

Inhalt

3 Editorial

FORSCHUNGSBERICHTE

- 5 Ein neues Verfahren für namensbasierte Zufallsstichproben von Migranten
Rainer Schnell et al.
- 35 Stichprobenziehung von Ehepaaren: Ein Praxisbericht zur Erhebung von seltenen Paarkombinationen
David Glowsky
- 59 Incentivierungsstrategien bei Minderheitenangehörigen: Ergebnisse eines Methodenexperiments
Patrick Fick und Claudia Diehl
- 89 Respondent Incentives in a National Face-to-Face Survey: Effects on Outcome Rates, Sample Composition and Fieldwork Efforts
Michael Blohm und Achim Koch
- 123 Empirische Forschungen an sozial- und politikwissenschaftlichen Lehrstühlen in Deutschland, der Schweiz und Österreich 2008 bis 2011
Rolf Hackenbroch
-

REZENSIONEN

- 129 Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. Udo Kuckartz, 2012
Cornelia Züll
- 132 Telephone Surveys in Europe: Research and Practice Sabine Häder, Michael Häder, Mike Kühne (Eds.), 2012
Volker Hüfken
- 138 Grundlagen sozialwissenschaftlichen Arbeitens. Eine anwendungsorientierte Einführung. Berninger, Ina et al. (2012)
Ralf Schultheiß
-

141 Information for Authors

Editorial

Liebe Leserin, lieber Leser,

dies ist die erste Ausgabe der mda, für die ich als geschäftsführender Herausgeber verantwortlich zeichne. Im Namen des Herausbergremiums bedanke ich mich herzlich bei Christof Wolf für seine ausgezeichnete Arbeit als Herausgeber in den letzten sieben Jahren. Ich bin sehr froh, dass er der Zeitschrift als Mitglied im wissenschaftlichen Beirat verbunden bleibt.

Im Zuge der Übergabe der Herausbergerschaft hat das Herausbergremium die zukünftige Entwicklung der mda intensiv diskutiert. Vor dem Hintergrund einer immer stärkeren Internationalisierung der Sozialwissenschaft haben wir uns entschlossen, die mda ab kommendem Jahr vollständig auf englische Sprache umzustellen. Den Titel der Zeitschrift passen wir leicht zu „methods, data, analyses“ an. Dem internationalen Charakter der Zeitschrift entsprechend haben wir außerdem eine weitere Herausgeberin aufgenommen. Wir sind sehr froh, dass wir Edith de Leeuw als neue Kollegin im Herausbergremium gewinnen konnten.

Die mda wird des Weiteren ihr Profil als Open Access Zeitschrift stärken, indem die Creative Commons Attribution License durchgängig verwendet wird. Im Sinne einer besseren Zitierbarkeit der Artikel bekommt jeder Artikel einen DOI zugewiesen. Diese Änderungen konnten wir bereits in der vorliegenden Ausgabe umsetzen. Wir sind überzeugt, mit der Einführung der thematischen Schwerpunktheftes im vorigen Jahr, der Internationalisierung ab 2014 sowie der Stärkung der Sichtbarkeit und Zitierfähigkeit die Zeitschrift für die deutsche und internationale Wissenschaftsgemeinschaft noch attraktiver zu machen.

In diesem Sinne freue ich mich auf Ihre Einreichungen, auf spannende Artikel und konstruktive Diskussionen im sich dynamisch entwickelnden Feld der Survey Methodology und der quantitativen Methoden.

Ein neues Verfahren für namensbasierte Zufallsstichproben von Migranten

A New Name-Based Sampling Method for Migrants

Rainer Schnell, Tobias Gramlich, Tobias Bachteler, Jörg Reiher, Mark Trappmann, Menno Smid, Inna Becher

Zusammenfassung

Zu den besten Verfahren für die Konstruktion von Migrantenstichproben gehören namensbasierte Stichproben. Hierfür werden bislang entweder ad-hoc-Listen oder namenskundliche Lexika für die Klassifikation von Namen verwendet. In diesem Beitrag wird ein neues Verfahren vorgeschlagen, das auf der automatischen Klassifikation eines Namens anhand der n -Gramme des Namens und der Anwendung des Bayes-Theorems basiert. Das neue Verfahren ist fehlertolerant gegenüber alternativen Schreibweisen und erlaubt auch die Klassifikation von Namen, die sich nicht in den Lexika finden. Das Verfahren wurde anhand der Namen der ca. 1.600 Ausländer im PASS-Panel und einer CATI-Studie in Hessen untersucht.

Abstract

The set of best methods for sampling migrant populations includes name-based sampling. So far this is done using either ad-hoc lists or onomastic dictionaries for the classification of names. This paper proposes a new name-based procedure, which uses a Bayes-classifier for the n -grams of the name. The new procedure is fault-tolerant of alternate spellings, and also allows the classification of names that are not found in dictionaries. It was tested using the names of about 1.600 foreigners in the PASS panel. Finally, a CATI survey based on the new method in Hesse is described.

Keywords: Bayes-Classification, n -gram, rare populations, onomastic

Schlüsselworte: Bayes-Klassifikation, n -Gramm, seltene Populationen, Onomastik



1 Einleitung¹

Ausländer- und Migrantenpopulationen² in Deutschland sind für viele Forschungsbereiche von zentraler Bedeutung, z. B. für die Bildungs- und Arbeitsmarktforschung. Die in der Sozialforschung üblichen Stichprobenverfahren greifen aus zwei Gründen bei solchen Populationen nicht: Erstens stellen einzelne Migrantengruppen nur einen geringen Bevölkerungsanteil, daher sind die meisten Verfahren zur Ziehung von Zufallsstichproben hier ineffizient. Zweitens enthalten die meisten in der Praxis verwendeten Auswahlgrundlagen (Telefonbücher, Einwohnermelderegister oder Patientenregister) häufig keine aktuellen oder zugänglichen Informationen über die Zugehörigkeit zur Zielpopulation. Vor allem Personen mit Migrationshintergrund, die jedoch die deutsche Staatsbürgerschaft besitzen, lassen sich mit den zur Verfügung stehenden Auswahlgrundlagen kaum bestimmen. Daher müssen für diese Populationen spezielle Auswahlverfahren verwendet werden.

2 Stichprobenverfahren für Ausländerstichproben

Für Migranten in Deutschland liegen in der Regel keine vollständigen Auswahlgrundlagen zur Stichprobenziehung vor.³ Daher müssen Stichprobenverfahren für

- 1 Die Autoren verdanken den beiden anonymen Gutachtern hilfreiche Anmerkungen. Die Autoren danken Andreas Humpert und Klaus Schneiderheine für hilfreiche Diskussionen früherer Manuskriptentwürfe. Das Verfahren und die Art des Tests wurde von Rainer Schnell vorgeschlagen und von Tobias Bachteler und Jörg Reiher umgesetzt. Die Klassifikation, die Berechnung der Tests und der erste Textentwurf erfolgte durch Tobias Gramlich; Mark Trappmann war an der Konzeption des Aufsatzes beteiligt, ermöglichte den Zugang zu den Daten und kommentierte unterschiedliche Fassungen des Manuskripts; Menno Smid leitete die CATI-Studie in Hessen. Inna Becher stellte eine Literaturdokumentation zur Verfügung. Die finale Textversion stammt von Rainer Schnell.
- 2 Selbstverständlich sind die Begriffe „Ausländer“ und „Migrant“ nicht deckungsgleich; zu den damit verbundenen konzeptuellen Problemen siehe z. B. Galonska et al. (2004) und Diefenbach und Weiß (2006). Entsprechend ist die empirische Umsetzung der Konzepte nicht trivial, vgl. z. B. Statistisches Bundesamt (2011a). Die Bezeichnungen „Ausländer“ und „Migranten“ werden im Folgenden trotz dieser Probleme synonym verwendet, da für die Beurteilung jedes empirischen Verfahrens mindestens ein Referenzwert benötigt wird. Das im Folgenden vorgestellte Verfahren wird mangels anderer verfügbarer Daten anhand der Staatsangehörigkeit evaluiert. Für das technische Verfahren selbst spielt das zugrunde liegende Konzept keine Rolle: Verfügt man über konzeptuell andere Referenzdaten zum Training des Verfahrens (Namen nach Migrationshintergrund), kann ein anderes Konzept eingesetzt werden.
- 3 Lediglich das Bundesamt für Migration und Flüchtlinge (BAMF) hat im Rahmen seines Auftrages das von ihm selbst geführte Ausländerzentralregister (AZR) in einzelnen Fäl-

seltene Populationen verwendet werden (Kalton 2009). Üblicherweise wird als obere Grenze einer „seltenen Population“ ein Anteil von maximal 10% an der Gesamtpopulation betrachtet (Kalton/Anderson 1986). Trotz eines Anteils an der Gesamtbevölkerung von insgesamt – je nach Definition – ca. 9,0% Ausländern oder 19,3% Personen mit Migrationshintergrund (Statistisches Bundesamt 2011a,b), stellen Ausländer in Deutschland in diesem Sinne eine „seltene“ Population dar.

2.1 Klumpenstichproben

Für seltene Populationen werden häufig Listen von Teilpopulationen als Auswahlgrundlagen verwendet. Dazu gehören z. B. Mitgliederlisten von Vereinen oder Organisationen. So konnte z. B. Rother (2010) für eine Ausländerstichprobe auf Verwaltungsdaten des Bundesamtes für Migration und Flüchtlinge (BAMF) zurückgreifen und eine Stichprobe aus allen Teilnehmern von Integrations- und Alphabetisierungskursen ziehen.

Oftmals klumpen Elemente der Zielpopulation geographisch stark: Dann ist die Dichte an Ausländern aus einzelnen Ländern in einzelnen geographischen Regionen so hoch („Little Italy“, „China Town“), dass sich Standardauswahlverfahren effizient auf diese Regionen anwenden lassen (siehe z. B. Blane 1977, Ecob/Williams 1991). So beschränkt sich z. B. die europäische Erhebung zu Minderheiten und Diskriminierung (EU-MIDIS) in einigen Mitgliedsländern auf Minderheitengruppen mit einem Anteil von mindestens 5% an der Gesamtbevölkerung und vornehmlich auf Gebiete mit „mittlerer bis hoher Konzentration“ der Zielgruppen (European Union Agency for Fundamental Rights 2009: 18f). Das Problem dieser Verfahren besteht vor allem darin, dass sich die Personen in solchen Gebieten von Personen in Gebieten mit geringeren Konzentrationen in Hinsicht auf abhängige Variable wie z. B. „Integration“ und „Diskriminierungswahrnehmung“ unterscheiden können.

2.2 Nicht auf Zufallsverfahren basierende Stichproben

Vor allem in Deutschland werden bei Stichprobenziehungen von Ausländern häufig Quotenstichproben verwendet, die die Auswahl mindestens auf einer Stufe letztlich dem Interviewer überlassen. Ein bekanntes Beispiel für diese Art der Ziehung stellt die Publikation „Zuwanderer in Deutschland“ (Bertelsmann Stiftung

Fortsetzung Fußnote 3

len zur Stichprobenziehung verwendet (Babka von Gostomski 2008). Faktisch stehen die Mikrodaten des AZR für keine praktischen Zwecke außerhalb des BAMF zur Verfügung.

2009) dar. Gerade durch eine solche Auswahl durch den Interviewer dürften vor allem besser und gut integrierte Migranten erfasst werden. Allgemein sollte aufgrund der unkontrollierbaren Effekte für wissenschaftliche Zwecke immer von Quotenauswahlen abgesehen werden (Schnell/Hill/Esser 2011: 296-298).

Weiterhin werden auch Netzwerk- bzw. Schneeballverfahren zur Stichprobenziehung von Ausländern und Migranten eingesetzt, so z. B. für eine Teilstichprobe des SOEP von Spätaussiedlern aus der ehemaligen Sowjetunion (Burkhauser et al. 1997). Netzwerkstichproben leiden unter zwei Problemen: Die Auswahlwahrscheinlichkeiten sind in der Praxis kaum korrekt berechenbar und sozial isolierte Personen verfügen über geringere Auswahlwahrscheinlichkeiten. Netzwerkstichproben sind daher eher als allerletztes Mittel denn als Methode mit reproduzierbaren Ergebnissen anzusehen.

2.3 Screening

Oftmals liegen nur Auswahlgrundlagen für eine „Allgemeinbevölkerung“ vor. Einige Verfahren für die Auswahl seltener Populationen basieren darauf, dass ausgehend von einer solchen Liste alle Elemente einer Stichprobe da-raufhin untersucht werden, ob sie zur Zielpopulation gehören. Dieses Vorgehen wird als „Screening“ bezeichnet. Für Stichproben von Ausländern können z. B. Einwohnermeldeamtsstichproben gezogen und dann Ausländer anhand ihrer registrierten Staatsangehörigkeit identifiziert werden (vgl. Granato 1999). Auch für die Ausländerstichprobe des SOEP (Stichprobe B) wurde aus Melderegistern gezogen (Haisken-DeNew und Frick 2005). Dieses Vorgehen eignet sich jedoch nicht für alle Zielgruppen. Für Personen, die die deutsche Staatsbürgerschaft besitzen, ist dieses Vorgehen des Screenens anhand der Nationalität nicht möglich. Salentin (2007) schlägt daher für Aussiedler vor, bei Einwohnermeldeamtsstichproben anstatt nach der Staatsangehörigkeit nach dem Geburtsort zu screenen.⁴ Dieses Vorgehen eignet sich prinzipiell auch für alle anderen naturalisierten Migranten der ersten Generation. Dies gilt aber nicht für spätere Migrantengenerationen, also in Deutschland geborene Kinder oder Enkel der eingewanderten Migrantengeneration. Zwar wäre es möglich, in einer Befragung nach dem Geburtsland der Elternteile oder gar der Großeltern zu

4 Dies wäre nur dann möglich, wenn die Weitergabe des Merkmals „Geburtsort“ im Rahmen einer „Gruppenauskunft“ zulässig wäre. Nach Prüfung der Meldegesetze aller Bundesländer durch die Arbeitsgruppe kann festgehalten werden, dass eine Weitergabe des Merkmals Geburtsort nach der geltenden Rechtslage in Deutschland nur im Rahmen von Einzelauskünften „namentlich bekannter Einwohner“ im Rahmen einer „erweiterten Melderegisterauskunft“ möglich ist. Stichprobenziehungen stellen aber Gruppenauskünfte über „eine Vielzahl namentlich nicht bekannter Einwohner“ dar.

fragen, die Validität der Angaben dürfte aber eher fraglich sein. Empirische Studien hierzu scheinen nicht vorzuliegen.

2.4 Namensbasierte Verfahren

Häufig ist das eigentliche Screening-Merkmal nicht Bestandteil der Auswahlgrundlage, so dass Hilfsmerkmale verwendet werden müssen. In vielen Fällen stehen lediglich Namenslisten zur Verfügung. Namenslisten werden weltweit bei vielen Forschungsprojekten zu Migrantenpopulationen verwendet, da Namen auf die regionale oder ethnische Herkunft des Namensträgers hinweisen können. In der Praxis werden insbesondere oft namenskundliche Lexika von Namen mit bekannter Zugehörigkeit (d. h. ausschließlichem oder sehr häufigem Auftreten in der entsprechenden Gruppe) zu einem bestimmten Herkunftsland verwendet oder für diesen Zweck aufwendig erstellt.

In Deutschland hat sich das von Humpert und Schneiderheinze (2000) entwickelte onomastische Verfahren als Standardverfahren für sozialwissenschaftliche Stichprobenziehungen bei Migranten etabliert. Dabei werden Namen aus öffentlich zugänglichen Verzeichnissen mit Hilfe von aus ihnen erstellten Namensdatenbanken verschiedener Nationalitäten verglichen und entsprechend klassifiziert. Neuere in der Migrationsforschung bekannte Beispiele für die Anwendung dieses Verfahrens sind der Jahresbericht des Sachverständigenrates deutscher Stiftungen für Migration und Integration (SVR 2010), der Integrationsurvey des Bundesinstituts für Bevölkerungsforschung BiB (Mammey/Sattig 2002, Haug/Swiaczny 2003) sowie die Studie „Muslime in Deutschland“ (MiD, Brettfeld/Wetzels 2007).

Prinzipiell ähnliche Verfahren werden in der Epidemiologie verwendet. Hierbei dominiert die Nutzung ad hoc zusammengestellter Listen häufiger oder typischer Namen einzelner Länder. Diese Listen werden in der Regel von muttersprachlichen Experten gesichtet und editiert. Die resultierenden Listen werden anschließend für das Screening einer allgemeineren Auswahlgrundlage verwendet. Ein entsprechendes Beispiel stellt die Arbeit von Halm und Sauer (2005: 43ff) dar. Anhand einer Liste mit rund 10.000 „typischen“ Nachnamen und rund 7.000 Vornamen klassifizieren sie Einträge im Telefonbuch. Ähnliche Verfahren sind in der Epidemiologie international verbreitet. So erstellen z. B. Schwartz et al. (2004) eine Liste häufiger arabischer Vor- und Nachnamen aus verschiedenen öffentlich

zugänglichen Quellen⁵ und sichten die resultierende Liste manuell, um schließlich eine Stichprobe aus einem Detroit-er Krebsregister zu ziehen. Ähnlich verwendet Lauderdale (2006) eine Namensliste aller sozialversicherten Personen zur Erstellung einer Liste und zum Screening nach arabischstämmigen Frauen in kalifornischen Geburtsregistern.

3 Beschreibung des neuen Verfahrens

Das im Folgenden vorgestellte Verfahren basiert im Gegensatz zu allen bisher verwendeten Verfahren nicht auf der Klassifikation vollständiger Namen oder Namensendungen, sondern auf der Klassifikation von Buchstabenfolgen (n -Gramme).⁶

Den Ausgangspunkt des Verfahrens bilden dabei Listen, die für verschiedene Nationalitäten die jeweiligen empirischen Namenshäufigkeiten getrennt nach Vor- und Nachnamen aller Personen mit dieser Nationalität enthalten. Am geeignetsten für solche Listen sind Zensus- oder Sozialversicherungsdatenbanken.⁷

Aus dem Anteil eines bestimmten Namens $P_{(Name)}$, dem Anteil aller Personen mit einer bestimmten Nationalität $P_{(Nationalität)}$ und dem bedingten Anteil eines Namens unter Personen mit einer bestimmten Nationalität $P_{(Name|Nationalität)}$ kann mit Hilfe des Bayes-Theorems die bedingte Wahrscheinlichkeit für eine bestimmte Nationalität gegeben einen bestimmten Namen $P_{(Nationalität|Name)}$ berechnet werden:

$$P_{(Nationalität|Name)} = \frac{P_{(Name|Nationalität)} * P_{(Nationalität)}}{P_{(Name|Nationalität)} * P_{(Nationalität)} + P_{(Name|\neg Nationalität)} * P_{(\neg Nationalität)}}$$

- 5 Als Quellen dienten verschiedene Register (z. B. Geburts- und Sterberegister), die Namen und die entsprechenden Geburtsländer enthielten, Mitgliederlisten entsprechender (z. B. Kultur-)Vereine und Vereinigungen; außerdem sammelten (arabisch sprechende) Mitarbeiter aus Telefonbüchern „typische“ arabische Namen.
- 6 Das hier beschriebene Verfahren wurde von den Autoren unter anderem auf der ESRA Konferenz in Lausanne 2011 und der ASA Konferenz 2011 in Tilburg vorgestellt. Das Verfahren wurde von Rainer Schnell, Tobias Bachteler und Jörg Reiher an der Universität Konstanz im November 2006 entwickelt und von diesen in einer Reihe von Stichprobenziehungen und Rekord-Linkage-Projekten eingesetzt. Eine erste Beschreibung findet sich bei Schnell (2009). Eine frühere Fassung des vorliegenden Artikels wurde im Juni 2012 als Diskussionspapier veröffentlicht (Schnell et al. 2012).
- 7 Shackleford (1998) und Lauderdale (2006) beschreiben solche Listen von Namenshäufigkeiten für verschiedene Nationalitäten nach Auszahlungen des US-amerikanischen Zensus oder aller sozialversicherungspflichtig beschäftigten Personen in den USA.

Der Nachteil einer solchen Klassifikation besteht in der Verwendung vollständiger und korrekt geschriebener Namen.⁸ In der Literatur zum Rekord-Linkage werden Schreibfehler von 15-25% der Namen berichtet (Winkler 2009). Diese dürften sich bei Namen von Migranten häufen, so dass hier mit besonders vielen Fällen zu rechnen ist, bei denen jedes Verfahren, das exakte Übereinstimmungen voraussetzt, Ausfälle produzieren wird. Entsprechend würden hierbei Anteilswerte systematisch unterschätzt. Daher ist ein fehlertolerantes Verfahren für die Namensklassifikation wünschenswert.

Eine Möglichkeit besteht darin, für die Klassifizierung nicht die vollständigen Vor- oder Nachnamen zu verwenden, sondern die Namen in Buchstabenfolgen (n -Gramme) zu zerlegen und diese zur Klassifizierung der Namen zu verwenden. Solche n -Gramme werden in der Informatik für viele Probleme der Verarbeitung natürlicher Sprachen verwendet, so z. B. bei der Konstruktion von Suchmaschinen oder Programmen zur Rechtschreibprüfung.

Im Folgenden verwenden wir also die bedingten Häufigkeiten aller n -Gramme eines Vor- oder Nachnamens zur Berechnung der bedingten Wahrscheinlichkeit für eine Nationalität.⁹ Neben der Fehlertoleranz besitzt das Verfahren einen weiteren Vorteil: Es muss kein vollständiges Namenslexikon vorliegen. Es können also auch Namen klassifiziert werden, die nicht in einem Lexikon verzeichnet sind.

Tabelle 1 zeigt nun die Klassifikation am Beispiel des Namens „Peter“. Zunächst wird der Name „Peter“ (Länge 5 Buchstaben) in seine n -Gramm-Menge (hier Bigramme, $n=2$) aufgespalten. Der Name „Peter“ hat 4 n -Gramme der Länge 2: {PE,ET,TE,ER}. Da oftmals Wort- oder Namensanfänge und -endungen cha-

- 8 Die Idee der Namensklassifikation durch das Bayes-Theorem scheint zum ersten Mal für eine Anwendung im US-Bureau of the Census benutzt worden zu sein: Passel und Word (1993) und Perkins (1993) beschreiben die Konstruktion einer „spanischen“ Namensliste für den US-amerikanischen Zensus 1980. Hierbei wurden mithilfe des Bayes-Theorems die bedingten Wahrscheinlichkeiten der Zugehörigkeit zu hispanischen Subpopulationen berechnet. Diese Zensus-Anwendung basiert auf vollständigen Namen; das Verfahren wurde in der Stichprobenliteratur für seltene Populationen nicht aufgenommen. Davon unabhängig finden sich in der Computerlinguistik einzelne Anwendungen, bei denen Texte durch n -Gramme nach Sprachen klassifiziert werden (z. B. Cavnar/Trenkle 1994). Noch seltener ist die automatische Klassifikation von Namen (Konstantopoulos 2007). Beide Arbeiten verwenden zwar n -Gramme, klassifizieren aber über n -Gramm-Profilähnlichkeiten, nicht über das Bayes-Theorem. Auch diese Arbeiten finden sich nicht in der Stichprobenliteratur.
- 9 Wie die Ergebnisse der vorliegenden Studie zeigen, kann die Staatsangehörigkeit als Trainingsmerkmal trotz hoher Anteile von deutscher Staatsangehörigkeit bei Gruppen mit Migrationshintergrund erfolgreich verwendet werden, solange sich die bedingten Wahrscheinlichkeiten für ein n -Gramm zwischen den jeweils interessierenden Gruppen überzufällig unterscheiden.

rakteristisch für verschiedene Sprachen sind,¹⁰ empfiehlt sich das Anfügen von einem Leerzeichen vor und hinter die Namen, so dass zusätzliche n -Gramme am Namensanfang bzw. -ende entstehen:

$$\{ _P,PE,ET,TE,ER,R_ \}$$

Für jedes Land wird nun die Anzahl des Auftretens jedes einzelnen Bigramms dieser Bigrammmenge durch die Anzahl aller Namen¹¹ von Personen einer bestimmten Nationalität dividiert, und daraus für jedes Land dann das Produkt aus diesen relativen Häufigkeiten der einzelnen n -Gramme des Namens „Peter“ gebildet.¹² Dieses Produkt wird anschließend mit dem Faktor w multipliziert.¹³

Ein Name wird nun schließlich in dasjenige Land klassifiziert, für das diese Klassifikationsgröße maximal ist.¹⁴

10 Dies gilt auch für Namen: z. B. enden deutsche männliche Namen in der Regel nicht auf -a. Deutsche männliche Namen enden in der Regel auch nicht mit -e, wohingegen in Italienischen Namen mit Endung -e oftmals die männliche Namensvariante sind (z. B. die italienischen Namen Simone (männlich) und Simona (weiblich)).

11 Hier wird durch die Zahl der Namen dividiert und nicht etwa durch die Anzahl der n -Gramme. Dadurch werden Unterschiede zwischen den Ländern nicht weiter durch Unterschiede in der Länge der Namen zusätzlich betont.

12 Der Beitrag eines „unbekannten“, nicht erfassten n -Gramms zur Klassifikation eines Namens wurde nicht auf 0 gesetzt, andernfalls würde der gesamte Ausdruck bei der Multiplikation zur Gesamtwahrscheinlichkeit natürlich auch 0. Bei unserer Anwendung wurde statt 0 der Faktor $\frac{1}{N}$ verwendet. An dieser Stelle würde sich bei einem elaborierteren Programm der Einsatz von n -Gramm-Glättungen anbieten (Jurafsky und Martin 2009). Für die hier beabsichtigte erstmalige Anwendung eines Verfahrens in der Stichprobenziehung erwies sich das hier gewählte Vorgehen aber als praxistauglich.

13 Wenn $n_{\text{Namen}[j]}$ die Anzahl Namen des Landes j und $N = \sum n_{\text{Namen}[j]}$ die Anzahl Namen aller Länder bezeichnet, ergibt sich $w_{[j]} = \frac{n_{\text{Namen}[j]}}{N}$; das ist die Priorwahrscheinlichkeit aus dem Bayes-Theorem.

14 Die programmtechnische Umsetzung ist mit einer Skriptsprache oder einem Datenanalyseprogramm mit wenigen Zeilen möglich.

Tabelle 1 Klassifikation des Namens „Peter“^a

Nationalität	Bigramm	Anzahl Bigramme	Anzahl alle Namen/Land	p_{Land}	$\prod p_{Land}$	Korrekturfaktor w	Klassifikationsgröße (Land Peter ^a)
Deutschland	ER	5.004.637		0.1768			
	TE	2.689.490		0.0950			
	R+	2.477.673	28.309.791	0.0875	$8.09 \cdot 10^{-8}$	0.9271	$7.50 \cdot 10^{-8}$
	ET	1.750.568		0.0618			
	+P	929.959		0.0329			
	PE	767.277		0.0271			
Osteuropa ^b	ER	17.968		0.0705			
	R+	14.392		0.0565			
	ET	12.271	254.788	0.0482	$8.42 \cdot 10^{-9}$	0.0083	$7.03 \cdot 10^{-11}$
	TE	10.936		0.0429			
	+P	10.828		0.0425			
	PE	6.134		0.0241			
eh. Jugoslawien	R+	49.621		0.0686			
	ER	37.698		0.0521			
	ET	32.209	723.561	0.0445	$2.79 \cdot 10^{-10}$	0.0237	$6.61 \cdot 10^{-12}$
	TE	9.597		0.0133			
	+P	9.154		0.0126			
	PE	7.567		0.0105			
Italien	ER	12.990		0.0600			
	PE	12.458		0.0575			
	+P	9.885	216.672	0.0456	$6.37 \cdot 10^{-10}$	0.0071	$4.53 \cdot 10^{-12}$
	ET	8.735		0.0403			
	TE	5.420		0.0250			
	R+	872		0.0040			

Nationalität	Bigramm	Anzahl Bigramme	Anzahl alle Namen/Land	p_{Land}	$\prod p_{Land}$	Korrekturfaktor w	Klassifikationsgröße (Land"Peter")
	ER	70.013		0.1157			
	ET	58.307		0.0964			
Türkei	R+	47.197	605.029	0.0780	$1.45 \cdot 10^{-10}$	0.0198	$2.88 \cdot 10^{-12}$
	TE	8.054		0.0133			
	+P	2.279		0.0038			
	PE	2.013		0.0033			
...
Summen		14.221.676	30.535.580				

a Die hier nicht ausgewiesenen Namen aus Griechenland und der ehemaligen Sowjetunion wurden in den Berechnungen berücksichtigt.

b Polen und osteuropäische Nachbarländer

4 Trainingsdaten für das neue Verfahren

Zentraler Punkt des Verfahrens ist eine Namensliste, aus der diese n -Gramm-Häufigkeiten gewonnen werden können, d. h. eine Liste, die Namen und Häufigkeiten dieses Namens für alle (berücksichtigten) Nationalitäten enthält. Solche Listen sind – vor allem außerhalb Deutschlands – aus verschiedenen Quellen und in unterschiedlicher Qualität verfügbar. In Deutschland böten sich Listen auf der Basis der Einwohnermelderegister an; aufgrund der erforderlichen Kooperation vieler Gemeinden erscheint dieses Vorgehen in der Praxis eher schwierig. Eine naheliegende Alternative bestünde in der Verwendung von Telefonverzeichnissen, aus denen sich Namen und Häufigkeiten, hingegen aber nicht die zugehörige Nationalität der Personen gewinnen ließen. Werden hierbei Telefonverzeichnisse aus vielen verschiedenen Ländern verwendet, könnte man alle dort verzeichneten Namen als „einheimisch“ behandeln und die insgesamt geringe Wahrscheinlichkeit der unvermeidlichen Fehlklassifikationen durch Migranten akzeptieren oder sich auf die häufigsten Namen oberhalb je nach Land verschiedener Schwellenwerte beschränken. Einfacher zu handhaben sind natürlich entsprechende Listen aus einer Datenquelle. Für die Entwicklung des hier vorgestellten Verfahrens wurden erstmalig nach Nationalität getrennte Listen von Vor- und Nachnamenshäufigkeiten aller in Deutschland sozialversicherungspflichtig beschäftigten Personen erstellt.¹⁵ Diese Listen enthalten Namen und Häufigkeiten getrennt nach Staatsangehörigkeit. Aus Datenschutzgründen lagen Vor- und Nachname also nicht gemeinsam (z. B. „D – Peter Müller – $n=5$ “), sondern nur getrennt („D – Peter – $n=40.000$ “ und „D – Müller – $n=1.000$ “) vor.

Insgesamt umfassten die Namenslisten für die hier klassifizierten Nationalitäten (bzw. Ländergruppen) 112.831 unterschiedliche Vor- und 493.974 unterschiedliche Nachnamen. Entsprechend gewichtet mit den jeweiligen Häufigkeiten entspricht dies jeweils rund 30 Millionen Personen. Die Tabelle 2 zeigt die Anzahl unterschiedlicher Namen innerhalb der Namenslisten in den bei der Klassifikation berücksichtigten Ländergruppen.

Unter „Russland“ wurden alle Einträge von Namen und Häufigkeiten von Staaten der ehemaligen Sowjetunion und der ehemaligen Russischen Föderation

15 Diese Liste wurde mit hohem Aufwand im Jahr 2005 auf Wunsch des Erstautors einmalig vom Forschungsdatenzentrum der Bundesagentur für Arbeit nach Rücksprache mit den Datenschützern der Bundesagentur unter der Auflage zur Verfügung gestellt, dass die Namenslisten nur getrennt für Vor- und Nachnamen erstellt wurden und jeder Name mindestens fünfmal in den Datenbanken erschien.

Tabelle 2 Anzahl der Vor- und Nachnamen in den Trainings-Namenslisten

Staatsangehörigkeit	Vornamen		Nachnamen	
	Namen	Personen	Namen	Personen
Deutschland	58.757	28.309.791	383.592	27.551.167
Jugoslawien ^a	10.494	313.193	21.973	262.425
Türkei	8.137	605.029	20.835	587.175
Osteuropa ^b	2.750	103.300	7.273	47.366
Italien	2.707	216.672	14.334	180.118
Griechenland	2.517	112.744	9.452	75.732
Russland ^c	1.952	47.638	2.476	13.187
restliche Welt	25.517	470.446	34.039	286.887
Insgesamt	112.831	30.178.813	493.974	29.004.057

a ehemaliges Jugoslawien und Nachfolgestaaten

b Polen und osteuropäische Nachbarländer

c Russland und Staaten der ehemaligen Sowjetunion

zusammengefasst.¹⁶ Unter „Jugoslawien“ wurden Namen aus allen jugoslawischen Nachfolgestaaten zusammengefasst.¹⁷ Um die Fallzahlen für osteuropäische Länder zu erhöhen, wurden Polen und mehrere osteuropäische Nachbarländer zu einer Ländergruppe „Osteuropa“ zusammengefasst.¹⁸ Für die im folgenden beschriebene Validierung des Verfahrens wurde auf einen Paneldatensatz zurückgegriffen, der trotz seiner Größe von nahezu 19.000 Personen pro Welle lediglich für die in Deutschland häufigeren Ausländergruppen ausreichend große Fallzahlen aufweist. Daher beschränken wir uns in dieser Arbeit auf die größten Gruppen von Ausländern in Deutschland.¹⁹ Die Abbildung 1 zeigt zusammenfassend das Verfahren zur Erstellung der Trainingsdatenbank.²⁰

16 Hierzu gehören in diesem Datensatz Estland, Lettland, Litauen, Sowjetunion, Russische Föderation, Ukraine, Weißrussland, Armenien, Aserbaidschan, Kasachstan, Kirgisistan, Tadschikistan und Turkmenistan.

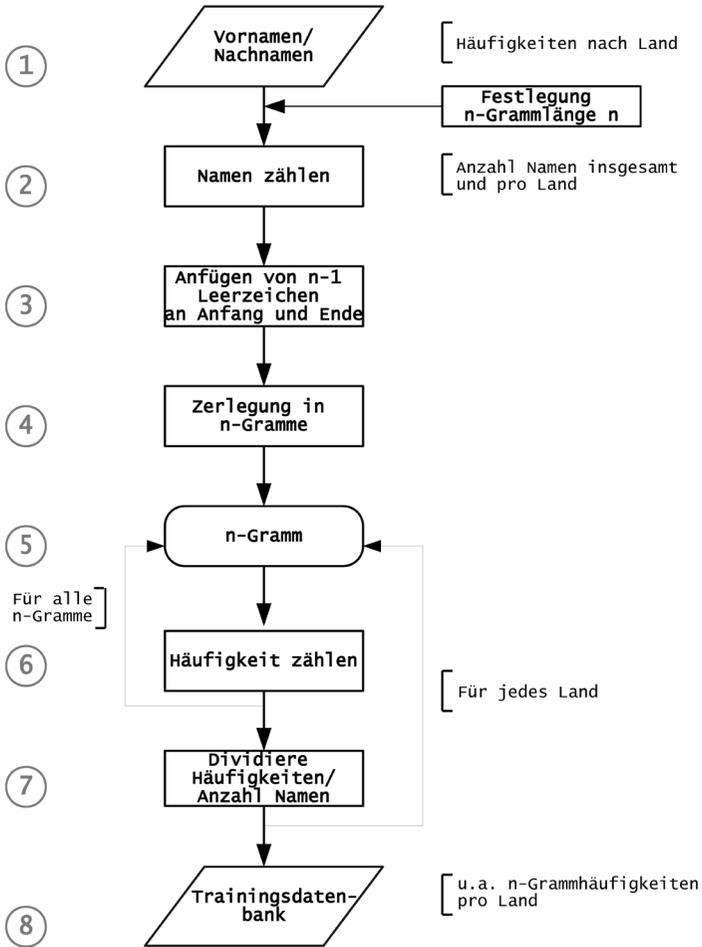
17 Hierzu gehören Bosnien und Herzegowina, Jugoslawien, Kroatien, Makedonien, Albanien und Slowenien. Die ehemalige Bundesrepublik Jugoslawien war zum Zeitpunkt der Erstellung der Namensliste noch nicht in Serbien und Montenegro zerfallen.

18 Hierzu gehören Polen, Bulgarien, Ungarn, Rumänien, ehemalige Tschechoslowakei, Slowakei und die Tschechische Republik.

19 Diese Beschränkung resultiert allein aus der Größe der Subgruppen im Validierungsdatsatz. In anderen Anwendungen kann eine andere – oder auch gar keine – Zusammenfassung der in den Trainingsdaten erfassten Länder zu Gruppen gewählt werden.

20 Da das Verfahren unter anderem für kommerzielle Zwecke (z. B. durch Adresshändler) verwendet werden kann, hat sich die Arbeitsgruppe entschlossen, weder die Trainingsdaten noch die Klassifikationsergebnisse öffentlich zur Verfügung zu stellen.

Abbildung 1 Schematische Darstellung des Verfahrens zur Erstellung der Trainingsdatenbank



5 Validierung des neuen Verfahrens

Das oben beschriebene Klassifikationsverfahren wurde anhand zweier Datensätze untersucht: Dem Paneldatensatz PASS des IAB und einer Stichprobe „Kriminalitätsfurcht in Hessen“ in Zusammenarbeit mit dem Institut für angewandte Sozialwissenschaft (infas, Bonn). Die Ergebnisse beider Studien werden im Folgenden vorgestellt.

