgl. ähnliche Ergebnisse bei Reuband 1999a) und 2) die sPT bietet eine gute Möglichkeit, *sowohl* Anonymität *als auch* effektive Rücklaufkontrolle in postalischen Befragungen zu kombinieren – und dies sogar mit der Zielgruppe älterer Menschen.

Ein häufig genannter Kritikpunkt gegenüber postalischen Befragungen ist, dass die Erhebungssituation nicht kontrolliert werden kann, es also z. B. nicht sicher ist, wer den Fragebogen ausgefüllt hat (z. B. Diekmann 1998: 439; Schnell/Hill/Esser 2008: 359). Um zumindest das Ausmaß gemeinschaftlicher Beantwortung abschätzen zu können, wurden die Respondenten am Ende des Bogens gefragt: „*Haben Sie den Fragebogen allein oder gemeinsam mit einer anderen Person ausgefüllt?*“ 7,9 % der Befragten gaben an, den Fragebogen gemeinsam mit einer anderen Person ausgefüllt zu haben, die restlichen 92,1 % beantworteten die Fragen allein (n=1.962). Geht man von der Wahrhaftigkeit dieser Angaben aus, sind diese Werte sehr befriedigend. Wie später noch gezeigt wird, nimmt der Anteil derjenigen, die den Fragebogen gemeinschaftlich ausfüllen, allerdings mit jeder Mahnaktion zu (vgl. Abschnitt 5.1).

5 Ergebnisse

Wie bereits in Abschnitt 3 angedeutet, werden die unterschiedlichen Auswirkungen von Mahnaktionen jeweils separat vorgestellt: zunächst wird der Einfluss auf den Gesamtrücklauf betrachtet sowie untersucht, inwiefern sich das Ausmaß an Item-Nonresponse in Abhängigkeit der Kontaktierungen verändert (Abschnitt 5.1). Anschließend werden die Auswirkungen von Nachfassaktionen auf die Stichprobenzusammensetzung und auf verschiedene weitere Merkmale der Befragten behandelt (Abschnitte 5.2 und 5.3).

21 Diese Quote liegt deutlich über bisher dokumentierten Erfahrungen, etwa den von Reuband (1999: 58) berichteten Raten für allgemeine Bevölkerungsbefragungen in Dresden (66 %) und Düsseldorf (60 %).

5.1 Auswirkungen auf den Gesamtrücklauf und den Item-Nonresponse

Die erste Zeile von Tabelle 1 zeigt erneut, was bereits durch zahlreiche Studien belegt wurde: Nachfassaktionen, wie in der Total/Tailored Design Method (Dillman 1978, 2000; Dillman/Smyth/Christian 2009) vorgeschlagen, sind eine erfolgreiche Maßnahme zur Rücklauferhöhung. Allerdings erscheint eine dritte oder gar vierte Nachfassung entgegen Dillman'scher Empfehlungen angesichts der deutlich sinkenden Beiträge der einzelnen Kontaktierungen zum Gesamtrücklauf kaum mehr rentabel (vgl. auch Blasius/Reuband 1996; Reuband 1999b: 86, 2001: 319). Die Netto-Ausschöpfungsquoten der einzelnen Erhebungswellen sind im Gegensatz zu der von Dillman (1978: 185f.) bzw. Hippler (1988: 245) proklamierten „Faustregel“ *nicht* in etwa gleich hoch, sondern nehmen linear ab (vgl. Tab. 1).

Tabelle 1 Auswirkungen von Maßnahmen zur Rücklauferhöhung

	Erstversand	1. Mahnung	2. Mahnung		
1) Gesamtrücklaufquote nach ...	33,4	50,1	58,3		
2) Netto-Ausschöpfungsquote	33,4 (ber. N=3423)	24,9 (ber. N=2286)	16,3 (ber. N=1734)		
	Erstversand (n=1144)	1. Mahnung (n=570)	2. Mahnung (n=283)	Gesamt (n=1997)	Spearman Rho
3) Item-Nonresponse-Rate gesamter Fragebogen	6,13	7,80	7,50	6,80	0,051*
4) Item-Nonresponse-Rate 14 Prävalenzen, Lebensspanne	1,59	3,32	2,60	2,23	0,048*
5) Item-Nonresponse-Rate 14 Inzidenzen, seit 50. Geburtstag	2,37	4,05	3,81	3,05	0,017
6) Item-Nonresponse-Rate 14 Inzidenzen, letzte 12 Monate	4,00	5,10	4,42	4,37	-0,014

Gesamtrücklaufquote nach Erhebungswelle (ber. N=3423; in Prozent), Netto-Ausschöpfungsquoten der Erhebungswellen (in Prozent), Item-Nonresponse-Raten nach Erhebungswellen (in Prozent; grau: die Item-Nonresponse-Rate des Erstversands ist sign. geringer als die der 1. und 2. Mahnung, $p \leq 0,05$).

* $p \leq 0,05$

zu 3) $F=7,734$; $p(2\text{-seit.})=0,00$; $df=2$; zu 4) $F=3,438$; $p(2\text{-seit.})=0,03$; $df=2$; zu 5) $F=4,026$; $p(2\text{-seit.})=0,02$; $df=2$; zu 6) $F=1,209$; $p(2\text{-seit.})=0,30$; $df=2$

Was das Ausmaß an Item-Nonresponse betrifft, so beträgt der durchschnittliche Anteil fehlender Antworten am gesamten Fragebogen in der Südbadener Befragung 6,8 %. Im Mittel fehlen damit ca. 14 Antworten (von 202) pro Fragebogen. Dies stellt einen unauffälligen Wert dar.²² Bemerkenswert ist, dass die durchschnitt-

22 Nach Hippler (1988) liegt die INR-Rate postalischer Befragungen üblicherweise zwischen 1 und 8 %. Das Ausmaß fehlender Antworten in der Südbadener Befragung liegt mit einer INR-Rate von 6,8 % also am oberen Ende des Durchschnitts. Es wird vermutet, dass das Alter der Befragten hierfür verantwortlich sein könnte, da zwischen dem Ausmaß fehlender Antworten und dem Alter der Befragten ein Zusammenhang in Höhe von $r=0,224$ ($p \leq 0,01$) besteht.

lichen INR-Raten sämtlicher Fragen zur Delinquenz der Befragten (auf Itemebene) und zum Netto-Einkommen lediglich zwischen 2 und 4,4 % betragen. Bei den heiklen Fragen sind im Vergleich zur gesamten Befragung also unterdurchschnittlich wenige Antwortausfälle zu verzeichnen. Tabelle 1 gibt für verschiedene Fragenkomplexe einen Überblick über das Ausmaß der Antwortausfälle in Abhängigkeit der einzelnen Erhebungswellen. Mittels einfaktorierter Varianzanalyse wurde jeweils geprüft, ob sich die Item-Nonresponse-Raten in Abhängigkeit des Erhebungszeitpunktes unterscheiden. Anhand des Student-Newman-Keuls-Tests wurde zudem ermittelt, *welche* Erhebungswellen sich signifikant voneinander unterscheiden.

Aus Tabelle 1 geht deutlich hervor, dass das Ausmaß fehlender Antworten mit den Mahnaktionen ansteigt. Während jedoch Hippler/Seidel (1985) und Reuband (1999b, 2001) einen kontinuierlichen Anstieg fehlender Antworten über die Mahnaktionen feststellen, zeigt sich hier ein kurvilinearere Zusammenhang. Sämtliche INR-Raten erreichen ihren Höhepunkt unter den Respondenten der 1. Mahnung, sinken dann leicht wieder ab, bleiben aber über dem Niveau der nach dem Erstversand retournierten Fragebögen.²³

Reuband (2001: 321f.) konnte zeigen, dass es durch die Mahnaktionen gelingt, auch Personen mit Zeitmangel und tendenziell nicht-kooperative Personen in die Befragung einzubeziehen. Dadurch erklärt sich seines Erachtens auch der unter den nachträglich rekrutierten Respondenten erhöhte INR (2001: 322). Mit den Worten Reubands (2001: 322) kann deshalb Folgendes festgehalten werden:

„Man mag dies [Anm. d. Autorin: den INR-Anstieg] einerseits als Nachteil für die Qualität der Daten ansehen: gilt doch, dass diese Personen weniger motiviert sind und den Fragebogen weniger sorgfältig beantworten. Andererseits bedeutet es auch, dass Personen mit wenig auskristallisierten Einstellungen und 'non-attitudes' (Converse 1964) erfasst werden. Widersprüche in den Einstellungen und Meinungslosigkeit sind Bestandteil der sozialen Realität und bedürfen der Berücksichtigung.“

Zudem erscheint der Trade-off zwischen verbesserter Stichprobenqualität und verschlechterter Datenqualität – sofern man Item-Nonresponse überhaupt als Beeinträchtigung der Datenqualität wertet – in Anbetracht der zwar meist signifikanten,

23 Eine mögliche Erklärung für die INR-Verteilung könnte sein, dass Personen ohne Schulabschluss bzw. mit Volks-/Hauptschulabschluss im Vergleich zum Erstversand und zur 2. Mahnaktion am stärksten auf die 1. Mahnung reagiert haben, während Personen mit höheren Bildungsabschlüssen anteilmäßig sowohl im Erstversand als auch in der 2. Mahnung stärker vertreten sind als in der 1. Mahnung. (Diese Ergebnisse sind hier nicht explizit dargestellt, können aber auf Anfrage zur Verfügung gestellt werden.) Aufgrund etwas geringerer kognitiver Kompetenzen bzw. weil längeres Lesen und Schreiben ungewohnt ist, könnten deshalb vermehrt Fehler beim Ausfüllen des Fragebogens aufgetreten sein. Eine alternative Erklärung wäre, dass Teilnehmer der 1. Mahnung eher „genervt“ und deshalb unkonzentrierter antworten als Befragte der 2. Mahnaktion, die eventuell aufgrund schlechten Gewissens der sehr späten Teilnahme wiederum etwas sorgfältiger antworten.

aber insgesamt recht geringen Steigerung der Antwortausfälle über die Nachfassaktionen durchaus akzeptabel.

Das Ausmaß fehlender Antworten kann als *ein* Kriterium der Qualität der Erhebungsdaten angesehen werden. Ein weiteres ist die gemeinschaftliche Beantwortung des Fragebogens. Wie bereits weiter oben erwähnt wurde, steigt mit den Mahnaktionen auch der Anteil derjenigen an, die den Fragebogen gemeinsam mit einer anderen Person, statt wie gewünscht allein, ausgefüllt haben: von 5,5 % im Erstversand auf 12,7 % in der 2. Mahnung. Dieser Zusammenhang zwischen gemeinschaftlicher Beantwortung und Beantwortungszeitpunkt ist signifikant (Chi-Quadrat=13,429; df=2; Cramers V=0,083; p(2-seitig)=0,001).

5.2 Auswirkungen auf die soziale Zusammensetzung der Stichprobe

Um zu überprüfen, inwiefern Mahnaktionen die soziale Zusammensetzung der Befragtenstichprobe – und damit ihre Güte – beeinflussen, wurden zunächst die sozio-demografischen Merkmale Geschlecht, Alter (gruppiert) und Wohnort der Befragten (kategorisiert nach Einwohneranzahl) den entsprechenden Verteilungen in der Grundgesamtheit gegenüber gestellt.²⁴ Um Veränderungen der Stichprobenkomposition bezüglich der einzelnen Merkmale zu erkennen, wurden zudem die Merkmalsverteilungen der über die einzelnen Erhebungswellen *akkumulierten* Netto-Stichproben betrachtet. Insgesamt konnte Folgendes festgestellt werden:

- a) Die gesamte Befragtenstichprobe bildet die Grundgesamtheit hinsichtlich des Alters, des Geschlechts und des Wohnorts der Personen recht gut ab. Abweichungen von den Verteilungen der Population liegen jeweils im minimalen Bereich.²⁵
- b) Die Befragten der verschiedenen Erhebungswellen unterscheiden sich hinsichtlich der Geschlechtszugehörigkeit sowie des Alters *nicht* signifikant voneinander. Tendenziell gilt jedoch auch hier, dass ältere Zielpersonen eher antworten als jüngere (vgl. Reuband 1999b; vgl. Reuband 2001; Petermann 2005). Die Art des Wohnortes der Befragten allerdings unterscheidet sich signifikant

24 Die Daten der Grundgesamtheit wurden der Bevölkerungsstatistik 2005 des Statistischen Landesamtes Baden-Württemberg entnommen. Da die Befragung anonym durchgeführt wurde, können sämtliche Abgleiche lediglich auf der Aggregatdatenebene vollzogen werden.

25 Tendenziell beteiligten sich Männer minimal häufiger an der Befragung als Frauen. Dies könnte daran liegen, dass Männer einen stärkeren Bezug zum Thema „Kriminalität“ haben. Personen im Alter zwischen 60 und 70 Jahren sind in der Südbadener Befragung etwas überrepräsentiert, während Angehörige der höchsten Altersgruppe leicht unterrepräsentiert sind. Die Bewohner von Groß- bzw. Mittelstädten sind in der Netto-Stichprobe der gesamten Befragung leicht stärker vertreten als in der Grundgesamtheit, während Dorfbewohner etwas unterrepräsentiert sind.

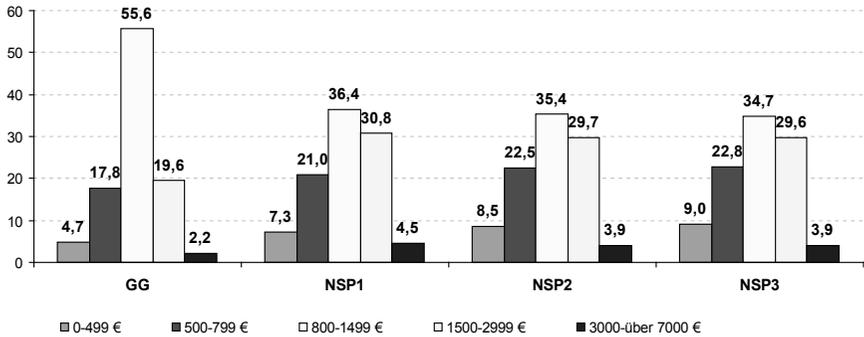
zwischen den Erhebungswellen: der Anteil der Dorfbewohner steigt mit zunehmender Zahl der Kontaktierungen an (Chi-Quadrat=13,050; df=4; Cramers V=0,059; p(2-seitig)=0,011).

- c) Zwar unterscheiden sich die Netto-Stichproben der einzelnen Erhebungswellen nicht wesentlich voneinander (ähnlich Blasius/Reuband 1996; Reuband 1999b, 2001; ähnlich Petermann 2005). Dennoch ist die *gesamte* Befragtenstichprobe im Vergleich zu den Stichproben nach Erstversand bzw. nach Erstversand und 1. Mahnung hinsichtlich der geprüften Merkmale den Verteilungen in der Grundgesamtheit jeweils am ähnlichsten. Dies bedeutet, dass die Durchführung von zwei Mahnaktionen die soziale Komposition der Stichprobe im Sinne einer Annäherung an die Verhältnisse in der Population positiv beeinflusst. Andererseits bedeutet es eben auch, dass bereits die Stichprobe nach dem Erstversand zumindest die Verteilung zentraler sozio-demografischer Merkmale in der Grundgesamtheit relativ gut repräsentiert.

Um weitere Hinweise auf eventuelle Verzerrungen der Stichprobe infolge von Mahnaktionen zu erhalten, wurden ferner das monatliche Personen-Nettoeinkommen, die Schulbildung sowie die Haushaltsgröße der Befragten mit offiziellen Daten der Grundgesamtheit – hier: Daten für den Regierungsbezirk Freiburg des Mikrozensus 2006²⁶ – verglichen. Aus Abbildung 1 geht hervor, dass sowohl Personen der untersten beiden Einkommenskategorien (0 bis 799 Euro) als auch der beiden oberen (1.500 bis über 7.000 Euro) in der gesamten Befragtenstichprobe (NSP3) im Vergleich zur Grundgesamtheit überrepräsentiert sind, während Personen mit einem mittleren Einkommen in Höhe von 800 bis 1.499 Euro deutlich unterrepräsentiert sind. Die Einkommensverteilung der Respondenten der einzelnen Erhebungswellen lässt die Tendenz erkennen, dass die Anteile derjenigen Personen, deren monatliches Netto-Einkommen in den unteren Kategorien liegt, mit zunehmenden Mahnaktionen etwas zunimmt, während der Anteil derjenigen mit Einkommen in den höheren Kategorien leicht abnimmt. Allerdings sind diese Unterschiede zwischen den Erhebungswellen minimal und *nicht* signifikant (ähnliche Ergebnisse vgl. Reuband 2001: 89f.; Petermann 2005: 72f.).