5.1 Das Panel „Arbeitsmarkt und Soziale Sicherung“ PASS

Das „Panel Arbeitsmarkt und Soziale Sicherung“ (PASS) (Promberger 2007, Trappmann et al. 2010) ist eine vom Institut für Arbeitsmarkt- und Berufsforschung (IAB) der Bundesagentur für Arbeit (BA) im Jahr 2007 neu gestartete Haushaltsbefragung.

Das PASS-Panel besteht aus zwei Teilstichproben: Stichprobe 1 besteht aus Haushalten aktueller Leistungsbezieher nach dem SGB II; sie wird zu jeder Welle um Neuzugänge in den Leistungsbezug aufgefrischt. Stichprobe 2 ist eine disproportional geschichtete allgemeine Bevölkerungsstichprobe. In den ausgewählten Haushalten waren alle Personen ab 15 Jahren zu interviewen, Personen über 65 Jahren erhielten einen verkürzten Fragebogen. Die Befragung erfolgte – falls möglich – telefonisch (CATI), bei Bedarf auch persönlich durch Interviewer vor Ort (CAPI).

In den Räumlichkeiten des IAB wurden die vom Datensatz getrennten Vor- und Nachnamen der Respondenten der ersten und zweiten Welle mit dem beschriebenen Verfahren klassifiziert.²¹ Insgesamt lagen aus den ersten beiden Wellen 18.795 Vor- und Nachnamen von Teilnehmern zur Klassifikation vor.

Tabelle 3 zeigt die Häufigkeiten der hier berücksichtigten Länder im PASS. Die insgesamt 1.610 Personen mit ausländischer Staatsangehörigkeit stammen vor allem aus der Türkei (31%), den Staaten der ehemaligen Sowjetunion (15%) und dem ehemaligen Jugoslawien (10%). Nur insgesamt 9% der PASS-Respondenten haben eine ausländische Staatsangehörigkeit; der Anteil der Personen, die außerhalb Deutschlands geboren sind, liegt mit 17% deutlich höher. Personen aus der ehemaligen Sowjetunion stellen hier mit 23% den größten Anteil an den im Aus-

21 Alle Arbeiten an und mit den Namen haben im IAB in Nürnberg stattgefunden; die Namen haben das IAB nicht verlassen, lediglich die aus der Klassifikation der Namen resultierende „geschätzte Nationalität“. Die Namen wurden lediglich zur Klassifikation verwendet; die Autoren hatten zu keinem Zeitpunkt Zugriff auf die Namen.

Tabelle 3 Anzahl Personen, Nationalitäten im PASS, Welle 1 und 2

	Staatsangehörigkeit		Geburtsland	
	Personen	in %	Personen	in %
Deutschland	17.140	91,2	15.757	83,8
Türkei	514	2,7	530	2,8
Russland ^a	246	1,3	697	3,7
Jugoslawien ^b	163	0,9	170	0,9
Osteuropa ^c	114	0,6	444	2,4
Italien	68	0,4	49	0,3
Griechenland	45	0,2	33	0,2
restliche Welt	460	2,5	1.091	5,8
unbekannt	45	0,2	24	0,1
Insgesamt	18.795	100,0	18.795	100,0

a einschließlich Staaten der ehemaligen Sowjetunion

b einschließlich Nachfolgestaaten

c Polen und osteuropäische Nachbarländer

land geborenen Personen dar, gefolgt von in der Türkei (rund 18%) oder in den osteuropäischen Nachbarländern geborenen Personen (15%).²²

5.2 Probleme der Klassifikation der Staatsangehörigkeit und des Migrationshintergrundes

Zur Evaluierung der Namensklassifikation wird ein Vergleichsstandard benötigt, mit dem die automatische Klassifikation verglichen werden kann. Dazu stehen aus dem PASS prinzipiell mehrere Alternativen zur Verfügung: die selbstberichtete Staatsangehörigkeit, das selbstberichtete Geburtsland, gegebenenfalls das Geburtsland der Eltern und Großeltern. Die Wahl jedes dieser Kriterien wäre mit Problemen behaftet (vgl. zusammenfassend Diefenbach/Weiß 2006). Im Zentrum des Inter-

22 Nur rund 44% der außerhalb Deutschlands geborenen Personen haben auch eine ausländische Staatsangehörigkeit, demgegenüber sind rund 83% der Personen mit ausländischer Staatsangehörigkeit nicht in Deutschland geboren. Die Übereinstimmung von Nationalität gegeben das Geburtsland ist für Personen aus dem ehemaligen Jugoslawien am höchsten (88%), gefolgt von Personen, die in Russland (78%) oder der Türkei (71%) geboren wurden. Die Übereinstimmung des Geburtslands gegeben die Nationalität ist für osteuropäische Staatsangehörige am höchsten (93%), gefolgt von griechischen (90%), italienischen und türkischen Staatsangehörigen (jeweils 74%).

esses steht hier aber lediglich das Ausmaß der empirischen Übereinstimmung zwischen diesen Kriterien für eine Klassifikation im Rahmen eines Screenings, dem detailliertere Analysen folgen könnten. Die Bewertung des Klassifikationsverfahren muss sich nach dem verwendeten Konzept der Trainingsdaten richten: Dieses ist hier in Ermangelung anderer Datenquellen die Staatsangehörigkeit.²³

Abgesehen von den unvermeidlichen inhaltlichen Problemen jedes einzelnen Kriteriums für Migrationshintergrund bzw. Staatsangehörigkeit entsteht ein Problem durch die Veränderung der Staatsangehörigkeit oder des Namens im Laufe der Zeit. In Deutschland werden mit sinkender Tendenz etwas mehr als 10% der Ausländer jährlich naturalisiert. Daher werden in PASS knapp 50 Naturalisierungen zwischen Welle 1 und Welle 2 beobachtet. Ebenso sind Nachnamensänderungen möglich: Vor allem bei Frauen sind durch Heirat mit deutschen Ehepartnern falsch negative Klassifikationen möglich.²⁴

Tabelle 4 zeigt die Übereinstimmungen einerseits zwischen Nationalität und Geburtsland sowie andererseits Nationalität bzw. Geburtsland mit dem Herkunftsland der Eltern.²⁵ Aufgrund des hohen Anteils gebürtiger Deutscher stimmen für fast 90% aller Personen die in PASS berichtete Nationalität und das Geburtsland überein. Bei den Ausländern insgesamt stimmt in über 75% der Fälle Nationalität und Geburtsland überein, wobei die Übereinstimmung je nach Geburtsland variiert. Betrachtet man lediglich, ob eine Person eine deutsche oder irgendeine ausländische Staatsangehörigkeit besitzt und in Deutschland oder generell im Ausland geboren wurde, beträgt die Übereinstimmung der beiden Merkmale knapp 90%.²⁶ Diese relativ großen empirischen Übereinstimmungen der Merkmale lassen die Verwendung der (selbstberichteten) Nationalität als Kriterium für die Namensklassifikation nicht unplausibel erscheinen.

23 Da elaboriertere Konzepte Messungen der Einzeldimensionen erfordern, bedürfte es eines Surveys, bei dem neben den Namen diese Einzeldimensionen an einer sehr großen Stichprobe erhoben werden müssten, wenn man auch seltene Populationen klassifizieren wollte. Datensätze dieser Größe können in der Bundesrepublik nur durch die amtliche Statistik erhoben werden, wobei dabei aber dann prinzipiell die Weitergabe „erhebungstechnischer Hilfsmerkmale“ allgemein und von Namen oder daraus abgeleiteten Größen auf nahezu unüberwindliche Datenschutzprobleme stoßen würde.

24 Ca. 10% der jährlich geschlossenen Ehen in Deutschland sind binationale Ehen zwischen deutschem und nicht deutschem Partner (Haug 2010); im PASS bestehen 8.4% der rund 8.700 Paare aus einem deutschen und einem ausländischem Partner.

25 Kappa korrigiert die prozentuale Übereinstimmung um den Anteil der allein aufgrund der Randverteilung zu erwartenden Übereinstimmungen und liegt daher unter dem Prozentsatz der Übereinstimmungen (Schnell/Hill/Esser 2011: 395).

26 Für knapp zwei Drittel der Ausländer stimmt die eigene Nationalität mit der der Mutter bzw. des Vaters überein. Die Übereinstimmung beim Geburtsland liegt bei 47%.

Tabelle 4 Übereinstimmung zwischen Nationalität, Geburtsland und Herkunftsland der Eltern

Übereinstimmung zwischen	Anteil Übereinstimmung	Kappa
Nationalität – Geburtsland (Alle)	0.88	0.54
Nationalität – Geburtsland (Ausland)	0.78	0.73
ausländische Nationalität – im Ausland geboren	0.89	0.53
Nationalität – Herkunft Eltern ^a (Alle)	0.88	0.32
Nationalität – Herkunft Eltern ^a (Ausland) ^b	0.66	0.59
Geburtsland – Herkunft Eltern ^a (Alle)	0.90	0.53
Geburtsland – Herkunft Eltern ^a (Ausland) ^b	0.47	0.34
Herkunftsland Vater – Mutter (Alle)	0.99	0.99
Herkunftsland Vater – Mutter (Ausland)	0.98	0.98

a gegebenfalls der Mutter

b bei ausländischer Herkunft mindestens eines Elternteils

5.3 Klassifikationskriterien

Nach einer Klassifikation der Namen sind vier Ergebnisse möglich (Tabelle 5): Als richtig positiv (*RP*) wird eine Klassifikation dann bezeichnet, wenn eine Person mit ausländischem Namen auch tatsächlich eine ausländische Staatsangehörigkeit besitzt.²⁷ Entsprechend bezeichnet eine richtig negative (*RN*) Klassifikation eine Person mit einem „deutschen“ Namen und deutscher Staatsangehörigkeit. Fehlerhafte, falsch positive Klassifikationen (*FP*) bezeichnen Personen mit deutscher Staatsangehörigkeit, die entweder einen ausländischen Namen tragen, oder deren Namen vom Verfahren fälschlicherweise als ausländisch klassifiziert wurden. Durch falsch positiv klassifizierte Personen entsteht bei Stichproben möglicherweise Overcoverage.²⁸ Falsch negative (*FN*) Klassifikationen bezeichnen Personen mit einer ausländischen Staatsangehörigkeit, die entweder keine ausländischen Namen tragen oder deren Namen fälschlicherweise nicht als ausländisch klassi-

27 Hier muss berücksichtigt werden, dass es genau genommen nicht um die Klassifikation „ausländischer“ Namen geht, sondern um die Klassifikation der Namen von Personen mit ausländischer Nationalität; Namen an sich sind weder „deutsch“ noch „ausländisch“. Das automatische Screeningverfahren beruht aber darauf, dass bestimmte Namen unter Personen deutscher Nationalität häufiger oder weniger häufig auftreten als unter Personen einer anderen Nationalität.

28 Zu den Fehlern durch Under- bzw. Overcoverage vgl. Lessler und Kalsbeek (1992).

Tabelle 5 Mögliche Ergebnisse einer Namensklassifikation

Staatsangehörigkeit	Namensklassifikation	
	deutsch	nicht deutsch
deutsch	Richtig negativ <i>RN</i>	Falsch positiv <i>FP</i>
nicht deutsch	Falsch negativ <i>FN</i>	Richtig positiv <i>RP</i>

fiziert wurde. Durch falsch negativ klassifizierte Personen besteht die Gefahr der Verzerrung der Auswahlgrundlage durch Undercoverage.

Erst mehrere dieser Kriterien gemeinsam erlauben eine Beurteilung der Güte der Klassifikation; jedes Kriterium allein beschreibt eine Klassifikation nur ungenügend. Ob ein Klassifikationsverfahren als gut oder weniger gut betrachtet werden kann, hängt von den Konsequenzen der Falschklassifikationen ab. In der hier beschriebenen Anwendung müssen die Effekte durch die Berücksichtigung fälschlicherweise als „ausländisch“ klassifizierte Namen und die Effekte durch den Ausschluss fälschlicherweise als „deutsch“ klassifizierte Namen gegeneinander abgewogen werden. Bei einer Screeninganwendung wird im Allgemeinen eher eine hohe Rate falsch Positiver akzeptiert, da diese im weiteren Verlauf der Analysen noch ausgeschlossen werden können. Wünschenswert wäre dabei, dass der Anteil der richtig Positiven an den positiv Klassifizierten insgesamt $\left(\frac{RP}{RP+FP}\right)$ möglichst hoch sein sollte.

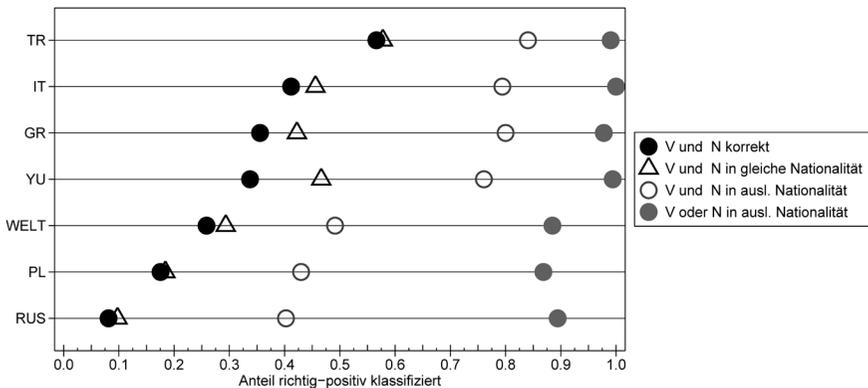
5.4 Kombinationsregeln für die Klassifikation von Vor- und Nachnamen

Das oben beschriebene neue Verfahren klassifiziert Vor- und Nachnamen getrennt. Daher muss eine Entscheidung getroffen werden, wie die separaten Klassifikationen von Vor- und Nachnamen für eine Gesamtklassifikation berücksichtigt werden sollen.²⁹

Bei der Entscheidung über die Art der Kombination von Vor- und Nachnamensklassifikationen muss hier nach den Kosten falscher Klassifikationen entschieden werden. Ein Name könnte als ausländisch gelten, wenn Vor- oder Nachname als ausländisch klassifiziert wird. Dieses Vorgehen erzeugt offensichtlich wenige falsch

29 Das Problem der Kombination von Vor- und Nachnamensklassifikation betrifft ebenso alle lexikonbasierten Ansätze, die – wie im Allgemeinen üblich – getrennte Lexika verwenden. Diese Schwierigkeit ließe sich vermeiden, stünde eine Trainingsdatenbank zur Verfügung, welche die gemeinsamen Häufigkeiten für Vor- und Nachnamen enthielte.

Abbildung 2: Anteil richtig positiver Klassifikationen nach Ländern und Entscheidungsregel. V=Vorname, N=Nachname. Datenbasis: PASS

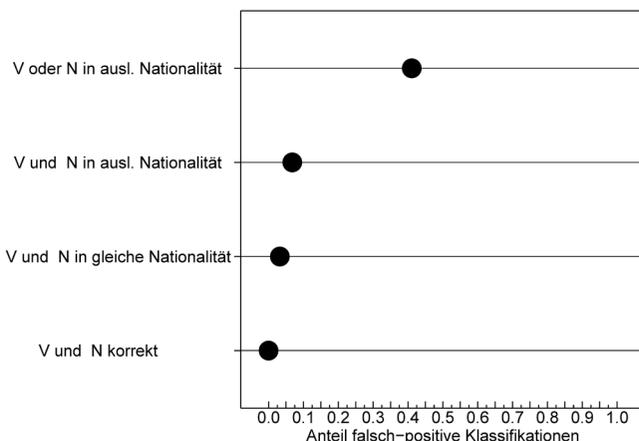


negative, dafür mehr falsch positive Klassifikationen. Klassifiziert man hingegen einen Namen erst dann als ausländisch, wenn Vor- und Nachname als ausländisch klassifiziert werden, dann sind entsprechend mehr falsch Negative und weniger falsch Positive erwartbar. Eine noch rigidere Regel würde fordern, dass Vor- und Nachname in die gleiche ausländische Nationalität klassifiziert werden müssen, damit ein Name als ausländisch gilt. Durch diese Regel werden viele falsch positive Klassifikationen vermieden, allerdings auf Kosten eines möglichen Undercoverage-Bias durch mehr falsch Negative. Die empirische Wirkung unterschiedlicher Klassifikationsregeln wird daher im nächsten Abschnitt untersucht.

5.5 Klassifikationsergebnisse des Verfahrens: Staatsangehörigkeit

Die Abbildungen 2 und 3 zeigen getrennt die Anteile der richtig und falsch positiv klassifizierten Namen anhand der Staatsangehörigkeit. Je nach Klassifikationsregel (Kombination von Vor- und Nachnamen) zeigen sich unterschiedliche Ergebnisse. Der Anteil der richtig positiv klassifizierten Namen liegt im Durchschnitt über alle berücksichtigten Länder bei über 90%, wenn es ausreicht, dass Vor- oder Nachname nicht als deutsch klassifiziert werden (und erzeugt also nur $100\% - 90\% = 10\%$ falsch negativ klassifizierte Namen). Der Anteil richtig positiver Klassifikationen sinkt auf 65%, wenn Vor- und Nachnamen als ausländisch klassifiziert werden. Die dritte Kombinationsregel (Klassifikation von Vor- und Nachnamen in dieselbe Nationalität) führt durchschnittlich zu einem Anteil richtig Positiver von über 35%.

Abbildung 3: Anteil falsch positiver (FP) Klassifikationen nach Entscheidungsregel. V=Vorname, N=Nachname. Datenbasis: PASS



Betrachtet man den Anteil der Fälle, bei denen Vor- und Nachnamen korrekt klassifiziert werden, dann werden immer noch 30% aller Ausländer korrekt klassifiziert.

Deutlich zeigen sich in Abbildung 2 Unterschiede zwischen den hier betrachteten Ländergruppen. Insbesondere für Personen aus der Türkei, aus Italien, Griechenland oder dem ehemaligen Jugoslawien zeigt das hier beschriebene automatische Klassifikationsverfahren sehr gute Ergebnisse. Insbesondere bei der strengeren Klassifikationsregel (Vor- und Nachname positiv klassifiziert) werden unter Angehörigen dieser Länder im Durchschnitt noch 4 von 5 Personen korrekt klassifiziert.

Abbildung 3 zeigt die Kosten der unterschiedlichen Klassifikationsregeln: Je schwächer die Klassifikationsregel wird, desto mehr falsch positive Klassifikationen ergeben sich. Die schlechteste Regel (bereits ein „ausländischer“ Vor- oder Nachname genügt zur Klassifikation als „ausländisch“) führt zu einem Anteil von fast 40% der Personen mit deutscher Staatsangehörigkeit, die fälschlicherweise als ausländisch klassifiziert werden. Jede andere Regel führt aber zu weniger als 10% falsch positiver Klassifikationen. Das hier vorgestellte Verfahren liegt in dieser Hinsicht auf gleicher Höhe wie eine Klassifikation anhand von Namenslisten.³⁰

Je nach beabsichtigter Zielpopulation lassen sich die Kombinationsregeln so wählen, dass der Anteil richtig positiver deutlich oberhalb von 75% und der Anteil falsch positiver Fälle unterhalb von 10% liegt. Sowohl für die klassischen Migra-

30 Simon und Kloppenburg (2007: 152) geben für eine onomastische Stichprobe türkischer Haushalte eine falsch-positiv-Rate von rund 7% an.

Tabelle 6 Übereinstimmungen zwischen Klassifikation und Migrationshintergrund

Übereinstimmung zwischen Klassifikation und Migrationshintergrund	Anteil Übereinstimmung	Kappa
V&N neg., kein Migrationshintergrund	0.67	0.32
V&N neg., selbst zugezogen	0.34	-0.26
V&N neg., min. 1 Elternteil zugezogen	0.45	-0.04
V&N neg., min. 1 Großelternanteil zugezogen ^a	0.47	-0.00
V N pos., kein Migrationshintergrund	0.33	-0.30
V N pos., selbst zugezogen	0.66	0.29
V N pos., min. 1 Elternteil zugezogen	0.55	0.04
V N pos., min. 1 Großelternanteil zugezogen ^a	0.53	0.00
V&N pos., kein Migrationshintergrund	0.19	-0.18
V&N pos., selbst zugezogen	0.86	0.43
V&N pos., min. 1 Elternteil zugezogen	0.85	0.11
V&N pos., min. 1 Großelternanteil zugezogen ^a	0.86	-0.02
V=N pos., kein Migrationshintergrund	0.21	-0.10
V=N pos., selbst zugezogen	0.85	0.28
V=N pos., min. 1 Elternteil zugezogen	0.90	0.11
V=N pos., min. 1 Großelternanteil zugezogen ^a	0.91	-0.02

a Eltern sind in Deutschland geboren

tionspopulationen in Deutschland als auch für die Migranten aus den Nachfolgestaaten der Sowjetunion lassen sich daher mit dem neuen Klassifikationsverfahren Screening-Stichproben konstruieren, die weit effizienter sind als das Screenen reiner Zufallsstichproben.

5.6 Klassifikationsergebnisse des Verfahrens: Migrationshintergrund

Tabelle 6 zeigt den Anteil der Übereinstimmungen zwischen der Klassifikation von Vor- und Nachnamen und dem Migrationshintergrund (unabhängig von der Nationalität). Der Anteil der Übereinstimmungen zwischen den Merkmalen „Klassifikation von Vor- und Nachname als deutsch“ und „kein Migrationshintergrund“ beträgt 0.67. Die Übereinstimmung zwischen einem als „deutsch“ klassifizierten Vor- und Nachnamen und eigenem Migrationshintergrund beträgt hingegen nur 0.34, dies ist deutlich weniger, als durch die Randverteilung allein erwartbar wäre

Tabelle 7 Migrationshintergrund nach Klassifikationsergebnis von Vor- bzw. Nachnamen aller PASS-Teilnehmer, Spaltenprozent.

Migrationshintergrund	V & N negativ	V N positiv	V & N positiv	V & N +identisch ^a
kein Migrationshintergrund	88,9	57,0	20,5	15,7
ausländische Staatsangehörigkeit	1,0	17,6	46,1	51,4
im Ausland geboren	3,4	15,6	21,6	20,3
2 Eltern im Ausland geboren	0,9	2,7	5,3	6,8
1 Elternteil im Ausland geboren	3,5	4,7	5,1	4,6
ausländische Sprache im Haushalt	0,2	0,3	0,5	0,5
alle Großeltern im Ausland geb.	0,1	0,1	0,1	0,0
ein Großelternanteil im Ausland geb.	2,1	2,1	0,9	1,5

a positiv und Vor- und Nachnamen in dasselbe Land klassifiziert

(deutlich durch ein negatives Kappa von -0.26). Sind Eltern oder Großeltern zugezogen, beträgt die Übereinstimmung zwar immer noch ca. 0.45, das Kappa nahe Null zeigt aber, dass diese Übereinstimmung allein durch die Randverteilung erklärbar ist.

Erwartungsgemäß verringert sich die Übereinstimmung zwischen der Klassifikation und dem Migrationshintergrund weiter, wenn nur Vor- oder Nachname als „ausländisch“ klassifiziert wird. Nur bei einem eigenen Migrationshintergrund sind die Übereinstimmungen nicht allein durch den Zufall erklärbar. Sind sowohl Vor- als auch Nachname als ausländisch klassifiziert bzw. Vor- und Nachname in die gleiche Auslandskategorie klassifiziert worden, liegt das Ausmaß der Übereinstimmung der Klassifikation mit dem Migrationshintergrund bei über 85%.

Betrachtet man die diesen Ergebnissen entsprechenden Anteile falsch positiver und falsch negativer Klassifikationen (Tabelle 7), zeigt sich ein vergleichbares Ergebnis: Je strenger die Kombination von Vor- und Nachnamensklassifikation, desto höher der Anteil der Personen mit Migrationshintergrund.

Werden Vor- und Nachnamen als „deutsch“ klassifiziert, haben lediglich 1% der so klassifizierten Personen eine ausländische Staatsangehörigkeit und nur rund 7% haben einen Migrationshintergrund im weiteren Sinne. Werden hingegen Vor- oder Nachnamen als ausländisch klassifiziert, haben bereits knapp die Hälfte der Personen auch tatsächlich einen Migrationshintergrund, darunter bereits rund 18% auch eine ausländische Nationalität. Weitere 16% haben zwar keine ausländische Nationalität, sind aber selbst im Ausland geboren. Die restlichen 10% haben

Tabelle 8 Anteile der Übereinstimmung der Namensklassifikation mit der Nationalität nach Länge der verwendeten n -Gramme.

Land (wahrer Wert)	Nachname		Vorname	
	Bigramme	Trigramme	Bigramme	Trigramme
Deutschland	0.90	0.84	0.87	0.68
Italien	0.69	0.79	0.42	0.51
Türkei	0.63	0.75	0.55	0.77
Griechenland	0.57	0.60	0.49	0.47
Jugoslawien ^a	0.48	0.60	0.31	0.52
Osteuropa ^b	0.31	0.36	0.23	0.42
Russland ^c	0.17	0.14	0.33	0.58
Insgesamt	0.87	0.81	0.83	0.67

a einschließlich Nachfolgestaaten;

b Polen und osteuropäische Nachbarländer

c einschließlich Staaten der ehemaligen Sowjetunion

keinen eigenen Migrationshintergrund, sondern sind Migranten der zweiten (7,5%) oder dritten Generation.

Werden Vor- und Nachnamen als „ausländisch“ klassifiziert, haben insgesamt vier von fünf der so positiv klassifizierten Personen einen Migrationshintergrund. Bei dieser Regel sind dann 46% ausländische Staatsangehörige. Wird das Kriterium zur Klassifikation weiter verschärft, so dass Vor- und Nachnamen nicht nur als ausländisch gelten, sondern beide auch in die gleiche Nationalität klassifiziert werden müssen, haben lediglich nur rund 16% der Personen keinen Migrationshintergrund und über 50% sind ausländische Staatsangehörige. Auch hier zeigt sich, dass bei Wahl der für die eigenen Zwecke geeigneten Klassifikationsregel effiziente Screeningkriterien durch das Verfahren realisiert werden können.

5.7 Zur Wahl von Bigrammen oder Trigrammen zur Klassifikation

Zusätzlich zur Frage, wie die separaten Klassifikationen von Vor- und Nachnamen kombiniert werden, muss eine Entscheidung getroffen werden, ob für die Klassifikation aus den Namen Buchstabenketten aus zwei oder aus drei Buchstaben (Bigramme bzw. Trigramme) verwendet werden sollen.³¹ Tabelle 8 enthält zum Ver-

31 Längere n -Gramme eignen sich für die Klassifikation kaum, da längere n -Gramme für die häufigen kurzen Namen einen exakten und nicht mehr fehlertoleranten Abgleich implizieren.

gleich der Klassifikation durch Bi- oder Trigramme den Anteil korrekter Klassifikationen. Deutlich wird, dass der Anteil korrekter Klassifikationen unter den Ausländern bei der Verwendung von Trigrammen in der Regel höher ist als bei der Verwendung von Bigrammen. Ausnahmen hiervon sind Vornamen bei Russen und Nachnamen bei Personen mit griechischer Staatsangehörigkeit.³²

6 Validierung des Verfahrens anhand einer prospektiven Studie mit einer Stichprobe türkischer Namen

Infas führte im Jahr 2009 im Bundesland Hessen eine telefonische Befragung unter anderem von Personen mit türkischem Migrationshintergrund durch. Für diese Studie wurde das hier beschriebene Verfahren erstmals für eine Stichprobenziehung verwendet. Eine große Stichprobe aus den Nachnamen einer Telefon-CD Hessens wurde mit dem Verfahren klassifiziert.³³ Mit den Telefonnummern, die zu den als „türkisch“ klassifizierten Namen gehörten, konnten 839 Interviews realisiert werden. Da in der Erhebung für die Befragten der Migrationshintergrund erhoben wurde, kann diese Studie für eine weitere teilweise Validierung des Verfahrens verwendet werden.

Tabelle 9 zeigt für alle als türkisch klassifizierten Personen der Stichprobe den Migrationshintergrund bis zur dritten Generation. 12% der als türkisch klassifizierten Personen haben keinen türkischen Migrationshintergrund, 43% besitzen tatsächlich die türkische Staatsangehörigkeit, weitere 13% sind zwar keine türkischen Staatsangehörigen (mehr), sind aber in der Türkei geboren. Weitere 13% haben keinen eigenen Migrationshintergrund, mindestens ein Elternteil ist jedoch in der Türkei geboren. Berücksichtigt man weiter die Herkunft der Großeltern und die Sprache, die im Haushalt gesprochen wird, haben von 839 als türkisch klassifizierte Personen mehr als zwei Drittel (70,8%) einen türkischen Migrationshintergrund im engeren Sinne (eigener Migrationshintergrund oder der Eltern).³⁴ Das Klassifikationsverfahren hat in dieser Stichprobe also erfolgreich die Zahl der

32 Bei Personen mit deutscher Staatsangehörigkeit erzeugt die Zerlegung der Vor- oder Nachnamen in Bigramme jeweils höhere Anteile korrekter Klassifikationen.

33 Die positiv klassifizierten Namen wurden in dieser Studie anschließend von einem Angehörigen der Zielpopulation gesichtet und um wenige „offensichtlich“ falsch-positive Fälle bereinigt. Diese zusätzliche Bereinigung war aufgrund des Auftraggebers unverzichtbar.

34 Von den rund 100 falsch Positiven ohne türkischen Migrationshintergrund haben rund ein Viertel eine andere ausländische Staatsangehörigkeit.

Tabelle 9 Vergleich der Klassifikationen der CATI-Studie in Hessen mit PASS

Klassifikation: türkisch	CATI Hessen	PASS
	%	%
türk. Staatsangehörigkeit	43,4	49,6
in der Türkei geboren	12,8	14,4
beide Eltern in der Türkei geboren	10,7	11,0
ein Elternteil in der Türkei geboren	2,0	0,5
Sprache im Haushalt: Türkisch	1,9	2,0
alle Großeltern NICHT in D geboren	16,9	10,4
kein türk. Migrationshintergrund	12,3	12,1
Anzahl als „türkisch“ klassifiziert	839	960

Kontakte, die notwendig waren um durch Screening eine ausreichend große und zufallsbasierte Stichprobe zu gewinnen, reduzieren können.

Zum Vergleich zeigt Tabelle 9 das Ergebnis im PASS, wenn Vor- oder Nachnamen als türkisch klassifiziert werden. Die Ergebnisse sind hier sehr ähnlich.³⁵ In beiden Stichproben ergeben sich nur rund 12% falsch Positive. Insgesamt bestätigt die hessische CATI-Studie die praktische Einsatzfähigkeit und die Effizienz eines Screenings mit dem beschriebenen Verfahren.³⁶

7 Zusammenfassende Beurteilung

Das hier vorgestellte Verfahren eignet sich zum effizienten Screenen von Ausländern und Migranten in Namenslisten. Je nach den jeweiligen Kosten von Falschklassifikationen lassen sich mehr oder weniger strenge Klassifikationsregeln wäh-

35 Das ist umso erstaunlicher, als für diese Studie anhand der PASS-Daten kein zusätzliches manuelles Review der positiv klassifizierten Fälle stattfand, sondern nur automatisch klassifiziert wurde. In PASS wurden alle türkischen Staatsangehörige vom Verfahren als „türkisch“ klassifiziert, d. h. es gab keine falsch negativen Fälle.

36 In einem vom Bundesministerium für Bildung und Forschung geförderten Projekt „Stichprobenziehung für seltene Populationen“ des Erstautors wurden verschiedene Stichprobenverfahren für Migranten miteinander verglichen (Schnell 2012). Die Ergebnisse sind Gegenstand einer anderen Publikation.

len und entsprechend der jeweiligen Aufgabenstellung eher falsch negative oder eher falsch positive Screeningergebnisse vermeiden.³⁷

Das hier beschriebene vollautomatische namensbasierte Verfahren eignet sich für einige Migrantengruppen und Länder besser als für andere. Insbesondere bietet es sich für die Suche nach Personen an, die aus den klassischen Migrationsländern der Bundesrepublik stammen (Türkei, Italien, Griechenland und Jugoslawien). Für andere Gruppen (z. B. Russen und Angehörige osteuropäischer Nachbarländer) ist das vorgestellte Verfahren auch geeignet, besitzt aber für diese Gruppen eine geringere Effizienz als für die klassischen Migrationsländer. Aber auch bei den weniger geeigneten Subpopulationen liegt die Effizienz des Verfahrens immer noch über dem der reinen Zufallsauswahl; ebenso können auch geringere Auswahleffekte erwartet werden als bei Quoten- oder Schneeballverfahren.

Das Verfahren ermöglicht mit geringem Aufwand in kurzer Zeit ein Screening nach Ausländer- oder Migrantenpopulationen in umfangreichen Auswahlgrundlagen, wie z. B. Einwohnermelderegistern, Telefonbüchern, Patientendateien usw. Ein weiterer Vorteil liegt darin, dass ein vorheriges manuelles Erstellen von Einträgen in einem Namenslexikon vermieden werden kann. Gegenüber den reinen Lexikonverfahren bietet das hier vorgestellte Verfahren durch die Zerlegung der Namen in n -Gramme den Vorteil, dass es fehlertoleranter gegenüber den in Registern häufigen minimalen Schreibvarianten insbesondere ausländischer Namen ist. Eine Erweiterung der grundsätzlichen Einsetzbarkeit des Verfahrens ist durch die Bereitstellung anderer Trainingsdaten zu erreichen. Schließlich ist das Verfahren ohne Modifikationen auch auf andere Länder anwendbar, falls dort entsprechende Trainingsdaten zur Verfügung stehen.

Literatur

Babka von Gostomski, C., 2008: Türkische, griechische, italienische und polnische Personen sowie Personen aus den Nachfolgestaaten des ehemaligen Jugoslawien in Deutschland. Erste Ergebnisse der Repräsentativbefragung „Ausgewählte Migrantengruppen in Deutschland 2006/2007“ (RAM). Working Paper 11, Bundesamt für Migration und Flüchtlinge, Nürnberg.

37 Die Frage der Konsequenzen falsch negativer Klassifikationen bei namensbasierten Verfahren im Allgemeinen oder im Vergleich dieses Verfahrens mit anderen Verfahren überschreitet den Rahmen dieser Arbeit. Eine entsprechende umfangreiche Analyse ist Gegenstand laufender Bemühungen der Arbeitsgruppe (Schnell/Gramlich/Trappmann 2011).

- Bertelsmann Stiftung (Hg.), 2009: Zuwanderer in Deutschland. Ergebnisse einer repräsentativen Befragung von Menschen mit Migrationshintergrund. Gütersloh: Bertelsmann Stiftung.
- Blane, H. D., 1977: Acculturation and Drinking in an Italian American Community. *Journal of Studies on Alcohol* 38 (7): 1324-1346.
- Brettfeld, K. und P. Wetzels, 2007: Muslime in Deutschland. Integration, Integrationsbarrieren, Religion sowie Einstellungen zu Demokratie, Rechtsstaat und politisch-religiös motivierter Gewalt. Berlin: Bundesministerium des Innern.
- Burkhauser, R. V., M. Kreyenfeld und G. G. Wagner, 1997: The Immigrant Sample of the German Socio Economic Panel. Aging Studies Working Paper 7, Maxwell Center for Demography and Economics of Aging, Syracuse, NY.
- Cavnar, W. B. und J. M. Trenkle, 1994: N-Gram-Based Text Categorization. S. 161-175 in: Proceedings of SDAIR-94: 3rd Annual Symposium on Document Analysis and Information Retrieval.
- Diefenbach, H. und A. Weiß, 2006: Zur Problematik der Messung von „Migrationshintergrund“. *Münchener Statistik* 3: 1-14.
- Ecob, R. und R. Williams, 1991: Sampling Asian Minorities to Assess Health and Welfare. *Journal of Epidemiology and Community Health* 45: 93-101.
- European Union Agency for Fundamental Rights, 2009: EU-MIDIS Technical report: methodology, sampling and fieldwork. European Union minorities and discrimination survey. Budapest: Elanders Hungary Kft.
- Galonska, C., M. Berger und R. Koopmans, 2004: Über schwindende Gemeinsamkeiten: Ausländer- versus Migrantenforschung. Technischer Bericht, Wissenschaftszentrum Berlin (WZB), Berlin.
- Granato, N., 1999: Die Befragung von Arbeitsmigranten: Einwohnermeldeamt-Stichprobe und telefonische Erhebung? *ZUMA-Nachrichten* 45 (23): 44-60.
- Haisken-DeNew, J. und J. R. Frick, 2005: Desktop Companion to the German Socio-Economic Panel (SOEP). Berlin: DIW Berlin.
- Halm, D. und M. Sauer, 2005: Freiwilliges Engagement von Türkinnen und Türken in Deutschland. Projektbericht, Stiftung Zentrum für Türkeistudien an der Universität Duisburg-Essen, Essen.
- Haug, S., 2010: Interethnische Kontakte, Freundschaften, Partnerschaften und Ehen von Migranten in Deutschland. Working Paper 33, Bundesamt für Migration und Flüchtlinge, Nürnberg.
- Haug, S. und F. Swiaczny, 2003: Migrations- und Integrationsforschung in der Praxis. Das Beispiel BiB-Integrationsurvey. *Zeitschrift für Angewandte Geographie* 27 (1): 16-20.
- Humpert, A. und K. Schneiderheinze, 2000: Stichprobenziehung für telefonische Zuwandererumfragen. Einsatzmöglichkeiten der Namensforschung. *ZUMA-Nachrichten* 24 (47): 36-64.
- Infas, 2009: Methodenbericht des Projekts Kriminalitätsfurcht in Hessen. infas: Bonn.
- Jurafsky, D. und J. H. Martin, 2009: *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. New Jersey: Pearson.
- Kalton, G., 2009: Methods for oversampling rare subpopulations in social surveys. *Survey Methodology* 35 (2): 125-141.
- Kalton, G. und D. Anderson, 1986: Rare Populations. *Journal of the Royal Statistical Society. Series A* 149 (1): 65-82.

- Konstantopoulos, S., 2007: What's in a Name? In: Proceedings of the Computational Phonology Workshop. 6th International Conference on Recent Advances in NLP. Bulgarien. <http://arxiv.org/abs/0710.1481>.
- Lauderdale, D. S., 2006: Birth Outcomes for Arabic-Named Women in California Before and After September 11. *Demography* 43 (1): 185–201.
- Lessler, J. T. und W. D. Kalsbeek, 1992: *Nonsampling Errors in Surveys*. New York: Wiley.
- Mammy, U. und J. Sattig, 2002: Determinanten und Indikatoren der Integration und Segregation der ausländischen Bevölkerung (Integrationsurvey). Projekt- und Materialdokumentation. Materialien zur Bevölkerungswissenschaft des Bundesinstituts für Bevölkerungsforschung, Nr. 105a. Wiesbaden: Bundesinstitut für Bevölkerungsforschung.
- Passel, J. S. und D. L. Word, 1993: Constructing the List of Spanish Surnames for the 1980 Census: An Application of Bayes' Theorem (Paper presented at the Annual Meeting of the Population Association of America) 1980. Technical Working Paper 4, Population Division, U.S. Bureau of the Census, Washington D.C.
- Perkins, C. R., 1993: Evaluating the Passel-Word Spanish Surname List: 1990 Decennial Census Post Enumeration Survey Results. Technical Working Paper 4, Population Division, U.S. Bureau of the Census, Washington D.C.
- Promberger, M. (Hg.), 2007: *Neue Daten für die Sozialstaatsforschung. Zur Konzeption der IAB-Panelerhebung „Arbeitsmarkt und Soziale Sicherung“*. IAB- Forschungsbericht (12). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung.
- Rother, N., 2010: *Das Integrationspanel. Ergebnisse einer Befragung von Teilnehmenden zu Beginn ihres Alphabetisierungskurses*. Working Paper 29, Bundesamt für Migration und Flüchtlinge, Nürnberg.
- Sachverständigenrat deutscher Stiftungen für Integration und Migration (SVR) (Hg.), 2010: *Einwanderungsgesellschaft 2010. Jahresgutachten 2010 mit Integrationsbarometer*. Berlin: Sachverständigenrat deutscher Stiftungen für Integration und Migration.
- Salentin, K., 2007: Die Aussiedler-Stichprobenziehung. *Methoden – Daten – Analysen* 1 (1): 25–44.
- Schnell, R., 2009: *Wie man Nadeln in Heuhaufen findet. Stichprobenverfahren für seltene und sehr seltene Bevölkerungsgruppen*, Vortragsmanuskript, Universität Duisburg-Essen, Duisburg.
- Schnell, R., 2012: *Empirische Untersuchung eines CATI-Stichprobendesigns für seltene Populationen. Endbericht für das Bundesministerium für Bildung und Forschung. Technischer Bericht*, Universität Duisburg-Essen, Duisburg.
- Schnell, R., T. Gramlich und M. Trappmann, 2011: *Potential Undercoverage and Bias in Name-based Samples of Foreigners*. Vortrag auf der ASA Spring Methodology Conference, Mai 2011, Tilburg.
- Schnell, R., P. B. Hill und E. Esser, 2011: *Methoden der empirischen Sozialforschung*, 9. Auflage, München: Oldenbourg.
- Schnell, R., T. Gramlich, T. Bachteler, J. Reiher, M. Trappmann, M. Smid und I. Becher, 2012: *Ein neues Verfahren für namensbasierte Zufallsstichproben von Migranten*. Working Paper Wp-grlc-2012-02, German Record Linkage Center, Nürnberg.
- Schwartz, K. L., A. Kulwicki, L. K. Weiss, H. Fakhouri, W. Sakr, G. Kau und R. K. Severson, 2004: Cancer among arab americans in the metropolitan Detroit area. *Ethnicity & Disease* 14 (1): 141–146.
- Shackleford, M., 1998: *Actuarial Note: Name Distributions in the Social Security Administration Area, August 1997*. Social Security Administration Actuarial Note 139, Social Security Administration. Office of the Chief Actuary, Baltimore, MA.

- Simon, E. und G. Kloppenburg, 2007: Das Fernsehpublikum türkischer Herkunft – Fernsehnutzung, Einstellungen und Programmerwartungen. Ergebnisse einer Repräsentativbefragung in Nordrhein-Westfalen. *Media-Perspektiven* 3: 142-152.
- Statistisches Bundesamt, 2011a: Bevölkerung und Erwerbstätigkeit. Bevölkerung mit Migrationshintergrund. Ergebnisse des Mikrozensus 2010. Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt, 2011b: Bevölkerung und Erwerbstätigkeit. Ausländische Bevölkerung. Ergebnisse des Ausländerzentralregisters. Wiesbaden: Statistisches Bundesamt.
- Trappmann, M., S. Gundert, C. Wenzig und D. Gebhardt, 2010: PASS: A household panel survey for research on unemployment and poverty. *Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften* 130 (4): 609-622.
- Winkler, W. E. (2009). Record linkage. S. 351-380 in: D. Pfeffermann und C. Rao (Hg.), *Handbook of Statistics. Sample Surveys: Design, Methods and Applications*. Bd. 29A. Amsterdam: Elsevier.

Anschrift des Autors

Rainer Schnell
Universität Duisburg-Essen
Institut für Soziologie
Lotharstraße 65
47057 Duisburg
E-Mail: rainer.schnell@uni-due.de

Stichprobenziehung von Ehepaaren

Sampling Couples

*Ein Praxisbericht zur
Erhebung von seltenen
Paarkombinationen*

*An Approach to Identifying
Rare Couple Combinations*

David Glowsky

Zusammenfassung

Sozialwissenschaftliche Fragestellungen beziehen sich häufig nicht auf Einzelpersonen, sondern auf Kombinationen von Personen im Familienverband. Für seltene Kombinationen, wie gleichgeschlechtliche Paare oder Familien mit Zwillingen erweist sich die Ziehung einer Zufallsstichprobe jedoch oft als problematisch, da herkömmliche Screeningverfahren hier an ihre Grenzen stoßen. In diesen Fällen ist es wirtschaftlicher und geht es schneller, entsprechende Personenkonstellationen im Melderegister zu identifizieren, indem man Daten des Familienverbandes verknüpft und entsprechende Kombinationen direkt aus dem Melderegister herausfiltert. Der vorliegende Beitrag stellt diesen Ansatz exemplarisch anhand einer Befragung von binationalen Ehepaaren vor. In einem ersten Teil werden Schwierigkeiten bei der Ziehung seltener Paarkombinationen diskutiert. Zweitens wird die Möglichkeit erläutert, Daten von Familienmitgliedern im Melderegister zu verknüpfen. Die Genauigkeit dieses Ansatzes wird anhand der Berliner Paarstudie ausgewer-

Abstract

Research questions in the social sciences often do not focus on individuals, but rather on combinations of persons within a family. For rare combinations, such as same sex couples and families with twins, it often proves difficult to draw random samples, since conventional screening methods reach their limits when the target population is too small. In such cases, another method promises to be faster and more cost-effective: combinations of family members can be identified in the population register by linking single entries, which enables us to filter certain combinations on the basis of attributes of several family members. The article presents this approach, using a survey of binational couples as an example. In a first step, difficulties of sampling rare couple combinations are discussed. Secondly, the paper presents the alternative approach, in which data of several family members are combined within the population register. The accuracy of this approach is evaluated, using the Berlin Couple Survey, in which name-based tech-



tet, bei der zusätzlich eine onomastische Filterung zum Einsatz kam. Als Ergebnis zeigt sich, dass sich die meisten Kombinationen von Ehepartnern deutscher und ausländischer Herkunft treffsicher ermitteln lassen, während Beifang vor allem durch Einbürgerungen und deutsche Namensformen bedingt ist. Abschließend werden weitere Anwendungsmöglichkeiten des Verfahrens diskutiert.

niques supplemented the selection process. The results show that most combinations of couples with an ethnic German partner and a foreign-born partner can be identified accurately, with false identification being mainly due to naturalization and Germanized names. Finally, further possible applications of this sampling approach are discussed.

1 Einleitung¹

Quantitative Sozialforschung lebt von verallgemeinerungsfähigen und unverzerrten Datensätzen. Sofern zu einer Fragestellung keine amtlichen Daten vorliegen, werden die benötigten Informationen meist mittels einer Bevölkerungsbefragung erhoben. Dazu gilt es zunächst, eine Zufallsstichprobe aus der Zielpopulation zu gewinnen. Für den Großteil der Forschungsfragen lassen sich relativ leicht Zufallsstichproben gewinnen, da sie sich auf die Gesamtbevölkerung eines bestimmten Territoriums beziehen. Eine Zufallsauswahl aus der Gesamtbevölkerung lässt sich beispielweise über eine zufällige Auswahl von Telefonnummern, das Random-Route-Verfahren oder eine einfache Melderegisterstichprobe realisieren. Die Kunst besteht in diesen Fällen nicht darin, eine besonders große Stichprobe zu erreichen, sondern eine möglichst kleine Auswahl von wenigen tausend Befragten so zusammenzusetzen, dass sie stellvertretend für die Gesamtbevölkerung stehen kann. Für die Bearbeitung allgemeiner Fragestellungen, die sich auf die Gesamtbevölkerung beziehen, ist es häufig auch nicht notwendig, jeweils eine neue Befragung durchzuführen. Stattdessen kann man oft auf die existierende Dateninfrastruktur zurückgreifen, die umfangreiche Befragungen der Bevölkerung bereithält, wie das Sozio-oekonomische Panel (Wagner/Frick et al. 2007), den Mikrozensus (Lüttinger/Riede 1997) oder den ALLBUS (Terwey 2000).

Mit zunehmender Ausdifferenzierung der Sozialwissenschaften wächst aber auch die Zahl differenzierter Fragestellungen, die sich nicht auf die Gesamtbevöl-

1 Dieser Artikel ist aus einem Projekt entstanden, das von der Fritz Thyssen Stiftung und der Studienstiftung des deutschen Volkes finanziert wurde. Teile des Beitrags wurden Kapitel 4 in Glowsky (2011) entnommen. Ich danke Jürgen Schupp, Lars Breuer, den Herausgebern der MDA und zwei anonymen Gutachtern für hilfreiche Kommentare. Sabine Hübgel, Lisette Schneider und Teresa Zeckau haben das Projekt als Hilfskräfte unterstützt.

kerung, sondern auf bestimmte Teilgruppen beziehen. Abhängig davon, wie klein diese Teilgruppe ist, kann sich die Ziehung einer Zufallsauswahl als problematisch erweisen. Dies ist beispielsweise dann der Fall, wenn die Zielpopulation nicht aus Einzelpersonen, sondern aus seltenen Kombinationen von Personen besteht. In der Familiensoziologie kommen häufiger Fragestellungen auf, die sich auf Personenkonstellationen innerhalb eines Familienverbandes beziehen. So interessiert sich die Scheidungsforschung für Paare, die sich getrennt haben oder bei denen eine Trennung bevorsteht. Für Studien zur Fertilität sucht man möglicherweise ältere kinderlose Paare oder kinderreiche Familien. Aber auch andere Forschungszweige haben Bedarf an Stichproben mit seltenen Personenkonstellationen. So dienen beispielsweise gleichgeschlechtliche Paare als Forschungsobjekte der Genderforschung, die Jugendforschung kann hingegen Adressen von Familien mit Stiefeltern suchen. Die Zwillingsforschung wiederum benötigt Stichproben von Familien mit Kindern, die am gleichen Tag geboren wurden.

Diese Populationen sind häufig so klein, dass Screening-Verfahren an ihre Grenzen stoßen. In diesen Fällen ist es wirtschaftlicher und geht es vermutlich auch schneller, entsprechende Personenkonstellationen im Melderegister zu identifizieren, indem man die Daten des Familienverbandes verknüpft. Auf diese Weise können Eigenschaften von Partnern bzw. Eltern und Kindern im Familienverband bei der Stichprobenziehung gemeinsam einbezogen und direkt aus dem Melderegister herausgefiltert werden.

Der vorliegende Beitrag präsentiert diesen Ansatz exemplarisch anhand der Berliner Paarstudie, einer Befragung von binationalen Ehepaaren. Der folgende Abschnitt (2) diskutiert Schwierigkeiten bei der Stichprobenziehung von seltenen Populationen. Der Hauptteil (3) beschreibt zunächst das übliche Vorgehen bei einer Melderegisterstichprobe und erläutert dann die Möglichkeit, Daten im Familienverband zu verknüpfen. Daran anschließend geht der Beitrag näher auf methodische Aspekte bei der Stichprobenziehung der betreffenden Studie ein. Ein Fazit (4) beurteilt schließlich den Mehrwert des hier präsentierten Vorgehens und wirft einen Ausblick auf weitere Anwendungsmöglichkeiten.

2 Schwierigkeiten der Stichprobenziehung bei seltenen Paarkombinationen

Die Grundlage für diesen Bericht bildet ein von der Fritz Thyssen Stiftung gefördertes Projekt, das von 2006 bis 2010 am Institut für Soziologie der Freien Universität Berlin durchgeführt wurde. Das Projekt war im Forschungsfeld über Heiratsmigra-

tion angesiedelt. Ziel war es, Mechanismen der Partnerwahl bei Eheschließungen zwischen deutschen Männern und Heiratsmigrantinnen aus dem ärmeren Ausland zu untersuchen. Eine Sekundäranalyse eines existierenden Datensatzes war ausgeschlossen, da in Deutschland keine Datensätze mit ausreichender Fallzahl und partnerwahlspezifischen Informationen existierten. Für das Projekt war es deshalb notwendig, einen neuen Datensatz zu erstellen. Da eine mündliche interviewergestützte Befragung aus Kostengründen ausschied, sollte stattdessen eine schriftliche Befragung von zusammenlebenden Ehepaaren durchgeführt werden, in denen der Mann in Deutschland geboren war oder seit Geburt die deutsche Staatsbürgerschaft innehatte. Ferner sollten der Mann und mindestens ein Elternteil Deutsch als Muttersprache haben. In der deutschen Unterstichprobe sollte die Frau in Deutschland geboren sein oder seit Geburt die deutsche Staatsbürgerschaft haben. In den anderen Gruppen sollte die Frau im Ausland geboren sein und eine ausländische Staatsbürgerschaft innehaben. Um diese Befragung durchführen zu können, sollte möglichst eine Zufallsstichprobe dieser Grundgesamtheit gezogen werden.

Allerdings umfassen Paare mit deutsch-ausländischer Herkunft nur einen sehr kleinen Teil der deutschen Bevölkerung. Im Mikrozensus 2003 finden sich 3106 Männer, die in Deutschland geboren sind und die deutsche Staatsangehörigkeit innehaben und deren Frau im Ausland geboren ist. Diese 3106 Männer entsprechen 0,62 % der Gesamtpopulation. Im umgekehrten Fall sind es sogar nur 2893 Frauen (0,58 % der Grundgesamtheit), die mit einem im Ausland geborenen Mann verheiratet sind. Bei einem derart geringen Anteil der Zielgruppe an der Bevölkerung handelt es sich um eine *seltene Population* (Schnell/Hill et al. 2008: 294). Seltene Populationen sind oft nur mit großem Aufwand zu ermitteln, insbesondere Zufallsstichproben gestalten sich hierbei schwierig. Im glücklichsten Fall existiert eine vollständige Liste der Zielpopulation, aus der sich dann eine Zufallsstichprobe ziehen lässt.

Alternativ kann man auf Screeningverfahren zurückgreifen, die in der Gesamtpopulation systematisch nach Angehörigen der Zielpopulation suchen und eine Zufallsstichprobe gewährleisten. Ein Telefonscreening mit dem Random-Digit-Dial-Verfahren (Schnell/Hill et al. 2008: 294ff.; Gabler/Häder 1997; Häder/Gabler 1998) sowie das Random-Route-Verfahren (Hoffmeyer-Zlotnik 2003; 1997; Alt/Bien et al. 1991) kamen für die Berliner Paarstudie nicht infrage. Bei diesen Verfahren wäre im vorliegenden Fall eine sehr hohe Zahl an Kontaktversuchen notwendig gewesen, um ausreichend Paare mit deutschem Mann und ausländischer Frau zu erreichen. Der dabei entstehende Zeit- und Kostenaufwand hätte die zur Verfügung stehenden Mittel überschritten. Eine zeit- und kostensparende Methode, die insbesondere für die Stichprobenziehung bei Migrantenforschungen

eingesetzt wird, ist das onomastische Verfahren (Humpert/Schneiderheinze 2000; 2002; Salentin 2002). Bei diesem Verfahren werden typische Vor- und Nachnamen in einem öffentlichen Verzeichnis – z. B. dem Telefonbuch – gesucht. Dieses Verfahren ließ sich für die Berliner Paarstudie jedoch nicht anwenden, da Paare mit einem deutschen Ehemann gesucht wurden. Ehefrauen nehmen auch heute noch größtenteils den Nachnamen des Mannes an. Eine Analyse des Melderegisters der Stadt Berlin zeigt, dass 95 % der Ehefrauen thailändischer Nationalität, 30 % der brasilianischen Ehefrauen, 79 % der polnischen Ehefrauen und 68 % der russischen Ehefrauen den gleichen Namen tragen wie der Ehemann. Da die Frauen fast ausschließlich den deutschen Namen des Mannes annehmen, würden zu viele Paare nicht erfasst, weshalb auch dieses Verfahren ausgeschlossen wurde.²

Zwei weitere Alternativen zur Stichprobenziehung bei seltenen Populationen sind das Schneeballverfahren (Gabler 1992; Diekmann 2008: 400; Schnell/Hill et al. 2008: 300) und die Suche in speziellen Kontexten. Beide Verfahren führen jedoch zu erheblichen Verzerrungen. Beim Schneeballverfahren besteht die Stichprobe schließlich aus einem oder mehreren persönlichen Netzwerken. Da Menschen in der Regel soziale Kontakte mit Anderen pflegen, die über ähnliche Eigenschaften verfügen, kommt es in der Folge zu einer systematischen Überrepräsentation bestimmter Merkmale (Salentin 1999: 124). Zudem lassen sich seltene Populationen häufig in speziellen Kontexten antreffen. Der Rückgriff auf derartige *foci of activity* (Feld 1981) geht jedoch in aller Regel mit einer Verzerrung mehrerer Merkmale einher. Niesner (2000) befragt lediglich ausländische Ehefrauen, die in Beratungszentren Hilfe gesucht haben; Paare, die gut miteinander auskommen, fallen aus dem Raster heraus.

Um eine Stichprobe aus der hier anvisierten seltenen Population zu gewinnen, sind die genannten Verfahren ineffektiv bzw. ungeeignet. Es bleibt deshalb zu fragen, ob es Listen gibt, in denen diese Paare verzeichnet sind. Im Ausländerzentralregister (AZR) sind alle dauerhaft in Deutschland lebenden Personen verzeichnet, die nicht die deutsche Staatsangehörigkeit besitzen (Babka von Gostomski/Pupeter 2008). Das Verzeichnis speichert Angaben zu Geschlecht, Staatsangehörigkeiten und Familienstand, womit verheiratete Frauen bestimmter Staatsangehörigkeiten ermittelt werden könnten. Allerdings sind im AZR keine Informationen zum Ehegatten gespeichert, weshalb sich aus dem Verzeichnis weder ermitteln lässt, ob der Ehegatte noch lebt, noch ob beide Partner zusammenleben, noch welche

2 Die Fallzahl beträgt 1.000 Paare pro Gruppe, mit Ausnahme der brasilianischen Frauen (451). Fast ausschließlich übernehmen die Frauen den Namen des Mannes. In den ausländischen Gruppen ließen sich nur zwei Fälle ermitteln, in denen ein Mann den Namen der Frau angenommen hat.

Staatsangehörigkeit der Ehepartner hat.³ Einer Befragung müsste also wieder ein aufwendiges Screening vorhergehen.

Es wurde deshalb entschieden, eine Melderegisterstichprobe zu ziehen. Die Melderegister der Gemeinden bilden eine solide Datenbasis für die hier anvisierte Stichprobe, da sie über die wichtigsten Merkmale für die Identifizierung verfügen und eine Suche nach Kombinationen von Ehepartnern ermöglichen. Der folgende Abschnitt stellt ein Verfahren vor, das mit einer Stichprobe aus dem Melderegister treffsicher Anschriften von binationalen Ehepaaren ermittelt.

3 Stichprobe aus dem Melderegister

3.1 Eine Einführung in Melderegisterstichproben

Die Melderegister sind aufgrund der gesetzlichen Meldepflicht das umfassendste Verzeichnis der deutschen Wohnbevölkerung mit relativ hoher Aktualität. Theoretisch sind im Melderegister alle in der Bundesrepublik lebenden Personen erfasst. Aus mehreren Gründen stimmen die Melderegister nicht vollständig mit der Wohnbevölkerung überein. Sie beinhalten einerseits Anschriften, die nicht mehr gültig sind (u. a. wegen Tod, Umzug), andererseits sind nicht alle Personen an ihrem tatsächlichen Wohnort registriert. Der Zensus 2001 (Statistische Ämter des Bundes und der Länder 2004) hat die Qualität der Melderegister der Gemeinden geprüft. Es zeigte sich, dass bundesweit 4,1 % der in den Melderegistern verzeichneten Haushalte nicht mehr existieren; die Zahl der Haushalte, die umgekehrt nicht im Melderegister enthalten sind, liegt bei 1,7 %. Die Zahl der Karteileichen kann um den Anteil „temporärer“ Karteileichen bereinigt werden, indem die Registerdaten zu einem späteren Zeitpunkt noch einmal abgeglichen werden. Durch dieses Verfahren können Fälle identifiziert werden, in denen z. B. eine Person zum Stichtag verstorben oder während eines Umzuges für kurze Zeit an zwei Orten mit Hauptwohnsitz gemeldet ist. Die bundesweite Karteileichenrate lag nach der Bereinigung bei 2,3 %. Sachsen (0,9 %), Thüringen (1,1 %) und Mecklenburg-Vorpommern (1,2 %) weisen die niedrigsten Raten an verbleibenden Karteileichen auf, in Berlin (5,8 %), dem Saarland (6,3 %) und Hamburg (5,4 %) liegen diese Raten am höchsten. Stichprobenziehungen aus dem Melderegister führen prinzipiell zu höheren Fehlerquoten in großen Städten, während sie in kleinen Gemeinden eine bessere Datenqualität erwarten lassen. Das Melderegister weicht damit vom Ideal eines

3 Telefonat mit dem Ansprechpartner für das Ausländerzentralregister beim Bundesamt für Migration und Flüchtlinge, 30.06.2011.

Vollverzeichnisses aller in Deutschland lebenden Personen ab, weist aber insgesamt eine hohe Übereinstimmung mit den tatsächlich existierenden Haushalten auf.⁴

Die rechtliche Grundlage für Stichprobenziehungen bildet das Melderechtsrahmengesetz (MRRG) des Bundes, auf dem die jeweiligen Meldegesetze der 16 Bundesländer beruhen. In der Fassung vom 19.04.2002 ermöglicht § 21 Abs. 3 eine sogenannte „Gruppenauskunft“: „Melderegisterauskunft über eine Vielzahl nicht namentlich bezeichneter Einwohner (Gruppenauskunft) darf nur erteilt werden, wenn sie im öffentlichen Interesse liegt.“ Dieselbe Formulierung findet sich auch in den Meldegesetzen der Länder. Universitäten und andere öffentliche Forschungseinrichtungen erfüllen die Anforderung des öffentlichen Interesses. Ferner sind Universitäten öffentliche Stellen, denen noch etwas weitere Auskunftsrechte eingeräumt werden.

Die Melderegister sind auf der Gemeinde- bzw. Landkreisebene angesiedelt. Um eine Stichprobe aus einem Melderegister zu ziehen, muss jede Einwohnermeldebehörde jeweils einzeln kontaktiert werden. Dies führt zu einem erheblichen Administrationsaufwand auf der Seite der Forscher. Zudem entstehen bei jeder Ziehung neue Kosten. Der Aufwand und die Kosten lassen sich erheblich verringern, indem nur eine kleine Zahl von Einwohnermeldeämtern in die Stichprobenziehung einbezogen wird. Umfragen, die sich auf die gesamte Bundesbevölkerung beziehen, verwenden deshalb in der Regel zweistufige, geschichtete Stichproben (Albers 1997). In der ersten Stufe wird zunächst eine Zufallsstichprobe von Gemeinden gezogen.⁵ In der zweiten Stufe wird dann aus den Registern dieser Gemeinden eine Zufallsstichprobe von Personen gezogen.

Der ALLBUS 2008 z. B. wählte in der ersten Stufe bundesweit 162 „Sample-Points“ aus (diese entfielen auf 148 Gemeinden, wobei Städte wie Berlin und Ham-

- 4 Personen mit ausländischer Staatsbürgerschaft werden überproportional als Karteileichen in den Melderegistern geführt. So kann z. B. Granato (1999: 58) 93 % der Adressen der deutschen Stichprobe verifizieren, doch nur 78 % der türkischen Stichprobe. Es ist zu vermuten, dass dies auch für die hier untersuchte Population gilt, jedoch in deutlich geringerem Maße, da die Ehefrauen zwar ausländische Staatsangehörigkeit haben, die Ehemänner jedoch in Deutschland geboren sind und deutsche Staatsangehörigkeit haben. Sie weisen vermutlich ein sehr ähnliches Meldeverhalten wie die allgemeine deutsche Bevölkerung auf.
- 5 Mit der Verringerung der Gemeinden geht eine Erhöhung des Klumpeneffektes einher. Die Einwohner einer Gemeinde (Klumpen/Cluster) neigen dazu, einander ähnlich zu sein und sich von den Einwohnern anderer Gemeinden zu unterscheiden. Je homogener die Cluster und je größer die Unterschiede zwischen den Clustern sind, umso größer ist der Genauigkeitsverlust (Kish 1965). Zudem steigt der Designeffekt mit der Zahl der Elemente pro Cluster. Bei einer sehr kleinen Anzahl von Gemeinden ist eine Aussage über die Grundgesamtheit „Bevölkerung der Bundesrepublik Deutschland“ äußerst unpräzise.

burg aufgrund ihrer relativen Größe mit mehreren Sample-Points vertreten waren⁶) und zog in der zweiten Stufe aus jedem Sample-Point 51 Personen (Wasmer/Scholz et al. 2010). Eine Stichprobenziehung in diesem Umfang ist allerdings sehr kostspielig. In der Praxis entscheiden sich Forscher deshalb häufig dafür, die Ziehung auf deutlich weniger Gemeinden zu beschränken. Dies gilt für Pilotstudien (z. B. Granato 1999: Mannheim; Diehl 2007: Essen und München; Brüderl/Castiglioni et al. 2007: Bremen, Mannheim, München, Chemnitz) sowie für kleinere Projekte mit eingeschränkter finanzieller Förderung (z. B. Reuschke 2010: München, Stuttgart, Düsseldorf, Berlin). Wenn die Stichprobe nur wenige Gemeinden enthalten soll oder kann, werden bevorzugt große Städte gewählt, wofür sich generell drei Gründe nennen lassen. Erstens lässt sich so bei der kleinstmöglichen Gemeindezahl die größtmögliche Zahl von Adressen realisieren, was kostensparend wirkt (Diehl 2007). Zweitens kann die Beschränkung auf Großstädte theoriegeleitet sein, z. B. wenn sich die Fragestellung auf Berufspendler in Metropolen bezieht (Reuschke 2010). Drittens handelt es sich häufig um Städte, in denen die Forscher selbst lokalisiert sind, was die Rekrutierung und Kontrolle von Interviewern erleichtert (Brüderl/Castiglioni et al. 2007). Bei diesem Vorgehen wird gleichzeitig ein deutlicher Nachteil hingenommen: Die Stichprobe hat nur beschränkte Aussagekraft, da sie nur Rückschlüsse auf die Bevölkerung in großen Städten zulässt. Über andere Siedlungsgebiete, z. B. ländliche Regionen, sind damit keine Aussagen möglich.

3.2 Verknüpfung von Daten im Familienverband

Die Datenstruktur der Melderegister basiert auf der Erfassung von Einzelpersonen. Diese Datenstruktur reicht für eine Zufallsstichprobe der gesamten Bundesbevölkerung aus. Entsprechend dokumentieren Erfahrungsberichte zur Nutzung des Melderegisters in aller Regel die Stichprobenziehung von Einzelpersonen, ohne Informationen über weitere Familienmitglieder zu berücksichtigen (vgl. u. a. Albers 1997; Brüderl/Castiglioni et al. 2007; Granato 1999; Diehl 2007; Reuschke 2010).

Für die hier anvisierte Stichprobe sind jedoch mehrere Informationen über *beide Ehepartner* notwendig. Im Melderegistereintrag einer Person sind neben den persönlichen Daten auch Daten eventueller Ehegatten und Lebenspartner (Vor- und Familiennamen, Doktorgrad, Tag der Geburt, Anschrift und Sterbetag)

6 Die mehrfache Ziehung von Großstädten in der ersten Stufe ging auf die Gewichtung der Gemeinden nach ihrer Bevölkerungszahl zurück. Diese bestimmte das Gewicht einer Gemeinde, woraus sich z. B. ein Gewicht von 0,33 für Pinneberg (ein oder kein Sample-Point) und ein Gewicht von 2,16 für München (zwei oder drei Sample-Points) ergab (Wasmer/Scholz et al. 2010).

sowie die Daten minderjähriger Kinder (Vor- und Familiennamen, Tag der Geburt und Sterbetag) gespeichert (vgl. § 2, Melderechtsrahmengesetz). Abhängig von der verwendeten Datenbanksoftware sind die Melderegister deshalb in der Lage, Verknüpfungen zwischen Ehegatten und gemeinsam wohnenden Lebenspartnern sowie minderjährigen Kindern herzustellen. Ist diese Verknüpfung hergestellt, so lassen sich die in den einzelnen Einträgen vorhandenen Informationen gemeinsam für die Auswahl verwenden. Auf diese Weise ist es also möglich, bestimmte Kombinationen von Ehegatten und ihren Kindern aus dem Melderegister zu filtern. Die Vorteile dieses Vorgehens liegen auf der Hand. Visiert eine Befragung eine spezielle Population an, die sich aus bestimmten Kombinationen von Ehegatten oder Kindern zusammensetzt, können mittels einer Verknüpfung diejenigen Fälle ausgewählt werden, die auch tatsächlich in die Grundgesamtheit fallen. Unnötiger Beifang kann so vermieden werden.

Dieser Vorteil lässt sich anhand der Anforderungen der Berliner Paarstudie illustrieren: Die Studie zielte auf eine sehr seltene Population ab, die aus Kombinationen von Ehepartnern bestand. Eine Ziehung von Einzelpersonen aus dem Melderegister hätte zu großem Beifang geführt. Zwar wäre in einer Stichprobe mit verheirateten Frauen einer bestimmten Staatsangehörigkeit vermutlich ein großer Prozentsatz „richtiger“ Fälle enthalten gewesen. In Deutschland lebende verheiratete Frauen mit brasilianischer oder thailändischer Staatsbürgerschaft haben vermutlich häufiger einen deutschen Ehemann als einen Mann mit der gleichen Staatsbürgerschaft. Doch neben der Staatsbürgerschaft des Mannes bliebe bei dieser Methode auch die Frage ungeklärt, ob das Ehepaar zusammenlebt. Der Anteil „falscher“ Fälle wäre also nicht unerheblich.

Für die Berliner Paarstudie ließ sich hingegen die Möglichkeit nutzen, eine Stichprobe nach kombinierten Eigenschaften der Ehepartner zu ziehen. Dies ermöglichte eine deutlich genauere Eingrenzung der Fälle. Auf diese Weise konnten *beide* Partner nach ihrer Staatsangehörigkeit bestimmt werden, was die Ermittlung aller denkbaren Kombinationen erlaubt. Weiterhin war es möglich, nur jene Paare zu berücksichtigen, die unter der gleichen Anschrift gemeldet waren. Getrennt lebende Paare wurden so vermieden.⁷ Abschnitt 3.4 geht auf die Vorgehensweise

7 Eine noch bessere Eingrenzung der Stichprobe ist theoretisch über die Filterung nach dem Geburtsort beider Partner möglich. So könnte man eine Stichprobe auf Ehepaare begrenzen, bei denen ein Partner in Deutschland geboren ist und die deutsche Staatsbürgerschaft besitzt, der andere Partner dagegen in einem anderen Land geboren ist und dessen Staatsangehörigkeit trägt. Bei diesem Verfahren ist es jedoch notwendig, möglichst vollständige landesspezifische Ortslisten anzulegen, die alle Schreibvarianten der Orte beinhalten und danach im Melderegister zu suchen (Salentin 2007). Dieses Verfahren ist mit großem Codieraufwand verbunden und erhöht so die Kosten der

genauer ein. Zuvor diskutiert Abschnitt 3.3 jedoch mögliche Ungenauigkeiten, die auftreten, wenn deutsch-ausländische Ehepaare aus dem Melderegister gezogen werden.

3.3 Mögliche Untererfassung der Zielpopulation

Mit einer Melderegisterstichprobe von Ehepaaren gehen mehrere Nachteile einher, die zu einer möglichen Untererfassung der Zielpopulation führen können (Groves/Fowler et al. 2004: 70ff.; Diekmann 2008: 376ff.; Schnell/Hill et al. 2008: 271ff.):

Erstens werden keine Paare erfasst, die im Ausland leben und keinen Wohnsitz in Deutschland gemeldet haben. Es ist möglich, dass sich diese Paare systematisch von Paaren unterscheiden, die sich nach der Heirat entscheiden, in Deutschland zu wohnen. Vermutlich sind deutsche Partner, die der Kultur des ausländischen Partners gegenüber aufgeschlossener sind, auch eher bereit, in dessen Herkunftsland zu leben. Dies verzerrt die Stichprobe auch für andere Merkmale, die mit kultureller Offenheit korrelieren.

Zweitens sind bei diesem Verfahren Paare mit eingebürgerten Partnern ausgeschlossen. Auch hier kann es zu einer Verzerrung kommen, wenn ein Zusammenhang zwischen der Bereitschaft zur Einbürgerung und hypothesenbezogenen Merkmalen besteht.

Drittens sind geschiedene Paare im Melderegister nicht verzeichnet. Die Scheidungswahrscheinlichkeit hängt vom Grad der kulturellen Differenz zwischen den Partnern ab (Kalmijn/Graaf et al. 2005; Roloff 1998). Paare mit besonders großen Differenzen in Werteinstellungen dürften deshalb zwischenzeitlich häufiger geschieden und in der Melderegisterstichprobe unterrepräsentiert sein.

Wie problematisch diese Einschränkungen für die Befragungsergebnisse tatsächlich sind, hängt von der Fragestellung der Studie ab. Wenn sich die Fragestellung z. B. nur auf verheiratete, in Deutschland lebende Paare mit unterschiedlicher Staatsbürgerschaft bezieht, führen die genannten Einschränkungen nicht zu Undercoverage. Wenn die Studie jedoch untersuchen möchte, welche deutschen Staatsbürger Partner aus dem Ausland heiraten, dann sollte man auch Paare befragen, die im Ausland leben, bei denen der ausländische Partner eingebürgert ist oder

Fortsetzung Fußnote 7

Stichprobenziehung erheblich. Personen mit Geburtsorten, die in der Ortsliste nicht aufgeführt sind, werden bei diesem Verfahren zudem fälschlicherweise nicht einbezogen, was bei einer so seltenen Stichprobe wie den hier untersuchten deutsch-ausländischen Ehen unbedingt vermieden werden sollte.

die bereits geschieden sind – diese Paare sind allerdings nur mit sehr großer Mühe bzw. nicht mittels einer Zufallsstichprobe zu ermitteln.

3.4 Stichprobenziehung der Berliner Paarstudie

Die Berliner Paarstudie beschränkte sich auf eine Stichprobe aus dem Berliner Melderegister. Das geht vor allem auf die Kosten für eine Gruppenauskunft zurück, die in vielen Städten relativ hoch sind. Salentin (2007: 37) schätzt die durchschnittlichen Gebühren auf 200 bis 1000 Euro je Kommune, mit Ausreißern von über 10000 Euro. Da die Studie mit einem eingeschränkten Budget auskommen musste, war es leider notwendig, die Stichprobe auf ein Melderegister zu beschränken. Die Wahl fiel auf Berlin, da es sich hierbei mit 3,4 Mio. Einwohnern (Amt für Statistik Berlin Brandenburg 2009) um das größte Verzeichnis in Deutschland handelt. Zudem liegt aufgrund des hohen Ausländeranteils eine hohe Dichte der gesuchten Ehepaare vor. Darüber hinaus hat Berlin den Vorteil, dass hier langjährige BRD-Bürger und ehemalige DDR-Bürger gemeinsam anzutreffen sind.