26 Nähere Informationen zum Mikrozensus: <http://www.forschungsdatenzentrum.de/bestand/Mikrozensus/index.asp> (07.06.2010). Da für die Nettovalidierung tief regionalisierte Daten (Regierungsbezirksebene) des Mikrozensus benötigt wurden und diese nicht im Scientific Use File enthalten sind, wurden die Analysen mittels sog. kontrollierter Datenfernverarbeitung durchgeführt.

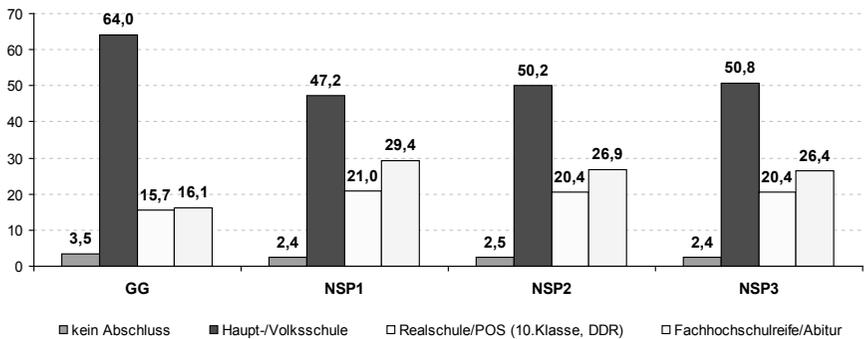
Abbildung 1 Rücklaufquoten nach Personen-Nettoeinkommen



Monatliches Personen-Nettoeinkommen (kategor.) in Grundgesamtheit (GG), Netto-Stichprobe nach Erstversand (NSP1), Netto-Stichprobe nach Erstversand und 1. Mahnung (NSP2) und gesamter Netto-Stichprobe (NSP3) (in Prozent).

Auch in Bezug auf den höchsten Schulabschluss fällt auf, dass die Befragtenstichprobe gegenüber der Population verzerrt ist: Personen ohne Abschluss bzw. mit Haupt- oder Volksschulabschluss sind unterrepräsentiert, während v. a. Personen mit höheren Bildungsabschlüssen (Fachhochschulreife bzw. Abitur) deutlich überproportional vertreten sind (vgl. Abbildung 2). Wie bereits weiter oben erwähnt, ist dies in Bezug auf den postalischen Befragungsmodus ein häufiger Befund.

Abbildung 2 Rücklaufquoten nach Schulbildung



Schulbildung in Grundgesamtheit (GG), Netto-Stichprobe nach Erstversand (NSP1), Netto-Stichprobe nach Erstversand und 1. Mahnung (NSP2) und gesamter Netto-Stichprobe (NSP3) (in Prozent).

Die Schulbildung unterscheidet sich ferner signifikant zwischen den Erhebungswellen²⁷ (ähnliche Ergebnisse vgl. Reuband 2001: 89f.). Der Anteil derjenigen Personen, die als höchsten Schulabschluss einen Haupt- bzw. Volksschulabschluss oder den Abschluss der Polytechnischen Oberschule/10. Klasse (DDR) aufweisen, steigt mit zunehmenden Mahnaktionen an, während sich die Anteile der Personengruppen mit anderen Abschlüssen verringern.

Die Haushaltsgröße hingegen wird durch die gesamte Befragungsstichprobe recht gut abgebildet. Es ergeben sich lediglich kleinere Abweichungen von der Verteilung in der Grundgesamtheit.²⁸

Zusammenfassend kann Folgendes festgehalten werden: ein Vergleich der Einkommens-Mittelwerte in der Grundgesamtheit und allen Netto-Stichproben zeigt, dass der Mittelwert von NSP3 demjenigen der GG noch am stärksten angenähert ist²⁹. Hinsichtlich des höchsten Schulabschlusses sowie der Haushaltsgröße bildet ebenfalls die NSP3 die Verteilung in der GG am besten ab. Die Unterschiede zwischen den einzelnen Netto-Stichproben (v. a. zwischen NSP2 und NSP3) sind allerdings stets marginal. Mit zunehmenden Mahnaktionen ergibt sich demnach keine wesentliche Verbesserung der Stichprobenzusammensetzung in Bezug auf die drei untersuchten Merkmale der Befragten.

Abschließend wurde für drei weitere sozio-demografische Merkmale der Befragten – Berufsbildung, Erwerbstätigkeit³⁰ und subjektiver Sozialstatus³¹ – unabhängig von deren Verteilung in der Grundgesamtheit geprüft, ob diese sich in Abhängigkeit der Erhebungswelle verändern. Während die Erwerbstätigkeit in *keinem* signifikanten Zusammenhang zum Erhebungszeitpunkt steht, unterscheiden sich die

27 Chi-Quadrat=20,838; df=10; Cramers V=0,074; p(2-seitig)=0,022.

28 Während Zwei-Personen-Haushalte überproportional häufig erfasst wurden, sind Ein- und Mehrpersonenhaushalte (ab drei Haushaltsmitglieder) geringfügig unterrepräsentiert.

29 Zur besseren Übersicht wurden in Abb. 1 einige Kategorien zusammengefasst dargestellt. Das Personen-Nettoeinkommen ist ursprünglich wie folgt kodiert: (1) 0-499 €; (2) 500-799 €; (3) 800-999 €; (4) 1.000-1.499 €; (5) 1.500-1.999 €; (6) 2.000-2.999 €; (7) 3.000-über 7.000 €. Es ergeben sich entsprechend folgende Mittelwerte: Mean GG=3,55; Mean NSP1=3,77; Mean NSP2=3,65; Mean NSP3=3,64.

30 „Sind Sie zurzeit erwerbstätig? Unter Erwerbstätigkeit wird jede bezahlte bzw. mit einem Einkommen verbundene Tätigkeit verstanden, egal welchen zeitlichen Umfang sie hat.“ – voll erwerbstätig (min. 35 h/Woche); teilzeitbeschäftigt (15-34 h/Woche); geringfügig erwerbstätig (bis max. 14 h/Woche); gelegentlich oder unregelmäßig beschäftigt; nicht erwerbstätig (Arbeitslose, Vorruheständler/Rentner/Pensionäre ohne Nebenverdienst).

31 „In unserer Gesellschaft gibt es Bevölkerungsgruppen, die eher oben stehen und solche, die eher unten stehen. Sie sehen hier eine Skala, die von 1 bis 10 reicht, wobei 1 das untere Ende und 10 das obere Ende darstellt. Wenn Sie an sich selbst denken, wo auf dieser Skala würden Sie sich einordnen?“ – 1 (ganz unten)...10 (ganz oben). (Vgl. V450/ALLBUS 1990; V724/ALLBUS 2000. Nähere Informationen zum ALLBUS: <http://www.gesis.org/dienstleistungen/daten/umfragedaten/allbus/> (07.06.2010).

Befragten jeder Erhebungswelle hinsichtlich ihrer Berufsbildung³² und ihrer selbst eingeschätzten Stellung innerhalb der Gesellschaft³³ voneinander. Mit zunehmender Kontaktierung werden vorrangig Personen in die Stichprobe rekrutiert, die eine Berufsbildung auf weniger hohem Niveau absolviert haben und sich stärker im unteren Bereich der Gesellschaft einordnen. In Übereinstimmung mit Blasius/Reuband (1996), Reuband (1999b, 2001) sowie Petermann (2005) konnte also auch hier festgestellt werden, dass durch Mahnaktionen besonders Personen unterer sozialer Schichten bzw. Personen mit geringerer Bildung in die Befragung einbezogen werden. Geht man davon aus, dass postalische Befragungen üblicherweise Verzerrungen in Richtung höher Gebildeter aufweisen (vgl. Schnell/Hill/Esser 1995: 334; Reuband 2001), bedeutet die Rekrutierung dieser Gruppierungen durch Mahnaktionen eine Verbesserung der Stichprobenqualität.

5.3 Auswirkungen auf die Teilnahmebereitschaft von Menschen mit verschiedenen Einstellungen, Eigenschaften, Erfahrungen und Verhaltensweisen

Die im vorangegangenen Abschnitt überwiegend durchgeführten „Repräsentanzabgleiche“ mit offiziellen Daten (Nettovalidierung) können selbstverständlich lediglich als Indiz, nicht aber als Beweis für eine gelungene Abbildung der Grundgesamtheit durch die Zufallsstichprobe gelten (vgl. auch Schnell/Hill/Esser 1995: 287). Zahlreiche weitere Merkmale, die in keinem oder nur sehr schwachem Zusammenhang mit den geprüften Sozio-Demografika stehen, sind für die Charakterisierung der Grundgesamtheit bzw. der Stichprobe von Bedeutung. Deshalb sollen in einem nächsten Schritt die Befragten der einzelnen Erhebungswellen hinsichtlich verschiedener weiterer Merkmale miteinander verglichen werden. Da diese Variablen nicht extern validiert werden können, ist zwar aus einem Vergleich der Befragungsteilnehmer „[...] nicht abzulesen, ob bestehende Verzerrungen ausgeglichen, abgeschwächt, verstärkt oder umgekehrt werden“ (Petermann 2005: 70). Es kann aber festgestellt werden, ob sich die Verteilungen der einzelnen Merkmale in Abhängigkeit des Erhebungszeitpunktes verändern. Dies gibt Hinweise darauf, inwiefern die Ausschöpfungsquote bzw. das Durchführen von Mahnaktionen Konsequenzen für die Verteilung der Merkmale bzw. die Generalisierung der Ergebnisse

32 Die Anteile derer, die eine Fachhochschule (Chi-sq=12,405; df=4; Cramers V=0,056; p(2-seitig)=0,015) oder Hochschule (Chi-sq=5,922; df=2; Cramers V=0,054; p(2-seitig)=0,052) absolviert haben, sinken mit zunehmenden Mahnaktionen.

33 Zusammenhang zwischen subjektivem Sozialstatus und Antwortzeitpunkt: F=4,642; p(2-seitig)=0,010, df=2. Die Korrelation ist schwach negativ (Spearman's Rho=-0,078, p≤0,01).

hat. Ein weiterer Vorteil dieser Vorgehensweise besteht ferner darin, dass mehr als nur die wenigen sozio-demografischen Variablen wie Alter, Geschlecht, Bildung etc. zur Analyse potenzieller Stichprobenveränderungen herangezogen werden können.

Zunächst werden Zusammenhänge zwischen verschiedenen Merkmalsausprägungen und den Nachfassaktionen auf bivariater Ebene betrachtet. Die entsprechenden Analyseresultate sollten aufgrund fehlender Drittvariablenkontrolle allerdings unter Vorbehalt und mit Zurückhaltung interpretiert werden.

Tabelle 2 gibt Hinweise darauf, inwiefern sich aus der Durchführung von Mahnaktionen hinsichtlich delinquenzbezogener Angaben Veränderungen in der Stichprobe ergeben.

Tabelle 2 Delinquenzbezogene Antworten nach Erhebungswellen

	Antworten nach Erstversand	Antworten nach 1. Mahnung	Antworten nach 2. Mahnung	Gesamt	Cramers V/ Spearmans Rho
Delinquentes Verhalten					
1) Versatilität, Lebensspanne (5-14 Deliktarten)	16,2	13,4	12,8	14,9	-0,12**
2) Prävalenz Delinquenz, Lebensspanne (2)	72,1	60,1	59,0	66,8	0,13***
3) Prävalenz Delinquenz, seit 50. Geburtstag (2)	51,1	43,3	39,6	47,3	0,09***
4) Opfer v. Diebst./Raub/Betrug, letzte 5 Jahre (2)	7,8	9,6	10,6	8,7	0,04
5) Kriminalitätsfurcht, Standarditem (3 & 4)	11,8	12,3	10,3	11,7	0,03
6) Allgemeine Gesetzestreue (arithm. Mittel)	8,1	8,1	8,2	8,1	0,02

Anteile in Prozent bzw. arithmetisches Mittel; grau: $p \leq 0,01$ in Klammern: dargestellte Antwortkategorien; ** $p \leq 0,01$; *** $p \leq 0,001$; Cramers V bzw. Spearmans Rho beziehen sich stets auf alle Antwortkategorien der Variablen; zu 6) $F=0,280$; $p(2\text{-seit.})=0,76$; $df=2$

Fragewortlaut/Variablen: 1) Es wurde für 14 Deliktarten abgefragt, ob diese im Leben jemals begangen wurden und ein Summenindex zur Anzahl begangener Deliktarten über die Lebensspanne gebildet. 2) Es wurde für 14 Deliktarten abgefragt, ob diese im Leben jemals begangen wurden und über alle Delikte die Lebenszeitprävalenz berechnet: 1) keines der 14 Delikte jemals begangen, Lebenslang; 2: mindestens eines der 14 Delikte min. einmal verübt, Lebenslang. 3) Es wurde für 14 Deliktarten abgefragt, ob diese seit dem 50. Geburtstag begangen wurden und über alle Delikte die Prävalenz berechnet: 1) keines der 14 Delikte begangen seit dem 50. Geburtstag; 2: mindestens eines der 14 Delikte min. einmal verübt, seit dem 50. Geburtstag. 4) Sind Ihnen die folgenden Ereignisse in den letzten 5 Jahren, also seit Beginn des Jahres 2004, passiert? Ich wurde bestohlen, beraubt oder anderweitig um mein Eigentum betrogen. 1: nein, 2: ja. 5) Wie sicher fühlen Sie sich oder würden Sie sich fühlen, wenn Sie hier in Ihrem Wohngebiet nachts draußen alleine sind? 1: Sehr sicher, 2: Ziemlich sicher, 3: Ziemlich unsicher, 4: sehr unsicher. 6) Bitte geben Sie an, wie sehr Sie persönlich den folgenden Aussagen zustimmen! An Gesetze muss man sich immer halten, auch wenn die eigenen Interessen dabei zu kurz kommen. 1: Stimme überhaupt nicht zu ...10: Stimme voll und ganz zu.

Zwar unterscheiden sich die Befragten in Bezug auf die weniger sensitiven Informationen zu Viktimisierung, Kriminalitätsfurcht und allgemeiner Gesetzestreue nicht signifikant voneinander.³⁴ Angaben zum Schwerpunktthema der Befragung – eigenes kriminelles Verhalten – werden allerdings von den Mahnaktionen beein-

flusst. Es wird offensichtlich, dass sowohl die Versatilität (Deliktbreite) als auch das delinquente Verhalten über die Lebensspanne bzw. seit dem 50. Geburtstag mit zunehmender Kontaktierung abnehmen.³⁵ Unter der Annahme, dass das Auftreten sozial erwünschten Antwortverhaltens vom Antwortzeitpunkt unabhängig ist,³⁶ können die schwächer ausgeprägten Delinquenzangaben als geringerer Bezug der Respondenten späterer Kontaktierungen zum Thema der Befragung interpretiert werden. Damit würde sich hier erneut der Befund früherer Studien bestätigen, wonach thematisch weniger stark interessierte Personen erst durch Nachfassaktionen zur Teilnahme bewegt werden (Reuband 1999b, 2001; Petermann 2005). Die Konsequenzen für das Sample sind eindeutig: hätte man die Befragten lediglich einmal angeschrieben, würde man das kriminelle Verhalten der Population um einige Prozentpunkte³⁷ überschätzen. Zudem liefern die Ergebnisse ein Indiz dafür, dass die Befolgung der Reziprozitätsnorm nicht zwangsläufig mit der Befolgung anderer sozialer (Konformitäts-)Normen zusammenhängt. Anders ausgedrückt: die Vermutung, dass „kriminellere“ Personen in postalischen Befragungen weniger kooperationsbereit sind, also z. B. erst nach (mehrmaliger) Erinnerung antworten, erscheint empirisch unbegründet.³⁸

Tabelle 3 zeigt für weitere „weiche“ Merkmale der Befragten, ob diese von den Erhebungswellen beeinflusst werden. Es ist erkennbar, dass sich die Befragten verschiedener Erhebungswellen hinsichtlich ihres Bedürfnisses nach einfachen Aufgaben (simple tasks), ihrer hedonistischen Orientierung, anomischer Gefühle sowie in Bezug auf ihre Hilfsbedürftigkeit, ihre politische Interessiertheit und Aktivität bzw. Mobilität signifikant voneinander unterscheiden. Die Zusammenhänge sind jedoch stets sehr schwach und teilweise nicht linear. So sind etwa weniger hedonistisch orientierte Befragte sowie der höchste Anteil anomischer Personen in der Stichprobe der 1. Mahnung vertreten. *Keine* Auswirkungen haben die Mahnaktionen auf die Befragtenmerkmale Sorgenanzahl, Lebenszufriedenheit und eigener Lebensstandard³⁹ (vgl. ähnlich Reuband 1999b).