Die Stichprobenziehung in Berlin hat zur Folge, dass direkte Schlussfolgerungen aus dem Datensatz nur für die Berliner Bevölkerung möglich sind. Verallgemeinerungen auf andere Teile der Bundesbevölkerung müssen von einer starken Ähnlichkeit mit der Berliner Bevölkerung ausgehen. Dies dürfte für andere Großstädte z. B. viel eher der Fall sein als für ländliche Regionen.

Die Berliner Paarstudie zog eine Stichprobe von Ehepaaren mit deutschen Männern und Frauen aus Polen, Russland, Thailand, den Philippinen und Brasilien. Ehepaare mit deutschen Frauen dienten als Kontrollgruppe.⁸ Es war vorgesehen, für die sechs genannten Herkunftsländer jeweils 1000 Frauen mit deutschen Ehemännern anzuschreiben. Dazu war ein mehrstufiger Auswahlprozess notwendig. Im Oktober 2007 wurde eine nach Herkunftsland der Ehefrauen geschichtete Stich-

8 Diese Länderauswahl orientierte sich an empirischen und theoretischen Kriterien. Polen ist seit zwei Jahrzehnten das wichtigste Herkunftsland ausländischer Ehefrauen. Russland und Thailand nahmen im gleichen Zeitraum alternierend den zweiten und dritten Platz ein. Die Philippinen waren in den 1980er Jahren das Hauptherkunftsland ausländischer Ehefrauen außerhalb Europas. Brasilien ist in den letzten Jahren als einziges südamerikanisches Herkunftsland unter den zehn wichtigsten Herkunftsländern vertreten gewesen. Zudem weisen alle diese Länder einen niedrigeren Human Development Index auf als Deutschland. Die Türkei ist seit den 1970er Jahren regelmäßig unter den zehn wichtigsten Herkunftsländern ausländischer Ehefrauen zu finden, allerdings geht der Großteil dieser Heiraten auf türkischstämmige Deutsche zurück (vgl. Straßburger 2000: 29f.; 2003).

probe aus dem Melderegister Berlins gezogen.⁹ Zunächst wurde eine Zufallsauswahl von jeweils 2.000 Adressen von Ehepaaren mit Männern deutscher Staatsangehörigkeit und Frauen deutscher, thailändischer, polnischer und russischer Staatsangehörigkeit vorgenommen. Von deutsch-brasilianischen Paaren lagen im Melderegister nur 682 Adressen vor, von den deutsch-philippinischen Paaren nur 377 Adressen, die jeweils vollständig übermittelt wurden. Technisch handelte es sich dabei um eine Ziehung von Einzelpersonen: verheiratete Frauen, die eine der genannten Staatsangehörigkeiten innehatten. Diese wurden jedoch nur dann ausgewählt, wenn ihr Ehemann ausschließlich die deutsche Staatsangehörigkeit besaß. Die übermittelten Daten umfassten Name, Geburtsjahr und Anschrift der Frau sowie des Mannes.

Im ersten Schritt wurde die Anschrift der Partner von der Meldebehörde noch nicht automatisch für die Auswahl einbezogen. Deshalb mussten im zweiten Schritt aus den übermittelten Daten jene Paare gefiltert werden, die unter der gleichen Adresse gemeldet waren. Dabei stellten sich große Unterschiede zwischen den Unterstichproben heraus (Tabelle 1).

Deutsche Frauen sind am häufigsten unter der gleichen Adresse wie ihre Ehemänner gemeldet (89,0 %). Unter den Paaren mit ausländischen Ehefrauen zeigt sich eine Zweiteilung: Von den Frauen aus Thailand, den Philippinen und Brasilien wohnen etwa 80 % mit ihren Männern zusammen. Bei den Frauen aus Osteuropa – Polen und Russland – sind es nur weniger als 70 %. Aus der Tabelle geht ebenfalls hervor, dass die Zahl der fehlenden Angaben teilweise erheblich ist. Bei Ehefrauen aus Polen und Russland fehlt in bis zu 20,4 % bzw. 16,3 % der Fälle die Anschrift des Mannes.

Es sollten weiterhin nur jene Paare angeschrieben werden, bei denen der Mann in Deutschland geboren war oder von Geburt an die deutsche Staatsbürgerschaft innehatte, zudem sollten er und mindestens ein Elternteil Deutsch als Muttersprache haben. Über den Zeitpunkt des Erwerbs der deutschen Staatsbürgerschaft und Sprachkenntnisse liegen im Melderegister jedoch keine Angaben vor; das Geburtsland ließe sich über den Geburtsort nachvollziehen, dieses Verfahren ist jedoch sehr aufwendig und kostenintensiv (Salentin 2007). Die Entscheidung fiel deshalb auf ein onomastisches Verfahren, bei dem alle Männer ausgeschlossen wurden, deren Kombination des Vor- und Nachnamens auf eine ausländische Herkunft schließen ließ. Jeder Name wurde von einer Hilfskraft gelesen und als

9 Dankenswerterweise zeigte sich das zuständige Berliner Landesamt für Bürger- und Ordnungsangelegenheiten (LABO) trotz des notwendigen Programmieraufwands sehr kooperativ.

Tabelle 1 Initialstichprobe aus dem Berliner Melderegister (%)

	Deutschland	Thailand	Philippinen	Brasilien	Polen	Russland
n	2.000	2.000	377	682	5.000	3.440
Wohnung						
Zusammenlebend	89,0	81,6	81,2	80,2	65,7	69,1
Getrennt lebend	9,5	14,0	10,3	14,5	13,9	14,6
Anschrift Mann fehlt/ unvollständig	1,5	4,4	8,5	5,3	20,4	16,3
Mann						
Deutscher Name	87,8	86,3	73,8	79,9	42,4	49,3
Ausländischer Name	11,7	10,4	23,3	18,0	48,9	42,8
Name fehlt	0,5	3,3	2,9	2,1	8,7	7,9
Zusammenlebend, Mann trägt deutschen Namen	87,4 (1.756)	73,8 (1.476)	64,0 (241)	66,1 (451)	33,7 (1685)	36,6 (1260)

Quelle: Melderegister der Stadt Berlin (2007), eigene Berechnungen. Stichprobe von 2.000 Paaren für deutsche und thailändische Gruppe, Stichprobe von 5000 Paaren für polnische Gruppe. Vollerhebung für philippinische, brasilianische und russische Gruppe.¹⁰

deutsch oder ausländisch eingestuft.¹¹ Für Paare mit Frauen aus Deutschland, Thailand, den Philippinen und Brasilien war diese Aufgabe leicht zu lösen, da die Namen der Männer in den meisten Fällen eindeutig einer deutschen oder ausländischen Herkunft zuzuordnen sind. Eine Auswertung der rückgesendeten Fragebögen bestätigt dies: In den genannten Gruppen waren 95,2 % bis 99,0 % der angeschriebenen Männer in Deutschland geboren (Tabelle 3). Deutlich schwieriger gestaltete sich die Namensfilterung für Männer, deren Frauen die polnische oder russische Staatsbürgerschaft hatten. Polnische Familiennamen enden u. a. häufig auf ek, czyk, ski und ihren orthografischen Varianten. Diese Namen sind aber auch in Deutschland bei Familien geläufig, die seit mehreren Generationen in Deutschland leben, deutsch sprechen und keine Verbindung zum Herkunftsland ihrer Ahnen mehr haben. Um dies zu umgehen, gingen in diesen Fällen verstärkt

10 Der Anteil der zusammenlebenden Paare, bei denen der Name des Mannes als deutsch eingestuft wurde, lag bei der polnischen und russischen Unterstichprobe bei jeweils nur etwa einem Drittel, weshalb die Stichprobe aus dem Melderegister von 2.000 Paaren nicht ausreichte, um 1.000 Paare anzuschreiben. Deshalb wurde erneut eine Zufallsstichprobe von 5.000 Paaren mit polnischer Ehefrau gezogen; bei den Paaren mit russischer Ehefrau wurden alle 3.440 gemeldeten Paare aus dem Melderegister zur Verfügung gestellt.

11 Um die Intercoderreliabilität einschätzen zu können, ist es meist üblich, derartige Ratings von zwei oder mehr Ratern vornehmen zu lassen. Aus Kapazitätsgründen musste hierauf jedoch verzichtet werden.

die Vornamen in die Entscheidung mit ein.¹² Lediglich 42,4 % der Ehemänner polnischer Frauen wurden anhand ihres Namens als deutschstämmig eingestuft. In dieser Gruppe ist die Trefferquote etwas schlechter. Nur 87,6 % der angeschriebenen Männer mit polnischen Frauen sind in Deutschland geboren (Tabelle 3). Die größten Schwierigkeiten bereiteten jedoch die Namen von Männern mit russischen Frauen. Hierbei handelt es sich offenbar in vielen Fällen um Aussiedler, die seit der Einwanderung über die deutsche Staatsangehörigkeit verfügen. Sie nehmen häufig auch deutsch klingende Namen an, was dazu führt, dass sie anhand dieser Kriterien nicht mehr als russischstämmig zu erkennen sind. Immerhin konnten 42,8 % der Männer mit russischen Frauen anhand ihres Namens mit hoher Wahrscheinlichkeit als Migranten identifiziert werden, doch in dieser Gruppe liegt die Trefferquote am niedrigsten: Lediglich 67,7 % der angeschriebenen Männer russischer Frauen sind in Deutschland geboren, die übrigen 32,3 % stammen aus Russland oder einem anderen Staat der ehemaligen Sowjetunion. Sie waren über ihre Namen nicht identifizierbar gewesen.

Nach dieser Filterung blieben 1.756 Paare mit deutscher Frau, 1.476 Paare mit thailändischer Frau, 241 Paare mit philippinischer Frau, 451 Paare mit brasilianischer Frau, 1.685 Paare mit polnischer Frau und 1260 Paare mit russischer Frau übrig. Die philippinische Unterstichprobe wurde wegen einer zu niedrigen Fallzahl verworfen, wohingegen die brasilianische Gruppe vollständig in die Befragung einging. In den anderen Gruppen wurde aus den verbleibenden Fällen eine Zufallsauswahl von jeweils 1.000 Paaren gezogen.

Die Studie strebte vorrangig einen Vergleich zwischen zwei Gruppen von Männern an: jenen, die eine Frau aus dem ärmeren Ausland geheiratet hatten und jenen, die eine deutsche Frau geheiratet hatten. Die Altersverteilung zwischen Männern mit deutschen Frauen (Mittelwert des Geburtsjahres 1949, SD = 13,4) und Männern mit ausländischen Frauen (1958, SD = 12,2) differierte jedoch stark. Um zu verhindern, dass die Altersdifferenz von knapp zehn Jahren andere Unterschiede verdeckte, wurden in der deutsch-deutschen Stichprobe die Männer aus den Geburtsjahrgängen disproportional ausgewählt. Mit der Anpassung konnte eine Halbierung der Differenz erreicht werden. In der angepassten 1.000er-Stichprobe mit deutschen Ehefrauen liegt das durchschnittliche Geburtsjahr der Männer bei 1954 (SD = 12,4).

12 Migranten vergeben auch in der zweiten Generation meist noch Vornamen, die im Herkunftsland üblich sind (Gerhards/Hans 2009). Deshalb ist davon auszugehen, dass nach diesem Kriterium auch Migranten der zweiten oder dritten Generation herausgefiltert werden.

Bei der anschließenden Befragung wurden 4.451 Paare angeschrieben: jeweils 1.000 Paare mit deutschen Männern und Frauen aus Deutschland, Thailand, Polen und Russland sowie 451 Paare mit Frauen aus Brasilien.

3.5 Rücklauf der Fragebögen und mögliche Ausfälle

Im Februar 2008 wurden die Fragebögen versendet, gefolgt von einer schriftlichen und einer telefonischen Erinnerung. Die Bruttoreücklaufquote aus den einzelnen Stichproben beträgt 31,1 % für Deutschland, 20,4 % für Thailand, 22,8 % für Brasilien, 26,7 % für Polen und 25,4 % für Russland. Insgesamt kamen 25,6 % aller versendeten Fragebögen ausgefüllt zurück. Es war leider nicht möglich, die Zahl der falschen Adressen zu ermitteln, da die Fragebögen aus Kostengründen per *Infopost* versendet wurden und bei dieser Frankierart nicht zustellbare Sendungen vernichtet werden. Nach den Ergebnissen des Zensus-tests (Statistische Ämter des Bundes und der Länder 2004) sowie den Erfahrungen von Granato (1999) und Brüderl et al. (2007) kann man aber von etwa 8 % falscher Adressen ausgehen, was einen korrigierten Rücklauf von 28 % bedeuten würde. Darüber hinaus gab es auch Fälle, bei denen die Anschrift eines Partners zwar noch korrekt war, die Partnerschaft aber seit mehreren Jahren nicht mehr bestand. Die Nettorücklaufquote aus der Grundgesamtheit dürfte also noch etwas höher liegen.

Eine Antwortquote in dieser Größenordnung lässt nichtsdestotrotz vermuten, dass die Ausfälle zu systematischen Verzerrungen führen. Dies lässt sich überprüfen, indem man die Liste der angeschriebenen Personen mit den Eigenschaften der antwortenden Personen vergleicht. Die Melderegisterstichprobe enthält die Geburtsjahre der angeschriebenen Männer und Frauen. In Tabelle 2 ist die Altersverteilung der angeschriebenen sowie der antwortenden Personen abgetragen.

Die Abweichungen zwischen der angeschriebenen und der realisierten Stichprobe liegen meist unter einem Prozentpunkt pro Zehnjahresgruppe und sind damit als niedrig einzuschätzen. Bei den Frauen fällt die leichte Unterrepräsentation von 40 bis 59-jährigen zugunsten der 20 bis 29-jährigen Frauen auf. Doch auch wenn man die Unterstichproben einzeln betrachtet, zeigt sich ein ähnliches Ergebnis: Der relative Ausfall in Bezug auf das Alter hält sich in einem durchaus annehmbaren Rahmen. Weitere Ausfallanalysen, z. B. zum Bildungsstand, lassen sich nicht realisieren, da das Melderegister hierzu keine Angaben enthält.

Tabelle 2 Relativer Ausfall nach Alter und Geschlecht (%)

Alter	Männer		Frauen	
	Angeschrieben	Rücklauf	Angeschrieben	Rücklauf
20 bis 29 Jahre	3,3	3,9	9,7	11,4
30 bis 39 Jahre	16,5	17,7	31,1	31,1
40 bis 49 Jahre	32,9	31,7	30,4	29,0
50 bis 59 Jahre	23,5	22,8	18,2	17,3
60 bis 69 Jahre	16,7	16,8	7,5	7,9
70 bis 79 Jahre	5,7	6,0	2,6	2,8
80 bis 95 Jahre	1,4	1,1	0,5	0,4
n	4.451	1.135	4.451	1.134

Quelle: Melderegister der Stadt Berlin; Berliner Paarstudie.

Ehepaare mit deutschen Männern und deutschen Frauen haben am häufigsten geantwortet, Paare mit Frauen ausländischer Staatsbürgerschaft bis zu zehn Prozentpunkte seltener. Worin ist dies begründet? Im Datensatz liegen hierzu keine Angaben vor, mehrere Vermutungen liegen jedoch nahe. Einerseits sind die Adressdaten von Paaren mit einem ausländischen Partner möglicherweise häufiger veraltet als jene von deutsch-deutschen Ehepaaren. Dies lässt sich hier nicht überprüfen, doch aus anderen Studien ist bekannt, dass die im Melderegister gespeicherten Adressdaten von Ausländern häufiger veraltet sind als jene von Personen mit deutscher Staatsangehörigkeit. Dies kann z. B. darauf zurückzuführen sein, dass Personen, die ins Ausland umziehen, sich nicht abmelden (Salentin 1999: 117f.). Das erklärt aber nicht, warum der Rücklauf bei Paaren mit Frauen aus Thailand und Brasilien niedriger ist als bei Paaren aus Polen und Russland. Es ist zu vermuten, dass der größte Bias auf die Intimität des Themas und die moralisch aufgeladene öffentliche Diskussion binationaler Ehen zurückgeht, von denen thailändisch-deutsche und brasilianisch-deutsche Paare aufgrund der offensichtlich fremden Herkunft der Frauen möglicherweise häufiger betroffen sind als Paare mit Frauen aus Polen und Russland. Eine starke Ablehnung mehrerer Items kann dann, gerade bei einer schriftlichen Befragung ohne Incentives, auch zu einem Unit-Nonresponse führen. Der Fragebogen erhebt relativ intime Informationen, aus denen sich u. a. der Body Mass Index der Befragten und der Altersabstand des Paares berechnen lässt. Personen, die einen besonders auffälligen BMI oder einen ungewöhnlichen Altersabstand aufweisen, werden möglicherweise von der Beantwortung absehen. Das Gleiche gilt für Personen, die z. B. traditionelle Geschlechterrolleneinstellungen

vertreten oder häusliche Gewalt für akzeptabel halten. Tatsächlich weisen die Teilstichproben mit den niedrigsten Rücklaufquoten empirisch die stärksten Tendenzen in diese Richtungen auf.

Die Anrede beider Partner mit vollem Namen im Anschreiben, die Zweisprachigkeit des Fragebogens für die Frau und die Intimität einiger Fragen machten viele Probanden misstrauisch. Nach Versendung der Fragebögen erreichten uns innerhalb von zwei Wochen etwa 150 Anrufe und E-Mails, wovon der Großteil aus Beschwerden und Absagen bestand. Die Beschwerden der Befragten lassen sich in zwei Punkten zusammenfassen. Erstens waren sie über mangelnden Datenschutz bei der Herausgabe ihrer Anschriftendaten und die Verwendung ihrer Antworten besorgt. Einige wendeten sich an den Datenschutzbeauftragten des Landes Berlin, der eine Anfrage an uns stellte, welche Daten weitergegeben worden waren und wann sie gelöscht würden. Wir gaben an, welche Daten aus dem Melderegister übermittelt worden waren und konnten versichern, dass sie direkt nach der Befragung gelöscht wurden. Rechtlich ist gegen das Vorgehen nichts einzuwenden, doch die Besorgnis der Befragten sollten Sozialwissenschaftler ernst nehmen.¹³ Das Anschreiben teilte bereits mit, dass die Daten aus dem Melderegister stammten und ein beigelegtes Datenschutzblatt erklärte, wie die Antworten anonymisiert werden. Bei Anfragen über die Herkunft der Daten klärten wir die Befragten ausführlicher auf. Zweitens waren einige Befragte über den Inhalt des Fragebogens verärgert. Manche fanden die gestellten Fragen zu persönlich; andere erklärten, dass sie mit der öffentlichen Diskussion des Themas binationale Ehen bereits schlechte Erfahrungen gemacht hätten und sich nicht weiter dazu äußern wollten. Damit gehen Äußerungen einher, der Fragebogen enthalte „zu viele Klischees“ und diene lediglich der „Beschreibung und Entwicklung von Stereotypen“. Aus der rein wissenschaftlichen Warte heraus sind diese Beschwerden nicht resonanzfähig. Wissenschaftler sind bestrebt, die Realität zu beschreiben. Wenn sich ein unangenehmes Klischee im Zuge einer Untersuchung als „wahr“ herausstellt, ist das ein auf legitime Weise erzieltes und berichtenswertes Ergebnis. Soziologen sollten sich aber stets bewusst machen, welche möglichen Wirkungen mit ihrer Forschung verbunden sind. Der Ethik-Kodex der Soziologie erklärt: „Soziologinnen und Soziologen [tragen] soziale Verantwortung. Ihre Empfehlungen, Entscheidungen und Aussagen

13 Neben den Datenschutzgesetzen betont auch der Ethik-Kodex der Soziologie (Deutsche Gesellschaft für Soziologie 1992) die Bedeutung des Datenschutzes. So ist die Anonymität von Untersuchten zu wahren und „von untersuchten Personen erlangte vertrauliche Informationen müssen entsprechend behandelt werden.“ Obwohl nicht explizit ausgewiesen, gilt dies selbstverständlich auch für Daten, die gesammelt werden, um Probanden zu kontaktieren.

können das Leben ihrer Mitmenschen beeinflussen. Sie sollten sich der Situation und immanenten Zwänge bewusst sein, die zu einem Missbrauch ihres Einflusses führen könnten" (Deutsche Gesellschaft für Soziologie 1992). Insbesondere bei stigmatisierten Gruppen, wie den in der Berliner Paarstudie untersuchten deutschen Männern mit Frauen aus dem deutschen Ausland, sind Soziologen in der Pflicht, bei der Veröffentlichung von Ergebnissen behutsam vorzugehen. Aussagen über stigmatisierte Gruppen sollten nur dann erfolgen, wenn sie wissenschaftlich abgesichert sind; nicht abgesicherte Spekulationen sollten weitgehend vermieden werden, um benachteiligten Randgruppen nicht noch weiter zu schaden.

3.6 Effektivität der Stichprobenfilter

Die Studie zielte auf eine sehr spezielle Population ab. Gesucht wurden zusammenlebende Ehepaare, in denen der Mann in Deutschland geboren war oder seit Geburt die deutsche Staatsbürgerschaft innehatte. Zudem sollten der Mann und mindestens ein Elternteil Deutsch als Muttersprache haben. Die ausländischen Frauen sollten im Ausland geboren sein und eine ausländische Staatsbürgerschaft besitzen. Bei der Stichprobenziehung wurde großer Wert darauf gelegt, die anvisierte Grundgesamtheit mittels verschiedener Filter bereits so weit wie möglich zu ermitteln. Aufgrund verschiedener Ungenauigkeiten war ein gewisser Beifang aber nicht zu vermeiden. Anhand des Rücklaufes läßt sich bestimmen, wie groß der Anteil der korrekt ausgewählten Paare ist (Tabelle 3).

Bei den Kriterien Geburtsland und Staatsbürgerschaft bei Geburt der Frau wurden hohe Trefferquoten erzielt. In der deutschen Unterstichprobe sind 95,2 % der Frauen in Deutschland geboren, in den polnischen, thailändischen und brasilianischen Unterstichproben sind zwischen 98,9 % und 100 % der Frauen im anvisierten Land geboren. Lediglich in der russischen Stichprobe liegt die Trefferquote niedriger, dort sind 6,9 % der Frauen in einem anderen Land geboren. Es handelt sich dabei ausschließlich um Republiken der ehemaligen Sowjetunion (Weißrussland, Usbekistan, Kasachstan etc.). Alle Frauen in den ausländischen Unterstichproben besitzen ihre ausländische Staatsangehörigkeit bereits von Geburt an. Bei den Frauen in der deutschen Stichprobe haben 3,5 % die deutsche Staatsangehörigkeit erst später angenommen.

Bei den Männern liegt die Trefferquote bezüglich des Geburtslandes deutlich niedriger. In der deutschen, thailändischen und brasilianischen Stichprobe ist sie mit 95,2 % bis 99,0 % zufriedenstellend. Doch trotz der Namensfilterung sind in der polnischen Stichprobe nur 87,6 % aller Männer in Deutschland geboren, in der russischen Stichprobe sind es sogar nur 67,7 %. In dieser Gruppe wurden 7,9 % in

Tabelle 3 Anteil der korrekt identifizierten Paare (%)

	Deutschland	Thailand	Brasilien	Polen	Russland
Basis (Rücklauf)	311	204	103	267	254
Frau					
Im anvisierten Land geboren	95,2	99,5	100,0	98,9	91,5
Staatsbürgerschaft bei Geburt Deutsch	96,5	0,0	0,0	0,0	0,0
Mann					
In Deutschland geboren	96,5	99,0	95,2	87,6	67,7
Staatsbürgerschaft bei Geburt Deutsch	97,7	99,0	97,1	88,4	76,0
Muttersprache Deutsch	98,7	99,5	97,2	89,9	79,8
Ein Elternteil Muttersprache Deutsch	95,0	93,5	93,6	89,1	81,2
Effektivität der Stichprobenfilter					
Paar erfüllt alle Kriterien	93,9	82,4	89,3	82,8	63,0
Mann verletzt Kriterien*	4,5	3,4	5,8	15,0	29,1
Frau verletzt Kriterien*	3,5	14,2	4,9	2,6	9,8

Quelle: Berliner Paarstudie. Für jedes Merkmal ist nur die gesuchte Kategorie angegeben, um die Übersichtlichkeit zu erhöhen.

* Diese beiden Kategorien schließen sich gegenseitig nicht aus. Wenn in einem Paar beide Partner die Kriterien verletzen, steigt die Gesamtprozentzahl auf über 100.

einem dritten Land geboren, wobei es sich wieder um ehemalige Sowjetrepubliken handelt (Kasachstan, Ukraine, Tadschikistan). Entsprechend ist auch die Verteilung der Staatsbürgerschaft bei Geburt: Während die Männer in der deutschen, thailändischen und brasilianischen Strichprobe überwiegend in Deutschland zur Welt gekommen sind, liegt diese Quote in den anderen beiden Gruppen entsprechend niedriger. Analog dazu stellt sich auch die Verteilung von Deutsch als Muttersprache dar.

Der Erfolg der Namensfilterung ist somit für die fünf Gruppen unterschiedlich. In der deutschen, thailändischen und brasilianischen Stichprobe wurde auf diese Weise eine sehr hohe Genauigkeit der Stichprobenziehung erzielt. In der polnischen und russischen Gruppe war der Erfolg hingegen schwächer. Zwar konnte bei der Namensfilterung bereits etwa die Hälfte der gezogenen Adressen ausgeschlossen werden, die Zahl der „falsch“ gewählten Männer ist jedoch weiterhin beträchtlich.

Abschließend kann man die einzelnen Informationen kombinieren und so diejenigen Paare ermitteln, auf die alle Kriterien zutreffen.¹⁴ An dieser Erfolgsbilanz

14 Dabei sind alle Kriterien mit *und* verknüpft, mit folgender Ausnahme: Der Mann muss in Deutschland geboren sein *oder* bei Geburt die deutsche Staatsangehörigkeit gehabt haben. Gleiches gilt für Frauen in der deutschen Gruppe.

lässt sich die Güte des Vorgehens bei der Stichprobenziehung ablesen, denn wenn alle Paare alle Kriterien erfüllten, wäre die Stichprobenziehung dieser seltenen Population perfekt verlaufen. Eine besonders gute Auswahl wurde in der deutschen Unterstichprobe mit 93,9 % erzielt. Auch in der brasilianischen Stichprobe gehören neun von zehn Paaren zu der anvisierten Population. Bei den thailändischen Paaren fällt dieser Wert etwas schlechter aus, was meist auf fehlende Angaben bei den Frauen zurückgeht. In dieser Gruppe haben viele Frauen entweder ihr Geburtsland oder ihre Staatsbürgerschaft bei Geburt nicht angegeben und konnten deshalb nicht eindeutig zugeordnet werden. In der polnischen und der russischen Gruppe erfüllen 82,8 % bzw. nur 63,0 % alle Kriterien, was meist auf die Angaben der Ehemänner zurückgeht. Sie stammen, wie bereits besprochen, teilweise aus dem Herkunftsland der Frau. In der russischen Gruppe könnte ein etwas höherer Wert erzielt werden, wenn man auch ehemalige Mitgliedsstaaten der Sowjetunion als Herkunftsland der Frau zulässt.

Insgesamt ist das Ergebnis sehr zufriedenstellend. Mit der Stichprobenziehung ist es gelungen, die anvisierte seltene Stichprobe nahezu exakt einzugrenzen. Während lediglich 0,5 % der Grundgesamtheit der deutschen Wohnbevölkerung zu der Population deutscher Männer mit im Ausland geborenen Ehefrauen gehören, ist es hier gelungen, diese Paare mit einer Genauigkeit zwischen 63,0 % und 89,3 % aus dem Melderegister der Stadt Berlin herauszufiltern. Das ist eine beachtlich genaue Stichprobenziehung, bei der sich der Beifang zwischen 10,7 % und 17,6 % verschmerzen lässt. Lediglich in der russischen Unterstichprobe liegt mit 37 % ein hoher Beifang vor.

4 Schlussbemerkungen

Ausgangspunkt dieses Beitrags war ein Forschungsvorhaben zur Partnerwahl bei binationalen Ehepaaren. Hierzu wurde erstens eine ausreichend große Stichprobe benötigt, zweitens mussten sich die abgefragten Informationen auf den Zeitpunkt der Eheschließung beziehen. Binationale Ehepaare stellen jedoch eine äußerst seltene Population dar, sodass sie nicht in ausreichend großer Zahl in bereits existierenden Datensätzen repräsentiert sind. Es galt also, für das Forschungsprojekt eine eigene Erhebung durchzuführen, wofür eine Zufallsstichprobe aus dieser seltenen Population benötigt wurde. Die Wahl fiel auf ein mehrstufiges Filterverfahren, das zuerst eine Stichprobe aus dem Melderegister mit verknüpften Partnerdaten und zweitens eine onomastische Filterung der Männer beinhaltete. Auf diese Weise konnten in den einzelnen realisierten Unterstichproben Trefferquoten von 63,0 % bis 89,3 % erzielt werden. Das Filterverfahren zeigte seine Stärken vor allem bei

Paaren, in denen die Frau aus Thailand oder Brasilien stammt, ferner auch bei Paaren mit Frauen aus Polen. Bei deutsch-russischen Paaren stellte sich das Verfahren als etwas schwächer heraus, da aus Russland stammende Männer, die vermutlich als Aussiedler die deutsche Staatsangehörigkeit angenommen haben, häufig nicht an ihren Namen erkennbar waren. Dies ließe sich mit dem von Salentin (2007) vorgeschlagenen Verfahren noch verbessern, bei dem der Geburtsort für die Auswahl verwendet wird. Doch ist dieses Vorgehen ungleich kostenintensiver, hier wäre je nach Zielgruppe zwischen Kosten und Nutzen abzuwägen.

Die Verknüpfung von Daten im Familienverband, wie sie das hier vorgestellte Projekt genutzt hat, ermöglicht schließlich noch eine Reihe von anderen Anwendungen, die für die Familiensoziologie und andere Forschungsgebiete von Interesse sein können. Spezielle Stichproben lassen sich auf diese Weise um ein Vielfaches punktgenauer ziehen und Beifang vermeiden. Für die Scheidungsforschung ist es auf diese Weise z. B. möglich, getrennt lebende Ehepaare zu ermitteln, indem man die Anschrift beider Partner vergleicht. Über die Einbeziehung des Geschlechts der Partner ließen sich auch gleichgeschlechtliche Paare finden, die nach dem Lebenspartnerschaftsgesetz verheiratet sind. Über die Verknüpfung der Daten von Eltern und Kindern wäre es u. a. auch möglich, Familien mit Zwillingen sowie kinderlose oder kinderreiche Familien auszuwählen und zudem noch die Alterskohorte der Eltern bzw. der Kinder zu bestimmen. Über die Kombination der Familiennamen der Eltern und der Kinder sowie den zeitlichen Abstand zwischen Geburtsdatum der Kinder und dem Heiratsdatum der Eltern wäre es zudem möglich, mit einer gewissen Wahrscheinlichkeit Stieffamilien zu ermitteln.¹⁵ Derzeit gibt eine Reihe von aufwendigen Studien, die sich mit diesen und ähnlichen Themen befassen, wie Pairfam und den Gender and Generations Survey. Für eventuelle Erweiterungen dieser Studien oder andere Projekte, die sich z. B. mit binationalen Partnerschaften befassen,¹⁶ ermöglicht dieses Vorgehen sehr spezielle Stichproben. Inwiefern Meldeämter diese Datenverknüpfung öffentlichen Stellen wie Universitäten ermöglichen, hängt vermutlich jeweils vom Einzelfall ab. Wenn es umsetzbar ist, kann dieses Vorgehen bei der Stichprobenziehung sehr hilfreich sein.

15 Ein linguistisches Projekt zur Entwicklung von Zweisprachigkeit bei Babys hat kürzlich eine Stichprobe aus dem Berliner Melderegister gezogen, bei dem das Geburtsdatum des Kindes und die Staatsangehörigkeiten der Eltern kombiniert wurden („Sprachvergleichende Untersuchungen zur Entwicklung von Rhythmuspräferenzen“, <http://www.uni-potsdam.de/babylab/projects/language.html>).

16 So beispielsweise das internationale Forschungsprojekt „Toward a European Society: Single Market, Binational Marriages, and Social Group Formation in Europe (EUMARR)“ (<http://www.ibei.org>).

Literatur

- Albers, I., 1997: Einwohnermelderegister-Stichproben in der Praxis. Ein Erfahrungsbericht. S. 117-125 in: S. Gabler und J. Hoffmeyer-Zlotnik (Hg.): Stichproben in der Umfragepraxis. Opladen: Westdeutscher Verlag.
- Alt, C., W. Bien und D. Krebs, 1991: Wie zuverlässig ist die Verwirklichung von Stichprobenverfahren? Random route versus Einwohnermeldeamtsstichprobe. ZUMA-Nachrichten 28: 65-72. (PID): <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-209813>
- Amt für Statistik Berlin Brandenburg, 2009: Bevölkerungsstand. <http://www.statistik-berlin-brandenburg.de> (02.07.2009).
- Babka von Gostomski, C. und M. Pupeter, 2008: Zufallsbefragung von Ausländern auf Basis des Ausländerzentralregisters. Erfahrungen bei der Repräsentativbefragung „Ausgewählte Migrantengruppen in Deutschland 2006/2007“ (RAM). Methoden – Daten – Analysen 2(2): 149-177. (PID): <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-126673>
- Brüderl, J., L. Castiglioni, U. Krieger und K. Pforr, 2007: Design und Feldphase des Mini-Panels. http://www.mzes.uni-mannheim.de/publications/papers/Minipanel-Methodological_Report_070711.pdf (23.02.2009).
- Deutsche Gesellschaft für Soziologie, 1992: Ethik-Kodex. Fassung vom 27.11.1992. <http://www.sociologie.de/index.php?id=19> (09.04.2010).
- Diehl, C., 2007: Materialband und Endbericht zur Neuzuwandererbefragung – Pilotstudie. Erste und zweite Welle. Wiesbaden: Bundesinstitut für Bevölkerungsforschung.
- Diekmann, A., 2008: Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. 19. Auflage. Reinbek bei Hamburg: Rowohlt.
- Feld, S. L., 1981: The Focused Organization of Social Ties. *American Journal of Sociology* 86(5): 1015-1035.
- Gabler, S., 1992: Schneeballverfahren und verwandte Stichprobendesigns. ZUMA-Nachrichten 31: 47-69. (PID): <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-210848>
- Gabler, S. und S. Häder, 1997: Überlegungen zu einem Stichprobendesign für Telefonumfragen in Deutschland. ZUMA-Nachrichten 41(41): 7-18. (PID): <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-208339>
- Gerhards, J. und S. Hans, 2009: From Hasan to Herbert: Name-Giving Patterns of Immigrant Parents between Acculturation and Ethnic Maintenance. *American Journal of Sociology* 114(4): 1102-1128.
- Glowsky, D., 2011: Globale Partnerwahl. Soziale Ungleichheit als Motor transnationaler Heiratsentscheidungen. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Granato, N., 1999: Die Befragung von Arbeitsmigranten. Einwohnermeldeamt-Stichprobe und telefonische Erhebung? ZUMA-Nachrichten 23(45): 44-60. (PID): <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-208119>
- Groves, R. M., F. J. Fowler, M. P. Couper, J. M. Lepkowski et al., 2004: *Survey Methodology*. Hoboken, NJ: Wiley-Interscience.
- Häder, S. und S. Gabler, 1998: Ein neues Stichprobendesign für telefonische Umfragen in Deutschland. S. 69-88 in: S. Gabler, S. Häder und J. Hoffmeyer-Zlotnik (Hg.): *Telefonstichproben in Deutschland*. Opladen: Westdeutscher Verlag.
- Hoffmeyer-Zlotnik, J., 1997: Random-Route-Stichproben nach ADM. S. 33-42 in: S. Gabler und J. Hoffmeyer-Zlotnik (Hg.): *Stichproben in der Umfragepraxis*. Opladen: Westdeutscher Verlag.
- Hoffmeyer-Zlotnik, J., 2003: New Sampling Designs and the Quality of Data. S. 205-217 in: A. Ferligoj und A. Mrvar (Hg.): *Developments in Applied Statistics*. Ljubljana: FDV.

- Humpert, A. und K. Schneiderheinze, 2000: Stichprobenziehung für telefonische Zuwandererumfragen. Einsatzmöglichkeiten der Namensforschung. ZUMA-Nachrichten 24(47): 36-64. (PID): <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-208052>
- Humpert, A. und K. Schneiderheinze, 2002: Stichprobenziehung für telefonische Zuwandererumfragen – Praktische Erfahrungen und Erweiterung der Auswahlgrundlage. S. 187-208 in: S. Gabler und S. Häder (Hg.): Telefonstichproben – Methodische Innovationen und Anwendungen in Deutschland. Münster: Waxmann.
- Kalmijn, M., P. M. d. Graaf und J. P. G. Janssen, 2005: Inter-marriage and the Risk of Divorce in the Netherlands: The Effects of Differences in Religion and in Nationality, 1974-94. *Population Studies* 59(1): 71-85.
- Kish, L., 1965: *Survey Sampling*. New York: Wiley.
- Lüttinger, P. und T. Riede, 1997: Der Mikrozensus. Amtliche Daten für die Sozialforschung. ZUMA-Nachrichten 41: 19-44. (PID): <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-208344>
- Niesner, E., 2000: Mythos und Wirklichkeit auf einem bikulturellen Heiratsmarkt. S. 163-181 in: E. Friebe-Blum (Hg.): *Wer ist fremd? Ethnische Herkunft, Familie und Gesellschaft*. Opladen: Leske + Budrich.
- Reuschke, D., 2010: *Multilokales Wohnen: Raum-zeitliche Muster multilokaler Wohnarrangements von Shuttles und Personen in einer Fernbeziehung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Roloff, J., 1998: Eheschließungen und Ehescheidungen von und mit Ausländern in Deutschland. *Zeitschrift für Bevölkerungswissenschaft* 23(3): 319-334.
- Salentin, K., 1999: Die Stichprobenziehung bei Zuwandererbefragungen. ZUMA-Nachrichten 23(45): 115-135. (PID): <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-208159>
- Salentin, K., 2002: Zuwandererstichproben aus dem Telefonbuch: Möglichkeiten und Grenzen. S. 164-186 in: S. Gabler und S. Häder (Hg.): *Methodische Innovationen und Anwendungen in Deutschland*. Münster: Waxmann.
- Salentin, K., 2007: Die Aussiedler-Stichprobenziehung. *Methoden – Daten – Analysen* 1(1): 25-44. (PID): <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-126439>
- Schnell, R., P. B. Hill und E. Esser, 2008: *Methoden der empirischen Sozialforschung*. München/Wien: Oldenbourg.
- Statistische Ämter des Bundes und der Länder, 2004: Ergebnisse des Zensusstests. *Wirtschaft und Statistik* 8/2004: 813-833.
- Straßburger, G., 2000: Das Heiratsverhalten von Personen ausländischer Nationalität oder Herkunft in Deutschland Sachverständigenkommission 6. Familienbericht (Hg.): *Familien ausländischer Herkunft in Deutschland: Empirische Beiträge zur Familienentwicklung und Akkulturation*. Materialien zum 6. Familienbericht. Opladen: Leske + Budrich.
- Straßburger, G., 2003: Heiratsverhalten und Partnerwahl im Einwanderungskontext. Eheschließungen der zweiten Migrantenmigration türkischer Herkunft. Würzburg: Ergon.
- Terwey, M., 2000: ALLBUS: A German General Social Survey. *Schmollers Jahrbuch* 120(1): 151-158.
- Wagner, G. G., J. R. Frick und J. Schupp, 2007: The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. *Schmollers Jahrbuch* 127(1): 139-169.
- Wasmer, M., E. Scholz und M. Blohm, 2010: Konzeption und Durchführung der „Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS) 2008. http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2010/TechnicalReport_10-4.pdf (27.05.2011).

Anschrift des Autors

David Glowsky
Humboldt-Universität zu Berlin
Präsidialbereich, Stabsstelle Internationalisierung
Unter den Linden 6
10099 Berlin
E-Mail: david.glowsky@hu-berlin.de

Incentivierungsstrategien bei Minderheitenangehörigen

*Ergebnisse eines
Methodenexperiments*

Incentive Strategies for Minorities

Results of an experiment

Patrick Fick und Claudia Diehl

Zusammenfassung

Schriftliche Befragungen stellen unter bestimmten Bedingungen eine kostengünstige Form der Datenerhebung unter Angehörigen ethnischer Minderheiten dar, die ansonsten finanziell sehr aufwändig sind. Ein Vorteil ist, dass eingesparte Mittel in gezielte Maßnahmen zur Steigerung der Ausschöpfungsquote investiert werden können. Hiervon ausgehend werden in dem Beitrag unterschiedliche Formen der Incentivierung und deren generelle und subgruppenspezifische Wirkung auf die Teilnahmebereitschaft und die Stichprobenqualität untersucht. Zunächst werden allgemeine theoretische Überlegungen zum Einsatz von Incentives vorgestellt. Darauf aufbauend wird eine bislang wenig diskutierte „doppelte Incentivierung“ mit einem unbedingten und einem konditionalen Anteil vorgeschlagen. Sie zielt darauf ab, gleichermaßen die Kosten der Nicht-Teilnahme und den Nutzen der Teilnahme für die Befragten zu erhöhen. In einem Methodenexperiment unter deutsch-türkischen jungen Erwachsenen werden diese theoretischen Überle-

Abstract

Mail surveys are under certain conditions a cost-effective mode of data collection in the generally cost-intensive research on ethnic minorities. Compared to other survey modes they make financial resources available that can be invested into strategies to increase survey response rates. Based on these assumptions, we study general as well as sub-group specific effects of various forms of incentives on response rates and sampling bias. We start out by presenting theoretical considerations about the use of different forms of incentives. It follows an introduction of a hitherto little-discussed „double incentive“ strategy consisting of an unconditional and a conditional part. The logic of such a strategy is to increase the costs of non-participation and the benefits of participation for participants at the same time. Empirical evidence is presented on the basis of an experiment with a specific population of young German-Turkish adults. After all, the double incentive strategy has positive effects on the response rate and therefore seems to be a reasonable option.



gungen empirisch überprüft. Dabei erweist sich die doppelte Incentivierung als sinnvolle Strategie mit positiven Effekten auf die Ausschöpfungsquote und die Stichprobenqualität. So können durch den Einsatz materieller Anreize offenbar benachteiligte oder – im Fall von Minderheitenangehörigen – weniger integrierte Subgruppen besser erreicht werden als bei einem Verzicht auf Incentives.

In accordance with prior findings incentives seem to reduce sample bias. Obviously, they lead to a decrease in sampling bias of respondents from disadvantaged or less integrated subgroups.

1 Einleitung¹

Die standardisierte Befragung von Personen mit Migrationshintergrund stellt aus verschiedenen Gründen eine Herausforderung dar. Zwar liegt ihr Bevölkerungsanteil bei rund 20 Prozent (Statistisches Bundesamt 2012), die Heterogenität dieser Population macht jedoch häufig subgruppenspezifische Analysen etwa auf Grundlage der Nationalität notwendig, für die ausreichend hohe Fallzahlen benötigt werden. Die Praxis, Migrantinnen und Migranten lediglich entsprechend ihres Anteils an der Wohnbevölkerung „mitzubefragen“, geht zudem oft damit einher, ausschließlich deutschsprachiges Erhebungsmaterial einzusetzen. Personen ohne ausreichende Deutschkenntnisse können bei dieser Vorgehensweise nicht erreicht werden. Um Fragestellungen aus dem Bereich der Integrationsforschung zu bearbeiten, wird daher meist ein *Oversampling* der interessierenden Gruppen oder ein gesondertes Sample von Subgruppen ähnlicher ethnischer oder nationaler Herkunft benötigt.

Im Fall von „Migrantenbefragungen“ ist ein melderegisterbasiertes Sampling anhand der Nationalität mit anschließender *Face-to-Face* Befragung gängige Praxis. Zum Einsatz kommen aber auch onomastische, d. h. namenbasierte Auswahlverfahren zur Stichprobenziehung aus Telefonverzeichnissen (vgl. Haberland et al. 2011), auf deren Grundlage telefonische Interviews durchgeführt werden.² Beide Methoden sind nicht unproblematisch. Das erstgenannte Verfahren schließt Eingebürgerte aus, so dass bei vielen Fragestellungen der Migrationsforschung eine entlang der abhängigen Variablen verzerrte Stichprobe die Validität der Ergebnisse

1 Wir bedanken uns bei Peter Preisendörfer und Jan Mewes sowie den beiden anonymen Gutachtern für die hilfreichen Hinweise zu einer früheren Version des Manuskripts.

2 Bisweilen werden auch andere Verfahren eingesetzt, etwa eine Ziehung aus dem der Forschung nicht grundsätzlich zur Verfügung stehenden Ausländerzentralregister (vgl. Babka von Gostomski/Pupeter 2008) oder ein Random-Walk-Verfahren nach der BIK Methode wie beim Integrationsurvey des BiB (vgl. Haug/Diehl 2005).

beeinträchtigen kann. Selbst wenn bei der Auswahl der Adressen onomastische Verfahren angewendet werden, bleibt das Problem der gerade bei Migranten sehr hohen Kosten persönlicher Befragungen bestehen (vgl. für *Face-to-Face* Befragungen allgemein Dillman et al. 2009). Diese sind u. a. darauf zurückzuführen, dass es bei vielen Gruppen empfehlenswert ist, bilingual-koethnische Interviewer einzusetzen, um die Befragungsfähigkeit insbesondere von weniger integrierten Befragten zu erhöhen und so die Stichprobenqualität zu verbessern (vgl. Blohm/Diehl 2001; Feskens et al. 2006). Da die wenigsten Erhebungsinstitute einen entsprechenden Interviewer-Pool vorhalten, müssen zumindest für größere Erhebungen neue Interviewer rekrutiert werden. *Telefonisch* sind Migrantinnen und Migranten häufig schwer bzw. eher mobil zu erreichen, so dass diese Erhebungsart mit besonders niedrigen Ausschöpfungsquoten einhergeht (vgl. Granato 1999; Feskens et al. 2006). Diese und verwandte Probleme führen dazu, dass in vielen großen Erhebungen keine gesonderte Migrantenstichprobe gezogen wird (z. B. im Beziehungs- und Familienpanel *pairfam*; für eine Vorstudie siehe: Baykara-Krumme 2010).

Vor diesem Hintergrund erscheint es lohnenswert, über die Alternative der schriftlichen Befragung nachzudenken. Denn zumindest für Fragestellungen von begrenzter Komplexität oder für die Befragung bestimmter Subgruppen wie etwa jüngere Minderheitenangehörige, bietet sie augenscheinlich viele Vorteile. Zwar sind auch schriftliche Befragungen von sinkenden Ausschöpfungsquoten betroffen (vgl. Aust/Schröder 2009), sie sind allerdings preisgünstig und ihre Feldzeit ist in der Regel kürzer als die von *Face-to-Face* Erhebungen (vgl. Schnell 2012).³ Im direkten Vergleich zu anderen Erhebungsmodi werden Ressourcen frei, die in die Anfertigung von mehrsprachigen Erhebungsmaterialien sowie Maßnahmen zur Erhöhung der Ausschöpfungsquote investiert werden können, etwa in eine gezielte Incentivierung der Zielpersonen. Durch solche Anreize können möglicherweise gerade ansonsten schwer zu erreichende Subgruppen, z. B. weniger gut integrierte Personen, eher erreicht werden.

- 3 Die Feldzeit telefonischer Interviews von ethnischen Minderheiten kann unter bestimmten Bedingungen kürzer sein als die von schriftlichen Erhebungen. Dabei kann jedoch das (in Deutschland) dazu häufig notwendige onomastische Auswahlverfahren im Zusammenhang mit dem nötigen Screening sowie die Zahl verfügbarer bilingualer Interviewer die Feldzeit beeinflussen. Dies spiegelt sich in den unterschiedlichen Feldzeiten aktuellerer telefonischer Migrantenbefragungen wider. Ebenso wie die Einbürgerungsstudie von Weinmann/Becher/Babka von Gostomski (2012) berichten Haberland et al. (2011) eine Feldzeit von knapp 35 Tagen, die beide damit mehr als doppelt so lang im „Feld“ waren wie die vom Zentrum für Türkeistudien (2012) jährliche durchgeführte Mehrthemenbefragung mit 14 Tagen.

Schriftliche Befragungen, so unser Ausgangsargument, stellen – in Kombination mit einer gezielten Incentivierung – angesichts der enormen Schwierigkeiten und Kosten, die mit alternativen Erhebungsmethoden einhergehen, bei Migrantenbefragungen unter bestimmten Bedingungen eine gute Wahl dar. Wir möchten uns mit der Frage beschäftigen, ob der Einsatz von Incentives tatsächlich zu höheren Beteiligungsraten führt – und ob sich die Qualität der Stichprobe im Vergleich zur Befragung ohne Incentives verändert, möglicherweise sogar verbessert. Dazu werden im Anschluss an eine Diskussion des Forschungsstands zur generellen und subgruppenspezifischen Wirkung von Incentives (Abschnitt 2) theoretische Argumente zur Art der Incentivierung vorgestellt (Abschnitt 3). Diese sind allgemeiner Natur und keineswegs „migrantenspezifisch“. Sie münden in das Argument, dass eine „doppelte Incentivierung“, die einen bedingungslosen und einen an die Teilnahme geknüpften materiellen Anreiz kombiniert, eine effektive und effiziente Incentivierungsstrategie darstellt. Der darauf folgende Abschnitt 4 beschreibt ein Methodenexperiment unter 500 jungen türkischstämmigen Erwachsenen, in dessen Verlauf Anreize in Form von Bargeld bis zu einer Höhe von 15 Euro eingesetzt wurden. In der Analyse der Ergebnisse (Abschnitt 5) werden die verschiedenen Incentivierungsstrategien hinsichtlich ihrer quantitativen (Ausschöpfungsquote) und qualitativen (Stichproben- und Datenqualität) Auswirkungen untersucht. Besonderes Augenmerk gilt der Frage, ob strukturell, sozial und identifikativ weniger integrierte Subgruppen durch materielle Anreize möglicherweise besonders gut erreicht werden.

2 Auswirkungen der Incentivierung auf die generelle und subgruppenspezifische Teilnahmebereitschaft: Ein Überblick

Incentives, d. h. materielle Befragungsanreize, werden schon lange als Methode zur Erhöhung der Beteiligung an Umfragen eingesetzt und untersucht (vgl. Wotruba 1966; Schnell 2012).⁴ Ihr – insbesondere in Nordamerika – verbreiteter Einsatz geht mit einer kaum mehr zu überschauenden Literatur einher, die unterschiedliche

4 In seiner *Tailored-Design-Method* weist auch Dillman (1978; 2009) auf Incentives als eine mögliche Form zur Steigerung der Umfragebeteiligung hin. Daneben beschreibt er weitere Faktoren (z. B. ein Ankündigungsschreiben, die Zahl der Kontaktversuche, die Art und Farbe des Papiers, der Tinte oder des Briefumschlags, die Nennung des Sponsors oder die Verwendung von Briefmarken) die sich auf die Beteiligung auswirken können (vgl. für entsprechende Evidenzen die Metastudie von Edwards et al. 2002).

Formen der Incentivierung sowie deren Effektivität und Effizienz analysiert und beschreibt (vgl. aber den Überblick bei Singer 2002; siehe auch: www.websm.org). Im Folgenden werden zunächst Forschungsergebnisse präsentiert, die grundsätzlich für oder gegen einen Einsatz unterschiedlicher Incentivierungsformen sprechen. Im Anschluss daran werden Befunde zu etwaigen subgruppenspezifischen Auswirkungen dargestellt.

2.1 Incentivierung und generelle Teilnahmebereitschaft

Bereits eine ältere Metastudie von Yu und Cooper (1983) konnte auf die insgesamt positive Wirkung von Incentives als Anreiz zur Teilnahme an standardisierten Befragungen hinweisen, spätere Metastudien bestätigen dies grundsätzlich (vgl. Church 1993; Singer et al. 1999; Jobber/Saunders/Mitchell 2004; Edwards et al. 2002). Yu und Cooper zeigen, dass mit monetären Anreizen *ceteris paribus* höhere Rücklaufquoten zu erwarten sind (51 vs. 35 Prozent bei Verzicht auf Incentives). Auch für den deutschsprachigen Raum liegen einige Studien zur Wirkung (geldnaher) Incentives vor (vgl. Arzheimer/Klein 1998; Harkness et al. 1998; Diekmann/Jann 2001; Berger et al. 2005; Becker/Imhof/Mehlkop 2007; Mehlkop/Becker 2007; Stadtmüller 2009). Diese bestätigen die positiven Effekte auf die Rücklaufquote; beispielhaft genannt sei hier Stadtmüller (2009), der bereits durch eine dem Anschreiben beigefügte Euromünze eine signifikante Steigerung der Rücklaufquote erzielt (43 vs. 30 Prozent). Allerdings kommen in der Praxis verschiedene Incentivierungsstrategien zum Einsatz, für die unterschiedliche Auswirkungen beschrieben werden.

Monetäre vs. nicht-monetäre Incentivierung

In einer Reihe von Studien wird darauf hingewiesen, dass mit Bargeld bzw. geldnahen Incentives eine höhere Rücklaufquote erreicht werden kann als mit nicht-monetären Anreizen. So konnte im Vergleich zur Auszahlung eines US-Dollars mit einer Lotterieteilnahme keine bessere Rücklaufquote erzielt werden als in der nicht-incentivierten Kontrollgruppe (vgl. Paolillo und Lorenzi 1984). Auch Warriner et al. (1996) können diesen Befund bestätigen (siehe auch: Porter/Whitcomb 2003; für gegenteilige Befunde bei Online-Befragungen: Bosnjak/Tuten 2003). Während Lotterielose also keinen Effekt entfalten, bewirken versprochene Spenden für wohltätige Zwecke sogar einen geringeren Rücklauf (vgl. Warriner et al. 1996). Auch Geschenke wie Kugelschreiber oder Schlüsselanhänger tragen, selbst unter Kontrolle der Höhe ihres materiellen Werts, weniger zur Steigerung der Rücklaufquote

bei als monetäre Incentives (vgl. Edwards et al. 2002; Singer et al. 1999). Ryu/Couper/Marans (2005) zeigen, dass in einer schriftlichen Befragung beim Einsatz eines Gutscheins trotz seines mehr als doppelt so hohen Wertes die Beteiligung niedriger ist als bei Einsatz eines monetären Incentives (45 vs. 67 Prozent).

Die deutschsprachige Forschung konzentriert sich folglich auf die Wirkung geldnaher (vgl. Arzheimer/Klein 1998; Harkness et al. 1998; Porst 1999; Diekmann/Jann 2001) oder bar ausgezahlter Incentives (vgl. Berger et al. 2005; Stadtmüller/Porst 2005; Becker/Imhof/Mehlkop 2007; Mehlkop/Becker 2007; Stadtmüller 2009).

Konditionale vs. unbedingte Incentivierung

Die Strategie *unkonditionaler Incentivierung* (UI) zielt darauf ab, die Befragten mit Hilfe eines vorab und ohne weitere Bedingung überreichten Geschenks zur Teilnahme an der Befragung zu motivieren. Die Logik *konditionaler Incentivierung* (KI) besteht hingegen darin, den Nutzen der Teilnahme mit Hilfe der angebotenen Belohnung für eine erfolgreiche Teilnahme zu erhöhen. Yu und Cooper (1983: 39) zeigen, dass – verglichen mit der Situation ohne Incentives – sowohl monetäre unbedingte („prepaid“) als auch konditionale („promised“) Incentives zu signifikant höheren Rücklaufquoten führen. Singer et al. (1999) bestätigen dies: Beide, KI und UI, bewirken höhere Rücklaufquoten; der Effekt bedingungslos gezahlter Incentives ist dabei tendenziell (aber nicht signifikant) stärker. Church (1993) indes unterstreicht, dass der Einsatz monetärer wie nicht-monetärer Anreize dann wirkungslos bleibt, wenn diese nur *konditional* ausgezahlt werden. Die vorliegenden Metastudien kommen somit zu unterschiedlichen Ergebnissen bezüglich des allgemeinen Nutzens der KI gegenüber einem völligen Verzicht auf Incentives. Eindeutig ist hingegen der Befund einer Gegenüberstellung von konditionalen und unbedingten Strategien: Vorab und bedingungslos gemachte (monetäre) „Geschenke“ haben einen stärkeren Effekt auf die Bereitschaft zur Teilnahme an Befragungen als lediglich versprochene Geschenke.

Im deutschsprachigen Raum konnten sowohl Diekmann und Jann (2001) als auch Becker/Imhof/Mehlkop (2007) diesen Befund bestätigen und zudem auf mögliche negative Effekte einer KI hinweisen. So liegt in beiden Experimenten die Ausschöpfungsquote der KI leicht unter derjenigen der nicht-incentivierten Kontrollgruppe (KG).⁵

5 Die Ausschöpfungsquote der nicht-incentivierten KG liegt im Experiment von Diekmann und Jann (2001) bei 77 Prozent, der konditionalen Gruppe bei 74 Prozent und der unbedingten Gruppe bei 81 Prozent (jeweils 10 CHF als Telefonkarte). Becker/Imhof/Mehlkop (2007) erreichen in der KG einen Rücklauf von 39 Prozent, in der kondi-

Wie viel hilft viel? – Hohe vs. niedrige Incentivierung

Yu und Cooper (1983) haben auf einen starken linearen Zusammenhang zwischen Incentivehöhe (von 0,1 bis 50 Dollar) und Teilnahmebereitschaft hingewiesen. Singer et al. (1999), bei deren Studien Incentives zwischen 1 bis 100 Dollar eingesetzt wurden, sowie Jobber/Saunders/Mitchell (2004) (0,1 bis 5 Dollar) kommen zu ähnlichen Ergebnissen. Weniger eindeutig erscheinen diese Befunde, wenn einzelne Methodenexperimente betrachtet werden, welche die Höhe der Incentivierung unter sonst gleichen Bedingungen variieren und so deren Wirkung direkter erfassen.

Trussell und Lavrakas (2004) liefern in einem Experiment mit schrittweiser Erhöhung des Incentives um einen bis maximal zehn Dollar, einen Nachweis für den annähernd linearen Effekt materieller Anreize. Bolstein und James (1990) testeten andererseits vier unterschiedlich hohe unkonditionale Incentives (0,25, 0,5, 1 und 2 Dollar) gegen eine nicht incentivierte KG. Ihren Befunden zufolge können sowohl mit einem als auch mit zwei Dollar signifikante Steigerungen der Rücklaufquote gegenüber der Nicht-Incentivierung erreicht werden; dabei macht es keinen messbaren Unterschied, ob nun mit einem oder zwei Dollar „incentiviert“ wurde. Auch eine spätere Studie spricht gegen einen linearen Effekt der Anreizhöhe auf die Teilnahmebereitschaft (vgl. Bolstein/James 1992). Sie erzielten mit einem unkonditional versandten Scheck über 20 Dollar die höchste Rücklaufquote (79 vs. 52 Prozent in der KG). Mit den niedrigeren Beträgen (1, 5 und 10 Dollar) wurden dann zwar geringere Rückläufe erreicht (aber immer noch höhere als in der KG), ihre jeweiligen Abstände zueinander sind jedoch klein und nicht überzufällig. Bei einer Verdoppelung der jeweils unkonditional gezahlten 20 auf schließlich 40 Dollar stellt sich überraschend eine um 10 Prozentpunkte geringere Rücklaufquote ein. Ein noch höherer konditionaler Anreiz von 50 Dollar führt dazu, dass die Rücklaufquote unter das Niveau der KG fällt (siehe auch: Singer/Van Hoewyk/Maher 2000). Als Gegenbeispiel sei hier allerdings noch auf die Arbeit von Dykema et al. (2011) hingewiesen. Hier kann eine besonders schwer zu befragende Gruppe (Mediziner) durch außerordentlich hohe materielle Anreize sehr gut erreicht werden: Mit konditional angebotenen 50 Dollar nahmen 15 Prozent (vs. 6 Prozent in der KG) und mit 100 Dollar schließlich 25 Prozent der Mediziner an der Befragung teil.

Berger et al. (2005) haben Ergebnisse einer Untersuchung vorgelegt, bei der, abgesehen von einem bislang nicht weiter veröffentlichten Methodenexperiment im Rahmen des SOEP (vgl. Schupp et al. 2010), erstmalig systematisch die Wir-

Fortsetzung Fußnote 5

tionalen Gruppe 34 Prozent und in der unkonditionalen Gruppe 63 Prozent (jeweils 10 CHF in bar).

kung unterschiedlich hoher Incentives im deutschsprachigen Raum getestet wird. Mit steigendem Incentive (10 bzw. 20 DM un konditional sowie 50 DM konditional) gehen keine signifikanten Veränderungen der Rücklaufquote einher, tendenziell sinkt diese sogar.

Diese Ergebnisse machen deutlich, dass Incentives offenbar keine lineare Wirkung auf die Teilnahmebereitschaft haben. Berger et al. argumentieren, dass höhere materielle Anreize durchaus höhere Rücklaufquoten bewirken können, allerdings nur bis zu einer bestimmten Grenze. Die Befunde hierzu sind allerdings uneinheitlich. In einigen Studien finden sich sogar negative Auswirkungen besonders hoher konditionaler Incentives.

2.2 Incentivierung und subgruppenspezifische Teilnahmebereitschaft

Die Leverage-Saliency Theorie weist darauf hin, dass sich jeder einzelne Faktor eines Erhebungsdesigns auf das Teilnahmeverhalten bestimmter Gruppen der Zielpopulation unterschiedlich auswirken kann und dass „the activation of the potential leverage depends on whether the attribute is made salient to the sample person during the survey request“ (Groves/Singer/Cornig 2000: 307). Für jegliche Attribute des Erhebungsdesigns und damit auch für Incentives ergibt sich daraus die Konsequenz, dass deren Einsatz „maßgeschneidert“ an die jeweilige Zielpopulation angepasst (vgl. Dillman 1978) und ihr Einsatz getestet werden sollte. Die Frage, ob und durch welche Incentivierungsstrategie die Teilnahmebereitschaft insgesamt erhöht werden kann, ist primär von forschungspraktischer Relevanz. Entscheidend ist die Frage, ob bestimmte gesellschaftliche Subgruppen *unterschiedlich* auf die Incentivierung reagieren. In diesem Fall bestünde die Gefahr eines Stichprobenbias – oder aber die Chance, durch eine gezielt eingesetzte Incentivierung Ausfälle bei ansonsten schwer zur Teilnahme zu bewegendenden Bevölkerungsgruppen zu verringern.

In vielen Studien werden die zu befragenden Zielpersonen zunächst zufällig unterschiedlichen Incentivierungsbedingungen zugewiesen (Bruttostichproben); anschließend werden die jeweiligen tatsächlich befragten Zielpersonen (Nettostichproben) miteinander verglichen. Durch einen solchen Vergleich kann allerdings keine Aussage darüber getroffen werden, welche „Versuchsgruppe“ am wenigsten Verzerrungen gegenüber der jeweiligen Bruttostichprobe aufweist. Dies gilt etwa für die Studie von Stadtmüller (2009), der auf einen geringeren Anteil an Hauptschülern in der Gruppe stößt, die ein Incentive erhalten hat. Welche Incentivierungsstrategie diesbezüglich die besten Ergebnisse liefert, kann nur anhand eines Vergleichs von Brutto- und Nettostichprobe ermittelt werden. Die Befunde

der entsprechenden Studien sind eher uneinheitlich, wie wir im folgenden Abschnitt zeigen werden.

Incentivierung und Teilnahmebereitschaft von ökonomisch Benachteiligten und ethnischen Minderheiten

Singer et al. (1999) weisen darauf hin, dass bei einem Großteil der untersuchten Studien keine Stichprobenverzerrungen von Incentivierungen ausgehen (vgl. auch James/Bolstein 1990; Warriner et al. 1996; Arzheimer/Klein 1998; Shettle/Mooney 1999). In einigen erhöhen allerdings materielle Anreize die Bereitschaft zur Teilnahme gerade solcher Gruppen, die gemeinhin eine geringere Teilnahmebereitschaft aufweisen („e.g., low income or nonwhite race“). Die Qualität der Stichprobe konnte durch Incentives hier also erhöht werden. Wetzels et al. (2008) zeigen in einem Methodenexperiment des niederländischen Labour Force Surveys, dass materielle Anreize keine verzerrenden Auswirkungen auf die sozioökonomische Zusammensetzung der realisierten Stichprobe haben. Bereits in einer älteren Studie hat Goyder (1994) Informationen u. a. über Geschlecht, Familienstand und Vermögen einbeziehen und zeigen können, dass diese Merkmale Einfluss auf die Teilnahmebereitschaft haben und z. B. Arbeitslose seltener an der Befragung teilnahmen. Der verzerrende Effekt des jeweiligen Merkmals weist in den incentivierten Gruppen zwar in die gleiche Richtung, fällt aber in der Gruppe mit einem un konditional gezahlten Dollar (vs. KG sowie 1 bzw. 10 Dollar konditional gezahlt) am schwächsten aus. Incentives konnten mithin den jeweiligen Bias nicht eliminieren, schwächten ihn im Vergleich mit der KG aber ab. In einer australischen Studie können Barón et al. (2009) die Information über das Ausmaß wohlfahrtsstaatlicher Bezüge der zurückliegenden zwölf Jahre auf Haushaltsebene rekonstruieren und diese Information in einem Experiment nutzen. Dabei zeigt sich ein Bias dergestalt, dass Familien, die nie wohlfahrtsstaatliche Leistungen erhalten haben, deutlich häufiger an der Befragung teilnahmen als Familien, die wiederholt Empfänger von Transferleistungen gewesen sind. Durch den Einsatz materieller Incentives konnte dieser Bias reduziert werden. In einer deutschen Incentivierungsstudie werden Kontextinformationen über das durchschnittliche Einkommen des Ortsteils einbezogen, signifikante Einflüsse der Incentivierung auf die Repräsentativität der Stichprobe werden dabei jedoch nicht gemessen (vgl. Arzheimer/Klein 1998).

Eine der wenigen Studien zu Auswirkungen einer Incentivierung bei Migranten verweist darauf, dass in den Niederlanden ethnische Minderheiten aus nicht westlichen Ländern auf den angebotenen Anreiz kaum reagieren (Wetzels et al. 2008). Zwar steigt die Kooperationsrate der Ausländer von 79 Prozent in der KG auf 93 Prozent unter der Bedingung der höchsten un konditionalen Incentivierung

(autochthone Bevölkerung: 74 vs. 84 Prozent). Dieser Unterschied ist bei den Ausländern – vermutlich aufgrund der sehr kleinen Stichprobengröße – jedoch nicht signifikant. So waren in der höchsten Incentivierungsgruppe lediglich 41 ausländische Personen (gegenüber 355 Autochthonen) enthalten. Die Autoren liefern keine weitere Erklärung dieses Befundes, der in einer weiteren Veröffentlichung detaillierter beschrieben wird (vgl. Feskens et al. 2008). Im Anschluss an die Vertrauensforschung, die zeigt, dass Vertrauen mit steigender sozialer Distanz abnimmt (vgl. Goto 1996), könnte eine denkbare Erklärung sein, dass die für die Teilnahmebereitschaft wichtige Reziprozität zwischen Mehrheits- und Minderheitenangehörigen, die sich als Befragende bzw. Befragte gegenüberstehen, schwerer zu aktivieren ist als zwischen gleichen Gruppen.

Insgesamt betrachtet scheinen Befürchtungen, dass materielle Anreize in standardisierten Befragungen die gewonnene Stichprobe negativ verzerren könnten, auf Basis der vorliegenden Forschungsbefunde unbegründet. Wenn Incentives überhaupt einen Effekt auf die Stichprobenqualität haben, dann eher einen positiven im Sinne einer höheren Teilnahmebereitschaft gerade bei solchen Gruppen, die häufig schwer zur Teilnahme an Surveys zu bewegen sind.

3 Warum und wie wirken Incentivierungsstrategien auf die Teilnahmebereitschaft?

Vor dem Hintergrund der teils widersprüchlichen empirischen Befunde zu den Folgen des Incentiveeinsatzes stellt sich die Frage, wie man unterschiedliche Arten der Incentivierung theoretisch begründen kann, und wie eine „theoriegeleitete Incentivierungsstrategie“ aussehen könnte. Dazu lohnt zunächst ein genauerer Blick auf die Frage, was Befragte dazu bewegt, *ohne* jeglichen materiellen Anreiz an Befragungen teilzunehmen, obwohl diese in der Regel auf freiwilliger Basis stattfinden. Dies ist voraussichtlich dann der Fall, wenn der subjektiv erwartbare Nutzen der Teilnahme den der Nicht-Teilnahme übersteigt (vgl. Esser 1986). Dieser wiederum variiert mit dem Glauben an die Wichtigkeit des Befragungsthemas, dem Gefühl einer „ausgewählten“ Gruppe anzugehören, der Verbundenheit mit oder dem Respekt vor der befragenden Person oder der jeweiligen Institution und dem Gefühl, an etwas Sinnstiftendem beteiligt zu sein (vgl. Dillman 1978; Porst 1999).

Zudem wird bereits durch die Kontaktaufnahme eine Austauschbeziehung eröffnet, in der die Reziprozitätsnorm mit ihrem universellen Charakter Wirkung entfaltet (vgl. Gouldner 1960). Die Reziprozitätsthese geht davon aus, dass Individuen bemüht sind, freiwillig erhaltene Vorteile durch eine Gegenleistung zu erwi-

dern. Die hier relativ abstrakte – und daher sehr schwache – Form von Reziprozität wird erzeugt, indem den Befragten die gesamtgesellschaftliche Notwendigkeit vor Augen geführt und unterstrichen wird, „dass zu guter und aktiver Staatsbürgerschaft einer offenen demokratisch verfassten Gesellschaft auch die Bereitschaft zählt, die Forschung zu unterstützen“ (Schupp 2012: 20). „Die Interviewteilnahme stiftet so durch die Erfüllung der [...] Reziprozitätsnorm dem Befragten einen zusätzlichen Nutzen“ (Arzheimer/Klein 1998: 8). Eine reziproke Situation kann in interviewbasierten Erhebungen auch zwischen Interviewern und Teilnehmern entstehen, wenn sich letztere verpflichtet fühlen, den um ein Interview Bittenden zu helfen.

Es gibt nur wenige Möglichkeiten den geringen Nutzen der Interviewteilnahme zu erhöhen, jedenfalls wenn er in der intrinsisch motivierten „Teilnahme um der Sache willen“ besteht und zudem auf dem Gefühl beruht, eine Reziprozitätsnorm zu erfüllen. In schriftlich und persönlich durchgeführten Erhebungen wird zwar schon durch ein Begleitschreiben bzw. eine Vorankündigung versucht, die Befragten von der Wichtigkeit des Themas zu überzeugen und dabei die freiwillige Teilnahme als wichtigen Bestandteil des sozialen Austauschs („Unterstützung der Forschung“) darzustellen. Kostenseitig steht der Teilnahme aber einerseits die Unsicherheit über die Verwendung der preisgegebenen Informationen und andererseits die zu investierende Zeit gegenüber. Der einzig zu veranschlagende Kostenpunkt der Teilnahmeverweigerung bleibt hingegen psychischer Natur und besteht in der Verletzung von „Höflichkeitsnormen“ (Esser 1986: 39).

Die oben dargestellten empirischen Befunde, denen zufolge ohne Incentivierung offenbar ökonomisch bessergestellte gesellschaftliche Subgruppen leichter zu befragen sind, könnten in dieser Anreizstruktur ihre Ursache haben: Es scheint durchaus plausibel, dass gerade die Unterstützung wissenschaftlicher Forschung einen Nutzenterm darstellt, der für diese meist höher gebildeten Gruppen eher wirkt als für Bildungsferne. Die Unsicherheit über die verwendeten Daten könnte bei dieser Gruppe gleichzeitig höher sein.

Unkonditionale Incentivierung

Angesichts der im letzten Absatz dargestellten Anreizstruktur stellt sich die Frage, wie die Wahrscheinlichkeit der Wahl der Handlungsalternative „Teilnahme“ erhöht werden kann, wenn die Möglichkeiten ausgereizt sind, den perzipierten Nutzen der Interviewteilnahme zu beeinflussen. An dieser Stelle setzt die unkonditionale Incentivierung (UI) an, die versucht, die Kosten der Nicht-Teilnahme zu erhöhen. Dabei geht es darum, durch eine vorab und bedingungslos geleistete „Gabe“ die Bindung an die Reziprozitätsnorm und damit an das – ohne Incentivierung schwa-

che – Gefühl der Verpflichtung zur angemessenen Gegenleistung zu verstärken: „Sending the money with the questionnaire is a goodwill gesture that puts the sponsor and questionnaire in a positive light and sets the stage for the respondent to reciprocate with an appropriate gesture of completing the questionnaire“ (Dillman 2007: 168). Das gezahlte Incentive ist dabei keine Kompensation für die investierte Zeit, sondern vielmehr ein symbolischer Akt, der den Nutzen der Teilnahme nicht berührt. Es stellt daher kein direktes „motive or incitement to action“ (Grant/Sugarman 2004: 720) dar, sondern verteuert vielmehr die Alternative „Nicht-Teilnahme“.

Vor diesem Hintergrund erscheint auch der oben dargestellte Befund plausibel, dass ein besonders hohes unkonditionales Incentive zu einer sinkenden Beteiligung führt: Eine allzu große „Gabe“ dürfte als nicht erwidierbar und deshalb unangemessen wahrgenommen werden und nicht als rein symbolischer Akt, der frei von der impliziten Erwartung einer Gegenleistung ist (vgl. Mauss 1968). Durch eine zu große Vorleistung rückt die Frage nach dem Grund für die großzügige Entlohnung in den Vordergrund und erzeugt Reaktanz, eine Abwehrreaktion auf die „Nötigung“ durch die Forschenden (vgl. Becker/Imhof/Mehlkop 2007). Ab einem bestimmten – individuell aber variierenden – Schwellenwert dürfte die Teilnahmebereitschaft also sinken, weil nun der Nutzen der Nicht-Teilnahme steigt (vgl. Esser 1986). Die Nicht-Teilnahme ist wohl die einzig mögliche Handlung, um die Kontrolle über die soziale Situation wiederzugewinnen und auf die „Nötigung“ zu reagieren.⁶ Sie bleibt zudem völlig folgenlos, weil die Teilnahme an der Befragung freiwillig ist und das Incentive bedingungslos versandt wurde.