35 Vom Erhebungsdesign unbeeinflusst sind die Anteile derjenigen Personen, die innerhalb der letzten 12 Monate mindestens eines der 14 Delikte verübt haben sowie die Häufigkeiten delinquenten Handelns seit dem 50. Geburtstag bzw. in den letzten 12 Monaten.

36 Die Möglichkeit, dass mit zunehmenden Mahnaktionen der Anteil sozial erwünscht antwortender Personen steigt, kann hier nicht geprüft und somit auch nicht ausgeschlossen werden.

37 Je nach Kriminalitätsmaß würde hier eine Überschätzung um 1,3 bis 5,5 % erfolgen. Diese Werte ergeben sich aus den Differenzen der jeweiligen Anteile „nach Erstversand“ und „gesamt“.

38 Vgl. auch die Ergebnisse der multivariaten Analyse in Tabelle 4.

39 Hinsichtlich drei weiterer reflexiver Wertorientierungen (normorientierte Leistungsethik, religiöse- und subkulturell-materialistische Orientierung), die laut Hermann (2003) in Zusammenhang mit kriminellem Verhalten stehen, sowie hinsichtlich ihres Ausmaßes an Selbstkontrolle, welches nach Gottfredson/Hirschi (1990) kriminalitätsrelevant ist, unterscheiden sich die Befragtenstichproben der einzelnen Erhebungswellen ebenfalls *nicht* signifikant voneinander.

Tabelle 3 Eigenschaften, Wertorientierungen, Einstellungen und Verhaltensweisen nach Erhebungswellen

	Antworten nach Erstversand	Antworten nach 1. Mahnung	Antworten nach 2. Mahnung	Gesamt	Cramers V/ Spearmans Rho
1) Selbstkontrolle: Simple-tasks (arithm. Mittel)	2,6	2,6	2,5	2,6	-0,06**
2) hedonistische Orientierung (arithm. Mittel)	3,6	3,4	3,6	3,6	-0,05*
3) wenige Sorgen (0 & 1)	72,4	74,1	72,5	72,8	0,00
4) starke Anomia (1 & 2)	69,1	76,6	69,8	71,4	0,06*
5) hohe Lebenszufriedenheit (1 & 2)	85,0	81,6	80,8	83,4	0,03
6) guter Lebensstandard (1&2)	87,3	85,7	84,4	86,4	0,04
7) Hilfsbedürftigkeit vorhanden (1)	20,9	26,5	30,7	23,8	0,09***
8) politisches Interesse vorhanden (1-4)	42,8	37,2	36,0	40,2	0,06**
9) hohe Mobilität bzw. Aktivität (1 & 2)	75,0	69,2	65,7	72,0	0,07**

Anteile in Prozent bzw. arithmetisches Mittel; grau: $p \leq 0,05$ in Klammern: dargestellte Antwortkategorien; * $p \leq 0,05$; ** $p \leq 0,01$; *** $p \leq 0,001$; Cramers V bzw. Spearmans Rho beziehen sich stets auf alle Antwortkategorien der Variablen; zu 1) $F=3,241$; $p(2\text{-seit.})=0,04$; $df=2$; zu 2) $F=6,221$; $p(2\text{-seit.})=0,002$; $df=2$.

Fragewortlaut/Variablen: 1) Es wurde die Subskala „Simple tasks“ der Self-Control-Skala nach Grasmick et al. (1993) in der dt. Übersetzung von Seipel/Eifler (2004) anhand von vier Items erhoben. Werte der Antwortkategorien liegen zwischen 1 und 4. Nach gleichsinniger Polung der Items bedeuten niedrigere Werte ein höheres Bedürfnis nach einfachen und weniger komplexen Aufgaben. 2) Es wurde die hedonistische Orientierung als individuell reflexive Wertorientierung (genauer: als Wertdimension innerhalb einer Wertorientierung) nach Klages/Gensicke (1993) und Herbert (1993) (zitiert nach Hermann 2003: 58) anhand von vier Items erhoben und zu einem Index zusammengefasst. Werte der Antwortkategorien liegen zwischen 1 und 6. Niedrigere Werte bedeuten eine schwächere hedonistische Orientierung. 3) Es wurde für acht verschiedene Dinge, die Menschen Sorgen bereiten können abgefragt, wie oft die Befragten daran denken. 1: Nie, 2: Selten, 3: Manchmal, 4: Oft, 5: Sehr oft. Diese Variable wurde so rekodiert, dass die Anzahl der Dinge, an welche Personen oft bzw. sehr oft denken, kategorisiert erfasst wurde: 0: keine Sorge; 1: 1 bis 2 Sorgen, 3: 3 bis 4 Sorgen, 4: 5 bis 8 Sorgen. 4) Stimmen Sie diesen Aussagen (eher) zu oder lehnen Sie diese (eher) ab? Heutzutage ändert sich alles so schnell, dass man oft nicht weiß, woran man sich halten soll. 1: Stimme voll zu, 2: Stimme eher zu, 3: Lehne eher ab, 4: Lehne stark ab. 5) Wie beurteilen Sie Ihr Leben insgesamt? Schätzen Sie ein, inwieweit die folgenden Aussagen auf Sie zutreffen! Die wichtigsten Dinge, die ich im Leben will, habe ich weitestgehend erreicht. 1: Trifft genau zu, 2: Trifft eher zu, 3: Trifft eher nicht zu, 4: Trifft gar nicht zu. 6) Wie bewerten Sie Ihren derzeitigen Lebensstandard? 1: Sehr gut, 2: eher gut, 3: eher schlecht, 4: sehr schlecht. 7) Bräuchten Sie mehr Hilfe bei Arbeiten im Haushalt? 1: Ja, 2: Nein. 8) Wenn Sie einmal an die letzten 12 Monate denken, wie oft gehen Sie da in der Regel den folgenden Tätigkeiten nach? Wie oft besuchen Sie politische Veranstaltungen, z. B. von Parteien, Gewerkschaften oder Bürgerinitiativen? 1: Mehrmals in der Woche, 2: Einmal in der Woche, 3: 1-3mal im Monat, 4: Seltener; 5: Nie. 9) Wie oft sind Sie in Ihrem täglichen Leben, egal aus welchen Gründen, auch außerhalb Ihres unmittelbaren Wohngebietes unterwegs? 1: Täglich, 2: 3-6mal pro Woche, 3: 1-2mal pro Woche, 4: 1-3mal im Monat, 5: seltener.

Da die Effekte der Mahnaktionen auf Einstellungen, Wertorientierungen und ähnliche Variablen schwach oder nicht signifikant sind, unterscheiden sich die Befragten des Erstversands nur minimal von denen der gesamten Befragungsstichprobe.⁴⁰ Insgesamt kann daher festgehalten werden, dass die psychosozialen Befindlichkeiten der Befragten durch mehrmalige Kontaktierungen annähernd reproduziert werden (vgl. auch Reuband 1999b: 96).

Allerdings gibt es Zusammenhänge zwischen Mobilität/Aktivität bzw. Anomia und dem Bildungsniveau: je geringer die Schulbildung ist, desto seltener sind die Personen unterwegs und desto häufiger fühlen sie sich anomisch.⁴¹ Ferner ist die Schulbildung mit dem subjektiven Sozialstatus korreliert ($r=0,32$). Um nun genauer zu erkennen, inwiefern die „weichen“ signifikant unterschiedlichen Merkmale untereinander zusammenhängen bzw. von sozio-demografischen Variablen beeinflusst werden und um zu prüfen, inwiefern sich die einzelnen Befragtenstichproben voneinander unterscheiden, werden nun Berechnungen auf multivariater Ebene durchgeführt. Tabelle 4 zeigt die Ergebnisse von drei logistischen Regressionsmodellen der Merkmale von Befragten verschiedener Erhebungszeitpunkte.⁴²

Auffällig ist zunächst, dass einige Zusammenhänge, die in den bivariaten Analysen signifikant sind, in der multivariaten Betrachtung nicht mehr nachgewiesen werden können: dies trifft auf die Schulbildung,⁴³ sowie die Versatilität und Prävalenz kriminellen Handelns seit dem 50. Geburtstag zu. Ein Alterseffekt, der sich auf bivariater Ebene *nicht* zeigt, ist hingegen nun zu beobachten: entsprechend den Befunden bisheriger Untersuchungen und gemäß der Tendenz, die zumindest auch in vorliegender Studie bereits auftritt (vgl. Abschnitt 5.2), ist die Wahrscheinlichkeit späterer Teilnahme für jüngere Personen höher als für ältere.

In Übereinstimmung mit den bivariaten Analysen zeigen sich in jeweils mindestens einem Modell, welchen Einfluss die Wohnortgröße, der subjektive Sozialstatus, das politische Interesse, die Aktivität sowie die frühere Kriminalität eines Befragten auf die Wahrscheinlichkeit ausüben, dass er erst nach ein- bzw. mehrmaliger Mahnung kooperiert.

Des Weiteren geht aus Tabelle 4 hervor, dass sämtliche Modelle, gemessen anhand des Pseudo- R^2 , eine sehr geringe Erklärungskraft aufweisen. Entsprechend geringfügig sind die systematischen Stichprobenverzerrungen, die sich aufgrund der geprüften Variablen ergeben. Mit anderen Worten: die Teilnehmer verschiedener Erhebungszeitpunkte unterscheiden sich hinsichtlich dieser Merkmale nicht wesentlich voneinander. Modell 2 besitzt dabei die höchste Erklärungskraft, d. h. die Teilnehmer des Erstversands und der 2. Mahnaktion unterscheiden sich noch

41 Spearman's Rho (Mobilität bzw. Aktivität/Bildung)=-0,12** und Spearman's Rho (Anomia/Bildung)=0,33***. Letztere Korrelation stellt zugleich die stärkste aller möglichen Zusammenhängebeziehungen aller in die bivariaten Betrachtungen eingeschlossenen Variablen dar.

42 Um die Modelle nicht zu überladen, wurden lediglich einige bedeutsame Variablen in die Berechnung der logistischen Regressionen einbezogen. Nicht aufgenommene Variablen vorangegangener Analysen sind: Haushaltsgröße, Berufstätigkeit, Anomia, hedonistische Orientierung und Simple Tasks.

43 Scheinbar unterscheiden sich die Befragten verschiedener Mahnaktionen stärker bezüglich der subjektiven Einschätzung ihres Sozialstatus als im Hinblick auf ihr formales Bildungsniveau voneinander.

am stärksten – nämlich in Bezug auf sechs Merkmalsverteilungen – voneinander. Allerdings unterscheiden sich auch die Befragten des Erstversands von der gesamten Befragungsstichprobe in fünf Merkmalen voneinander (Modell 3): die Befragten der gesamten Stichprobe nach Durchführung der Mahnaktionen sind im Vergleich zu Personen, die bereits auf das erste Anschreiben reagieren, jünger, wohnen verstärkt im ländlichen Raum, haben geringeres politisches Interesse (bzw. nehmen zu einem größeren Anteil „nie“ an politischen Veranstaltungen teil) und sind im Verlauf ihres Lebens deutlich weniger kriminell. Am ähnlichsten sind sich die Teilnehmer des Erstversands und der 1. Mahnung (Modell 1).

Tabelle 4 Logistische Regressionen der Respondentenmerkmale

Variablen	Teilnahme nach Erstversand vs. nach 1. Mahnung	Teilnahme nach Erstversand vs. nach 2. Mahnung	Teilnahme nach Erstversand vs. nach Erstversand, 1. oder 2. Mahnung
Geschlecht (Ref.: w)	-0,036 (0,123) [0,965]	0,043 (0,158) [1,044]	-0,012 (0,107) [0,988]
Alter (in Jahren)	-0,137* (0,066) [0,872]	-0,192* (0,084) [0,825]	-0,160** (0,058) [0,852]
Schulbildung (Ref.: Fachabitur/Abitur)			
kein SA/Haupt-/Volksschule	0,250 (0,156) [1,284]	0,057 (0,203) [1,059]	0,190 (0,137) [1,209]
Realschule/ 10. Klasse POS (DDR)	0,022 (0,182) [1,022]	-0,068 (0,235) [0,934]	0,009 (,158) [1,009]
Wohnort (Ref.: Dorf)			
Groß-/Mittelstadt	-0,338* (0,157) [0,713]	-0,605* (0,194) [0,546]	-0,423** (0,136) [0,655]
Kleinstadt	-0,230 (0,152) [0,794]	-0,422* (0,187) [0,656]	-0,290* (0,133) [0,748]
subjektiver Sozialstatus	-0,060 (0,062) [,942]	-0,130! (0,078) [0,878]	-0,083 (0,055) [0,920]
politisches Interesse (Ref.: nie)			
oft/manchmal	-0,465 (0,353) [0,628]	-0,393 (0,430) [0,675]	-0,443 (0,296) [0,643]
selten	-0,112 (0,124) [0,894]	-0,317! (0,165) [0,729]	-0,183! (0,109) [0,833]
Aktivität/Mobilität	0,021 (0,065) [1,021]	0,174* (0,078) [1,190]	0,077 (0,056) [1,080]
Versatilität, Lebensspanne	-0,007 (0,079) [0,993]	-0,135 (0,112) [0,874]	-0,047 (0,070) [0,954]
Prävalenz Delinquenz, Lebensspanne (Ref.: min. ein Delikt verübt, Lebensspanne)	0,597*** (0,195) [1,816]	0,391 (0,242) [1,478]	0,522** (0,168) [1,685]
Prävalenz Delinquenz, seit 50. Geburtstag (Ref.: min. ein Delikt verübt, seit 50. Geburtstag)	-0,187 (,164) [,829]	0,016 (,207) [1,016]	-0,117 (0,141) [0,890]
Konstante	-0,741*** (0,164) [,829]	-1,120*** (0,259) [,326]	-0,228 (0,179) [0,796]
Pseudo-R ² Nagelkerke	0,04	0,07	0,05
Chi-Quadrat (df)	38,20*** (13)	49,83*** (13)	62,01*** (13)
n	1411	1205	1643

grau: $p \leq 0,01$

metrische Variablen z-standardisiert; ! $p \leq 0,1$; * $p \leq 0,05$; ** $p \leq 0,01$; *** $p \leq 0,001$

Angegeben sind die unstandardisierten Koeffizienten; in runden Klammern: Standardfehler; in eckigen Klammern: stand. Effektkoeffizienten.

6 Zusammenfassung und Schlussfolgerungen

Der vorliegende Beitrag untersuchte auf der Grundlage einer regionalen postalischen Befragung älterer Menschen die Auswirkungen von Mahnaktionen auf das Antwortverhalten der Respondenten und die soziale Zusammensetzung der Stichprobe.

Die Untersuchung konnte erneut zeigen, dass die postalische Befragung – sofern bei der Konzeption und Durchführung die grundlegenden Empfehlungen der Total/Tailored Design Method (Dillman 1978, 2000; Dillman/Smyth/Christian 2009) beachtet werden – hinsichtlich realisierbarer Teilnahmequoten und der Datenqualität ähnlich erfolgreich wie andere Erhebungsmodi sein kann. Entgegen teilweise bestehender Skepsis stellt sie somit durchaus eine funktionierende Datenerhebungsmethode dar. Ferner scheint der postalische Modus auch für die Befragung älterer Menschen gut geeignet zu sein⁴⁴ (vgl. ähnlich Brune/Werle/Hippler 1991; Reuband 2006, 2008b, 2008a).