Im Gegensatz zur Nicht-Incentivierung dürften bei der Strategie der UI also nicht nur diejenigen an der Befragung teilnehmen, die eine intrinsische Motivation zeigen, die Forschung zu unterstützen, und für die durch die Befragungssituation als solche bereits eine reziproke soziale Beziehung hergestellt werden kann. *Zusätzlich* sollten auch diejenigen teilnehmen, bei denen diese Reziprozität erst durch einen geeigneten materiellen Anreiz erzeugt bzw. verstärkt wird und somit die Kosten der Nicht-Teilnahme gesteigert werden. Dies dürfte besonders für diejenigen Befragten relevant sein, die weniger ausgeprägte Affinitäten zur Wissenschaft aufweisen, z. B. weil sie diesem System aufgrund ihrer eigenen Bildungsferne distanziiert gegenüberstehen. Auch Angehörige ethnischer Minderheiten, die sich nur schwach mit dem Zielland identifizieren, könnten die befragenden Wissenschaftler stärker als out-group Mitglieder wahrnehmen (Johnson et al. 2002)

6 Eine noch schärfere Form der Reaktion auf diese „Nötigung“ wäre es, wenn die Teilnahme verweigert und das Incentives zurückgeschickt wird (immerhin zwei Befragte wählten in dem hier vorgestellten Experiment diese Variante).

und daher weniger geneigt sein, auf den impliziten Appell zu reagieren, der nicht-incentivierten Studien zugrunde liegt: dass die Unterstützung der Forschung eine staatsbürgerliche Pflicht darstelle.

Konditionale Incentivierung – Die rationale Wahl?

Die Strategie der UI geht aus einer Effizienzperspektive mit einer doppelten Ressourcenverschwendung einher. Der Sponsor investiert in diejenigen, die auch ohne Incentive an der Befragung teilgenommen hätten – und belohnt selbst die hartnäckigsten Verweigerer. Bereits Goyder wirft die Frage auf, ob bei *Face-to-Face* Erhebungen „the psychological ploy implied by prepayment might seem provocative“ (1994: 360), weshalb für diese Erhebungsform wohl häufig konditionale Incentives eingesetzt werden. Die rational nahe liegende Strategie, die Auszahlung des Incentives an die Teilnahme zu koppeln, scheint jedoch, wie oben gezeigt wurde, nicht die gewünschten Resultate zu bringen. Angesichts der finanziellen Vorteile, die mit der konditionalen Incentivierung (KI) verbunden sind, soll hier noch einmal die Frage beleuchtet werden, wie dies erklärt werden kann, um anschließend eine modifizierte KI vorzustellen, die gleichzeitig effektiv *und* effizient ist.

Die Logik der KI besteht anders als die der UI darin, den Nutzen der Alternative „Teilnahme“ zu erhöhen. Entsprechend muss der materielle Anreiz hinreichend attraktiv sein, um die oben beschriebenen Kosten der Teilnahme zu übertreffen, oder zugespitzt „attractive enough to tempt people to participate in a research study 'against their better judgement'“ (Grant/Sugarman 2004: 734). Bei genauerer Betrachtung wird schnell klar, warum die Strategie der KI wenig geeignet ist, die Kosten-Nutzen-Kalkulation ins Positive zu wenden. So ist die Befragungssituation für die Befragten ohnehin mit Unsicherheit (z. B. in Bezug auf die Sicherheit ihrer Daten) verbunden (vgl. Esser 1986). Das Versprechen eines bestimmten Geldbetrags ist ein weiterer Unsicherheitsfaktor; es besteht nun zusätzliche Unsicherheit darüber, ob das versprochene „Dankeschön“ für die Teilnahme tatsächlich ausgezahlt wird. Auch mit steigendem Incentive lässt sich lediglich der *potentiell* erwartbare Nutzen erhöhen. Die Unsicherheit über die Wahrscheinlichkeit der Zielerreichung kann so aber nicht beseitigt oder auch nur reduziert werden, möglicherweise steigt diese sogar mit der Höhe des angekündigten konditionalen Incentives, weil dieses Versprechen zunehmend unglaubwürdig wirkt.⁷ Dieses Argument trifft allerdings

7 Dykema et al. (2011) erzielen mit einem besonders hohen konditionalen Anreiz allerdings eine deutliche Verbesserung der Rücklaufquote. Bei einem extrem hohen Betrag erscheint die Möglichkeit, hereingelegt zu werden, nicht unbedingt kleiner, der ausgelegte „Köder“ ist jedoch schlicht zu verlockend, um dieses Risiko nicht einzugehen. Der subjektiv erwartbare Nutzen steigt also, jedoch nicht, weil der „p-Wert“, d. h. die

weniger auf *Face-to-Face* Erhebungen zu. Zwar dürfte die physische Präsenz der Interviewer die Unsicherheit bezüglich der Auszahlung des Incentives erheblich reduzieren, die eigentliche Interviewsituation, in der eine fremde Person Geld anbietet, um Einlass bittet und persönliche Informationen erfragen möchte, bleibt jedoch – abhängig vom „Charme“ der Interviewer – mit einiger Unsicherheit behaftet.⁸

Ein weiteres theoretisches Argument gegen die KI ist, dass die Befragten durch die Ankündigung einer Gratifikation als Gegenleistung für die Interviewteilnahme extrinsisch motiviert agieren – was ihre intrinsische Motivation untergraben kann – und rational kalkulieren. Da in den meisten dargestellten Studien die Höhe des eingesetzten konditionalen Incentives eher gering ist, fällt eine solche rationale Abwägung von Aufwand und finanziellem Ertrag aber offenbar meist zuungunsten der Teilnahme aus. Vor diesem Hintergrund lässt sich leicht erklären, warum eine KI in vielen Studien keine oder sogar eine kontraproduktive Wirkung zeigt: So dürften sich sowohl diejenigen an der Befragung beteiligen, die auch ohne Incentive dazu bereit gewesen wären – *abzüglich* derer, deren intrinsische Motivation durch die Ankündigung der Bezahlung negativ beeinflusst wird und *zuzüglich* derjenigen, die den Nutzen der Befragung aufgrund der KI höher einschätzen als deren Kosten.

Dies dürfte aber eine verhältnismäßig kleine Gruppe sein, da erstens auch in *Face-to-Face* realisierten Befragungen die Unsicherheit der Befragungssituation bestehen bleibt und weil zweitens ein KI häufig zu niedrig ist, um als angemessen und attraktiv wahrgenommen zu werden.

Die Logik doppelter Incentives

Angesichts der bislang dargelegten Argumente für und gegen die unterschiedlichen Incentivierungsformen liegt es nahe, über eine Incentivierungsstrategie nachzudenken, die *gleichermaßen* die Kosten der Nicht-Teilnahme und den Nutzen der Teilnahme erhöht (vgl. Abbildung 1). Eine solche Strategie soll im Folgenden unter

Fortsetzung Fußnote 7

subjektiv wahrgenommene Wahrscheinlichkeit, dass die Teilnahme zum gewünschten Ziel führt, ansteigt, sondern weil der Nutzenterm (U) selbst größer wird.

- 8 Vor diesem Hintergrund ist der Befund bei Goyder (1994) von Bedeutung, der die positiven Befunde der UI in schriftlichen Erhebungen zum Anlass nimmt, diese Ergebnisse für persönlich geführte Interviews zu validieren. Gegen eine nicht-incentivierte Kontrollgruppe testet er ein UI von einem Dollar sowie zwei KI in Höhe von 1 bzw. 10 Dollar. Bemerkenswert ist dabei, dass die von den Interviewern angebotenen 10 Dollar zu keiner signifikant besseren Beteiligung im Vergleich zur Kontrollgruppe geführt haben. Lediglich die zuvor versandte Dollarmünze sorgte für eine gut 12 Prozentpunkte signifikant höhere Beteiligung (vgl. auch Singer et al. 1999).

Abbildung 1 Auswirkungen unterschiedlicher Incentivierungsstrategien auf die Anreizstruktur der Interviewteilnahme

	Zusätzlicher Nutzen der Teilnahme...	Zusätzliche Kosten der Nicht-Teilnahme...
	... im Vergleich zur rein intrinsisch motivierten Teilnahme	
Unkonditionale Incentivierung (UI)	0	+
Konditionale Incentivierung (KI)	(+) ¹	0
Doppelte Incentivierung (DI)	+	+

1 Nur bei Vertrauen in die tatsächliche Auszahlung des konditionalen Anreizes

dem Stichwort der „doppelten Incentivierung“ (DI) diskutiert werden. Ziel ist, mit Hilfe des im ersten Schritt ausgezahlten unkonditionalen Anreizes die bindende Kraft der Reziprozität zu nutzen, d. h. die Kosten der Nicht-Teilnahme zu erhöhen. Das wesentliche Merkmal einer reziproken Situation ist der Aufbau von Vertrauen, insofern wirkt die UI hier auch als ein „symbol of trust“ (Dillman 1978 zitiert nach: Arzheimer/Klein 1998: 10). Dies gilt auch für persönlich durchgeführte Befragungen, bei denen der unkonditionale Anreiz vorab und damit vor der Teilnahmeentscheidung mit einem Ankündigungsschreiben versandt wird und die Kontaktaufnahme durch die Interviewer für die Befragten so in einem anderen, positiven Kontext stattfindet (vgl. Groves/Singer/Cornig 2000). Die UI schafft damit eine wichtige Grundlage für eine KI, denn das Versprechen einer materiellen Belohnung findet in einer sozialen Situation statt, in der die Befragten Vertrauen aufgebaut haben und außerdem hinreichend sicher sein können, dass die Ankündigung auch umgesetzt wird. Dann allerdings sollte ein höheres Incentive eine größere Wirkung zeigen als ein niedriges: Da der konditionale Anteil der DI direkt an eine zuvor erbrachte Leistung gekoppelt ist, sollte diese umso besser wirken, je angemessener und attraktiver der gegebene Anreiz erscheint.

An einer Befragung mit DI Strategie müssten folglich diejenigen – intrinsisch motivierten Befragten – teilnehmen, die auch ohne Incentive teilgenommen hätten, *zuzüglich* derjenigen, bei denen Reziprozität durch ein Incentive hergestellt wird, *zuzüglich* derjenigen, bei denen die Norm der Reziprozität trotz des unkonditionalen Incentives nicht wirkt, die aber auf den zusätzlichen Anreiz reagieren, mit sehr hoher Wahrscheinlichkeit eine (angemessene) Belohnung für ihre Teilnahme zu erhalten.⁹

9 Denkbar wäre zwar auch, dass die intrinsische Motivation einiger Befragter, die „im Dienste der Wissenschaft“ an der Befragung auch ohne jegliche Incentivierung teil

Was die Auswirkungen der unterschiedenen Strategien auf die Teilnahme anbelangt, lassen sich nun folgende Hypothesen formulieren: *Erstens müsste durch die UI die Teilnahme gegenüber der nicht incentivierten, rein intrinsisch motivierten KG steigen (H1)*, da die Reziprozitätsnorm durch eine kleine Vorabgabe gestärkt wird. *Zweitens dürfte eine Erhöhung der UI keine darüber hinausgehenden positiven Auswirkungen auf die Teilnahmebereitschaft zeigen (H2)*, da die Aktivierung der Reziprozitätsnorm keinen Nutzenerwägungen folgt. Diese könnte sogar ausbleiben, wenn die unverlangte Gabe so hoch ist, dass die Befragten glauben, keine Wahl mehr zu haben und mit Reaktanz reagieren.¹⁰ *Drittens dürfte unter der Bedingung einer Kombination aus UI und KI die Teilnahmebereitschaft höher sein als bei der reinen UI (H3)*. Denn durch DI werden nun zusätzlich diejenigen zur Teilnahme bewegt, die weder intrinsisch noch normativ motiviert sind, aber um der angekündigten Belohnung willen teilnehmen, die zu erhalten sie aufgrund der UI als wahrscheinlich einschätzen dürften. *Viertens sollte eine Erhöhung des KI positive Auswirkungen auf die Teilnahmebereitschaft haben (H4)*, da eine nachträglich gezahlte „Belohnung“ umso stärker wirkt, je näher diese an eine angemessene und attraktive Entlohnung des Aufwands heranreicht. *Fünftens erwarten wir, dass durch den Einsatz von Incentives eine Verzerrung zugunsten der ökonomisch bessergestellten Zielpersonen verringert werden kann (H5)*. Diese Hypothese beruht auf dem Argument, dass gerade privilegierte (und häufig höher gebildete) Personen am ehesten ein intrinsisches Interesse an einer wissenschaftlichen Befragung zeigen werden. Und wie oben argumentiert wurde, gehen wir *sechstens* davon aus, *dass Minderheitenangehörige, die starke Bindungen an ihre eigene ethnische Gruppe aufweisen, seltener auf den Appell reagieren, ohne jegliche Incentivierung an einer Befragung teilzunehmen, als andere Minderheitenangehörige (H6)*.

Fortsetzung Fußnote 9

genommen hätten, durch die versprochene Belohnung zerstört wird. Allerdings kann hier spekuliert werden, dass dies nicht unmittelbar geschieht, sondern sich als mittelfristiger Prozess negativ auf die Bereitschaft auswirkt auch ohne Incentivierung an künftigen Befragungen teilzunehmen. Positive Effekte von Incentives in Panelstudien, im Sinne einer geringeren Panelmortalität, sowie ausbleibende Effekte anfänglicher Incentivierung in Folgewellen sprechen allerdings gegen diese Überlegung (vgl. Castiglioni/Pforr/Krieger 2008; Arzheimer/Klein 1998).

10 Allerdings ist fraglich, ob diese Reaktion bereits durch den in der vorliegenden Studie eingesetzten moderaten Incentive ausgelöst wird.

4 Untersuchungsanlage und methodisches Vorgehen

Die Effekte unterschiedlicher Incentivierungsstrategien auf die Ausschöpfungsquote und die Zusammensetzung der realisierten Stichprobe wurde im Rahmen der Pilotstudie „Deutsche/r bleiben?“ untersucht. Die Pilotierung verfolgte das Ziel, die Befragungsbereitschaft junger Erwachsener mit doppelter (deutscher und türkischer) Staatsbürgerschaft zu ermitteln. Aufgrund der geringen Größe dieser Gruppe war eine hohe Beteiligung besonders wichtig, daher wurden verschiedene Maßnahmen zur Steigerung der Ausschöpfungsquote getestet.

Stichprobe und Erhebungsinstrument

Die aus inhaltlichen Gründen interessierende „seltene Population“ besteht aus türkischstämmigen „Optionspflichtigen“, die sowohl die deutsche als auch die türkische Staatsbürgerschaft besitzen (vgl. Diehl/Fick 2012).¹¹ Mit einer registerbasierten Zufallsstichprobe wurden die Adressen von 500 Personen der Jahrgänge 1990 bis 1993 gezogen, dazu wurde auf die Melderegister der Städte Augsburg, Bielefeld, Bremen, Bremerhaven, Dortmund, Düsseldorf und Hannover zurückgegriffen. Neben den bekannten „klassischen“ Maßnahmen zur Erhöhung der Ausschöpfungsquote (vgl. Dillman/Smyth/Christian 2009; Schnell 2012) wurde zumindest ansatzweise auf ein Mixed-Mode-Design zurückgegriffen: Dahinter steht die Idee, die Befragungsteilnahme so angenehm wie möglich zu gestalten, indem der Befragungsmodus frei gewählt werden kann (vgl. Millar/Dillman 2011). Da nur postalische Adressinformationen und keine Telefonnummern der Zielpersonen vorlagen, fanden sowohl der Erstkontakt als auch die Erinnerungsschreiben schriftlich statt, die Teilnahme selbst konnte schriftlich, telefonisch oder via Internet erfolgen, wobei die überwiegende Mehrheit den beigelegten schriftlichen Fragebogen von ca. 15 Minuten Länge nutzte.¹²

11 Optionspflichtige gehören zur zweiten Migrantengeneration, sind also in Deutschland geboren. Daher wurde auf den Versand zweisprachiger Erhebungsmaterialien verzichtet, weil sprachliche Schwierigkeiten bei der Beantwortung nicht zu erwarten waren. Der Methodenbericht der Einbürgerungsstudie von Weinmann/Becher/Babka von Gostomski (2012) bestätigt dies; so musste keines der Face-to-Face durchgeführten Interviews mit den türkischstämmigen Optionspflichtigen aufgrund von Sprachproblemen abgebrochen werden.

12 Da mehr als 75 Prozent der Interviews schriftlich erfolgten und keine signifikanten Unterschiede hinsichtlich der Incentivierungsbedingung bestehen, wird auf eine Differenzierung in der Ergebnisdarstellung verzichtet.

Experimentelles Design und Ablauf der Befragung

Zum Start der Befragung (Juni 2011) erhielten alle Befragten ein erstes postalisches Anschreiben.¹³ Die Ausgangsstichprobe wurde dabei zufällig in fünf Subgruppen à 100 Zielpersonen eingeteilt, bestehend aus einer Kontrollgruppe sowie vier unterschiedlichen Incentivegruppen. Aus pragmatischen Gründen und aufgrund der geschilderten recht eindeutigen Befunde existierender Studien wurde dabei auf eine rein konditionale Incentivierung verzichtet.

- I. Der *Kontrollgruppe* (KG) wurde kein materieller Teilnahmeanreiz gewährt. Wie in den übrigen Gruppen auch, lag dem Erhebungsmaterial ein Anschreiben bei, welches mit folgendem Satz abschloss: „Wir bedanken uns im Voraus für Ihre Mitarbeit und Ihren Beitrag zur sozialwissenschaftlichen Forschung.“
- II. Die zweite Gruppe erhielt ein *unkonditionales Incentive in Höhe von 5 EUR* (UI 5). Dazu wurde in der unteren freien Ecke des Anschreibens ein Fünfeuroschein aufgeklebt, darunter der Hinweis „5 € für Ihre Mühe als kleines Dankeschön!“.
- III. Um die Wirkung der Höhe des Anreizes messen zu können, erhielt die dritte Gruppe ein *unkonditionales Incentive* in Höhe von 10 EUR (UI 10).
- IV. Schließlich sollte die Strategie *doppelter Incentivierung mit 5 + 5 EUR* (DI 5+5) getestet werden. Analog zur Gruppe zwei war hier ein Fünfeuroschein als Dankeschön beigelegt. Abweichend von dem sonst völlig identischen Anschreiben wurde das Anschreiben allerdings um den folgenden – optisch hervorgehobenen – Satz ergänzt: „Sollten Sie an unserer Befragung teilnehmen, werden wir Ihnen zusätzlich zu dem beigelegten Geschenk eine Aufwandsentschädigung in Höhe von 5 € zukommen lassen.“
- V. Mit der Strategie der *doppelten Incentivierung mit 5 + 10 EUR* (DI 5+10) wurde die Wirkung der Höhe des versprochenen Anreizes getestet. Analog zur vierten Gruppe wurden hier zehn statt fünf Euro versprochen. Ausgehend von einer Interviewdauer von unter einer halben Stunde können 10 EUR gerade bei der jungen Zielgruppe als eine angemessene Entlohnung betrachtet werden.

Sieben Tage nach dem Start der Befragung erfolgte eine erste Nachfassaktion mittels Postkarte. Die zweite Nachfassaktion fand drei Wochen nach dem Start statt

13 Diese erste Aussendung bestand aus einem Anschreiben, in dem auf Datenschutzaspekte, die Herkunft der Adressdaten sowie die Freiwilligkeit der Teilnahme hingewiesen wurde, einem selbstauszufüllenden Fragebogen inkl. eines freigemachten Rückumschlags, einer Postkarte zur Übermittlung einer Telefonnummer und etwaiger Rückruftermine sowie dem Link zur Onlineumfrage inkl. Kennwort. Um die Teilnahme möglichst interessant zu gestalten, wurde das Anschreiben mit einer persönlichen Anrede, einem farbigen Briefkopf der Georg-August-Universität Göttingen sowie einer sich ebenfalls farblich abhebenden Unterschrift versehen.

und bestand aus einer Sendung, die identisch mit der ersten war. Während das bereits ausgezahlte Incentive in den Nachfassaktionen gegenüber keiner der Gruppen Erwähnung fand, wurden die beiden doppelt incentivierten Gruppen in beiden Nachfassaktionen an die versprochene „Belohnung“ für die Teilnahme mit einer Hervorhebung im Text erinnert und so eine entsprechende Salienz des Incentives hergestellt.

Anreicherung der Ausgangsstichprobe mit Geodaten

Die Ausgangsstichprobe wurde mit Geodaten der Firma microm¹⁴ angereichert, um Aussagen darüber machen zu können, ob bestimmte Subgruppen von Zielpersonen durch die unterschiedlichen Incentivierungsbedingungen eher zur Teilnahme motiviert werden können als andere. Auf der Grundlage dieser Daten stehen – wenn auch auf Aggregatebene – jenseits der im Melderegister verfügbaren Angaben Informationen für die gesamte Ausgangsstichprobe und nicht nur für die realisierte Stichprobe zur Verfügung. Damit können mögliche Stichprobenverzerrungen in den unterschiedlichen Incentivegruppen identifiziert werden. Entsprechend unseren theoretischen Argumenten gilt unser Interesse vor allem den ökonomisch schlechter gestellten bzw. weniger gebildeten Zielpersonen sowie denjenigen, die starke Bindungen an ihre ethnische Gruppe aufweisen. Als indirekte Indikatoren für diese beiden Individualmerkmale werden Brutto- und Nettostichprobe im Hinblick auf die lokale *Kaufkraft* und den *Ausländeranteil* miteinander verglichen. Diese zugespielten microm-Indikatoren bilden das nächste Wohnumfeld der Befragten auf der Ebene der Straßenabschnitte ab. Die Straßenabschnittsebene ist eine relativ kleine Untersuchungseinheit, in unserem Datensatz liegen hier pro Individuum aggregierte Informationen von durchschnittlich 130 Privathaushalten vor, die in einem Haus oder einem Straßenabschnitt wohnen.

Der *Kaufkraftindex* bildet die durchschnittliche Kaufkraft des Straßenabschnitts ab, in dem der oder die Befragte wohnt. Der microm Bundesdurchschnitt ist dabei auf den Wert 100 festgesetzt, darunter liegende Werte stehen somit für eine unterdurchschnittliche Kaufkraft des Straßenabschnittes. Der *Ausländeranteil* liegt als ordinale Variable vor (von 0 „niedrigster Anteil“ bis 8 „höchster Anteil“). Auf Straßenabschnittsebene lässt die Variable eine Einschätzung über die Zusammensetzung der Nachbarschaft des Haushaltes zu, dem der oder die jeweilige Befragte angehört.

14 Weitere Informationen über die hauptsächlich im Bereich des Konsumentenmarketings verwendeten Geodaten unter www.microm-online.de. Mit einem ähnlichen Vorgehen wurde der spezielle MICROM-SOEP-Datensatz erstellt (vgl. Goebel et al. 2007).

Für einen Vergleich der unter verschiedenen Incentivierungsbedingungen realisierten Nettostichproben wird schließlich auf Informationen aus der eigentlichen Befragung zurückgegriffen. Zur Untersuchung etwaiger Stichprobenverzerrungen wird hier der *Abiturientenanteil*, der Grad der *Identifikation als Deutsche/* sowie die *Identifikation als Türke/Türkin* herangezogen. Zur Einschätzung der Datenqualität wird der *Item-Nonresponse* gemessen und schließlich auch der Einfluss der Incentivierung auf die *Wiederbefragungsbereitschaft* untersucht. Schließlich werden auch die *Kosten pro realisiertem Interview* ausgewiesen.¹⁵

5 Ergebnisse: Der Einfluss der Incentivierung auf die Teilnahmebereitschaft und die Stichprobenqualität

Die Ergebnisse werden in Abbildung 2 zunächst grafisch, mit Hilfe der aus der Ereignisdatenanalyse bekannten Kaplan-Meier-Methode, in Form der failure function (kumulierte Ratenfunktion) $H(t)$ dargestellt (vgl. Blossfeld/Golsch/Rohwer 2007: 72f.):

$$\hat{H}(t) = -\log(\hat{G}(t))$$

Die failure function ist die komplementäre Überlebensfunktion $G(t)$, beschrieben als:

$$\hat{G}(t) = \prod_{l: \tau_l < t} \left(1 - \frac{E_l}{R_l}\right)$$

mit:

$$\hat{G}(0) = 1$$

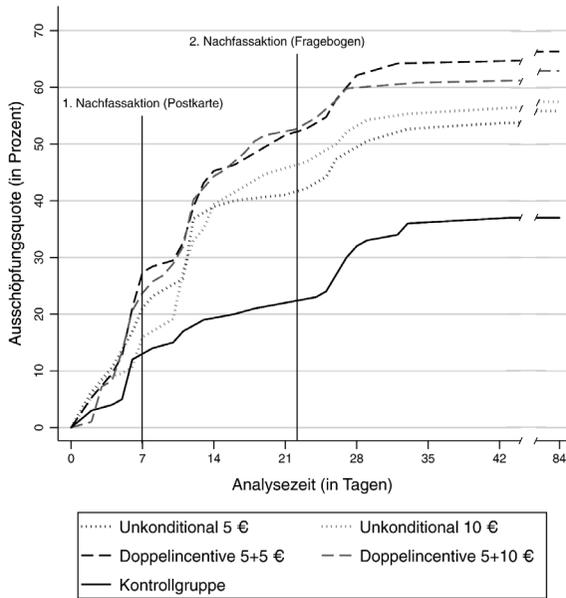
E_l = Zahl der Personen, die zum Zeitpunkt τ_l bereits geantwortet haben

R_l = Zahl der Personen, die zum Zeitpunkt τ_l noch antworten könnten

Dabei können die oben formulierten Erwartungen zum Teil bestätigt werden (vgl. Tabelle 1). Die Teilnahmebereitschaft ist, entsprechend der ersten formulierten Hypothese, bei einer UI deutlich höher als in der Kontrollgruppe, in der kein materieller Anreiz angeboten wurde (37 vs. 55 bzw. 57 Prozent). In der grafischen Darstellung ist zudem auffällig, dass sich die Teilnahmebereitschaft in der KG nach der

15 In die Kostenrechnung sind die Ausgaben für Druck und Versand der bis zu drei Aussendungen, das Incentive inkl. der Versandkosten des konditionalen Incentiveanteils sowie die Portokosten der Fragebogenrücksendung eingegangen. Diese Summe wurde anschließend durch die Anzahl der realisierten Interviews dividiert.

Abbildung 2 Failure-function über die Dauer der Feldphase nach Versuchsgruppen



Quelle: Deutsche/r bleiben? (n=483)

ersten Nachfassaktion deutlich verlangsamt und hier die zweite Nachfassaktion offenbar besonders wichtig war, um zur Teilnahme zu motivieren. Hinsichtlich der zweiten Hypothese zeigt sich, dass es zwischen den beiden vorab incentivierten Gruppen kaum Unterschiede in der Teilnahmebereitschaft in Abhängigkeit von der Höhe des unkonditionalen Anreizes gibt (55 vs. 57 Prozent). Ebenfalls erwartungsgemäß (H3) steigt die Teilnahmebereitschaft bei DI nochmals an (auf 66 Prozent bei DI 5+5). Allerdings führt *entgegen* der oben in H4 formulierten Erwartung eine Erhöhung des konditional ausgezahlten Incentiveanteils von 5 auf 10 EUR nicht zu einer weiteren Steigerung der Teilnahmebereitschaft (66 vs. 63 Prozent). Eine moderate Kombination von unkonditionaler und konditionaler Incentivierung scheint also, zumindest bei der hier untersuchten Gruppe, die besten Resultate zu liefern. Wenngleich die Unterschiede zwischen den vier Incentivierungsgruppen teilweise recht deutlich sind, sind sie – wohl vor allem bedingt durch die geringe Fallzahl – nicht signifikant voneinander verschieden.

Da die entscheidende Erfolgsgröße allerdings in der Regel die Stichprobenqualität und nicht die Ausschöpfungsquote ist, wird in einem zweiten Schritt genauer untersucht, ob bestimmte Subgruppen besonders stark auf eine bestimmte

Tabelle 1 Ausschöpfungsquoten und Kosten pro Interview unter verschiedenen Incentivierungsstrategien (n in Klammern)

	KG	UI 5	UI 10	DI 5+5	DI 5+10	Gesamt
Bruttostichprobe (in %)	100 (100)	100 (100)	100 (100)	100 (100)	100 (100)	100 (500)
Stichprobenneutrale Ausfälle (in %)	0,0 (0)	4,0 (4)	5,0 (5)	5,0 (5)	3,0 (3)	3,4 (17)
Nettostichprobe (in %)	100 (100)	100 (96)	100 (95)	100 (95)	100 (97)	100 (483)
Ausschöpfungsquote (in %) ^{1,2}	37,0 (37)	55,2** (53)	56,8** (54)	66,3** (63)	62,9** (61)	55,5 (268)
Preis pro Interview (in Euro)	14,43	19,07	27,07	21,53	27,04	22,43

1 Ausschöpfungsquote = Interviews / (Versandte Fragebögen – Ausfälle); entspricht RR5 der AAPOR (2011) Standard-Definition

2 Chi-Quadrat-Test (jeweils vs. KG)

** $p < .05$

KG: Kontrollgruppe; UI 5: Unkonditional 5 EUR; UI 10: Unkonditional 10 EUR; DI 5+5: Doppelte Incentivierung 5+5 EUR; DI 5+10: Doppelte Incentivierung 5+10 EUR

Quelle: Deutsche/r bleiben? (n=500)

Incentivierung reagieren. Dazu werden unter Einbezug der zugespielten Makrodaten zunächst Brutto- und Nettostichprobe miteinander verglichen (vgl. Tabelle 2). Die Informationen über beide Stichproben sind naturgemäß begrenzt, daher werden schließlich auch die unterschiedlichen Nettostichproben miteinander verglichen, um Zusammenhänge zwischen der Befragungsteilnahme und individuellen Merkmalen zu untersuchen. Je ähnlicher die jeweiligen Nettostichproben der Kontrollgruppe sind, desto eher kann davon ausgegangen werden, dass durch den Einsatz von Incentives kein bzw. nur ein geringer Stichprobenbias verursacht wurde.

In Tabelle 2 zeigt sich zunächst – entsprechend der in H5 formulierten Erwartung – dass Zielpersonen, die ohne Incentives (KG) um eine Teilnahme gebeten wurden und dieser Bitte nachgekommen sind, tendenziell (wenngleich nicht statistisch signifikant) häufiger in bessergestellten Wohngebieten leben (operationalisiert über den Kaufkraftindex) als diejenigen, deren Teilnahme incentiviert wurde. Bezüglich des Ausländeranteils in der Wohngegend zeigt sich keine systematische Verzerrung der Netto- gegenüber der Bruttostichprobe, H6 kann also nicht bestätigt werden. Der Blick auf den Vergleich des Frauenanteils in Brutto- und Nettostichprobe zeigt zudem, dass der Anteil der teilnehmenden Frauen in der KG erhöht ist: Ohne Incentivierung haben offenbar überdurchschnittlich viele Frauen an der Befragung teilgenommen, die diesbezüglich eher intrinsisch motiviert zu sein scheinen.

Tabelle 2 Vergleich von Brutto- und Nettostichprobe unter verschiedenen Incentivierungsstrategien

	KG	UI 5	UI 10	DI 5+5	DI 5+10	Gesamt
	Mittelwertdifferenzen (= Teilnehmer – Nicht-Teilnehmer) ¹					
Kaufkraftindex (BRD=100) ²	5,2**	0,3**	1,6**	-1,9**	-2,5**	0,3**
Ausländeranteil (Min: 0; Max: 8) ²	-0,1**	0,3**	0,1**	0,2**	-0,2**	0,1**
Frauenanteil ³	23,4**	5,3**	6,8**	0,7**	-2,0**	7,9**

1 Beispiel: Mittelwertdifferenz des Frauenanteils in der KG: 23,4 Prozentpunkte (= Anteil der Teilnehmerinnen: 56,7 % - Anteil der Nicht-Teilnehmerinnen: 33,3 %)

2 zweiseitiger t-Test

3 Chi-Quadrat-Test

** $p < .05$; * $p < .10$

KG: Kontrollgruppe; UI 5: Unkonditional 5 EUR; UI 10: Unkonditional 10 EUR; DI 5+5: Doppelte Incentivierung 5+5 EUR; DI 5+10: Doppelte Incentivierung 5+10 EUR

Quelle: Deutsche/r bleiben? (n=483)

Dies entspricht einem Ergebnis bei Lesser et al. (2001), auch bei ihnen ist das Geschlechterverhältnis nur in den incentivierten Gruppen ausgeglichen. Insgesamt bestätigen unsere Befunde die oben dargestellten Studien insofern als deutlich wird, dass durch eine Incentivierung der Stichprobenbias eher reduziert als vergrößert wird. Dieses Resultat zeigt, wie wichtig ein Vergleich von Brutto- und Nettostichprobe ist: Es kann keinesfalls – wie bei einem Vergleich der realisierten Stichproben untereinander – davon ausgegangen werden, dass die KG die „richtige“ Referenzgröße darstellt.

Ein Nachteil des Brutto-Netto-Vergleichs ist allerdings, dass bei diesem Vorgehen die konstatierten Zusammenhänge nur indirekt untersucht werden können, liegen doch mit den Makroindikatoren Kaufkraftindex und Ausländeranteil in der Nachbarschaft sehr crude Proxies für die theoretisch interessierenden Individualmerkmale vor. Abschließend soll deshalb untersucht werden, ob sich die unter unterschiedlichen Incentivierungsbedingungen realisierten Stichproben im Hinblick auf die theoretisch relevanten Individualmerkmale der Befragten unterscheiden (vgl. Tabelle 3). Auch die Datenqualität, operationalisiert über die Informationen zu Item-Nonresponse, sowie die Bereitschaft für eine erneute Befragung soll in den Blick genommen werden. Häufig wird angenommen, dass Teilnehmerinnen und Teilnehmer einer Befragung „ihre Arbeit“ besser erledigen und sorgfältiger antworten, wenn sie dafür einen Anreiz erhalten haben (vgl. Stadtmüller/Porst 2005).

Der Brutto-Netto-Vergleich hat bereits gezeigt, dass – entsprechend unserer in H5 formulierten Erwartung – Zielpersonen, die nicht incentiviert wurden, in Nachbarschaften mit einem tendenziell höheren Kaufkraftindex leben. Wir waren

Tabelle 3 Ausgewählte Items in der realisierten Stichprobe nach
Versuchsgruppen

	KG	UI 5	UI 10	DI 5+5	DI 5+10	Gesamt
Abiturientenanteil (in %) ¹	43,2	37,7	33,3	22,2**	26,2*	31,3
Identität als Deutsche/r (Min: 0; Max: 4) ²	2,2	1,6**	2,0	1,9	1,9	1,9
Identität als Türke/in (Min: 0; Max: 4) ²	2,7	3,0	3,2*	3,0	2,8	2,9
Erneute Befragung erwünscht (in %) ¹	69,4	76,9	64,2	65,6	83,6	72,2
Item-Nonresponse (Min: 0; Max: 35) ³	1,7	1,1	1,3	1,4	1,2	1,3

1 Chi-Quadrat-Test

2 Tau-b

3 zweiseitiger t-Test

** $p < .05$; * $p < .10$ (jeweils vs. KG)

KG: Kontrollgruppe; UI 5: Unkonditional 5 EUR; UI 10: Unkonditional 10 EUR; DI 5+5: Doppelte Incentivierung 5+5 EUR; DI 5+10: Doppelte Incentivierung 5+10 EUR

Quelle: Deutsche/r bleiben? (n=268)

davon ausgegangen, dass ökonomisch bessergestellte Zielpersonen häufig auch eine höhere Bildung und deshalb eine höhere intrinsische Motivation zur Teilnahme an den Tag legen. Sehr eindrücklich bestätigt dies nun ein Blick auf die Abiturientenquote auch auf der individuellen Ebene. In allen incentivierten Gruppen befinden sich deutlich weniger Personen mit (Fach-)Hochschulreife als in der Kontrollgruppe, deren Anteil ist insbesondere in der mit zwei mal fünf Euro incentivierten Gruppe (DI 5+5) signifikant von der KG verschieden und nur fast halb so hoch. Die „Überrepräsentation“ ökonomisch Bessergestellter in der KG zeigt sich auch bei einem hier nicht ausgewiesenen Vergleich der Kaufkraftindizes in den realisierten Stichproben: So befinden sich in der Gruppe DI 5+5 signifikant mehr Befragte aus ökonomisch benachteiligten Wohngebieten als in der KG.

Darüber hinaus zeigt Tabelle 3 zumindest tendenziell, dass die Befragten, die nicht incentiviert wurden, sich entsprechend unserer Erwartung (H6) weniger stark mit der eigenen ethnischen Gruppe identifizieren als die incentivierten Zielpersonen (beim Vergleich von UI 10 zur KG ist dieser Unterschied signifikant). Umgekehrt ist in der KG die Identifikation als Deutsche/r besonders hoch, was ebenfalls in Richtung des oben formulierten Arguments weist, dass der Appell an die „staatsbürgerschaftliche Pflicht“ zur Unterstützung der Forschung (vgl. Schupp 2012) durch eine Identifikation mit dem Zielland gefördert wird. Angesichts der geringen Unterschiede zwischen den Gruppen und der relativ kleinen Stichprobengröße sollte dieses Ergebnis allerdings nicht überbewertet werden, zumal wir keinen konsistenten Unterschied zwischen den unterschiedlichen Incentivierungsbedingungen feststellen können.