Was die Auswirkungen von Mahnaktionen betrifft, konnte Folgendes festgestellt werden:

1. Mahnaktionen erhöhen die Ausschöpfungsquote einer Befragung beträchtlich. Die von Dillman (1978: 185f.) und Hippler (1988: 245) proklamierte „Faustregel“, dass jede Nachfassaktion in etwa die gleiche Ausschöpfungsrate erzielt, konnte allerdings nicht bestätigt werden (ähnliche Befunde vgl. Kunz 2006: 111f.). Die einzelnen Kontaktierungen erbrachten deutlich unterschiedliche, mit dem Zeitpunkt der Mahnung sinkende Rücklaufraten. Drei Erhebungszeitpunkte erscheinen optimal, eine vierte Kontaktierung würde kaum zur Verbesserung der Stichprobenqualität und zur Steigerung des Rücklaufs beitragen.
2. Entsprechend den Ergebnissen bisheriger Studien zeigte sich auch hier, dass das Ausmaß an Item-Nonresponse (INR) mit den Nachfassaktionen ansteigt. Allerdings wurde kein kontinuierlicher Anstieg der Ausfälle einzelner Antworten, sondern ein kurvilinearere Zusammenhang zwischen Mahnung und INR beobachtet: die nach der 1. Mahnung retournierten Bögen weisen mit 7,8 %

44 Je nach Befragungsthema und Umfang der interessierenden Fragen dürfte die postalische Befragungsvariante aufgrund ihrer Spezifika für ältere Menschen sogar besser als andere Erhebungsmodi geeignet sein. Im Rahmen des Pretest der Südbadener Befragung gaben einige Personen (v. a. die älteren) an, dass sie für die Beantwortung des langen Bogens insgesamt mehrere Stunden benötigt hätten. In solchen Fällen „verteilen“ die Befragten ihre Auskunftserteilung auf verschiedene Tagesabschnitte bzw. sogar auf mehrere Tage. Mittels persönlicher oder telefonischer Befragung hätte die hier angestrebte Datenerhebung wahrscheinlich eine massive Überforderung der Befragten dargestellt und wäre entsprechend schwerer zu realisieren gewesen. Zudem enthielt die Befragung zahlreiche Erinnerungsfragen sowie sensitive Fragen zur eigenen Delinquenz – auch hier hätten die beiden letztgenannten Erhebungsmodi sicher zu einer Einschränkung der Antwortvalidität geführt.

den höchsten INR auf. Der Anteil derjenigen Befragten, die den Fragebogen gemeinsam mit einer anderen Person ausfüllen, nimmt mit zunehmender Anzahl der Kontaktierungen linear zu. Der Anstieg der Antwortausfälle und der gemeinschaftlichen Beantwortung ist jedoch jeweils nicht sehr ausgeprägt. Einerseits deuten diese Befunde darauf hin, dass durch Mahnaktionen Personen, die eher unmotiviert sind und solche, deren Fähigkeit zur selbständigen Teilnahme aus verschiedenen Gründen eingeschränkt ist, verstärkt rekrutiert werden. Dies dürfte zu einer verbesserten Stichprobenrepräsentativität führen. Zum anderen wird klar, dass Mahnaktionen eine geringfügige Verminderung der Datenqualität der Befragungsstichprobe bedeuten.

3. Die soziale Zusammensetzung der Stichprobe verändert sich geringfügig mit jeder Nachfassaktion. Vor allem das Alter, der Wohnort, die Schul- und Berufsbildung sowie der subjektive Sozialstatus der Befragten variieren in Abhängigkeit des Teilnahmezeitpunktes. Durch Mahnaktionen werden überwiegend Gruppierungen in die Befragungsstichprobe rekrutiert, die nach einmaligem Anschreiben unterrepräsentiert sind – nämlich Personen, die jünger sind, im ländlichen Raum wohnen, einen schwächeren Sozialstatus aufweisen sowie geringeres politisches Interesse haben. Mahnaktionen tragen damit zu einer Verbesserung der sozialen Zusammensetzung der Stichprobe, d. h. zur Steigerung der Güte der Erhebungsdaten bei. Zwar bildet die gesamte Befragungsstichprobe im Vergleich zur Stichprobe nach dem Erstversand bzw. nach dem Erstversand und nach der 1. Mahnung die Grundgesamtheit hinsichtlich sozialer Merkmale stets am besten ab. Die Unterschiede zwischen der gesamten Befragungsstichprobe und derjenigen nach dem Erstversand sind allerdings insgesamt recht gering.
4. Die Befragten der einzelnen Erhebungszeitpunkte unterscheiden auch bezüglich verschiedener weiterer Merkmale, wie etwa Einstellungen, Verhaltensweisen, Persönlichkeitseigenschaften und Lebenserfahrungen voneinander. Allerdings sind diese Unterscheide überwiegend schwach. Eine Ausnahme stellt das kriminelle Handeln der Befragten über die Lebensspanne dar: hier zeigte sich, dass mit zunehmenden Kontaktierungen deutlich weniger delinquente Personen in die Befragung einbezogen werden. Unterstellt man, dass die später antwortenden Personen (etwa aufgrund eines geringeren Bezugs zum Befragungsthema) in der Befragungsstichprobe des Erstversands unterrepräsentiert sind, würde die Berechnung kriminalitätsbezogener Parameter der Population auf der Basis der Befragten des Erstversands entsprechend zu spürbaren Überschätzungen führen. Auch hier wird noch einmal deutlich, inwiefern Mahnaktionen zur Verbesserung der Stichprobenqualität beitragen können.

Zusammenfassend deuten die Befunde darauf hin, dass auch postalische Befragungen mit geringen Ausschöpfungsquoten (zumindest im hier untersuchten Bereich, d. h. in Höhe von ca. 30 %) bzw. ohne Mahnaktionen eine hinreichend gute Stichprobenqualität im Sinne der Abbildhaftigkeit sozio-demografischer Merkmale der

Grundgesamtheit aufweisen können. Allerdings unterscheiden sich die Befragten ohne und nach weitere(n) Kontaktierungen in Bezug auf mehr oder weniger bedeutsame Variablen mehr oder weniger stark voneinander. Im günstigsten Fall sind bedeutsame Variablen nicht oder nur in geringfügigem Ausmaß von Antwortdifferenzen betroffen. Um die für statistische Analysen benötigte Fallzahl zu erreichen, wäre es unter diesen Umständen theoretisch gerechtfertigt, eine hinreichend große Brutto-Stichprobe einmalig anzuschreiben, d. h. auf Mahnaktionen zu verzichten. Im ungünstigeren Fall – und dies trifft auf die hier untersuchte Erhebung zu und wurde auch für andere Studien bereits nachgewiesen – stehen die Antworten der Respondenten zum Schwerpunktthema der Befragung in deutlichem Zusammenhang zum Erhebungszeitpunkt. Hier können Nachfassaktionen zur Abschwächung anfänglicher Verzerrungen beitragen.

Da im Vorfeld einer Befragung kaum bestimmbar ist, ob die angeschriebenen Personen unabhängig ihres Interesses am Erhebungsthema nach einmaliger Kontaktierung zu je etwa gleichen Anteilen antworten werden, *erscheint es empfehlenswert, Mahnaktionen prinzipiell in das Erhebungsdesign postalischer Befragungen einzuplanen*. Die Ergebnisse vorliegender Studie legen dabei nahe, dass bereits *eine* Nachfassaktion deutlich zur Verbesserung der Stichprobenzusammensetzung und damit zur Steigerung der Befragungsgüte beiträgt. *Zwei* Mahnaktionen sind in dieser Hinsicht optimal. *Drei* Mahnaktionen scheinen in Anbetracht von Kosten-Nutzen-Abwägungen wenig sinnvoll zu sein.

Indem eine geringere Anzahl von Personen kontaktiert wird, diese jedoch besser ausgeschöpft werden kann, wird zugleich der Pool potenziell zu befragender Personen in der Bevölkerung geschont. Ein solches Design stellt eine sozial verantwortungsbewusste Vorgehensweise dar und ist – da keine nachteiligen Auswirkungen von Mahnaktionen festgestellt werden konnten – somit auch aus forschungsethischer Sicht zu bevorzugen.

Literatur

- Arzheimer, K. und M. Klein, 1998: Die Wirkung materieller Incentives auf den Rücklauf einer schriftlichen Panelbefragung. *ZA-Information* 43: 6-43. http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/za_information/ZA-Info-43.pdf (9.9.2010).
- Atteslander, P., 2008: *Methoden der empirischen Sozialforschung*. Berlin: Schmidt.
- Becker, R., R. Imhof und G. Mehlkop, 2007: Die Wirkung monetärer Anreize auf den Rücklauf bei einer postalischen Befragung und die Antworten auf Fragen zur Delinquenz. *Methoden Daten Analysen* 1 (2): 131-160. www.gesis.org/download.php?url=/fileadmin/upload/forschung/publikationen/zeitschriften/mda/Vol.1_Heft_2/MDA2_07_Becker_Imhof_Mehlkopf.pdf (9.9.2010).

- Biemer, P. P. und L. E. Lyberg, 2003: *Introduction to Survey Quality*. Hoboken, New Jersey: John Wiley & Sons.
- Blasius, J. und K.-H. Reuband, 1996: Postalische Befragungen in der empirischen Sozialforschung. Ausschöpfungsquoten und Antwortqualität. *Planung & Analyse* 96 (1): 35-41.
- Blau, P., 1964: *Exchange and Power in Social Life*. New York: Wiley.
- Brambilla, D. J. und S. M. McKinlay, 1987: A Comparison of Responses to Mailed Questionnaires and Telephone Interviews In a Mixed Mode Health Survey. *American Journal of Epidemiology* 126 (5): 962-971.
- Brune, M., M. Werle und H.-J. Hippler, 1991: Probleme bei der Befragung älterer Menschen. Methodische Erfahrungen aus einer schriftlichen Umfrage zu Tätigkeitsformen im Ruhestand. ZUMA-Nachrichten 28: 73-91. http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten/zn_28.pdf (9.9.2010).
- Converse, P., 1964: The Nature of Belief-Systems in Mass Publics. S. in: D. A. Apter (Hg.): *Ideology and Discontent*. New York, London: The Free Press.
- De Leeuw, E. D., 1992: *Data Quality in Mail, Telephone and Face-to-Face Surveys*. Amsterdam: TT-Publikaties.
- Diekmann, A., 1998: *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*. Reinbek bei Hamburg: Rowohlt.
- Diekmann, K., 2007: *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*. Reinbek bei Hamburg: Rowohlt.
- Dillman, D. A., 1978: *Mail and Telephone Surveys*. New York: Wiley.
- Dillman, D. A., 2000: *Mail and Internet Surveys: The Tailored Design Method*. New York: Wiley.
- Dillman, D. A., J. D. Smyth und L. M. Christian, 2009: *Internet, Mail, and Mixed-Mode Surveys. The Tailored Design Method*. Hoboken; New Jersey: John Wiley & Sons.
- Dommeyer, C. J., 1985: Does Response to an Offer of Mail Survey Results Interact with Questionnaire Interest? *Journal of Market Research Society* 27 (1): 27-38.
- Friedrichs, J., 1990: *Methoden empirischer Sozialforschung*. Opladen: Westdeutscher Verlag.
- Gottfredson, M. R. und T. Hirschi, 1990: *A General Theory of Crime*. Stanford: Stanford University Press.
- Goudy, W., 1978: Interim Response to a Mail Questionnaire: Impacts on Variable Relationships. *Sociological Quarterly* 19 (2): 253-265.
- Grasmick, H., C. Tittle, R. Bursik und B. Arneklev, 1993: Testing The Core Empirical Implications of Gottfredson and Hirschi's General Theory of Crime. *Journal of Research in Crime and Delinquency* 30 (1): 5-29.
- Groves, R., 1989: *Survey Errors and Survey Costs*. New York: John Wiley & Sons.
- Guadagnoli, E. und S. Cunningham, 1989: The Effects of Nonresponse and Late Response on a Survey of Physician Attitudes. *Evaluation & the Health Professions* 12: 318-328.
- Häder, M., 2006: *Empirische Sozialforschung. Eine Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Herbert, W., 1993: *Wandel und Konstanz von Wertstrukturen*. Frankfurt Main u. a.: Peter Lang.
- Hermann, D., 2003: *Werte und Kriminalität. Konzeption einer allgemeinen Kriminalitätstheorie*. Wiesbaden: Westdeutscher Verlag.
- Hippler, H.-J., 1988: Methodische Aspekte schriftlicher Befragungen: Probleme und Forschungsperspektiven. *Planung & Analyse* 88 (6): 244-248.
- Hippler, H.-J. und K. Seidel, 1985: Schriftliche Befragungen bei allgemeinen Bevölkerungstichproben - Untersuchungen zur Dillmanschen „Total Design Method“. ZUMA-Nachrichten 16: 39-56. www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten/zn_16.pdf (9.9.2010).
- Homans, G. C., 1961: *Social Behavior. Its Elementary Forms*. New York: Harcourt, Brace & World.
- Kanuk, L. und C. Berenson, 1975: Mail Surveys and Response Rates: A Literature Review. *Journal of Marketing Research* 12 (Nov.): 440-453.
- Klages, H. und T. Gensicke, 1993: *Erläuterung der Speyerer Ziele und Methodik der Wert-erfassung*. Speyer. unveröff. Manuskript.

- Koch, A., 1998: Wenn „mehr“ nicht gleichbedeutend mit „besser“ ist: Ausschöpfungsquoten und Stichprobenverzerrungen in allgemeinen Bevölkerungsumfragen. ZUMA-Nachrichten 42: 66-93. http://www.gesis.org/download.php?url=/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten/zn_42.pdf (9.9.2010).
- Koenig, D., G. Martin und L. Seiler, 1977: Response Rates and Quality of Data: A Reexamination of the Mail Questionnaire. *Canadian Review of Sociology and Anthropology* 14: 432-438.
- Kunz, F., 2006: Täter-Opfer-Ausgleich im Jugendbereich - wie erleben und bewerten ihn die Konfliktbeteiligten? Eine schriftliche Befragung von Tätern und Opfern aus drei sächsischen Schlichtungsstellen. Institut für Soziologie, TU Dresden, unveröff. Diplomarbeit.
- Kunz, F., 2007: Im Osten was Neues: Täter-Opfer-Ausgleich aus Sicht der Beteiligten. Ergebnisse einer Befragung von jugendlichen bzw. heranwachsenden Tätern und ihren Opfern. *Monatsschrift für Kriminologie und Strafrechtsreform* 90 (6): 466-483.
- Lamnek, S. und R. Trepl, 1991: Die Nichtteilnahme an sozialwissenschaftlichen Befragungen. *Planung & Analyse* 6: 205-211.
- Leslie, L. L., 1972: Are High Response Rates Essential to Valid Surveys? *Social Science Research* 1 (3): 323-334.
- Mehlkop, G. und R. Becker, 2007: Zur Wirkung monetärer Anreize auf die Rücklaufquote in postalischen Befragungen zu kriminellen Handlungen. *Methoden Daten Analysen* 1 (1): 5-24. www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/mda/Vol.1_Heft_1/MDA1_Mehlkop_Becker.pdf (9.9.2010).
- Niedermayer, O., 1989: Zur Übertragung der „Total Design Method“ auf schriftliche Befragungen in der Bundesrepublik. S. 332-348, in: J. W. Falter, H. Rattinger und K. G. Troitzsch (Hg.): *Wahlen und politische Einstellungen in der Bundesrepublik Deutschland*. Frankfurt/Main: Lang.
- Petermann, S., 2005: Rücklauf und systematische Verzerrungen bei postalischen Befragungen. Eine Analyse der Bürgerumfrage Halle 2003. *ZUMA-Nachrichten* 29 (57): 56-78. http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten/zn_57.pdf (9.9.2010).
- Platek, R. und C.-E. Särndal, 2001: Can a Statistician Deliver? *Journal of Official Statistics* 17 (1): 1-20.
- Prüfer, P. und M. Rexroth, 2000: Zwei – Phasen – Pretesting. *ZUMA-Arbeitsbericht 2000/08*. www.gesis.org/download.php?url=/fileadmin/upload/forschung/publikationen/gesis_reihen/zuma_arbeitsberichte/00_08.pdf (9.9.2010).
- Reuband, K.-H., 1999a: Anonyme und nicht-anonyme postalische Bevölkerungsbefragungen. *Planung und Analyse* (1): 56-59.
- Reuband, K.-H., 1999b: Postalische Befragung in den neuen Bundesländern. Durchführungsbedingungen, Ausschöpfungsquoten und soziale Zusammensetzung der Befragten in einer Großstadtstudie. *ZA-Informationen* 45: 71-99. http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/za_information/ZA-Info-45.pdf (9.9.2010).
- Reuband, K.-H., 2001: Möglichkeiten und Probleme des Einsatzes postalischer Befragungen. *KZfSS* 53 (2): 307-333.
- Reuband, K.-H., 2006: Postalische Befragung alter Menschen. Kooperationsverhalten, Beantwortungsstrategien und Qualität der Antworten. *ZA-Informationen* 59: 100-127. www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/za_information/ZA-Info-59.pdf (9.9.2010).
- Reuband, K.-H., 2008a: Einsam und unglücklich im Alter? Eine empirische Studie zur Lebenssituation alter Menschen im großstädtischen Kontext. S. 354-374, in: A. Groenemeyer und S. Wieseler (Hg.): *Soziologie sozialer Probleme und sozialer Kontrolle* Wiesbaden: VS Verlag Sozialwissenschaften.
- Reuband, K.-H., 2008b. *Mail Surveys Among the Very Old. A Study of Retrospective Accounts of the Nazi Past*. Paper präsentiert auf der 62. Jahrestagung der American Association for Public Opinion Research (AAPOR), New Orleans.

- Schneekloth, U. und I. Leven, 2003: Woran bemißt sich eine „gute“ allgemeine Bevölkerungsumfrage? Analysen zu Ausmaß, Bedeutung und zu den Hintergünden von Nonresponse in zufallsbasierten Stichprobenerhebungen am Beispiel des ALLBUS. ZUMA-Nachrichten 53: 16-57. http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten/zn_53.pdf (9.9.2010).
- Schnell, R., 1997: Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen. Opladen: Leske + Budrich.
- Schnell, R., P. B. Hill und E. Esser, 1995: Methoden der empirischen Sozialforschung. München, Wien: Oldenbourg.
- Schnell, R., P. B. Hill und E. Esser, 2008: Methoden der empirischen Sozialforschung. München, Wien: Oldenbourg.
- Scott, C., 1961: Research on Mail Surveys. *Journal of the Royal Statistical Society*. 124 (2): 143-205.
- Seipel, C. und S. Eifler, 2004: Gelegenheiten, Rational-Choice und Selbstkontrolle. Zur Erklärung abweichenden Handelns in High-Cost- und Low-Cost-Situationen. S. 288-316, in: D. Oberwittler und S. Karstedt (Hg.): *Soziologie der Kriminalität*. Wiesbaden: VS Verlag Sozialwissenschaften.
- Sobal, J. und K. C. Ferentz, 1989: Comparing Physicians' Responses to the First and Second Mailings of a Questionnaire. *Evaluation & the Health Professions* 12 (3): 329-339.
- Stoop, I., 2005: *The Hunt for the Last Respondent. Nonresponse in Sample Surveys*. The Hague: Social and Cultural Planning Office.
- Toops, T. A., 1926: The Returns for Follow-Up Letters to Questionnaires. *Journal of Applied Psychology* 10: 92-110.
- Yammarino, F. J., S. J. Skinner und T. L. Childers, 1991: Understanding Mail Survey Response Behavior. *Public Opinion Quarterly* 55: 613-639.
- Yu, J. und H. Cooper, 1983: A Quantitative Review of Research Design Effects on Response Rates to Questionnaires. *Journal of Marketing Research* 20: 36-44.

Anschrift der Autorin Franziska Kunz
Max-Planck-Institut für ausländisches und
internationales Strafrecht
Kriminologische Forschungsgruppe
Günterstalstr. 73
79100 Freiburg i. Br.
f.kunz@mpicc.de

Rezensionen



MARTIN WEICHBOLD,
JOHANN BACHER &
CHRISTOF WOLF
(Hg.), 2009:
Umfrageforschung.
Herausforderungen
und Grenzen.
Sonderheft 9/2009
der Österreichischen
Zeitschrift für
Soziologie, ISBN
978-3-531-16319-2,
579 Seiten, 59,90 EUR.

Der Titel des Bandes „Umfrageforschung. Herausforderungen und Grenzen“ legt nahe, dass das Buch ein Kompendium des state-of-the-art der Umfrageforschung darstellt. Dies ist jedoch nicht der Fall. Die Herausgeber legen vielmehr eine umfangreiche Sammlung von Beiträgen zu Problemen und Neuentwicklungen in der Umfrageforschung aus zwei Tagungen vor. Auf 580 (!) Seiten sind 26 (!) Aufsätze versammelt, die in die sieben Blöcke Methodeneffekte, computerbasierte Befragungen, Stichproben, spezielle Zielgruppen und ihre Inklusion, Herausforderungen der Großforschung, alternative Befragungstechniken und methodologische Grundlagen gegliedert wurden. Trotz der Tatsache, dass es sich ‚nur‘ um einen (Doppel-) Tagungsband handelt, finden sich einige Beiträge im Sinne der oben formulierten Erwartung, die in systematisierender Absicht spezielle Problemstellungen der Umfrageforschung diskutieren und mit Analysen belegen.

Im Block *Methodeneffekte* werden die Einflüsse unterschiedlicher neuerer Erhebungsmodi, insbesondere der (Mobil)telefon- und der web-basierten Erhebung, auf die Teilnahmebereitschaft und Ergebnisverzerrungen untersucht.

Spannend sind die Untersuchungen zu unterschiedlichen innovativen Gestaltungsformen *computerbasierter Befragungen*. Beispielsweise werden Effekte des Einsatzes eines virtuellen Interviewers im Vergleich zu einem audio-unterstützten und einem klassischen selbstadministrierten computerbasierten Interview (CASI) diskutiert. In anderen Beiträgen wird mit unterschiedlichen multimedialen Gestaltungsformen von Antwortskalen experimentiert und die Möglichkeit erörtert, aus einem Online-Panel eine geschichtete Zufallsstichprobe zu ziehen.

Im Block *Stichproben* werden u. a. die Besonderheiten des innovativen Stichprobenplans des European Social Survey erörtert. Außerdem finden sich Analysen zu den rückläufigen Ausschöpfungsquoten von Bevölkerungsumfragen und zur Teilnahme an drop-off-Erhebungen sowie eine Konzeption zur Ergänzung einer Festnetz-Telefonstichprobe um eine Mobiltelefonstichprobe. Besonders wichtig erscheint mir der abschließende Beitrag von Johann Bacher in diesem Block: Er macht darauf aufmerksam, dass Umfragedaten in der Regel komplexe Zufallsstichproben darstellen; die inferenzstatistische Datenauswertung erfolgt aber meist unter der Annahme einer einfachen Zufallsauswahl, obwohl zum Teil bereits geeignete Verfahren existieren.

Im Block *spezielle Zielgruppen und ihre Inklusion* werden u. a. die Besonderheiten der Einbeziehung von Migranten diskutiert. Wenig erkenntnisreich ist dabei leider ein Beitrag zur Vorgehensweise im österreichischen Mikrozensus, wohingegen ein weiterer Beitrag systematisch die besonderen Problemlagen der quantitativen Befragung von Migranten untersucht. Ein Beitrag zur Inklusion von Personen mit gesundheitlichen Beeinträchtigungen diskutiert dieses Thema eher oberflächlich auf der Basis einer Studierendenumfrage. Die Analyse des Item-Nonresponse bei Alten und Hochbe-

tagten stellt dagegen ein Musterbeispiel für gelungene Methodenforschung dar, die systematisch an den Forschungsstand anschließt und diesen erweitert.

Im Block *Herausforderungen der Großforschung* wird zunächst – angesichts der Thematik recht langatmig – über den Stand der Harmonisierung europäischer Betriebsbefragungen über offene Stellen informiert. Spannend, aber gerade hinsichtlich der methodischen Aspekte zu knapp konzipiert ist ein Beitrag zur Messung von Globalisierung und seiner Anwendung in der Umfrageforschung; insbesondere bleibt die Frage offen, inwieweit sich Einflüsse der Globalisierung von anderen zeitveränderlichen Einflüssen trennen lassen. Außerdem wird in einem weiteren Beitrag gezeigt, dass die Verteilung von Proxy-Interviews im deutschen Mikrozensus in engem Zusammenhang zu Merkmalen der Sozialstruktur stehen, so dass die Wahrscheinlichkeit systematischer Fehler durch Proxy-Interviews für verschiedene Bevölkerungsteile unterschiedlich sein wird.

Im Block *Alternative Befragungstechniken* zeigt der erste Beitrag, dass die subjektive Wahrnehmung der Zeichen öffentlicher Unordnung durch Befragte dann stärker vom Ergebnis einer systematischen Beobachtung abweicht, wenn die Befragten selbst bereits direkt oder indirekt Opfer einer Straftat wurden. Weniger alternative Befragungsdenn Erhebungstechnik mit in erster Linie psychologischen Anwendungsmöglichkeiten stellt das sogenannte ambulatoische Assessment dar, dem ein weiterer Beitrag gewidmet ist. Hierbei wird die Messung bspw. subjektiver Befindlichkeiten mit Hilfe von Handheld-Computern in den Alltag integriert; eine andere Möglichkeit ist die kontinuierliche Erhebung von Körperfunktionen zur Messung physischer Aktivität. Zu kurz kommen in dem Beitrag allerdings die forschungsethischen Aspekte der Methode. In einem weiteren Beitrag soll die Leistungsfähigkeit der quasi-experimentellen Frageform herausgestellt werden. Interessant,

aber formal anspruchsvoll liest sich ein Aufsatz über randomisierte Befragungsdesigns zur Verringerung von Falschantworten und Antwortausfällen bei heiklen Fragen.

Im abschließenden Block werden *Methodologische Grundfragen* behandelt, wie z. B. die Frage nach der Abhängigkeit der Ergebnisse der Umfrageforschung von situativen Gegebenheiten; die Autoren diskutieren theoretisch und empirisch die Auswirkung von räumlichen und zeitlichen Aspekten sowie der individuellen Befindlichkeit auf Einschätzungs- und Bewertungsfragen. In weiteren Beiträgen wird die demokratiepolitische Funktion des Eurobarometers kritisch hinterfragt und ein systematisierender Überblick über die Möglichkeiten zur Messung der Qualität von Umfragen gegeben.

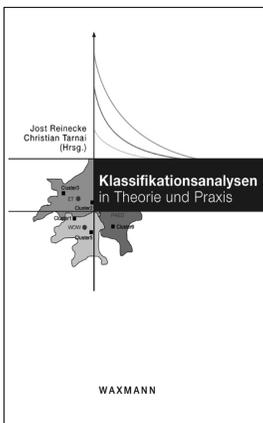
Wie bereits die Herausgeber in ihrer Einleitung bemerken, wird nicht nur im Umfang, sondern auch in den abgehandelten Themen die Vielfalt der Aspekte deutlich, unter denen Umfrageforschung zu diskutieren ist. Aufgrund dieser Vielfalt hat die Platzierung als Zeitschriften-Sonderband in der Tat ihre Berechtigung. Allerdings wird die Vielfalt auch zur Folge haben, dass der Rezensent und die Herausgeber die einzigen sein dürften, die diesen Band „durchlesen“ werden.

Leider gilt die Variationsbreite nicht nur für die Themen, sondern auch für die Güte der Beiträge. Wie gesagt werden in einigen Beiträgen Themen von übergeordnetem Interesse sehr systematisch diskutiert. In den meisten Beiträgen wird sehr solide Methodenforschung berichtet, deren Ergebnisse in vielen Fällen auch über das spezielle Forschungsgebiet hinaus verallgemeinerbare Erkenntnisse liefern können. Viel zu oft findet man jedoch auch zum Zwecke der Publikation aufbereitete Methodenberichte spezieller Erhebungen, deren Ergebnisse nicht systematisch auf ein allgemeines Erkenntnisinteresse des Fachpublikums hin diskutiert werden. Der nahezu unlesbare Beitrag zur quasi-experimentellen Frageform zeigt auch, dass eine systematisierende Absicht

des Autors nicht unbedingt zu einem systematischen und verständlichen Argument führt. Viele Texte sind angesichts der Botschaft etwas zu lang geraten. Ärgerlich sind einige – von den jeweiligen Autoren zu verantwortende – redaktionelle Unzulänglichkeiten, wie z. B. unvollständige Literaturverzeichnisse oder überdimensionierte Tabellen, die durch unnötige Seitenumbrüche sehr schwer lesbar werden.

Insgesamt gesehen präsentieren die Herausgeber aber eine beachtliche Vielfalt mehrheitlich guter Beiträge auf der Höhe des jeweiligen Forschungsstands; eine restriktivere Auswahl und eine intensivere Qualitätskontrolle hätten dem Band aber gut getan.

JOHANNES STAUDER, HEIDELBERG



JOST REINECKE & CHRISTIAN TARNAI (Hg.), 2008: Klassifikationsanalysen in Theorie und Praxis. Münster u. a.: Waxmann. ISBN 978-3-8309-2083-0, 248 Seiten, 29,80 EUR.

Das vorliegende Buch ist der zweite Band der Herausgeber, der auf der Grundlage von oder zumindest der Anregung durch Arbeitstagungen zur „Angewandten Klassifikationsanalyse“ entstanden ist. Ziel dieser Tagungen war es, „...für ein Forschungsproblem verschiedene Verfahren alternativ einzusetzen, um so die Bezüge zwischen den Verfahren und ihre Bedeutung für die Bearbeitung der Fragestellung deutlich werden zu lassen“ (Vorwort der Herausgeber).

In der Mehrzahl der Artikel wird entsprechend über die Erfahrungen bei alternativen Analysen zum jeweils gleichen Forschungsproblem berichtet. Die Alternativen beim Vorgehen beziehen sich dabei nicht nur auf die Verfahrenswahl, sondern z. T. auch auf unterschiedliche Daten und – dies ist ein besonderer Schwerpunkt des vorliegenden Bandes – auf die Analyse von Mischverteilungs-Modellen, bei denen die jeweiligen Analysemodelle unter Beachtung möglicherweise verdeckter Heterogenität gesondert innerhalb gleichzeitig ermittelter latenter Klassen untersucht werden.

Mit dieser Zielsetzung ist auch die Vielfalt der Themen angedeutet, die sich in diesem Band finden. Sie umfassen nicht nur Klassifikationsanalysen im herkömmlichen Sinne, die sich mit der Suche nach einer Klasseneinteilung von Personen oder Objekten (Merkmalsträger) unter möglichst optimaler Zuordnung einander besonders ähnlicher Objekte zu denselben Klassen bzw. einander relativ unähnlicher Objekte zu verschiedenen Klassen befassen. In etwas unscharfer Verwendung werden vielmehr sehr unterschiedliche Verfahren unter demselben Begriff zusammengefasst, die nur strukturell ähnlichen Methoden folgen. Dabei geht es um deterministische (feste) wie auch probabilistische Zuordnungen der Objekte zu Klassen, um die Zuordnung der Merkmalsträger anhand der sie beschreibenden Merkmale wie auch die Zuordnung der Merkmale anhand ihrer Ausprägungen über alle durch sie beschriebenen Merkmalsträger (früher als R- bzw. Q-Analyse beschrieben). In einigen Fällen wird auch die Zuordnung der Objekte zu Klassen (zunächst) nur implizit verfolgt: Die Objekte werden z. B. durch eine multidimensionale Analyse mit metrischen (MDS) oder nicht-metrischen Daten (NMDS) in einem Raum mit wenigen (meist 2) Dimensionen dargestellt. Ihre grafische Darstellung erlaubt dann die Wahrnehmung mehr oder weniger deutlich voneinander abgegrenzter Punktwolken („Klassen“).

Das geschieht hier mit z.T. innovativen Mitteln der Visualisierung. So untersuchen Stefan Ryf und Damian Läge die Beurteilungen (a) von 200 vorgegebenen Musikinterpreten („Objekte“) durch 203 Teilnehmer einer Web-Befragung bzw. (b) von 15 Erfrischungsgetränken („Objekte“) durch 24 Personen. Die Unterschiede (Distanzen) zwischen den Beurteilungen aller Paare von Objekten werden einer NMDS unterzogen. Daraus wird eine grafische Darstellung der Objekte in 2 Dimensionen (als Punktwolke) erzeugt. Eine Klassifikation der Objekte ist darin nur indirekt und undeutlich über Bereiche mit einer relativ dichten Besetzung erkennbar. In einem zweiten Schritt färben (bzw. schattieren) sie die Umgebung der Punkte („Ausstrahlung“) deshalb nach Maßgabe ihrer positiven bzw. negativen Beurteilungen. Dieses erfordert allerdings extern zu begründende Festlegungen der Weite und der Form der Ausstrahlung, führt dann aber (bei hinreichender Gruppierung der Punkte) zu deutlich anhand der Farbe oder Schattierung hervorgehobenen Bereichen zusammengehörender Objekte („Klassen“) und ihrer Trennung von anderen Bereichen. Die meisten der übrigen Artikel folgen jedoch der oben zitierten Vorgabe der Herausgeber zur vergleichenden Analyse.

Hans-Georg Wolff und Johann Bacher untersuchen Befragungsskalen, bei denen zumindest einzelne Items verschiedenen Dimensionen gleichermaßen zuzuordnen sind. Sie gehen dabei zunächst von künstlich erzeugten Datensätzen mit „bekannter Struktur der Dimensionen“ aus. Sie vergleichen sodann, inwieweit die bekannte Struktur der Daten mit verschiedenen Verfahren wie Hauptkomponenten-Analyse, konfirmatorischer Faktorenanalyse und multidimensionaler Skalierung aufgedeckt werden kann.