Hinsichtlich der Datenqualität und der Wiederbefragungsbereitschaft bestätigen unsere Befunde existierende Studien (vgl. Singer et al. 1999). Die Auswirkungen auf die Datenqualität sind schwach und eher positiv – im Sinne eines geringeren Item-Nonresponse im Falle der Incentivierung. Hinsichtlich der Wiederbefragungsbereitschaft zeigt sich kein eindeutiger Befund im Vergleich von Kontroll- und Incentivegruppen. Dies entspricht im Kern dem Befund von Singer (1998), ihr zufolge haben Incentives keinen positiven Einfluss auf die *Absicht* an zukünftigen Befragungen teilzunehmen, wohl aber auf das *Handeln*, wenn mit Abstand eines halben Jahres eine erneute Befragung durchgeführt wird (siehe auch: Shettle/Mooney 1999).

Lässt man abschließend die geschilderten qualitativen Aspekte der gewonnenen Stichprobe außer Acht, so ist unter reinen Kostengesichtspunkten die KG mit gut 14 Euro pro erfolgreichem Interview (vgl. Fußnote 15) die günstigste Variante (vgl. Tabelle 1). Bemerkenswert ist jedoch, dass im Vergleich von KG zu UI 5 oder DI 5+5 die Kostendifferenzen geringer sind als die jeweils eingesetzten Incentives. Dies trifft auch auf die Kostendifferenz von UI 5 zu DI 5+5 sowie von DI 5+5 zu DI 5+10 zu. Der schnellere Rücklauf und der damit möglich gewordene Verzicht auf weitere Nachfassaktionen hat sich kostendämpfend ausgewirkt.

6 Zusammenfassung und Ausblick

Je nach interessierender Subgruppe stellt sich in der Integrationsforschung häufig das Problem sehr kleiner Ausgangspopulationen und einem entsprechend kleinen Pool von Adressen oder Telefonnummern (z. B. bei Neuzuwanderern oder Angehörigen der dritten Generation). Darüber hinaus werden für differenzierte Analysen im Bereich der Integrationsforschung ausreichend hohe Fallzahlen für einzelne ethnische Gruppen benötigt. In allgemeinen Bevölkerungsumfragen werden diese Fallzahlen ohne ein *Oversampling* häufig nicht erreicht. Spezifische Migrantenbefragungen sind, gerade wenn sie als *Face-to-Face* Interviews durchgeführt werden, sehr kostspielig. Daher wurde im vorliegenden Beitrag argumentiert, dass eine schriftliche Befragung von Personen mit Migrationshintergrund für bestimmte Fragestellungen eine sinnvolle und günstige Alternative darstellen kann. Die frei werdenden finanziellen Mittel können dann in Maßnahmen zur Erhöhung von Ausschöpfung und Datenqualität investiert werden, beispielsweise in eine gezielte Incentivierungsstrategie.

In diesem Beitrag wurden zunächst theoretische Argumente zusammengetragen, die für den Einsatz einer Kombination von *unkonditionaler* und *konditio-*

nalere Incentivierung sprechen. Eine solche Strategie *doppelter Incentivierung* zielt darauf ab, die Kosten der Nicht-Teilnahme zu erhöhen, indem Reziprozität hergestellt bzw. verstärkt wird. Zusätzlich wird der erwartete Nutzen der Teilnahme gesteigert, indem Vertrauen aufgebaut und den Befragten so eine hinreichende Sicherheit darüber gegeben wird, das versprochene Incentive nach der Teilnahme tatsächlich zu erhalten.

Bei der empirischen Überprüfung dieser Annahmen haben sich die Vorteile einer doppelten Incentivierung in der hier dargestellten Befragung deutsch-türkischer junger Erwachsener zumindest ansatzweise bestätigt. In der doppelt incentivierten Gruppe wurde eine Ausschöpfungsquote von bis zu 66 Prozent erreicht – gegenüber 37 Prozent in der nicht-incentivierten Kontrollgruppe und 55 bzw. 57 Prozent in den rein un konditional incentivierten Gruppen.

Wenngleich die Ergebnisse insgesamt vor dem Hintergrund der geringen Stichprobengröße mit Vorsicht zu interpretieren sind, scheinen Befürchtungen, materielle Anreize in standardisierten Befragungen könnten die gewonnene Stichprobe negativ verzerren oder die Datenqualität mindern, auf Basis der vorgestellten Befunde unbegründet. Wenn Incentives überhaupt einen Effekt auf die Stichprobe und die Qualität der Daten haben, dann eher einen positiven im Sinne einer Verringerung der Nichtteilnahme von „Problemgruppen“ und einer verbesserten Datenqualität: Mit den Geodaten lagen Makroindikatoren vor, die einen Vergleich von Brutto- und Nettostichprobe ermöglichen. Bei diesem Vergleich ist deutlich geworden, dass die Teilnehmerinnen und Teilnehmer der nicht-incentivierten Gruppe tendenziell aus wohlhabenderen Wohngebieten stammen, die, so das theoretische Argument, eine höhere Bildung und eine stärkere intrinsische Motivation zur Beteiligung an wissenschaftlichen Umfragen aufweisen. Der Brutto-Netto-Vergleich hat in Übereinstimmung mit existierenden Befunden gezeigt, dass Frauen signifikant häufiger auch ohne materielle Anreize an der Befragung teilnahmen. Bei dem weitergehenden Vergleich spezifischer Items in den unter unterschiedlichen Incentivierungsbedingungen realisierten Stichproben haben sich diese Befunde insofern bestätigt, als in den doppelt incentivierten Subgruppen weniger hoch Gebildete stärker vertreten sind als in der nicht incentivierten Kontrollgruppe. Im Hinblick auf die (identifikative) Integration zeigt sich dieser Zusammenhang nur tendenziell, zudem scheint diesbezüglich ausschlaggebend zu sein, dass überhaupt incentiviert wird, d. h. es lässt sich kein Unterschied zwischen un konditionaler und doppelter Incentivierung feststellen.

Mit Verweis auf die Leverage-Saliency Theorie wurde verdeutlicht, dass Incentives sowie alle weiteren Merkmale des Erhebungsdesigns auf verschiedene Gruppen unterschiedlich auswirken können, insofern können die hier vorgetrage-

nen Befunde nicht generalisiert werden. Es spricht unseres Erachtens aber wenig dafür, dass Bevölkerungsgruppen ohne Migrationshintergrund anders auf materielle Anreize reagieren. Wir betrachten hier allerdings junge Erwachsene und es erscheint daher durchaus möglich, dass die positive Wirkung des konditionalen Anteils der DI auch damit zu tun hat, dass es sich um eine Gruppe mit begrenzten finanziellen Mitteln handelt. Gegen dieses „Bedürftigkeitsargument“ spricht freilich der Befund, dass die Ausschöpfungsquote mit steigendem konditionalen Incentiveanteil nicht weiter zunimmt. Insgesamt betrachtet lassen unsere Befunde den Schluss zu, dass die doppelte Incentivierung als eine mögliche Alternative zur rein unkonditionalen Incentivierung künftig eine größere Aufmerksamkeit verdient.

Literatur

- AAPOR, 2011: Standard Definitions: The American Association for Public Opinion Research. Lenexa.
- Arzheimer, K. und M. Klein, 1998: Die Wirkung materieller Incentives auf den Rücklauf einer schriftlichen Panelbefragung. *ZA-Informationen* 43: 6-31.
- Aust, F. und H. Schröder, 2009: Sinkende Stichprobenausschöpfung in der Umfrageforschung – ein Bericht aus der Praxis. S. 195-212 in: M. Weichbold, J. Bacher und C. Wolf (Hg.), *Umfrageforschung. Herausforderungen und Grenzen*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Babka von Gostomski, C. und M. Pupeter, 2008: Zufallsbefragung von Ausländern auf Basis des Ausländerzentralregisters. *Methoden - Daten - Analysen* 2: 149-177.
- Barón, J., R. Breunig, D. Cobb-Clark, T. Gørgens und A. Sartbayeva, 2009: Does the Effect of Incentive Payments on Survey Response Rates Differ by Income Support History? *Journal of Official Statistics* 25: 483-507.
- Baykara-Krumme, H., 2010: Die Berliner pairfam Studie. Ein Methodenbericht. Arbeitspapier 16 des DFG-Schwerpunktes "Beziehungs- und Familienentwicklungspanel".
- 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: 131-159.
- Berger, F., U. Grob, H. Fend und W. Lauterbach, 2005: Möglichkeiten zur Optimierung der Rücklaufquote in postalischen Befragungen. Bericht über die Vorstudie zum Forschungsprojekt Life. *Zeitschrift für Soziologie der Erziehung und Sozialisation* 25: 99-107.
- Blohm, M. und C. Diehl, 2001: Wenn Migranten Migranten befragen. Zum Teilnahmeverhalten von Einwanderern bei Bevölkerungsumfragen. *Zeitschrift für Soziologie* 30: 223-242.
- Blossfeld, H., K. Golsch und G. Rohwer, 2007: *Event History Analysis with Stata*. New York: Erlbaum.
- Bosnjak, M. und T. Tuten, 2003: Prepaid and Promised Incentives in Web Surveys. In: *Social Science Computer Review* 21: 208-217.
- Castiglioni, L., K. Pffor und U. Krieger, 2008: The Effects of Incentives on Response Rates and Panel Attrition. *Survey Research Methods* 2: 151-158.

- Church, A., 1993: Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis. *Public Opinion Quarterly* 57: 62-79.
- Diehl, C. und P. Fick, 2012: Deutschsein auf Probe. *Soziale Welt* 63: 339-360.
- Diekmann, A. und B. Jann, 2001: Anreizformen und Ausschöpfungsquoten bei postalischen Befragungen. Eine Prüfung der Reziprozitätshypothese. *ZUMA-Nachrichten* 48: 18-27.
- Dillman, D., 1978: *Mail and Telephone Surveys*. New York: Wiley.
- Dillman, D. 2007: *Mail and Internet Surveys*. Hoboken: Wiley.
- Dillman, D., J. Smyth und L. Christian, 2009: *Internet, Mail, and Mixed-Mode Surveys. The Tailored Design Method*, Hoboken: John Wiley & Sons.
- Dykema, J., J. Stevenson, B. Day, S. Sellers und V. Bonham, 2011: Effects of Incentives and Prenotification on Response Rates and Costs in a National Web Survey of Physicians. *Evaluation & the health Professions* 24: 434-447.
- Edwards, P., I. Roberts, M. Clarke, C. DiGiuseppi, S. Pratap, R. Wentz und I. Kwan, 2002: Increasing Response Rates to Postal Questionnaires. *British Medical Journal* 324: 1183-1185.
- Esser, H., 1986: Über die Teilnahme an Befragungen. *ZUMA-Nachrichten* 18: 38-47.
- Feskens, R., J. Hox, G. Lensvelt-Mulders und H. Schmeets, 2006: Collecting Data among Ethnic Minorities in an International Perspective. *Field Methods* 18: 284-304.
- Feskens, R., J. Hox, H. Schmeets und W. Wetzels, 2008: Incentives and Ethnic Minorities. *Survey Research Methods* 2: 159-165.
- Goebel, J., K. Spieß, N. Witte und S. Gerstenberg, 2007: Die Verknüpfung des SOEP mit MICROM-Indikatoren. *Data Documentation* 26. Berlin: DIW.
- Goto, S., 1996: To Trust or Not to Trust: Situational and Dispositional Determinants. *Social Behavior and Personality* 24: 119-131.
- Gouldner, A., 1960: The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review* 25: 161-178.
- Goyder, J., 1994: An Experiment with Cash Incentives on a Personal Interview Survey. *Journal of the Market Research Society* 36: 360-366.
- Granato, N., 1999: Die Befragung von Arbeitsmigranten: Einwohnermeldeamt-Stichprobe und telefonische Erhebung? *ZUMA-Nachrichten* 45: 44-60.
- Grant, R. und J. Sugarman, 2004: Ethics in Human Subjects Research: Do Incentives Matter? *Journal of Medicine and Philosophy* 29: 717-138.
- Groves, R., E. Singer und A. Cornig, 2000: Leverage-Saliency Theory of Survey Participation. *Public Opinion Quarterly* 64: 299-308.
- Haberfeld, Y., Y. Cohen, F. Kalter und I. Kogan, 2011: Differences in Earnings Assimilation of Immigrants from the Former Soviet Union to Germany and Israel During 1994-2005. *International Journal of Comparative Sociology* 52: 6-24.
- Harkness, J., P. Mohler, M. Schneid und B. Christoph, 1998: Incentives in Two German Mail Surveys 1996/97 & 1997. S. 201-218 in: A. Koch und R. Porst (Hg.), *Nonresponse in Survey Research*. Mannheim: ZUMA.
- Haug, S. und C. Diehl, 2005: Aspekte der Integration. Wanderungsverhalten, Eingliederungsmuster und Lebenssituation türkisch- und italienischstämmiger Jugendlicher in Deutschland. *Schriftenreihe des Bundesinstituts für Bevölkerungsforschung*, Band 35.
- James, J. und R. Bolstein, 1992: Large Monetary Incentives and Their Effect on Mail Survey Response Rates. *Public Opinion Quarterly* 56: 442-453.
- James, J. und R. Bolstein, 1990: The Effect of Monetary Incentives and Follow-Up Mailings on the Response Rate and Response Quality in Mail Surveys. *Public Opinion Quarterly* 54: 346-361.
- Jobber, D., J. Saunders und V. Mitchell, 2004: Prepaid monetary incentive effects on mail survey response. *Journal of Business Research* 57: 21-25.

- Johnson, T., D. O'Rourke, J. Burris und L. Owens, 2002: Culture and Survey Nonresponse. S. 55-69 in: R. Groves, D. Dillman, J. Eltinge und R. Little (Hg.): Survey Nonresponse. New York: John Wiley & Sons.
- Lesser, V., D. Dillman, J. Carlson, F. Lorenz, R. Mason, F. Willits, 2001: Quantifying the Influence of Incentives on Mail Survey Response Rates and Their Effects on Nonresponse Error. Proceedings of the Annual Meeting of the American Statistical Association: 1-6.
- Mauss, M., 1968: Die Gabe. Frankfurt am Main: Suhrkamp.
- Mehlkop, G. und R. Becker, 2007: Zur Wirkung monetärer Anreize auf die Rücklaufquote in postalischen Befragungen zu kriminellen Handlungen. Methoden - Daten - Analysen 1: 5-24.
- Millar, M. und D. Dillman, 2011: Improving Response to Web and Mixed-Mode Surveys. Public Opinion Quarterly 75: 249-269.
- Paolillo, J. und P. Lorenzi, 1984: Monetary Incentives and Mail Questionnaire Response Rates. Journal of Advertising 13: 46-48.
- Porst, R., 1999: Thematik oder Incentives? Zur Erhöhung der Rücklaufquoten bei postalischen Befragungen. ZUMA-Nachrichten 45: 72-87.
- Porter, S. und M. Whitcomb, 2003: The Impact of Lottery Incentives on Student Survey Response Rates. Research in Higher Education 44: 389-407.
- Ryu, E., M. Couper und R. Marans, 2005: Survey Incentives: Cash vs. In-Kind; Face-to-Face vs. Mail; Response Rate vs. Nonresponse Error. International Journal of Public Opinion Research 18: 89-106.
- Schnell, R., 2012: Survey-Interviews. Methoden standardisierter Befragungen. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schupp, J., 2012: Die verborgenen Kosten monetärer Anreize – lohnt sich Motivierung durch Incentivierung? DIW Wochenbericht 79: 20.
- Schupp, J., N. Siegel, S. Huber und G. Wagner, 2010: Incentives and Response Rates. First Experiences from the SOEP-Innovation-Sample 2009. "2nd Panel-Survey-Methodology-Workshop", Mannheim (Vortragsfolien).
- Shettle, C. und G. Mooney, 1999: Monetary Incentives in U.S. Government Surveys. Journal of Official Statistics 15: 231-250.
- Singer, E., 1998: Incentives for Survey Participation: Research on Intended and Unintended Consequences. ZUMA-Nachrichten 42: 7-29.
- Singer, E., 2002: The Use of Incentives to Reduce Nonresponse in Household Surveys. S. 163-177 in: R. Groves, D. Dillman, J. Eltinge, R. Little (Hg.): Survey Nonresponse. New York: Wiley.
- Singer, E., J. Van Hoewyk, N. Gebler, T. Raghunathan und K. McGonagle, 1999: The Effects of Incentives on Response Rates in Interviewer-Mediated Surveys. Journal of Official Statistics 15: 217-230.
- Singer, E., J. Van Hoewyk und M. Maher, 2000: Experiments with Incentives in Telephone Surveys. Public Opinion Quarterly 64: 171-188.
- Stadtmüller, S., 2009: Rücklauf gut, alles gut? Zu erwünschten und unerwünschten Effekten monetärer Anreize bei postalischen Befragungen. Methoden - Daten - Analysen 3: 167-185.
- Stadtmüller, S. und R. Porst, 2005: Zum Einsatz von Incentives bei postalischen Befragungen. ZUMA How-to-Reihe 14.
- Statistisches Bundesamt, 2011: Fachserie 1 Reihe 2.2, Wiesbaden: StBA.
- Trussell, N. und P. Lavrakas, 2004: The Influence of Incremental Increases in Token Cash Incentives on Mail Survey Response. Public Opinion Quarterly 68: 349-367.

- Warriner, K., J. Goyder, H. Gjertsen, P. Hohner und K. McSpurren, 1996: Charities, No; Lotteries, No; Cash, Yes. Main Effects and Interactions in a Canadian Incentive Experiment. *Public Opinion Quarterly* 60: 542-562.
- Weinmann, M. I. Becher und C. Babka von Gostomski, 2012: Einbürgerungsverhalten von Ausländerinnen und Ausländern in Deutschland sowie Erkenntnisse zu Optionspflichtigen. Nürnberg: BAMF.
- Wetzels, W., H. Schmeets, J. van den Brakel und R. Feskens, 2008: Impact of Prepaid Incentives in Face-to-Face Surveys: A Large-Scale Experiment with Postage Stamps. *International Journal of Public Opinion Research* 20: 507-516.
- Wotruba, T., 1966: Monetary Inducements and Mail Questionnaire Response. *Journal of Marketing Research* 3: 398-400.
- 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.
- Zentrum für Türkeistudien, 2012: Integrationsprozesse türkeistämmiger Migrantinnen und Migranten in Nordrhein-Westfalen. Essen: ZfT.

Anschrift der Autoren

Patrick Fick
Universität Konstanz
Fachbereich Geschichte und Soziologie
Universitätsstr. 10
78457 Konstanz
E-Mail: patrick.fick@uni-konstanz.de

Claudia Diehl
Universität Konstanz
Fachbereich Geschichte und Soziologie
Universitätsstr. 10
78457 Konstanz
E-Mail: claudia.diehl@uni-konstanz.de

Respondent Incentives in a National Face-to-Face Survey

Effects on Outcome Rates, Sample Composition and Fieldwork Efforts

Der Einsatz von Befragten-Incentives in einer bundesweiten face-to-face-Umfrage

Effekte auf Ausschöpfung, Stichprobensammensetzung und Feldarbeitsaufwand

Michael Blohm und Achim Koch

Abstract

Nonresponse is an ongoing challenge for survey research. In the German General Social Survey (ALLBUS) 2010, an experiment was set up to test the effect of respondent incentives on outcome rates, sample composition and fieldwork efforts. A random subsample of target persons was offered a monetary incentive of €10 to be paid upon completion of the interview. The other part of the sample acted as a control group receiving no incentive. The incentive used in ALLBUS 2010 led to an increase in the response rate, mainly by improving the cooperation rate. It did not change the sample composition in a major way. Concerning fieldwork efforts, a slight reduction was observed: In the incentive condition, a given number of interviews was achieved with a lower number of contact attempts than in the no incentive condition.

Zusammenfassung

Nonresponse ist eine Herausforderung für die gesamte Umfrageforschung. In der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) wurde im Jahr 2010 ein Experiment durchgeführt, um die Effekte von Befragten-Incentives auf die Ausschöpfungsquote, die Stichprobensammensetzung und den in der Feldarbeit notwendigen Aufwand zu untersuchen. Einer Zufallsstichprobe der Zielpersonen wurde für die Teilnahme an der Umfrage ein monetäres Incentive in Höhe von 10 € angeboten. Der verbleibende Teil der Zielpersonen fungierte als Kontrollgruppe und erhielt kein Incentive. Das verwendete Incentive führte zu einer höheren Kooperationsrate und einer höheren Ausschöpfung in der Experimentalgruppe. Die Stichprobensammensetzung unterschied sich nicht wesentlich zwischen Experimental- und



Kontrollgruppe. In der Experimentalgruppe konnten die Interviewer eine gegebene Zahl von Interviews mit weniger Kontaktversuchen erzielen als in der Kontrollgruppe.

Keywords: Incentive, nonresponse, bias, sample composition, fieldwork effort

1 Introduction¹

Declining response rates are a continuing problem for household surveys in many Western countries (Atrostic, Bates, Burt, & Silberstein, 2001; de Leeuw & de Heer, 2002; Dixon & Tucker, 2010; Kohut, Keeter, Doherty, Dimock, & Christian, 2012). As a consequence, low response rates are no longer exceptional in household surveys in Europe. For instance, in the fifth round of the European Social Survey (ESS), which was fielded in 2010, 4 out of the 27 participating countries had a response rate below 50%.² In the fourth wave of the European Values Study (EVS), fielded in 2008, 11 out of 47 countries/regions had a response rate of 50% or less.³ In the third European Quality of Life Survey (EQLS), fielded in 2011, 17 out of 27 countries had a maximum response rate of 50%, among which 8 countries with a response rate even below 40% (Eurofound, 2012).

Various attempts have been made to increase response rates or at least to halt downward trends in response rates. These include, for instance, an increased number of contact attempts, the use of advance letters or the provision of incentives to sample persons to encourage survey participation (Groves et al., 2004). The use of respondent incentives in order to increase response has a long tradition in mail surveys. More recently, however, the use of incentives has also become more common in face-to-face surveys (Kulka, Eyerman, & McNeeley, 2005; Singer, 2002). The ESS is a case in point. In this repeated cross-national survey, the number of countries using some kind of a respondent incentive has been continuously increasing over the first five survey rounds. Whereas in the first round, which was fielded in the years 2002/2003, only 7 out of the 16 permanent countries used a respondent incentive, this number rose to 8, 10, 11 and 13 countries, respectively, in rounds 2 to 5.⁴

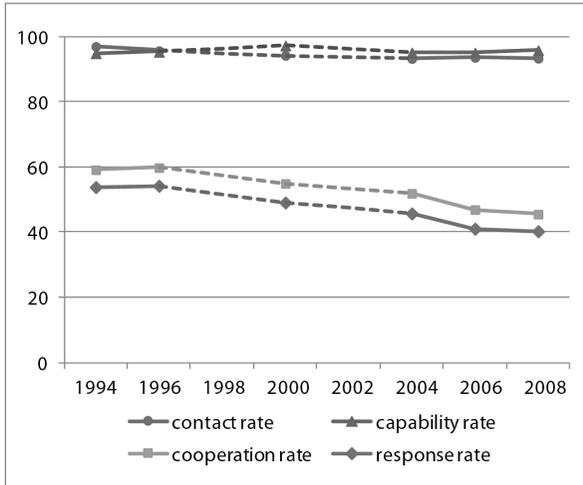
1 The authors would like to thank Dorothee Behr, Henning Best, Christof Wolf and the reviewers for providing constructive comments on a former draft of the paper.

2 Information from ESS website: <http://ess.nsd.uib.no/ess/round5/deviations.html> (03.04.2013).

3 Personal communication from Evelyn Brislinger, GESIS (28.03.2013).

4 Whether or not the use of an incentive actually increased the response rate in the relevant ESS countries (and, if yes, to what extent), unfortunately remains an unanswerable question, since apart from one exception (Phelps, 2008) the effects of introducing

Figure 1 Contact, capability, cooperation and response rates in ALLBUS 1994 – 2008 (in %)



The German General Social Survey (ALLBUS) has also faced an increase in nonresponse in the past years. The ALLBUS is a biennial face-to-face survey of the adult population, covering a wide range of topics and aiming at charting the long-term trends in attitudes and behaviour in Germany (<http://www.gesis.org/en/allbus>). Between 1994 and 2008 the response rate of ALLBUS decreased from 54% to 40%. The main reason for this decline was a drop in the cooperation rate (see Figure 1).⁵ In contrast, the contact rate remained broadly stable on a high level. Own analyses of data from the ALLBUS contact forms show that the latter result can be attributed to an increase in the number of contact attempts over time.

In this context, ALLBUS 2010 included an experiment that was set up to investigate whether or not the provision of a respondent incentive might help to counteract the downward trend in the response rate. Using a respondent incentive seemed to be a promising measure insofar as an incentive will primarily help

Footnote continued from previous page

incentives were not measured in the ESS (or at least, the results of relevant experiments were not (yet) published).

5 For a definition of the outcome rates, see Appendix 1. In order to ensure comparability over time, ALLBUS 1998 and 2002 were excluded, since in both years the survey was fielded by a different survey organisation.

to reduce refusals (Singer, 2002).⁶ The experiment in ALLBUS 2010 was designed to help to clarify three issues. First, the experiment should provide an answer to the question whether respondent incentives increase cooperation and response rates in a face-to-face survey like ALLBUS. Second, the experiment should provide evidence on whether respondent incentives affect sample composition. Third, the experiment should give some indication of whether the use of a respondent incentive affects fieldwork efforts and survey costs.⁷

Thus far, the literature on incentive effects is less comprehensive with regard to face-to-face surveys than with regard to mail surveys (see section 2). There is evidence that incentives help to increase response rates also in face-to-face surveys. However, there is less (and often only mixed) evidence as regards the effects of incentives on sample composition and survey costs. This is unfortunate since in the past decade a shift in the focus of nonresponse research has occurred (Singer, 2006). It is increasingly acknowledged that a concentration on response rates is inadequate; the focus has to be directed towards nonresponse bias. Regarding the use of incentives the crucial question is whether incentives reinforce or counteract tendencies to underrepresent certain subgroups of the population. Respondent incentives can help to decrease nonresponse bias if they disproportionately attract target persons whose response propensity is below average otherwise.⁸ Furthermore, the cost implications of various survey design features (like the number of contact attempts or refusal conversion activities) have gained more attention in recent years. Regarding respondent incentives the question is whether the obvious increase in costs associated with their use will (at least partly) be absorbed by a reduction in fieldwork efforts necessary to complete the interviews.

In the ALLBUS 2010 experiment, we wanted to test an incentive appropriate for the use in future survey rounds – given the specific design of the ALLBUS and the available budget. It was decided to investigate the effect of a €10 conditional cash incentive. A random subsample of the ALLBUS 2010 respondents was to receive a conditional monetary incentive of €10, and the rest of respondents was to act as a control group receiving no incentive. In the survey literature (see

6 Up to that point, ALLBUS had used a respondent incentive only once. In 2002, all ALLBUS respondents received a commemorative coin worth €10. At that time, no efforts were made to measure the effect of the incentive experimentally.

7 Note that in the present paper we are not concerned with the potential effects incentives may have on measurement, i.e., whether or not they change the response behaviour of respondents.

8 As far as the ALLBUS is concerned, for instance, an issue is whether an incentive might help to reduce the underrepresentation of lower educated people that had been observed in previous rounds of the survey (Wasmer, Scholz, & Blohm, 2010).

section 2), unconditional ('prepaid') incentives are often described as being preferable to conditional ('promised') incentives. In survey practice, however, conditional incentives play an important role in face-to-face surveys. In the fifth round of the ESS, for instance, six countries offered a conditional cash incentive or a shopping voucher to respondents, whereas only two countries provided an unconditional cash incentive to all target persons. Likewise, in the EVS 2008, eight countries used a conditional cash incentive and only one country an unconditional cash incentive.

Several reasons led us to the decision to use a conditional cash incentive in the ALLBUS experiment. First, in face-to-face surveys (in contrast to mail and telephone surveys) an interviewer is present who can pass the money to the respondent immediately after the interview has been completed. The presence of an interviewer and the short time span between the conduct of an interview and the handover of the incentive should induce respondents to trust that a promised incentive will actually be delivered to them (Singer, 2002). Second, the costs of a conditional incentive will be lower than the costs of an unconditional incentive of the same value, particularly if the expected response rate is rather low as it is in ALLBUS. Third, for government-financed surveys like ALLBUS, the use of a conditional incentive is easier to justify, both vis-à-vis the funders and the target persons of the survey. In contrast, the use of an unconditional incentive is more difficult to justify, as some people might see such an incentive as a waste of tax money.⁹

Implementing the experiment in the regular ALLBUS survey provides a solid base for the decision as to whether or not such an incentive should be used in future survey rounds of ALLBUS. In addition, the question is whether the present results will also be relevant for other surveys. Of course, the effect of an incentive (on response rates, sample composition and fieldwork efforts) can vary, depending on the concrete conditions in a survey, like the topic and target population of the survey, the survey mode, the burden imposed on respondents, the implemented call schedule and the refusal conversions efforts being made, etc.. According to Groves (2008), it has to be noted "... that the marginal effects of incentives are a function

9 In 2010, the German part of the Survey of Health, Ageing and Retirement in Europe (SHARE) fielded a large-scale experiment using unconditional monetary incentives worth €10, €20 and €40 (Börsch-Supan, Krieger, & Schröder, 2013). In each incentive condition, 6-7% of the target persons called the telephone hotline (compared to 1.5% of the target persons in the condition with no incentive). Most of them were unhappy with the inclusion of cash in the advance letter and questioned the legitimacy of the survey. The researchers responsible for the experiment also point out: "One complaint to a member of the Bundestag reached high levels at the Federal Ministry of Education and Research, the funders of SHARE in Germany." (Börsch-Supan et al., 2013: 11)

of the base protocol to which they are being compared." This limits the generalisability of the results of any single incentive experiment.¹⁰ As a consequence, we have to acknowledge that the results of the present experiment will be primarily relevant for face-to-face surveys similar to ALLBUS. ALLBUS represents a certain type of survey, namely cross-sectional high-quality face-to-face surveys of the general population covering a broad range of topics. This type of survey continues to play an important role in social sciences all over the world, both as national general social surveys (Smith, Kim, Koch, & Park, 2006) or as cross-national surveys like the ESS or the EVS.

In the following, we first summarise the previous research on incentives in face-to-face surveys and formulate our working hypotheses. We then describe the design and implementation of the ALLBUS 2010 incentive experiment. In the results section, the effect of the incentive on contact, capability, cooperation and response rates is analysed. In addition, we investigate whether the incentive differentially attracted certain demographic subgroups and whether the response distributions and survey results for selected variables differ between respondents in the incentive and no incentive condition. Finally, we touch upon the cost issues related to incentives and examine whether the incentive used reduced the amount of interviewer effort required to complete a case. We discuss our results against the backdrop of possible practical consequences for future surveys like ALLBUS.

2 Previous research and working hypotheses

Incentives have been used in mail surveys for a long time, and their positive effect on response rates is well documented (Church, 1993). Regarding telephone and face-to-face surveys one could expect incentives to be of lower importance than in self-administered mail surveys. In interviewer-administered surveys, the interviewer plays a vital role in persuading people to participate in the survey. This limits the potential effect of an incentive. Singer, van Hoewyk, Gebler, Raghunathan, and McGonagle (1999; see also Singer, 2002) summarise what is known from the experimental literature about the effects of respondent incentives in interviewer-mediated surveys. Their meta-analysis is based on 39 experiments conducted in telephone and face-to-face surveys in the United States and Canada. In the following, we describe their major findings and add results from several other face-

10 In order to achieve a solid base for the decision to use an incentive in a particular study, Singer and Kulka (2001) recommend to always pretest the specific incentive plan with the target population and the survey methods to be employed.

to-face studies that were either published only in recent years or conducted in other countries than the United States or Canada.

Singer et al. (1999) found that paying an incentive is an effective tool for increasing the response rate in face-to-face and telephone surveys. However, the effects were smaller than in mail surveys. On average, each dollar of incentive paid increased the response rate by about one third of a percentage point. Gifts were less effective in raising cooperation than money (even when the value of the incentive was controlled for). Both prepaid and promised incentives produced a significant increase in response rates (with prepaid incentives yielding higher response rates than promised incentives). Incentives were effective both in low and high burden surveys. The effect of incentives was larger in studies in which the response rate without an incentive was low. Incentives seem to work primarily by reducing refusals; there is no evidence that incentives are helpful in making contact with respondents (Singer, 2002).¹¹

The results of several other face-to-face studies are more or less in line with Singer's main conclusions. Both promised incentives (Eyerman, Bowman, Butler, & Wright, 2005; Lynn, Thomson, & Brook, 1998; Lynn, 2001; Nicolaas, 2004) and prepaid incentives (Börsch-Supan et al., 2013; Wetzels, Schmeets, van den Brakel, & Feskens, 2008) increase response rates, mainly by boosting the level of cooperation. Direct comparisons of the effects of unconditional and conditional incentives on response rates are rare. Only a few experiments were conducted outside the United States and Canada, and the evidence they provide is inconclusive. Scherpenzeel and Toepoel (2012) found that unconditional incentives worked better than conditional incentives in the recruitment for a new online household panel. In this study, contact was made by either telephone or face-to-face and then led to a short recruitment interview (CATI or CAPI, respectively). Contrary to this result, neither a prepaid nor a promised incentive raised the cooperation rate in the first wave of a face-to-face panel survey in Germany (Castiglioni, Pforr, & Krieger, 2008). However, having kept the incentive conditions the same in waves 2 and 3, it turned out that the cumulative response rate of the panel was lowest for the no incentive group and highest for the conditional incentive group. Similarly, Phelps (2008) found in an incentive experiment implemented in the United Kingdom in round 3 of the ESS that a conditional cash incentive (£10) led to a higher cooperation and response rate than an unconditional incentive in the form of stamps (worth £6).

11 Barón, Breunig, Cobb-Clark, Gørgens, and Sartbayeva (2009), however, report for their study that incentives improved the probability of making contact with individuals with heavy exposure to the income support system.

Regarding the effects of incentives on sample composition, the available information is limited and the results are mixed (Singer et al., 1999; Singer, 2002). Whereas in a number of studies no differences in sample composition between the incentive and the no incentive condition could be observed, there were also a few studies that found that incentives encouraged participation particularly among the groups who would otherwise be underrepresented in the survey, like low education, low income or minority respondents. All these groups have a lower socio-economic status, and it can be assumed that the perceived value of a given incentive will be higher for them than for groups who are better off. In addition, it was found that incentives can act as a motive to participate in a survey for groups in which other stimulating factors like interest in the survey topic or a sense of civic obligation are absent. At the same time, there was no indication that incentives reduce the motivation to participate among those groups who are more interested in the survey topic or more altruistic. In the Detroit Area Study, for instance, Groves, Singer, and Corning (2000) found that the effect of an incentive on the participation rate in a follow-up mail survey was higher for persons scoring low on an index of community/political involvement than for persons scoring high on the same index.¹² The results concerning differences for socio-economic groups as well as differences for persons with different levels of community and political involvement can be seen in accordance with leverage-saliency theory of survey participation (Groves et al., 2000). In both cases, we can assume that the leverage of incentives (i.e., the importance of incentives for the decision to participate) is higher for one group (with lower socio-economic status and lower civic involvement, respectively) than for its counterpart. Whether this actually translates into a difference in respondent cooperation will depend on the leverage of other survey features (like the survey topic, sponsor, etc.) and how salient the different survey features were made to the target persons during the presentation of the survey request.

More recent studies continue to provide mixed evidence regarding sample composition effects. Eyerman et al. (2005) report no effect of incentives on the demographic structure of their samples. In a slightly different analysis of the same data, Wright, Bowman, Butler, and Eyerman (2005) observed a larger effect of incentives for younger age groups and also an interaction effect with population density. Wetzels et al. (2008) found hardly any impact of incentives on response distributions of demographic variables (including age and household income). In a telephone survey, Barón et al. (2009) observed an equalising effect of incentives

12 It has repeatedly been shown for the United States that people doing volunteer work or, more generally, showing larger political and social engagement, are more willing to agree to participate in a survey (Abraham, Helms, & Presser, 2009; Kohut et al., 2012).

on response rates across different socio-economic groups. In a meta-analysis of studies, including self-administered, face-to-face and telephone surveys, Groves and Peytcheva (2008) found no evidence that incentives reduced the nonresponse bias in the statistics of interest.

Regarding fieldwork efforts and survey costs, Singer (2002) reports a number of examples in which incentives were cost-effective measures not only in mail but also in face-to-face surveys. For instance, in a field experiment in the National Adult Literacy Survey in the United States, the average costs per interview were lower when a \$20 incentive was used than in the no incentive condition (Berlin et al., 1992). Likewise, Scherpenzeel and Toepoel (2012) estimate that the costs per registered household for an online panel survey were lowest when a prepaid €10 incentive was used (compared both to the no incentive condition, and to all other conditions of prepaid/promised incentives). Several other face-to-face surveys not included in the summary of Singer looked at the mean number of interviewer visits that were required either to complete an interview or to achieve a final disposition code for all eligible addresses (Eyerman et al., 2005; Lynn et al., 1998; Nicolaas & Stratford, 2005). In each case it was found that less effort was required by interviewers to elicit respondent cooperation when an incentive was provided. However, as Singer (2002) points out, one has to be aware of the fact that the cost-effectiveness of an incentive will vary depending on the specific conditions in a survey.