Andere Beiträge untersuchen Eigenschaften oder Leistungen von Klassifikationsverfahren (im weiteren Sinne, s.o.) vor allem durch die vergleichende Anwendung auf unterschiedliche, denselben oder einen ähnlichen Themenbereich betreffende Datensätze:

Damian Läge und Katharina Schlatter vergleichen zwei Klassifikationen der etwa 200 in der Schweiz vorkommenden Vogelarten. (a) Die Daten der ersten Klassifikation beruhen auf den Urteilen von 19 Ornithologen, die anhand eines vorgegebenen Merkmalskatalogs zur Beschreibung von Nahrung und Habitat abgegeben wurden („paralleles Sortieren“). (b) Die Daten der zweiten Klassifikation – ebenfalls zu den Merkmalsbereichen Nahrung und Habitat – wurden durch weitgehend automatische Textanalyse aus 5 Sachbüchern zur Vogelkunde extrahiert und auf eine hier nicht zu diskutierende Weise kodiert (Trigramm-Kodierung). Beide Datensätze werden in diesem Fall gleichermaßen mit nonmetrischer multidimensionaler Skalierung (NMDS) analysiert und in 2 Dimensionen dargestellt. Die Vergleiche konzentrieren sich hier also in erster Linie auf die sehr unterschiedlichen Datensätze mit ihren ungleichen Chancen zur Erhebung sowie ihren Kosten in sehr unterschiedlicher Höhe.

Nadja Pfuhl und Christian Tarnai versuchen, die befragten Studierenden aus 3 Studiengängen (Elektrotechnik, Pädagogik, Wirtschaftswissenschaften) anhand ihrer Selbst- und Fremdbilder den in diesem Fall bekannten 3 Klassen (Studiengänge) zuzuordnen. Sie tun dies zunächst unter Vergleich zwischen (a) einer Diskriminanzanalyse und (b) einer logistischen Regression, wobei die Ergebnisse weitgehend übereinstimmen. Auf Basis der erstgenannten Analyse (a) wird dann nur mit den dort korrekt klassifizierten Studierenden eine weitere Diskriminanzanalyse durchgeführt. Schließlich werden die Funktionswerte der zweiten Diskriminanzanalyse auch noch einer hierarchischen Klassifikation mit 6 Klassen unterzogen. Es wird nicht ganz deutlich, inwieweit sich bei dieser sequentiellen Anwendung der Verfahren mögliche Fehler zwischen den Schritten neutralisieren oder verstärken. Eine Prüfung findet jedoch auf der Ebene inhaltlicher Plausibilität sowohl zwischen den ermittelten 6 Klassen und den

ursprünglichen 3 Studiengängen als auch zwischen den Klassifikationen auf der Basis von Selbst- und Fremdbildern statt.

Alina Pöge und Jost Reinecke untersuchen die Delinquenzverläufe von Jugendlichen über 4 Panelwellen. Als Datengrundlage dient einmal der Datensatz mit vollständigen Informationen über alle 4 Wellen, zum anderen werden zum Vergleich ergänzte Datensätze herangezogen, bei denen fehlende Werte der ersten Welle für die wichtigsten Analyse-Merkmale mithilfe multipler Imputation (10 ergänzte Datensätze) erzeugt wurden. Zusätzlich werden die Delinquenzmuster über die 4 Wellen auch mit einer Latent Class Analysis (LCA) auf mögliche systematische Unterschiede zwischen den latenten Klassen untersucht. Dazu aber mehr in den folgenden Abschnitten.

Wie bereits eingangs erwähnt, bilden Analysen mit Mischverteilungs-Modellen einen besonderen Schwerpunkt des vorliegenden Bandes. Dabei werden Mess- oder (in weiterem Sinne) Regressions-Modelle mit Hilfe der Latent Class Analysis (LCA) auf „unbeobachtete Heterogenität“ geprüft.

Wenn statistische Verfahren einheitlich auf ganze Populationen oder Auswahlen daraus angewandt werden, steht im Hintergrund immer die Annahme, dass die analysierten Populationen homogen seien im Hinblick auf die untersuchten Sachverhalte: Ansonsten würde ihre einheitliche Beschreibung durch Lage- oder Verteilungsparameter, durch die Parameter eines Regressionsmodells usw. möglicherweise in die Irre führen.

Bereits bestehende Vermutungen über Heterogenitäten und die Verfügbarkeit von Daten zu deren Beschreibung führen unter solchen Umständen regelmäßig entweder zur Schichtung der Population mit getrennter Auswertung der Teilpopulationen oder zur expliziten Einführung von Interaktionstermen, bei denen zumindest eines der beteiligten Merkmale die Trennung der Teilpopulationen kennzeichnet. Da man in solchen Fällen die vermutete Heterogenität

der Population theoriegeleitet und explizit anhand verfügbarer Daten behandelt, kann man auch von „beobachteter“ Heterogenität sprechen.

Bei der Analyse „unbeobachteter Heterogenität“ dagegen geht man a priori nicht von einer Kenntnis und/oder Verfügbarkeit von Merkmalen aus, mit denen die möglicherweise heterogene Gesamtpopulation in jeweils homogene Teilpopulationen zerlegt werden könnte. Statt dessen werden mit Hilfe von inzwischen auch allgemein verfügbaren Programmen (a) die theoretisch vorgegebenen Mess- oder Regressionsmodelle geschätzt und gleichzeitig (b) wird nach einer Klasseneinteilung der Population mit dem Ziel gesucht, dass die Modelle innerhalb der latenten Klassen die Eigenschaften der Mitglieder hinsichtlich der modellierten Zusammenhänge möglichst gut (und d. h. auch: einheitlich) erfüllen. Im vorliegenden Band wird über Erfahrungen bei der Anwendung solcher Mischverteilungs-Modelle aus verschiedenen Untersuchungsbereichen berichtet.

Nicht ganz bzw. nur unter methodisch-instrumentellen Gesichtspunkten gehört der Artikel von Rainer Alexandrowicz in diese Reihe, der mit Hilfe von Simulationsläufen Quasi-Wahrscheinlichkeits-Verteilungen für das „Bayesian Information Criterion“ (BIC) erzeugt. Beim BIC handelt es sich um ein häufig genutztes Maß zur Beurteilung von Mischverteilungs-Modellen. Mit den erzeugten „Wahrscheinlichkeits-Verteilungen“ wird die Möglichkeit eröffnet, auch zwischen Modellen mit relativ geringen Unterschieden zu unterscheiden. Dies erleichtert die Entscheidung zwischen alternativen Modellen und, so eine besonders häufige Verwendung, zwischen Mischverteilungs-Ansätzen mit unterschiedlicher Zahl latenter Klassen.

Im schon genannten Artikel von Pöge und Reinecke wird die Latent Class Analysis (LCA) dazu genutzt, systematische Unterschiede zwischen Teilpopulationen bei der Beschrei-

bung typischer Muster der Delinquenzverläufe Jugendlicher über die 4 Panelwellen aufzudecken. Die latenten Klassen werden also unter dem Gesichtspunkt ermittelt, dass sich die Mitglieder der jeweils selben Klasse im zeitlichen Muster ihres delinquenten Verhaltens möglichst ähnlich sein sollen.

Mit einem anderen Datensatz modelliert Jost Reinecke in einem weiteren Artikel den Delinquenzverlauf von Jugendlichen anhand von „Wachstumsmodellen“ und verwendet auch dort die LCA im Rahmen eines Mischverteilungs-Modells zur Ermittlung der (heterogenen) Teilpopulationen. Er verwendet dazu ein Mischverteilungs-Modell, bei dem für jede der latenten Klassen ein Wachstumsmodell mit festen Parametern geschätzt wird. Besondere Anpassung in den Verteilungsannahmen erfordert in diesem Zusammenhang der hohe Anteil nicht (bzw. nie) delinquenten Jugendlicher.

Einen anderen Mischverteilungs-Ansatz mit Wachstumsmodellen verfolgen Herbert Matschinger, Reinhold Kilian und Matthias C. Angermeyer. Hier geht es um die zeitliche Entwicklung der Zustände von schizophrenen Patienten. Die Zustände werden durch zwei Faktoren beschrieben, die soziale Funktionsfähigkeit (SOFAS) und die subjektive Lebensqualität (SF36), die beide durch bewährte Skalen gemessen werden. Zusätzlich zur Entwicklung über die Zeit wird ein Einfluss von SOFAS auf SF36 vermutet. Die Autoren diskutieren unterschiedliche Modelle der Entwicklung beider Eigenschaften, sowohl einzeln wie auch im Zusammenhang. Versuche der Aufklärung der durch starke Streuungen belasteten Modell-Parameter werden sowohl mit beobachteten (hier: demografischen) Merkmalen als auch durch latente Klassen unternommen. Vor allem durch die sequentielle Einführung der Modifikationen, ihre plausiblen Begründungen und die Diskussionen der Veränderungen im Ergebnis werden wertvolle Anwendungserfahrungen vermittelt.

Angelika Glöckner-Rist und Fred Rist nehmen in ihrem Artikel die teilweise inkonsistenten Ergebnisse zur dimensional Struktur eines häufig genutzten Messinstruments für hypochondrische Einstellungen und Überzeugungen auf (Whiteley-Index). Sie versuchen, diese Widersprüche einmal durch die Wahl geeigneter Messmodelle, zum anderen im Rahmen von Mischverteilungs-Ansätzen durch die getrennte Schätzung der Messmodelle in gleichzeitig ermittelten latenten Klassen zu lösen. Sehr informativ ist auch hier die mehrfache Modifikation des kombinierten Modells und die Diskussion der jeweiligen Veränderungen im Ergebnis.

Ebenfalls einer im Kern messtheoretischen Fragestellung folgen Ferdinand Keller, Thomas D. Meyer und Patrick Pössel in ihrem Aufsatz zur Allgemeinen Depressionsskala (ADS). In früheren Untersuchungen waren Polungsartefakte vermutet worden unter der Annahme, dass bei Fragen mit unterschiedlicher Polungsrichtung der Antworten ein solcher Wechsel von einigen Personen nicht erkannt wird. Die Autoren untersuchen dieses Problem mit unterschiedlichen Messverfahren und insbesondere auch mit Mischverteilungs-Modellen und dem Versuch, unter den latenten Klassen auch (mindestens) eine zu finden, deren Mitglieder aufgrund der nicht erkannten Umpolung als „nicht skalierbare Gruppe“ ausgeschlossen werden müssten. Das Problem erweist sich aber als (noch) komplexer, und es können selbst inhaltliche Gründe (für eine beginnende Depression) nicht völlig bei der Erklärung des abweichenden Antwortverhaltens zwischen den latenten Klassen ausgeschlossen werden.

Durch die kurze Aufzählung der Inhalte wird nochmals die Vielfalt der behandelten Themen wie auch die Schwerpunktsetzung im Bereich der Mischverteilungs-Modelle deutlich. So stellt sich abschließend die Frage nach den Zielgruppen, an die sich das Buch vor allem richtet. Negativ lässt sich diese Frage relativ leicht beantworten: Das Buch bietet keine Einführungen, nicht in die Klassifikationsanalyse, nicht in die zur

Aufklärung der jeweiligen Variablen-Zusammenhänge gewählten Analyse-Modelle und auch nicht in die mit ihrer Hilfe untersuchten inhaltlichen Problemstellungen. Andererseits wendet sich das Buch auch nicht ausschließlich an Spezialisten auf den jeweiligen Gebieten.

Viel breiter, als der relativ eng auf Klassifikationsanalysen konzentrierte Buchtitel vermuten lässt, vermittelt das Buch vielfältige Anwendungserfahrungen, die meist über die jeweils konkreten Problemstellungen und die dabei gewählten Verfahren hinausreichen. Wenn man sich auf diesen Gesichtspunkt der allgemeinen Vermittlung von Anwendungserfahrungen einlässt, dann wird man zunehmend fasziniert von einer Vorgehensweise, die sich über unterschiedliche Inhalte hinweg und mit verschiedenem methodischen Instrumentarium sehr sorgfältig um die Aufklärung von Leistungen, Beschränkungen und gegenseitigen Ergänzungen verschiedener Verfahren bemüht.

So werden Zusammenhangs-Modelle, seien sie nun mit eher „messtheoretischer“ oder eher „kausal-erklärender“ Absicht eingeführt, in der Mehrzahl aller Artikel Schritt für Schritt einer immer weitergehenden Prüfung unterzogen. Manchmal geschieht dies über den Vergleich der Ergebnisse unterschiedlicher Daten und/oder Verfahren, manchmal unter Modifikation der Modelle bzw. der jeweils spezifizierten Parameter. Manchmal werden im Rahmen der verfügbaren Daten die Modelle unter dem Verdacht der „beobachtbaren“ Heterogenität der Population in verschiedenen Situationen bzw. unter verschiedenen Bedingungs-Konstellationen miteinander verglichen. Und besonders häufig wird – dem Schwerpunkt des Bandes entsprechend – unter dem Verdacht nicht direkt beobachtbarer Heterogenität gleichzeitig mit der Prüfung der Modelle nach einer latenten Klassenstruktur gesucht, bei der durch relativ homogene Teilpopulationen innerhalb der latenten Klassen eine jeweils bessere Anpassung der Modelle an die Daten gelingen sollte. Durch

die Vermittlung von Anwendungserfahrungen bei der Analyse empirischer Daten in einem breiten Anwendungsfeld verdient das Buch u. E. besondere Beachtung, die keineswegs auf den Bereich der Klassifikationsanalysen beschränkt bleiben sollte.

WOLFGANG SODEUR, KÖLN



UDO KUCKARTZ,
STEFAN RÄDIKER,
THOMAS EBERT &
JULIA SCHEHL,
2010: Statistik.
Eine verständliche
Einführung. Wies-
baden: VS Verlag.
ISBN: 978-3-531-
16662-9/274, 274
Seiten Br., 14,95 EUR.

Ein Einführungsbuch zur Statistik, das bereits im Untertitel den Anspruch erhebt, verständlich zu sein, weckt bei Studierenden gleichermaßen Hoffnungen und Interesse. Dies gilt nicht zuletzt aufgrund der vielerorts diagnostizierbaren Aufwertung der Statistik- und Methodenausbildung im Rahmen der neuen Studiengänge und der bisweilen weit verbreiteten Skepsis und Ablehnung gegenüber deren Inhalten. Bei aller Skepsis ist sich eine Vielzahl der Studierenden nämlich durchaus bewusst – und dies betont Kuckartz auch in seinem Vorwort – dass eine fundierte Ausbildung in diesem Bereich nicht nur erhellend und für die Verfolgung des Forschungsstandes elementar ist, sondern sich auch begünstigend auf die Chancen am Arbeitsmarkt auswirkt. Da jedoch nicht Jedem ein intuitiver Zugang zu diesen Inhalten in die Wiege gelegt ist, wecken Lehrbücher, die vorgeben, den Stoff auf verständliche Weise zu präsentieren,

zwangsläufig Begehrlichkeiten. Um es vorweg zu nehmen: Kuckartz und seine Kollegen genügen diesem Anspruch vollstens.

Das Werk richtet sich explizit an Studierende der neuen Bachelor- und Masterstudiengänge und setzt „an dem Punkt [an], wo die auszuwertenden Daten bereits vorliegen“. Dieser Weg ist sinnvoll, trägt er doch der weithin vollzogenen Aufteilung der Statistik- und Methodenausbildung in separate Veranstaltungen Rechnung. Das erste Kapitel widmet sich der Datenaufbereitung und thematisiert die Erstellung eines Codeplans und den Umgang mit fehlenden Werten. Anschließend werden univariate Statistiken wie einfache Häufigkeitsverteilungen und deren grafische Darstellung (Kap. 2) sowie Lage- und Streuungsparameter (Kap. 3) vorgestellt, ehe sich Kapitel 4 mit der Analyse bivariater Zusammenhänge anhand von Kreuztabellen und Korrelationsmaßen für nominalskalierte Variablen beschäftigt. Die Kapitel 5 und 6 leiten von der deskriptiven Statistik zur Inferenzstatistik über und behandeln die Begriffe Wahrscheinlichkeit und Wahrscheinlichkeitsverteilungen (Kap. 5) und die Logik des statistischen Schließens (Kap. 6), in der die Verteilung von Stichprobenkennwerten sowie die Logik von Konfidenzintervallen und Hypothesentests besprochen werden. Die Durchführung eines t-Tests auf Mittelwertunterschiede steht im Zentrum von Kapitel 7, während sich Kapitel 8 mit der ein- und mehrfaktoriellen Varianzanalyse beschäftigt. Wie Zusammenhänge intervallskalierter Variablen analysiert werden können beschreibt Kapitel 9. Das zehnte Kapitel widmet sich der Skalenbildung, ehe der abschließende Abschnitt die bi- und multivariate Regression vorstellt und einen Ausblick auf die logistische Regression liefert. Die Autoren empfehlen zu Recht, die Kapitel chronologisch durchzuarbeiten, da die einzelnen Abschnitte aufeinander basieren. Dies gewährleistet einen sinnvollen und nachvollziehbaren Aufbau des Buches, obgleich Kapitel 10 zur Skalenbildung zunächst ein wenig deplatziert anmuten mag.

Bei näherer Betrachtung ist aber auch diese Anordnung nur logisch, stellt die Korrelationsanalyse in Kapitel 9 doch eine zwingende Voraussetzung zum Verständnis der Itemanalyse dar.

Zum Abschluss eines jeden Kapitels findet sich ein Abschnitt „So geht es mit SPSS/SYSTAT“, der dem Anspruch der Anwendungsorientierung des Werkes Rechnung tragen soll. Dem Leser wird hier vermittelt, wie die zuvor thematisierten Inhalte mit diesen recht intuitiven Softwarepaketen umgesetzt werden können. Sicherlich kann man hier monieren, dass sich die Anwendung ausschließlich über die Menüleiste vollzieht und nicht einmal auf die Möglichkeit verwiesen wird, diese Analysen auch syntaxbasiert durchzuführen, was dem wissenschaftlichen Postulat der Intersubjektivität eher genüge. Dennoch verzeiht man dies den Autoren schnell, da ein solches Vorgehen nicht recht zum Charakter des Buches passen würde und zudem die Anleitungen zur Durchführung einzelner Analysen und die zugehörigen Interpretationshilfen der Resultate vollauf überzeugen.

Als weitere Vorzüge des Buches sind zum einen das ausführliche Glossar und der Anhang zu nennen, der neben den üblichen Tabellen z. B. zur Standardnormalverteilung auch eine nützliche Interpretationshilfe enthält. Zum anderen werden eine Reihe prominenter sozial- und erziehungswissenschaftlicher Studien auf nationaler wie internationaler Ebene aufgeführt und deren Studiendesign skizziert. Schließlich haben die Autoren eine begleitende Webseite implementiert, die neben einer Formelsammlung und einem Auszug eines SPSS-Datensatzes auch verschiedene Links, z. B. zu Rechnern für die Normalverteilung oder die F-Verteilung, enthält. Zusammenfassend stellen auch die Begleitmaterialien zum Buch ein weiteres großes Plus dar.

Im Detail gestalten sich die ersten drei Kapitel in jeder Hinsicht überzeugend. Die univariate Analyse und Darstellung der Daten

wird anschaulich und verständlich präsentiert; lobend zu erwähnen ist zudem die kritische Auseinandersetzung mit (irreführenden) Grafiken und die ausführliche Erklärung von Boxplots als äußerst effiziente Methode zur Visualisierung der Verteilung einer Variablen. Mit Blick auf die bivariate Datenanalyse in Kapitel 5 ist jedoch (ebenso wie im Kapitel 9 zu Korrelationen) kritisch anzumerken, dass Maßzahlen für ordinale Zusammenhänge (z. B. Gamma oder Kendalls Tau-Maße) mit Ausnahme von Spearmans Rho nicht besprochen werden. Gleiches gilt für das PRE-Maß Lambda. Ihre Darstellung hätte sich als vorteilhaft erweisen können, da sich deren Logik als sehr eingängig darstellt und dem Leser womöglich ein tieferes Verständnis von der Grundidee statistischer Zusammenhänge beschert hätte. Zudem spielen ordinalskalierte Variable in den Sozialwissenschaften eine (mindestens) ebenso bedeutende Rolle wie Merkmale, die nominalskaliert sind oder metrisches Skalenniveau aufweisen.

In den folgenden Kapiteln geht es dann ans „Eingemachte“ und die bis dato vergleichsweise einfache Welt der uni- und bivariaten Statistik wird verlassen. Nun beginnt das Buch, seine eigentlichen Stärken auszuspielen. Das Kapitel zur Wahrscheinlichkeit und zu Wahrscheinlichkeitsverteilungen ist rundum überzeugend und selbst dann, wenn es um die abstrakten Normalverteilungen oder die Chi-Quadrat-Verteilung geht, in hohem Maße verständlich. Dies gilt auch für die Darstellung der Bedeutung der Stichprobenkennwerte – eine Materie, die erfahrungsgemäß vielen Studierenden enorme Schwierigkeiten bereitet, aber zum Verständnis der Inferenzstatistik essentiell ist. Den mit der Materie vertrauten Leser erfreut, dass die eigentlich korrekte Interpretation von Konfidenzintervallen zumindest in einer Fußnote Erwähnung findet, aber auch die gute und richtige Darlegung des Unterschieds zwischen statistischer Signifikanz und inhaltlicher Bedeutsamkeit. Kritisch ist allenfalls die etwas zu knappe

Einführung in die Logik von Konfidenzintervallen zu nennen. Auch überzeugt die Differenzierung zwischen Unterschieds- und Zusammenhangshypothesen einerseits und gerichteten und ungerichteten Hypothesen andererseits nur bedingt. Schließlich hätte sich bereits auf Seite 133 in Bezug auf den Standardfehler der Hinweis angeboten, dass dieser nicht nur von der Streuung des Merkmals, sondern auch von der Fallzahl abhängt. Dieser Hinweis folgt leider erst einige Seiten später.

Die Ausführungen zum t-Test für Mittelwertvergleiche und zur Varianzanalyse sind erneut gut, differenziert und verständlich erklärt. Einzig stören ein wenig die kleineren formalen Fehler, die sich bisweilen auch auf den Inhalt niederschlagen. So ist in einem Beispiel zum t-Test für abhängige Stichproben auf Seite 157 plötzlich von der Maßeinheit „Stunden“ die Rede, obwohl die Beispiele zuvor stets mit Kompetenzpunkten operieren. Ähnliches findet sich auch in Tabelle 9.4 auf Seite 198: Hier ist die Berechnung der einfachen Differenz der Rangplätze als Vorstufe zur Berechnung von Spearmans Rho nicht konsistent. Dessen ungeachtet besticht das Kapitel zur Korrelation erneut durch seine Verständlichkeit und stellt unter Beweis, dass Verständlichkeit und Gründlichkeit auch in der Aufbereitung von statistischen Lehrinhalten nicht notwendigerweise im Widerspruch stehen müssen. Besonders empfehlenswert ist zudem der Abschnitt zur Korrelation und Kausalität, der mit vielen Missverständnissen aufräumt. Auch das Kapitel zu Skalen weiß zu überzeugen und ist sehr ausführlich und anschaulich dargestellt. Diese Ausführlichkeit geht im abschließenden Part zur Regressionsanalyse allerdings ein wenig verloren. Zwar ist dieses Kapitel für Studierende, die sich einen ersten Zugang zu diesem Analyseverfahren eröffnen möchten, fraglos von hohem Wert. Ein wenig mehr Präzision wäre jedoch an der einen oder anderen Stelle wünschenswert gewesen. Um ein Beispiel zu nennen: Im Verlauf des Ka-

pitels erhält man schnell den Eindruck, der Determinationskoeffizient R^2 sei das einzige Maß zur Beurteilung der Güte von Regressionsmodellen. Auf den Root Mean Squared Error (RMSE) wird, obgleich dieser durchaus anschaulich zu vermitteln ist, nicht eingegangen. Weiterhin mag der Begriff des „Regressionsgewichts“ als Synonym für den b - bzw. Regressionskoeffizienten anschaulich sein – in der Forschung findet er jedoch kaum Verwendung. Schließlich sei noch kritisch auf den Ausblick auf die logistische Regression hingewiesen. Auf der einen Seite ist es lobenswert, diesen Ausblick zu liefern um darzustellen, welches Analyseverfahren sich für dichotome abhängige Variablen als adäquat gestaltet. Dennoch erscheint dies als eher halbherzige Lösung. Ein eigenes Kapitel zur logistischen Regression wäre die sinnvollere Variante gewesen, nicht zuletzt, da sich, wie die Autoren im Vorwort ausführen, die Auswahl der vorgestellten Verfahren an den methodischen Zugängen orientiert, die in den „Top-Fachzeitschriften“ auftreten. Hierzu zählen Kuckartz und Kollegen u. a. die Kölner Zeitschrift für Soziologie und Sozialpsychologie, die Zeitschrift für Erziehungswissenschaft, die Zeitschrift für Evaluation, die Zeitschrift für Soziologie und die Zeitschrift für Pädagogik. In diesen Periodika sind logistische Regressionen sicher ähnlich häufig anzutreffen wie lineare Regressionen, so dass ein eigenes Kapitel zur logistischen Regression eine gute Abrundung des Buches dargestellt hätte.

In der Zusammenschau stellt das Buch von Kuckartz und Kollegen ein wertvolles Einführungswerk insbesondere für jene Studenten dar, die über keinen intuitiven Zugang zur Statistik verfügen. Dies liegt in der durchweg – ganz im Sinne des Untertitels – verständlichen Darstellung der Inhalte und der Begrenzung auf die notwendigsten Formeln begründet. Trotz seiner Verständlichkeit schlägt der Vorwurf einer zu oberflächlichen Darstellung der Materie fehl. Der verständliche und forschungsorientierte Charakter des Buches äußert sich auch in

zahlreichen Beispielen, die sowohl der Illustration der Inhalte als auch der Anbindung an die aktuelle Forschung dienen. So stellt am Ende zahlreicher Kapitel ein eigener Abschnitt (z. B. „Die Regressionsanalyse in der Forschungsliteratur“) vor, wie das entsprechende Verfahren in aktuellen Forschungsbeiträgen eingesetzt und wie die Ergebnisse dargelegt werden. Diese Beispiele, die sich hauptsächlich auf den Kontext der Erziehungswissenschaft beziehen, gleichwohl aber für alle Sozialwissenschaftler nachvollziehbar sind, unterstreichen den hohen Wert des Buches für Studierende. Für in der Statistikausbildung tätige Dozenten dürfte – ganz im Sinne der Autoren – das Werk zudem eine sehr gute Grundlage für eine einführende Lehrveranstaltung darstellen, nicht zuletzt aufgrund seines überzeugenden Aufbaus.

SVEN STADTMÜLLER, MAINZ

Ankündigungen

Ausschreibung

ALLBUS-Preis 2011

ALLBUS-Preis 2011

Die seit 1980 alle zwei Jahre durchgeführte Allgemeine Bevölkerungsumfrage der Sozialwissenschaften – ALLBUS – ist ein zentraler Bestandteil der sozialwissenschaftlichen Infrastruktur in Deutschland. Mit dem ALLBUS wird der Profession eine qualitativ hochwertige Datenbasis für Sekundäranalysen zur Verfügung gestellt, die stark genutzt wird. Mittlerweile sind in der ALLBUS-Bibliographie über 1.800 Arbeiten verzeichnet, in denen ALLBUS-Daten verwendet werden.

Die Verleihung des *ALLBUS-Preises* soll diese besondere Bedeutung des ALLBUS als methodisch anspruchsvolle und inhaltlich reichhaltige Datenquelle für die Sozialwissenschaften unterstreichen. Die allen Wissenschaftlern offenstehende Auszeichnung wird in Zukunft alle zwei Jahre im Wechsel mit dem 2010 erstmals vergebenen *ALLBUS-Nachwuchspreis* verliehen.

Mit dem *ALLBUS-Preis* sollen herausragende wissenschaftliche Veröffentlichungen in Zeitschriften und Sammelbänden prämiert werden, in denen ALLBUS-Daten eine zentrale Rolle spielen. Über die Verleihung des Preises, der mit 1.000 € dotiert ist, entscheidet eine Jury, die aus den Mitgliedern des wissenschaftlichen Beirats des ALLBUS besteht. Die Originalität und Bedeutung der in der Arbeit behandelten – methodischen oder inhaltlichen – Fragestellung sowie das Niveau der Analyse sind maßgebliche Kriterien der Bewertung.

Eingereicht werden können alle in Deutsch oder Englisch verfassten sozialwissenschaftlichen Arbeiten mit ALLBUS-Daten, die in den letzten vier Jahren in Zeitschriften oder Sammelbänden publiziert wurden. Sowohl Eigenbewerbungen als auch Vorschläge durch Dritte werden berücksichtigt.

Einzureichen sind:

- die Publikation, in Papierform sowie als pdf-Datei
- eine maximal zweiseitige Kurzfassung der Arbeit
- ein Lebenslauf des Autors/der Autoren (mit Verzeichnis der - maximal 10 - wichtigsten Publikationen)

Einsendungen bitte bis spätestens zum **4. März 2011** an:

GESIS – Leibniz-Institut für Sozialwissenschaften
ALLBUS-Preis
Postfach 12 21 55
68072 Mannheim

Neue Klassifikation der Berufe 2010

KldB 2010

Die Bundesagentur für Arbeit (BA) hat eine neue Berufsklassifikation entwickelt: die „Klassifikation der Berufe 2010“ (KldB 2010). Die neue KldB wurde von einer Arbeitsgruppe bestehend aus den Bereichen Vermittlung/Beratung, Forschung (Institut für Arbeitsmarkt- und Berufsforschung) und Statistik der BA entwickelt und von einem Arbeitskreis „Klassifikation der Berufe 2010“ beratend begleitet, in dem neben dem Statistischen Bundesamt auch die betroffenen Bundesministerien sowie Experten der berufsfachkundlichen Forschung und der empirischen Sozialforschung vertreten waren.

Bei der KldB 2010 handelt es sich um eine Neuentwicklung, die die aktuelle Berufslandschaft in Deutschland realitätsnah abbildet und damit die in den letzten Jahrzehnten deutlich veränderten Berufsstrukturen zukünftig in Statistiken und Analysen angemessen erfassen kann. Mit Einführung der KldB 2010 wird zugleich das Nebeneinander zweier verschiedener nationaler Berufsklassifikationen beendet, da die Klassifikationen von 1988 (BA) und von 1992 (Statistisches Bundesamt) damit abgelöst werden. Ein weiterer Vorteil der KldB 2010 ist die hohe Kompatibilität zur internationalen Berufsklassifikation, der ISCO-08 (International Standard Classification of Occupations 2008), wodurch die internationale Vergleichbarkeit von Berufsinformationen in den amtlichen Statistiken und in der Forschung deutlich verbessert wird.

Das Statistische Bundesamt sowie die statistischen Ämter der Länder werden die KldB 2010 im nächsten Jahr einführen. In die Arbeitsmarktstatistiken der Bundesagentur für Arbeit wird die KldB 2010 im April 2011 eingeführt.

Beschreibung der KldB 2010: Die KldB 2010 ist als hierarchische Klassifikation mit fünf numerisch codierten Gliederungsebenen aufgebaut. Die KldB 2010 ist nach zwei Dimensionen gegliedert. Die strukturgebende Dimension ist die so genannte „Berufsfachlichkeit“. Das bedeutet, die Berufe sind in den obersten vier Ebenen anhand ihrer Ähnlichkeit der sie auszeichnenden Tätigkeiten, Kenntnisse und Fertigkeiten gruppiert. Auf der untersten Ebene erfolgt dann noch eine Untergliederung anhand der zweiten Dimension – dem „Anforderungsniveau“. Das Anforderungsniveau bezieht sich auf die Komplexität der ausübenden Tätigkeit und wird in vier Komplexitätsgraden – von 1 „Helfer- und Anlern-tätigkeiten“ bis 4 „hoch komplexe Tätigkeiten“ – erfasst.

Die KldB 2010 besteht in der obersten Hierarchieebene aus den folgenden zehn Berufsbereichen:

1. Land, Forst- & Tierwirtschaft und Gartenbau
2. Rohstoffgewinnung, Produktion & Fertigung
3. Bau, Architektur, Vermessung & Gebäudetechnik
4. Naturwissenschaft, Geografie & Informatik
5. Verkehr, Logistik, Schutz & Sicherheit
6. Kaufmännische Dienstleistungen, Warenhandel, Vertrieb, Hotel & Tourismus
7. Unternehmensorganisation, Buchhaltung, Recht & Verwaltung
8. Gesundheit, Soziales, Lehre & Erziehung
9. Sprach-, Literatur-, Geistes-, Gesellschafts- & Wirtschaftswissenschaften, Medien, Kunst, Kultur & Gestaltung
- 0 Militär

Auf der untersten Ebene (5-Steller) umfasst die Klassifikation 1.286 Einheiten. Rund 90 % dieser Positionen auf unterster Hierarchieebene lassen sich eindeutig einer der 436 Positionen der ISCO-08 auf unterster Hierarchieebene (unit groups) zuordnen. Die restlichen 10 % aller KldB-2010-Positionen teilen sich in der Regel auf zwei, in manchen Fällen auf mehr unit groups der ISCO-08 auf. Vor dem Hintergrund der Notwendigkeit einer adäquaten Abbildung der Berufswelt in Deutschland wurde damit die größtmögliche, zugleich aber auch eine hinreichende Kompatibilität zwischen der KldB 2010 und der ISCO-08 erreicht.

Methodisches Vorgehen: Die KldB 2010 wurde auf der Grundlage einer Gruppierung aktueller Berufe unter Einbeziehung von systematisch gepflegten berufsfachlichen Informationen und der Verwendung von statistischen Verfahren zur Klassifikation entwickelt. Damit konnte eine möglichst objektive Fundierung der Struktur der KldB 2010 erreicht werden. Im Unterschied zu den früheren nationalen Berufsklassifikationen, die deduktiv nach rein theoretischen Kenntnissen entwickelt worden sind, gründet die neue Klassifikation auf empirischen Analysen zur Ähnlichkeit von Berufen. Dieses Entwicklungsmodell stellt einen wesentlichen Fortschritt in der Konstruktion von Berufsklassifikationen dar. Es war erforderlich geworden, da die bisherigen Klassifikationen keinen geeigneten Anknüpfungspunkt für eine neue praxistaugliche Klassifizierung mehr boten und die Besonderheiten in den Strukturen des deutschen Arbeitsmarktes nicht mehr adäquat abgebildet wurden.

Die Klassifikation und die Dokumentationen zur Entwicklung sind zu finden unter:

<http://statistik.arbeitsagentur.de/Navigation/Statistik/Grundlagen/Klassifikation-der-Berufe/KldB2010/KldB2010-Nav.html>

Kontakt

Wiebke Paulus, Ruth Schweitzer, Silke Wiemer
Arbeitsgruppe „Klassifikation der Berufe 2010“
Bundesagentur für Arbeit
Regensburger Str. 104
90478 Nürnberg
Service-Haus.Statistik-KldB2010@arbeitsagentur.de

* * * * *

Drei neue Studiengänge für sozialwissenschaftliche Survey-Methoden

Studiengänge Survey-Methoden

Der durch das sozialwissenschaftliche Fachkolloquium der DFG genehmigte und seit 2008 bestehende DFG-Schwerpunkt „Priority Program for Survey Methodology“ (PPSM) widmet sich der Weiterentwicklung von Survey-Methoden. Im Zusammenhang mit dem Schwerpunktprogramm entstanden drei neue Masterstudiengänge, die alle zum Wintersemester 2010/2011 angelaufen sind. Es handelt sich hierbei um die Studiengänge „Survey Methodology“ (Duisburg-Essen), „Survey Statistics“ (Bamberg/Berlin/Trier) sowie „Statistische Wissenschaften“ (Bielefeld). Alle drei Programme sind viersemestrige Studiengänge, die mit jeweils unterschiedlichen inhaltlichen Schwerpunkten Themen aus dem Bereich „Survey Methodologie“ und Statistik für die Sozialwissenschaften behandeln.

(1) MA Survey Methodology der Universität Duisburg-Essen

Der Studiengang „Survey Methodology“ ist am Institut für Soziologie der Universität Duisburg-Essen angesiedelt und behandelt Survey Methodology im engeren Sinn. Es handelt sich um einen nicht-konsekutiven Studiengang für Sozial- und Wirtschaftswissenschaftler, der mit einem Master of Arts abschließt. Um eine enge und persönliche Betreuung der Studierenden gewährleisten zu können, ist die Anzahl der Studienplätze auf 16 begrenzt.

Neben dem Pflichtcurriculum, das außer spezifischen Inhalten wie beispielsweise Fragebogenkonstruktion und Stichprobenziehung auch sozialwissenschaftliche Veranstaltungen umfasst, bietet der Master durch Vereinbarungen mit anderen Fakultäten die Möglichkeit, auch Veranstaltungen aus benachbarten Fächern wie der Epidemiologie zu besuchen.

Voraussetzung für eine Zulassung ist ein Bachelor- oder ein äquivalenter Hochschulabschluss mit einem Leistungsumfang von mindestens 180 ECTS in den Sozial- bzw. Wirtschaftswissenschaften oder in Statistik. Die erforderliche Mindestnote beträgt 2,5. Innerhalb des ersten Studiums muss der Anteil der Methoden- und Statistikausbildung einen Umfang von mindestens 18 Credit Points haben.

Ansprechpartnerin ist Sabine von Thenen. Sie kann unter der E-Mail-Adresse surveymethodology@uni-due.de erreicht werden. Alle weiteren Informationen zum Studiengang „Survey Methodology“ finden sich unter www.uni-due.de/surveys.

(2) MSc Statistische Wissenschaften Universität Bielefeld

Der Masterstudiengang „Statistische Wissenschaften“ der Universität Bielefeld ist im Zentrum für Statistik an der Fakultät für Wirtschaftswissenschaften angesiedelt. Der Studiengang bietet ein fakultätsübergreifendes Programm, so dass die Studierenden die Möglichkeit erhalten, vertiefende Kenntnisse in den Bereichen Statistik und empirische Methoden aus Lehrveranstaltungen der Soziologie, Psychologie, Wirtschaftswissenschaften und der Mathematik zu erwerben.

Die Studierenden sollen im Rahmen dieses Masterstudiums mit den unterschiedlichsten Ausrichtungen der Statistik vertraut gemacht werden, so dass sie einerseits in der Lage sind, auf ein vielseitiges Methodenspektrum statistischer Verfahren zurückgreifen zu können und andererseits die speziellen Sichtweisen der einzelnen Anwendungsgebiete der Statistik zu verinnerlichen.

Zulassungsvoraussetzung ist der erfolgreiche Abschluss eines Hochschulstudiums mit mindestens sechssemestriger Regelstudienzeit. Einschlägig sind Studiengänge wie beispielsweise Wirtschaftswissenschaften, Mathematik, Soziologie, Psychologie, Epidemiologie, Public Health, Statistik oder Datenanalyse mit einem quantitativen Schwerpunkt im Umfang von 28 Leistungspunkten. Zur Erreichung der 28 Leistungspunkte können Angleichungsstudien im Umfang von max. 30 Leistungspunkten gemacht werden.

Ansprechpartner ist Prof. Dr. Jost Reinecke. Er kann unter der E-Mail-Adresse jost.reinecke@uni-bielefeld.de erreicht werden. Alle weiteren Informationen zum Studiengang „Statistische Wissenschaften“ finden sich unter <http://www.wiwi.uni-bielefeld.de/zest/studiengang>.

(3) MSc Survey Statistics

Der MSc „Survey Statistics“ ist ein gemeinsames Programm der Universitäten Bamberg und Trier (eigenständiger Survey-Statistik Master) und der Freien Universität Berlin (Schwerpunkt innerhalb des Masters in Statistik). Der Schwerpunkt des Masters liegt auf der Vermittlung moderner Erhebungs- und Schätzmethoden sowie auf der Behandlung von Nonresponse.

Das Programm bietet Studierenden die Möglichkeit, durch den Besuch von Veranstaltungen im Bereich der Volks- und Betriebswirtschaftslehre oder Soziologie auch ihre interdisziplinären Kenntnisse zu vertiefen. Die Studierenden haben so die Möglichkeit, sich durch individuelle Schwerpunktsetzung inhaltlich entweder auf die Forschungspraxis oder auf statistisch-methodische Inhalte zu konzentrieren. Die gemeinsamen Veranstaltungen der Survey-Statistik werden durch moderne Internet-Technologien an den drei Standorten simultan übertragen.

Für die Zulassung ist ein Bachelor- oder ein äquivalenter Abschluss in Mathematik, Wirtschaftsmathematik oder Statistik beziehungsweise Volkswirtschaftslehre, Betriebswirtschaftslehre oder Soziologie mit einer Mindestnote von 2,0 erforderlich. Innerhalb dieses Studiums müssen mindestens 24 Leistungspunkte in Mathematik, Statistik oder quantitativen Methoden erbracht worden sein.

Ansprechpartner ist Prof. Dr. Ralf Münnich. Er kann unter der E-Mail-Adresse muennich@uni-trier.de erreicht werden. Alle weiteren Informationen zum Studiengang „Survey Statistics“ finden sich unter <http://master.surveystatistics.net/>.

Autorinnen und Autoren Jahrgang 4 (2010)

- Gerrit Bauer, Mannheim
- Henning Best, Mannheim
- Michael Häder, Dresden
- Markus Klein, Mannheim
- Gerhard Krug, Nürnberg
- Michael Kühhirt, Mannheim
- Mike Kühne, Dresden
- Franziska Kunz, Freiburg
- Leander Steinkopf, Mannheim

Rezensentinnen und Rezensenten Jahrgang 4 (2010)

- Jürgen Schiener, Mainz
- Peter Schmidt, Braunsfels
- Wolfgang Sodeur, Köln
- Sven Stadtmüller, Mainz
- Johannes Stauder, Heidelberg
- Michael Wiedenbeck, Mannheim

Gutachterinnen und Gutachter Jahrgang 4 (2010)

Wir danken folgenden Kolleginnen und Kollegen, die für den Jahrgang 4 (2010) der Methoden – Daten – Analysen Manuskripte begutachtet haben:

- Jörg Blasius, Bonn
- Michael Blohm, Mannheim
- Michael Braun, Mannheim
- Andreas Diekmann, Zürich
- Nico Dragano, Düsseldorf
- Hermann Dülmer, Köln
- Stefanie Eifler, Bielefeld
- Frank Faulbaum, Duisburg
- Karlheinz Fleischer, Marburg
- Siegfried Gabler, Mannheim
- Michael Häder, Dresden
- Thomas Hinz, Konstanz
- Volker Hüfken, Düsseldorf
- Dagmar Krebs, Gießen
- Frauke Kreuter, Maryland
- Oliver Lüdtke, Tübingen
- Guido Mehlkop, Dresden
- Peter Preisendörfer, Mainz
- Wojtek Przepiorka, Zürich
- Werner Raub, Utrecht
- Ulrich Rosar, Köln
- Annette Schnabel, Umeå
- Volker Stocké, Bamberg
- Uwe Warner, Differdange
- Cornelia Weins, Trier

Hinweise für unsere Autorinnen und Autoren

Methoden – Daten – Analysen (MDA) veröffentlicht Beiträge aus dem Bereich der Empirischen Sozialforschung, insbesondere aus dem Bereich der Umfragemethodik. Im Vordergrund stehen Artikel, welche die methodischen und/oder statistischen Kenntnisse der Profession erweitern, sowie Beiträge, die sich mit der Anwendung der Methoden der Empirischen Sozialforschung in der Forschungspraxis beschäftigen, oder solche, in denen ein statistisches Verfahren exemplarisch angewandt wird. Obwohl der Schwerpunkt auf Umfragemethoden liegt, sind Beiträge zu anderen methodischen Bereichen willkommen. Die Artikel sollen für eine breite Leserschaft von Wissenschaftlern und Praktikern im Bereich der Empirischen Sozialforschung verständlich sein.

Manuskripte, die bereits an anderer Stelle veröffentlicht sind oder gleichzeitig anderen Publikationsorganen zur Veröffentlichung angeboten worden sind, werden grundsätzlich nicht berücksichtigt. Eine spätere Veröffentlichung eines in der MDA erschienenen Beitrages ist möglich, sofern an exponierter Stelle auf die Ersterscheinung des Beitrages in der MDA hingewiesen wird.

Jeder Beitrag, der zur Veröffentlichung in MDA eingereicht wird, wird zunächst von den Herausgebern danach bewertet, ob er für eine Veröffentlichung grundsätzlich in Frage kommt.

Falls die Herausgeber einer Veröffentlichung grundsätzlich ablehnend gegenüber stehen, werden die Autoren unter Angabe von Gründen für diese Entscheidung informiert.

Falls die Herausgeber zur Ansicht gelangen, dass der Beitrag grundsätzlich zur Veröffentlichung in Frage kommt, wird er anonymisiert an mindestens zwei unabhängige Gutachter verschickt, die um eine Stellungnahme gebeten werden. Im Zweifelsfalle wird ein drittes Gutachten eingeholt.

Wird ein Beitrag nach Beschluss der Herausgeber in das Begutachtungsverfahren gegeben, erfolgt die abschließende Entscheidung über ein Manuskript auf der Basis der Gutachten durch die Herausgeber. Im Falle einer Ablehnung erhalten die Autoren eine ausführliche Begründung für die Ablehnung. Wird eine Überarbeitung eines Beitrages für erforderlich gehalten, erhalten die Autoren detaillierte Überarbeitungshinweise.

Unabhängig vom Ergebnis des Begutachtungsverfahrens werden die Autoren von der Entscheidung durch die Redaktion per E-Mail informiert.

Die folgenden Regeln sind bei der Abfassung von Manuskripten zu beachten:

Manuskripte müssen per E-Mail (mda@gesis.org) eingereicht werden. Der Umfang der Manuskripte soll inklusive Leerzeichen alles in allem nicht mehr als 70.000 Zeichen betragen.

Den Beiträgen sind Abstracts in Deutsch und Englisch (jeweils ca. 15 Zeilen) voranzustellen. Auch der Titel des Beitrages ist in Deutsch und Englisch einzureichen.

Um die Anonymität der Beiträge zu wahren, darf in einem Manuskript nur der Titel des Beitrages enthalten sein, nicht aber Namen oder Anschriften der Autoren; Name und Anschrift der Autoren müssen, gemeinsam mit dem Titel des Beitrages, auf einer separaten Seite eingereicht werden.

Beiträge sind mit dem Dezimalklassifikationssystem zu untergliedern (1 - 2 - 2.1 - 2.2 - 3 usw.). Die Gliederungstiefe geht dabei höchstens auf *eine* Stelle nach dem Punkt.

Tabellen enthalten Tabellenummer und Titel im Tabellenkopf, Abbildungen werden analog behandelt.

Grafiken sind mittels gängiger Grafiksoftware zu erstellen. Ist eine spezielle Grafiksoftware erforderlich, übernimmt der Autor/die Autorin die endgültige Formatierung der Grafiken in eigener Regie.

Bei der Erstellung von Tabellen und Grafiken ist zu berücksichtigen, dass der Satzspiegel 11,5 cm (Breite) x 18,5 cm (Höhe) beträgt. Die Grafiken sind als jpeg- oder tif-Dateien in *Graustufen (CMYK)* mit einer Auflösung von mindestens 300 dpi zu liefern.

Die Beiträge sind unter Wahrung der gültigen Rechtschreiberegungen (neue Rechtschreibung) zu erstellen.

Werden in einem Beitrag empirische Daten verwandt, muss die Möglichkeit der Replikation bestehen. Im Falle einer Veröffentlichung in der MDA erklären sich die Autoren daher schriftlich bereit, Dritten auf deren Anfrage hin die Daten und Programmroutinen zur Verfügung zu stellen.

Anmerkungen und Fußnoten sind mit der Fußnotenfunktion des Schreibprogrammes (im Normalfalle Word) zu erstellen; bitte nicht gesondert formatieren. Fußnoten sind nur für inhaltliche Kommentare vorzusehen, nicht für bibliographische Hinweise.

Literaturhinweise im Text sind nach den folgenden Mustern aufzuführen: Müller (2002) – Schulze und Mayer (2003) – Müller, Mayer und Schulze (2004) – Müller et al. (2005) – Müller (2006: 75) – (vgl. Müller 2007: 75) – (Müller 2008; Mayer/Müller/Schulze 2009).

Das Literaturverzeichnis ist wie folgt zu gestalten:

Buchveröffentlichungen:

Strobl, R. und W. Kühnel, 2000: Dazugehörig und ausgegrenzt. Analysen zu Integrationschancen junger Aussiedler. Weinheim/München: Juventa.

Zeitschriftenbeiträge:

Becker, R., R. Imhof und G. Mehlkop, 2007: Die Wirkung monetärer Anreize auf den Rücklauf bei einer postalischen Befragung und die Antworten auf Fragen zur Delinquenz. Empirische Befunde eines Methodenexperiments. *Methoden – Daten – Analysen. Zeitschrift für Empirische Sozialforschung* 1 (2): 131-159.

Beiträge in Büchern:

Braun, M. und I. Borg, 2004: Berufswerte im zeitlichen und im Ost-West-Vergleich. S. 179-199 in: R. Schmitt-Beck, M. Wasmer und A. Koch (Hg.): Sozialer und politischer Wandel in Deutschland. Analysen mit ALLBUS-Daten aus zwei Jahrzehnten. Wiesbaden: VS-Verlag für Sozialwissenschaften.

Internetquellen:

Stadtmüller, S. und R. Porst, 2005: Zum Einsatz von Incentives bei postalischen Befragungen. *GESIS How-to-Reihe*, Nr. 14. Mannheim: GESIS. http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/howto/how-to14rp.pdf (1.12.2008).

Datenfile:

Forschungsgruppe Wahlen, Mannheim: Zur politischen Lage in Niedersachsen im Januar 2008. *GESIS Köln, Deutschland ZA Studie* Nr. 4863; doi: 10.4232/1.4863.

ISSN 1864-6956

4. Jahrgang 2010 © GESIS, Mannheim, Dezember 2010