All results taken together, we can conclude that the findings of field experiments on incentives in face-to-face surveys show that incentives help to increase response rates in face-to-face surveys. The increase in response is mainly achieved by increasing the cooperation rate. Regarding the effects of incentives on sample composition the evidence is mixed. Whereas some analyses found that incentives particularly attract certain subgroups of the population (like younger people, people with a lower socio-economic status or people with a lower community involvement), other analyses did not.¹³ Finally, the literature indicates that providing an

13 Finding out more about the reasons for these divergent results requires a closer look at the different surveys, the experiments and the incentives used. For example, it seems plausible that the higher effect of incentives on young people, which was observed in the study of Wright et al. (2005), might be a result of the fact that a \$20 or \$40 cash incentive is particularly attractive to young people. In contrast, in the study from Wetzels et al. (2008) no differential incentive effect for age was found. It is easily comprehensible that offering a few stamps (worth €2, €4 or €8), as it was done in this study, is not an incentive that particularly attracts young people. A profound review of incentive effects in face-to-face surveys would have to take into account all potential survey features which might interact with the way incentives work. Given the large number of these potential features and the small number of available empirical studies, such a review does not seem promising for the time being.

incentive in face-to-face surveys can help to reduce fieldwork efforts by reducing the number of contact attempts necessary to complete an interview.

Considering the main results of the relevant literature, we set up the following working hypotheses for the incentive experiment implemented in ALLBUS 2010:

Effects of incentives on outcome rates

H1. A higher cooperation and response rate will be achieved in the incentive condition than in the no incentive condition. We do not expect a difference in the contact and capability rate.

Effects of incentives on sample composition

H2. The sample of completed interviews will contain more people with lower socio-economic status in the incentive condition than in the no incentive condition.

H3. The sample of completed interviews will contain more people with low involvement or low interest in community, society and politics in the incentive condition than in the no incentive condition.

Effects of incentives on fieldwork efforts

H4. A lower number of contact attempts will be required to finalise a case / to complete an interview in the incentive condition than in the no incentive condition.

3 Study design

3.1 The sample

We use data from the German ALLBUS 2010 survey.¹⁴ The ALLBUS General Social Survey is a face-to-face survey, fielded every second year since 1980. Its goal is the long-term monitoring of attitudes, behaviour and social structure. Sampling and fieldwork is done by a commercial survey organisation. In 2010, TNS Infratest Social Research (Munich) was responsible for data collection (Wasmer, Scholz, Blohm, Walter, & Jutz, 2012).

14 GESIS – Leibniz Institute for the Social Sciences: ALLBUS 2010 – German General Social Survey. GESIS, Cologne, Germany, ZA4610 Data File version 1.0.0. (2011-05-30), doi:10.4232/1.10445.

ALLBUS uses a national area probability sample of non-institutionalised adults in Western and Eastern Germany, with some oversampling in Eastern Germany. In 2010, the sample was drawn in two stages. In the first stage, 149 communities (including 162 Primary Sampling Units, PSUs) were selected. In the second stage, 40 addresses of individuals were randomly selected from the lists of residents for every PSU. The gross sample of ALLBUS 2010 consisted of 6,480 addresses (162 x 40 addresses) in total.

ALLBUS 2010 was fielded as a CAPI survey. The average length of the interview was about 70 minutes. Fieldwork started at the end of May 2010 and ended at the beginning of November 2010. The incentive experiment was part of the main fielding period which lasted until the beginning of August 2010. In the main phase, 205 interviewers were deployed. The interviewers had to make at least four contact attempts to the target persons, spread over different days of the week and different times of the day. The response rate at the end of the main phase was 24.5%. At the end of the re-issue period a final response rate of 34.6% could be achieved.

3.2 The ALLBUS 2010 incentive experiment

When the use of an incentive is planned in a survey, many practical decisions have to be made (Kulka et al., 2005). These decisions refer, for instance, to the type (e.g., monetary vs. non-monetary) and the value of the incentive, to the timing of payment of the incentive (prepaid vs. promised) and to the groups of persons who should be offered the incentive (e.g., all target persons vs. only persons living in areas with low response rates).

For the ALLBUS 2010 experiment, it was decided to study the effect of a €10 conditional cash incentive offered to all target persons. €10 is a modest amount, thus limiting the risk of the incentive being perceived as a payment for time.¹⁵ To avoid confounding area effects, sample members were randomly assigned to treatments within PSUs. In each PSU, 16 addresses were assigned to the €10 incentive condition, and 24 addresses were assigned to the control group (no incentive).

15 In Germany, an incentive worth €10 might be characterised as a token of appreciation rather than a compensation for time spent for most target persons. For groups who are financially less well off, however, the incentive might be attractive also as a form of payment. In Germany, there is no uniform statutory minimum wage that applies to all employees. The various forms of industry-specific minimum wages are sometimes below €10 per hour (see the information provided by the German Federal Statistical Office:

<https://www.destatis.de/DE/ZahlenFakten/GesamtwirtschaftUmwelt/VerdiensteArbeitskosten/Mindestloehne/Tabellen/MindestlohnDeutschland.html?nn=50678> (02.04.2012)).

As a rule, all addresses from a PSU were allocated to a single interviewer. This means that each interviewer normally worked in both incentive conditions. Thus, the design also controls for interviewer effects (similar: Lynn et al., 1998; Lynn, 2001; Nicolaas, 2004).

The incentive was announced in the advance letter sent to all target persons by the fieldwork department of the survey organisation a few days before the fielding period started. Thus, all target persons were informed about the incentive in a standardised and direct way. The interviewers knew from the colour of the contact forms which addresses had been promised an incentive and which had not. There was no specific instruction for the interviewers with respect to how prominently they should mention the incentive when approaching the respondents.

3.3 Analysis plan

In order to test our hypotheses (see section 2), we first analyse differences in contact, capability, cooperation and response rates between the treatment conditions (H1).

In a second step, we analyse outcome rates for demographic subgroups using data from the sampling frame. The sampling frame of ALLBUS includes the following variables for each target person (i.e., for both respondents and non-respondents): age, gender, citizenship, size of community and region. We use these variables to examine whether incentives have a differential effect on certain population subgroups and may thus alter sample composition (no specific hypothesis formulated).

Subsequently, we use the ALLBUS data set of completed interviews and look at response distributions and survey results for indicators of low socio-economic status (H2) and low involvement in community, society and politics (H3). For both areas of interest the ALLBUS data set provides several indicators.

The next step we take is to look at different fieldwork efforts. We analyse whether the number of contact attempts is lower in the €10 incentive than in the no incentive condition (H4). The required data come from the contact forms the interviewers had to fill in.

All analyses refer to the results of the main fielding period. In all analyses, the sampling design is taken into account. This refers both to the geographical clustering of respondents (communities) and to the unequal selection probabilities for respondents from Eastern and Western Germany. The analyses were performed using stata 12.1 and IBM SPSS 20.

Table 1 Outcome rates ALLBUS 2010 by use of incentive

	N	Contact rate (%)	N	Capability rate (%)	N	Cooperation rate (%)	N	Response rate (%)
No incentive	3,585	82.3	2,951	91.2	2,690	30.8	3,585	23.1
€10	2,356	82.9	1,954	92.1	1,800	34.9	2,356	26.7
Total	5,941	82.6	4,905	91.5	4,490	32.4	5,941	24.5
$F_{\text{design-based}}$		0.431		0.966		7.092		8.197
(df 1, df 2)		(1, 149)		(1, 149)		(1, 149)		(1, 149)
p-value		.513		.327		.009		.005

Notes: The definition of the outcome rates is provided in Appendix 1. The columns labelled 'N' contain the sample size (= 100%) to which the contact, capability, cooperation and response rate refer. $F_{\text{design-based}}$ is a variant of the second-order Rao-Scott adjusted chi-square statistic.

4 Results

4.1 Effect of incentives on outcome rates

Table 1 shows outcome rates for the ALLBUS survey 2010 at the end of the main fielding period (before re-issuing took place). At this stage, the response rate was 24.5%.¹⁶ The use of the €10 incentive had a positive effect on the response rate in ALLBUS 2010. In the incentive group, the response rate is 3.6 percentage points higher than in the control group ($p=.005$). This is mainly achieved by a higher cooperation rate in the €10 condition: the cooperation rate is 34.9% in the incentive condition and 30.8% in the no incentive condition ($p=.009$). As expected, there were no significant differences in the contact and capability rate between the experimental conditions. Table 1 therefore shows clear support for our first hypothesis.

4.2 Differential effects for subgroups

With data from the sampling frame on respondents and non-respondents we can test whether incentives are equally effective in attracting different subgroups of the population. We ran logistic regression models with the four survey outcome variables as dependent (dichotomous) variables. The explanatory variables included

16 This is a rather low interim rate, even for the ALLBUS. In the three preceding survey rounds of ALLBUS, the respective rate had been 7 to 11 percentage points higher.

the frame variables gender (female vs. male), citizenship (German vs. non-German), region (Western vs. Eastern Germany), age (in years) and size of community (number of inhabitants, 7 categories), plus the treatment condition (€10 incentive vs. no incentive).

We started with a model including only the main effects of the frame variables and the treatment condition. Table 2 shows that, according to the values of the Pseudo R^2 , 'contact' and 'capability' can be better explained by the frame variables than 'cooperation' and 'response'. Turning to the effects of the individual frame variables, we find that females have a higher propensity of being contacted and a lower propensity of being capable to provide an interview than males. Older people are easier to contact than younger people. On the other hand they are less capable of providing an interview. German citizenship has a positive effect on 'capability' and 'response'. This is mainly due to the fact that an insufficient command of the German language is more widespread among non-Germans than among Germans. Regarding region, we do not observe an effect for any of the outcome variables. In contrast, size of community plays a significant role for several outcome variables. Target persons living in larger communities are more difficult to contact, and less cooperative than persons living in smaller communities. As a consequence, the chance of getting an interview is lower in larger communities than in smaller communities. These results are similar to results already observed in previous analyses (Koch, 1997). In addition to the effects of the frame variables, we also find a positive effect from the incentive. As in the bivariate case, the positive effect of the €10 incentive only refers to 'cooperation' and 'response' (and not to 'contact' and 'capability').

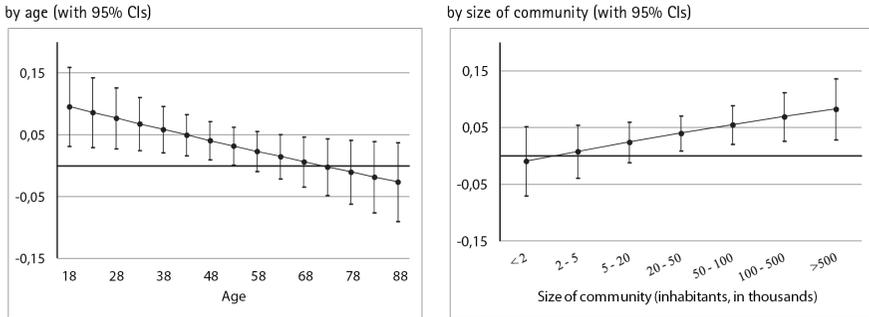
However, one has to be careful when reporting and interpreting parameter estimates of logistic regressions. Logit models are not very sensitive to explicitly specified interaction terms. The reason for this is that even without modelling interaction terms, logit models estimate model-inherent interactions (conditional effects) (Nagler, 1994; Best & Wolf, 2010). In addition, one has to bear in mind that the estimated levels of significance only refer to the latent variables; they do not directly indicate changes in probabilities, which is our main interest (Ai & Norton, 2003). As a consequence, Best and Wolf (2010, 2012) suggest analysing the relationship between independent and dependent variables using a graphic representation and to use the averaged effect as a measure (Average Marginal Effect,

Table 2 Logistic regression of contact, capability, cooperation and response on frame variables and incentives

	Contact			Capability			Cooperation			Response							
	Model 1			Model 1			Model 1			Model 1							
	β	SE		β	SE		β	SE		β	SE						
Female (<i>male</i>)	.278***	(.066)		-.246*	(.116)		.102	(.073)		.099	(.072)		.129+	(.068)		.126+	(.068)
German (<i>non-German</i>)	.169	(.144)		2.240***	(.185)		.087	(.154)		.083	(.155)		.462**	(.150)		.462**	(.150)
Western Germany (<i>Eastern Germany</i>)	.066	(.142)		.19	(.175)		.133	(.114)		.139	(.115)		.158	(.109)		.0163	(.109)
Age, centred	.020***	(.002)		-.047***	(.005)		-.003	(.002)		.000	(.003)		-.003+	(.002)		-.000	(.002)
Size of community, inhabitants	-.240***	(.044)		-.030	(.040)		-.067*	(.032)		-.097**	(.037)		-.119***	(.029)		-.141***	(.035)
€10 incentive (<i>no incentive</i>)	.033	(.072)		.174	(.128)		.186*	(.073)		-.112	(.166)		.190**	(.070)		-.024	(.153)
Age, centred*incentives										-.008*	(.004)					-.007+	(.003)
Size of community*incentives										.073+	(.037)					.052	(.034)
constant	2.330***	(.277)		0.783*	(.303)		-.777**	(.239)		-.655**	(.248)		-1.331***	(.227)		-1.245***	(.234)
N	5745	-2502.1	4749	-1187.7	4351	-2702.3	4351	-2698.1	5745	-3136.0	5745	-3136.0	5745	-3133.0			
AIC	0.873		0.503		1.245		1.244		1.094		1.093		1.093				
Pseudo R2	0.055		0.125		0.005		0.007		0.013		0.014		0.014				

Notes: Reference category in parentheses. + = $p < .10$, * = $p < .05$, ** = $p < .01$, *** = $p < .001$

Figure 2 Average Marginal Effects of incentives on the probability of 'cooperation'



AME¹⁷). In the present case, these were the AMEs of using a respondent incentive on the probability of 'cooperation' and 'response', respectively.

Figure 2 depicts the AMEs for the probability of 'cooperation' and the respective confidence intervals for age and size of community (again, the analyses take the sampling design into account). The positive effect of a respondent incentive for younger age groups can clearly be seen. The effect is significantly different from zero for persons aged 50 years or younger. In contrast, for persons aged 50 years or older the effect is no longer significant. Similarly, the effect of the incentive is dependent on the size of the community. Only for persons living in communities with more than 20,000 inhabitants we find a significant effect of the incentive on 'cooperation'. Appendix 2 includes the respective figures for 'response'. The structure of the results is quite similar to the results for 'cooperation'; the effects, however, are weaker. Generally, we can conclude that the interaction effects are not very strong; they improve the model fit only marginally (see the AIC in Table 2). The graphic representation, however, helps to identify the subgroups for which the use of an incentive has a positive effect that is significantly different from zero.

Taken together, our analyses provide some evidence for weak differential effects of incentives on 'cooperation'. The crucial point is that with respect to non-response bias the results work in opposite directions for the two variables 'age' and 'size of community'. Regarding urbanicity, the incentive seems to help to reduce an existing imbalance in cooperation propensities, whereas concerning age the incen-

17 To arrive at Average Marginal Effects (AMEs) for every observation, the marginal effects are computed and then averaged (Long, 1997). AMEs measure the mean change in $P(Y=1)$ as one independent variable X_k increases while all other variables are kept constant (Bartus, 2005).

Table 3 Socio-demographic variables and use of incentives

	N	F _{design based}	(df 1, df 2)	p-value
Gender	1,455	.044	(1, 147)	.835
Age	1,453	2.174	(4.8, 699.1)	.059
Size of community (inhabitants)	1,455	1.876	(5.4, 791.2)	.091
Citizenship (German/non-German)	1,454	.150	(1, 147)	.699
Region (Western/Eastern Germany)	1,455	.380	(1, 147)	.538

tive seems to introduce an imbalance in cooperation propensities in the first place. When it comes to 'response', however, both effects are attenuated.

4.3 Differences in response distributions in the realised sample

Another possibility to assess differential effects of incentives is to look at the response distributions in the realised sample. Since the incentives in our experiment were offered to a random subsample of the respondents, there should be no significant differences in variable distributions between respondents who received an incentive and those who did not. As long as we can preclude that the incentive had an effect on measurement (by altering the responses provided during the interview), any significant difference in variable distributions indicates that the incentive is more effective in recruiting certain subgroups than others.

The merit of this approach is that a wide array of variables can be examined for potential effects of incentives. In a first step, we repeated the analyses with the same variables as were used in the analyses of the sampling frame data (section 4.2). For each variable, we ran bivariate association tests. The present results (see Table 3; the detailed cross tabulations are provided in Appendix 3) follow the same pattern as the results for the analyses above. For age and size of community, we observe small differences between the incentive and the no incentive condition that reach borderline significance ($p=.059$ and $p=.091$). Gender, citizenship and region do not show any difference.

In a second step, we analysed differences for a range of variables indicating the socio-economic situation of the respondent in order to investigate whether H2 (more people with a lower socio-economic status in the incentive condition) comes true. The variables included both objective (e.g., education and income) and subjective measures (e.g., the subjective evaluation of the own economic situation or subjective social class). Table 4 contains the results of the bivariate association

tests for each variable (for the detailed cross tabulations see Appendix 3). The results provide only limited evidence that incentives bring people into the respondent pool who are socio-economically less well off. For two variables the difference between the treatment and the control condition is significant ($p < .050$). The proportion of people living in a rented flat is slightly higher in the €10 condition than in the control group (40.3% vs. 34.1%). Also, the proportion of students is higher in the subsample where the €10 incentive was offered (5.9% vs. 3.6%). However, in a multiple regression analysis with age and size of community as control variables both associations are attenuated and fail (student, $p = .138$) or nearly fail the usual p -level for significance (tenant of a flat, $p = .098$).¹⁸ The other objective and subjective measures of the socio-economic situation of the respondent are not related to the treatment condition. We find, for instance, no difference with regard to the educational level of respondents, their income or their subjective and objective social class placement.

The situation looks similar for all indicators of involvement in community, society and politics. We postulated a higher proportion of people with low involvement in the incentive condition (H3). However, the response distributions for both the subjective and objective variables in ALLBUS referring to this topic do not differ between the no incentive and the €10 incentive condition (see Table 5; for the detailed cross tabulations see Appendix 3). For instance, we find no indication that persons less interested in politics are more attracted by an incentive to participate in ALLBUS than persons highly interested in politics. Similarly, persons with a smaller personal network or non-members of associations cannot be found more often among the group who received an incentive.

Concluding this section we should note that, in addition to the abovementioned variables, we looked at a wide range of other variables for differences in response distributions between the two treatment conditions. The results of these analyses did not provide any further evidence for systematic differences between respondents who received an incentive and those who did not. This holds for both the demographic and the non-demographic variables included in ALLBUS.

18 We performed multiple logistic regressions with the treatment condition (no incentive (0) vs. €10 incentive (1)) as dependent variable. Tenant of a flat and student were both individually used as independent variables, and age and size of community were included as control variables.

Table 4 Socio-economic variables and use of incentives

	N	F _{design based}	(df 1, df 2)	p-value
Education	1,436	.369	(2.0, 288.4)	.688
HH equivalence income (quintiles, in €)	1,161	1.801	(3.8, 564.2)	.130
Personal income (quintiles, in €)	1,264	.408	(3.9, 575.3)	.799
Tenant of a flat	1,438	2.958	(2.9, 432.6)	.033
Student	1,451	4.123	(1, 147)	.044
Unemployed	1,451	1.269	(1, 147)	.262
Economic situation (subj.)	1,449	1.348	(3.8, 564.0)	.252
Social class (subj.)	1,426	.691	(3.0, 435.1)	.556
Social class (Goldthorpe/Terwey) ¹⁹	1,279	.650	(9.6, 1417.6)	.766

Table 5 Involvement in community, society and politics and use of incentives

	N	F _{design based}	(df 1, df 2)	p-value
Interest in politics	1,455	.551	(3.9, 566.7)	.692
Voted last election	1,451	.819	(1, 147)	.367
Size of personal network	1,382	.090	(2.9, 426.3)	.962
No. of association memberships	1,455	.053	(3.0, 435.3)	.983
Member of a party/a union	1,455	.174	(1, 147)	.678

4.4 Incentives and fieldwork efforts

The number of contact attempts interviewers have to make either to complete an interview or to achieve a final outcome status for the nonresponding cases is a strong determinant of fieldwork costs in face-to-face surveys. This assertion holds in particular for the number of in-person contact attempts. In our fourth hypothesis, we postulated that fewer contact attempts are needed when an incentive is promised. Table 6 provides no support for this hypothesis. There are no significant differences in the average number of contact attempts between the €10 incentive and the no incentive condition. Both for the gross sample and for the net sample of completed interviews, and both for the total number of contact attempts (in

19 For details see Terwey and Baltzer (2011).

Table 6 Average number of contact attempts by use of incentives

Mode of contact attempt		No incentive	€10
In person + by phone	Gross sample	3.06 (3,585)	3.04 (2,356)
	Net sample (interviews)	3.49 (827)	3.37 (628)
In person only	Gross sample	2.44 (3,585)	2.45 (2,356)
	Net sample (interviews)	2.62 (827)	2.58 (628)

Note: *N* in parentheses

person + by phone)²⁰ and the in-person contact attempts only, the results are more or less the same in the two treatment conditions.²¹ Thus, in this perspective we do not find an indication that a €10 incentive helps to reduce interviewer efforts.

The situation looks different when we take the higher response rate into account that was achieved in the incentive condition. When we compute the total number of contact attempts for the gross sample and divide it by the number of completed interviews in both treatment conditions, we find that the total effort exerted is lower when an incentive was used (Table 7). The average number of contact attempts per completed interview (including attempts on respondents and non-respondents) is 13.25 in the control group and 11.42 in the €10 treatment group. This means that for a given number of completed interviews on average 14% fewer contact attempts were needed when a €10 incentive was used. A reduction of a similar magnitude is observed when we look at personal contact attempts only (on average 10.57 vs. 9.20 attempts).

Unfortunately, we do not know what this reduction in efforts meant for the actual survey costs in ALLBUS. An estimation of the cost saving requires knowing about the costs of an interviewer visit. The survey organisation should be able to estimate these costs. However, a survey sponsor usually will not be in the position to receive this information. For this reason, a simple thought experiment might shed some light on this issue.

20 Interviewers were allowed to get in touch with the target person by telephone in order to make an appointment for an interview. However, the interviews had to be completed face-to-face. Since evidence exists that respondent cooperation can be more easily achieved in a face-to-face than in a telephone contact (Blohm, Hox, & Koch, 2007), the general recommendation for the ALLBUS interviewers was to make contact with target persons by a personal visit. As can be seen from Table 6, interviewers followed this recommendation: Most contact attempts were made in person.

21 Note that the results refer to the main fielding period only. The average number of contact attempts are higher when we also include the additional attempts interviewers made during the re-issue phase.

Table 7 Total number of contact attempts per completed interview by use of incentives

Mode of contact attempt		No incentive	€10
In person + by phone	Total # of contact attempts to all units of the gross sample	10,961	7,170
	# of completed interviews	827	628
	Average # of contact attempts per completed interview	13.25	11.42
In person only	Total # of contact attempts to all units of the gross sample	8,745	5,780
	# of completed interviews	827	628
	Average # of contact attempts per completed interview	10.57	9.20

The ALLBUS survey usually aims for 3,500 completed interviews. From the above analyses we know that in the condition with no incentive 10.57 personal contact attempts are made on average per completed interview (taking into account the total number of contact attempts on the complete gross sample). When we use a conditional incentive of €10, the same number of interviews can be completed with 9.20 contact attempts on average. The introduction of a €10 incentive for the whole sample of ALLBUS thus would mean a reduction in the total number of personal contact attempts by 4,795 attempts ($= 3,500 \times 1.37$). On the other hand, the use of a conditional monetary incentive of €10 would cause costs of €35,000 ($= 3,500 \times 10\text{€}$; not taking into account any handling costs). Accordingly, we might conclude that the introduction of the incentive would pay for itself if the average costs of a personal contact attempt amount to €7.30 or more ($= \text{€}35,000 / 4,795$).²²

5 Summary and conclusions

In ALLBUS 2010, an experiment was set up to test the effect of respondent incentives on outcome rates, sample composition and fieldwork efforts. A random subsample of target persons was offered a monetary incentive of €10 to be paid upon completion of the interview. The other part of the sample acted as a control group and received no incentive. In the following, we briefly summarise and discuss the

22 Note that we arrive at a very similar result when we use the total number of contact attempts in the main and re-issue phase for the calculation.

results of the experiment, separately for the three areas of outcome rates, sample composition and fieldwork efforts.

(1) Outcome rates

The monetary incentive used in ALLBUS 2010 led to an increase in the response rate (3.6 percentage points), mainly by improving the cooperation rate. This positive effect occurred at a rather low response rate level. In addition, we should point out that this response rate difference was observed at the end of the main fielding period, before re-issuing efforts were undertaken. Both factors suggest that the effect size we found is probably an upper bound for the size of an effect of a (modest) monetary incentive in a survey like ALLBUS. We might expect a smaller effect in surveys with higher response rates (Singer, 2002). We might also expect that in surveys implementing a re-issue phase, additional interviewer efforts in motivating previously reluctant target persons will tend to decrease the effect of an incentive. Fieldwork in ALLBUS 2010 included a re-issue phase. However, our data do not allow to assess in a rigorous manner whether the latter hypothesis is true or not, since in the re-issue period of ALLBUS 2010 an incentive of €20 was offered to all remaining target persons. After the finalisation of fieldwork, the response rate of ALLBUS 2010 was 34.6%, around 10 percentage points higher than at the end of the main fielding period. Despite the fact that the €20 incentive was offered to all sampling units in the re-issue phase, still a difference in the response rate of 2.9 percentage points could be observed between the original €10 incentive group and the control group. This difference is only a little smaller than the difference at the end of the main fielding period.

Our results refer to a conditional monetary incentive of a modest amount. They do not preclude that monetary incentives of a higher value lead to a more substantial increase in cooperation and response in surveys similar to ALLBUS. Eyerman et al. (2005) provide some empirical evidence that this is actually the case as regards conditional incentives. Likewise, Boersch-Supan et al. (2013) report that an increase in the value of an unconditional incentive (€10 - €20 - €40) led to higher cooperation rates in an almost linear fashion. In contrast to these results, Scherpenzeel and Toepoel (2012) found no difference in response rates when comparing an incentive worth €10 with an incentive worth €20 or €50 (including both promised and prepaid incentives). The different results illustrate that the effects of incentives may vary depending on the concrete survey context.

(2) Sample composition

The incentive used in ALLBUS 2010 did not change sample composition in a major way. We found that the incentive was somewhat more attractive to younger people and to people living in urban areas. Also, the incentive condition contained a slightly higher number of students and tenants of a flat. The latter results might be interpreted as a clue that a €10 incentive disproportionately attracts people who are financially less well off. At the same time, however, we have to acknowledge that other, more direct indicators of socio-economic status (like income) did not differ at all. One might speculate whether a larger monetary incentive might be followed by larger differences in sample composition regarding indicators of socio-economic status.

We also hypothesised that persons with a lower interest in politics and lower involvement in community and society would disproportionately be attracted by an incentive. The idea behind this was that interest in politics or involvement in community might indicate some kind of feeling of civic duty or a norm of cooperation that would motivate people with these predispositions to disproportionately participate in the ALLBUS survey. People with a low interest in politics or a low involvement in community, on the other hand, would be more attracted by an extrinsic benefit like an incentive. The results of our experiment provide no support for this hypothesis. A possible explanation might be that the variables we looked at are not well suited to capture the underlying concepts of 'civic duty' or 'norms of cooperation'. In addition, we might also ask whether 'norms of cooperation' are more relevant for motivating people to participate in surveys in the United States than in Germany. In any case, the widespread absence of differences in response distributions between the incentive and the no incentive condition in ALLBUS indicates that the incentive did not introduce any imbalance in sample composition.

(3) Fieldwork efforts

Decisions on survey design features typically have to be made against the backdrop of a limited project budget. The use of incentives, on the one hand, increases costs. These costs include both the pure costs of the incentives and the associated costs of their handling and administration. The crucial question is whether incentives, on the other hand, help to save costs by reducing the fieldwork efforts interviewers have to exert. This will be the case when an incentive makes (at least some) sample members more readily willing to participate.

In the ALLBUS experiment, we found that using an incentive allowed the survey organisation to achieve a certain number of interviews with a lower number

of contact attempts than when no incentive was used. From the outside it was not possible for us to provide a reliable estimation what this reduction in efforts meant for the actual survey costs in ALLBUS. Apart from that, one has to be aware that this cost saving first and foremost is a benefit for the survey organisation. In survey practice, it will often be questionable whether the survey organisation passes the cost savings associated with the use of a respondent incentive to the survey sponsor.

To conclude, the conditional monetary incentive of €10 used in ALLBUS 2010 had a small positive effect on the cooperation and response rate achieved. There is hardly any indication that the incentive altered the sample composition of the ALLBUS survey. Regarding survey costs, the incentive led to a small reduction in the fieldwork efforts the interviewers had to exert.

For the decision on the future use of a respondent incentive in ALLBUS, the willingness of the survey organisation to share the cost savings associated with the use of an incentive with the survey sponsor will play an important role. If the survey organisation is willing to pass on the cost savings, a higher response rate might be achieved for a similar price – which would be a clear argument in favour of using an incentive in upcoming survey rounds of ALLBUS. In case the survey organisation will not pass the cost savings to the survey sponsor, we might say that the decision will depend on the answer to the following question: Are you willing to accept an additional cost burden ($3,500 \text{ interviews} \times \text{€}10 = \text{€}35,000$) in order to achieve an increase in response by 3-4 percentage points? A well-considered answer to this question has to weigh the incentive costs against the costs of other survey features (like, e.g., better training of interviewers) that are also deemed adequate to increase response rates or to improve data quality in general (Rodgers, 2011).

In addition, for an informed choice we might also wish to take into account the potential effects of incentives on measurement. In the present paper, we did not investigate whether or not the €10 incentive changed the response behaviour of the ALLBUS 2010 respondents. Incentives can affect data quality, for instance, by changing the level of item nonresponse or by altering the expression of opinions. Up till now there exists only limited evidence on the effects of incentives on measurement. Some studies reveal a positive effect, others a negative effect. More recent studies often found no association between receiving an incentive and data quality (Singer & Ye, 2013). Given the rather modest value of the incentive used in the present study, large effects on measurement do not seem to be very likely.

Nevertheless, only future research can show whether this assumption is actually true or not.

Beyond ALLBUS, the present results should be relevant for other surveys as well. A very conservative approach would request that each survey considering the use of incentives should implement its own experiment to find out about the potential effects. We do not think that this claim has to be asserted in such a strict manner. The results of a particular experiment like the present one provide a reference point for other studies that follow a similar design and take place in a comparable social context. Thus, social surveys of the general population that are fielded face-to-face and cover a broad range of topics might profit from the present results. As we elaborated in the introduction, there is a large number of surveys all over the world for which this description applies.

References

- Abraham, K. G., Helms, S., & Presser, S. (2009). How Social Processes Distort Measurement: The Impact of Survey Nonresponse on Estimates of Volunteer Work in the United States. *American Journal of Sociology*, 114(4), 1129-1165.
- Ai, C., & Norton, E. C. (2003). Interaction Terms in Logit and Probit Models. *Economics Letters*, 80(1), 123-129.
- The American Association for Public Opinion Research (2011). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 7th edition. AAPOR.
- Atrostic, B. K., Bates, N., Burt, G., & Silberstein, A. (2001). Nonresponse in U.S. Government Household Surveys: Consistent Measures, Recent Trends, and New Insights. *Journal of Official Statistics*, 17(2), 209-226.
- Barón, J. D., Breunig, R. V., Cobb-Clark, D., Gørgens, T., & Sartbayeva, A. (2009). Does the Effect of Incentive Payments on Survey Response Rates Differ by Income Support History? *Journal of Official Statistics*, 25(4), 483-507.
- Bartus, T. (2005). Estimation of marginal effects using *margeff*. *The Stata Journal*, 5(3), 309-329.
- Berlin, M., Mohadjer, L., Waksberg, J., Kolstad, A., Kirsch, I., Rock, D., & Yamamoto, K. (1992). An Experiment in Monetary Incentives. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 393-398.
- Best, H., & Wolf, C. (2010). Logistische Regression. In C. Wolf & H. Best (Eds.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (pp. 827-854). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Best, H., & Wolf, C. (2012). Modellvergleich und Ergebnisinterpretation in Logit- und Probit-Regressionen. *Kölner Zeitschrift für Soziologie und Sozialwissenschaften*, 64(2), 377-395.
- Blohm, M., Hox, J., & Koch, A. (2007). The Influence of Interviewers' Contact Behavior on the Contact and Cooperation Rate in Face-to-Face Household Surveys. *International Journal of Public Opinion Research*, 19(1), 97-111.

- Börsch-Supan, A., Krieger, U., & Schröder, M. (2013). *Respondent Incentives, Interviewer Training and Survey Participation*. SHARE Working Paper 12-2013, Munich Center for the Economics of Aging (MEA).
- Castiglioni, L., Pforr, K., & Krieger, U. (2008). The Effect of Incentives on Response Rates and Panel Attrition: Results of a Controlled Experiment. *Survey Research Methods*, 2(3), 151-158.
- Church, A. H. (1993). Estimating the Effect of Incentives on Mail Survey Response Rates: A Meta-Analysis. *Public Opinion Quarterly*, 57(1), 62-79.
- de Leeuw, E., & de Heer, W. (2002). Trends in Household Survey Nonresponse. A Longitudinal and International Comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse* (pp. 41-54). New York: John Wiley & Sons.
- Dixon, J., & Tucker, C. (2010). Survey Nonresponse. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of Survey Research*. Second Edition (pp. 593-630). Bingley: Emerald.
- Eurofound (2012). *Third European Quality of Life Survey – Quality of life in Europe: Impacts of the crisis*. Publications Office of the European Union, Luxembourg.
- Eyerman, D., Bowman, K., Butler, C., & Wright, D. (2005). The differential impact of incentives on refusals: Results from the 2001 national household survey on drug abuse incentive experiment. *Journal of Economic and Social Measurement*, 30(2-3), 157-169.
- Groves, R. M. (2008). Verbal statement during a "Seminar On Survey Respondent Incentives: Research and Practice", held on March 10 by The Council of Professional Associations on Federal Statistics. Retrieved September 08, 2011, from http://www.copafs.org/seminars/research_and_practice.aspx
- Groves, R. M., & Couper, M. P. (1998). *Nonresponse in Household Interview Surveys*. New York: John Wiley & Sons.
- Groves, R. M., Fowler, F. J. Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey Methodology*. Hoboken, New Jersey: John Wiley & Sons.
- Groves, R. M., & Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias. A Meta-Analysis. *Public Opinion Quarterly*, 72(2), 167-189.
- Groves, R. M., Singer, E., & Corning, A. (2000). Leverage Saliency Theory of Survey Participation. *Public Opinion Quarterly*, 64(3), 299-308.
- Koch, A. (1997). Teilnahmeverhalten beim ALLBUS 1994. Soziodemographische Determinanten von Erreichbarkeit, Befragungsfähigkeit und Kooperationsbereitschaft. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 49(1), 98-122.
- Kohut, A., Keeter, S., Doherty, C., Dimock, M., & Christian, L. (2012). *Assessing the Representativeness of Public Opinion Surveys*. The Pew Research Center.
- Kulka, R. A., Eyerman, J., & McNeeley, M. E. (2005). The use of monetary incentives in federal surveys on substance use and abuse. *Journal of Economic and Social Measurement*, 30(2-3), 233-249.
- Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, California: Sage Publications.
- Lynn, P. (2001). The Impact of Incentives on Response Rates to Personal Interview Surveys: Role and Perceptions of Interviewers. *International Journal of Public Opinion Research*, 13(3), 326-336.
- Lynn, P., Thomson, K., & Brook, L. (1998). An Experiment with Incentives on the British Social Attitudes Survey. *Survey Methods Centre Newsletter*, 18(2), 12-14.
- Nagler, J. (1994): Scobit: An Alternative Estimator to Logit and Probit. *American Journal of Political Science*, 38(1), 230-255.
- Nicolaas, G. (2004). The use of incentives to motivate "hard to get" households on the National Travel Survey. *Survey Methods Newsletter*, 22(Spring), 19-27.

- Nicolaas, G., & Stratford, N. (2005). A Plea for the Tailored Use of Respondent Incentives. In C. van Dijkum, J. Blasius, & C. Durand (Eds.), *Recent developments and applications in social research methodology* (CD-ROM proceedings of the RC33 Sixth International Conference on Social Science Methodology, Workshop No. 321). Amsterdam: Budrich Verlag.
- Phelps, A. (2008). *UK Incentive Experiment*. Paper presented at the ESS Field Directors Meeting, Mannheim, January 2008.
- Rodgers, W. L. (2011). Effects of Increasing the Incentive Size in a Longitudinal Study. *Journal of Official Statistics*, 27(2), 279-299.
- Scherpenzeel, A. & Toepoel, V. (2012). Recruiting a Probability Sample for an Online Panel. Effects of Contact Mode, Incentives, and Information. *Public Opinion Quarterly*, 76(3), 470-490.
- Singer, E. (2002). The Use of Incentives to Reduce Nonresponse in Household Surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse* (pp. 163-177). New York: John Wiley & Sons.
- Singer, E. (2006). Introduction – Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, 70(5), 637-645.
- Singer, E., & Kulka, R. A. (2001). Paying Respondents for Survey Participation. In M. V. Ploeg, R. A. Moffitt, & C. F. Citro (Eds.), *Studies of Welfare Populations: Data Collection and Research Issues* (pp. 105-128). Committee on National Statistics, National Research Council.
- Singer, E., van Hoewyk, J., Gebler, N., Raghunathan, T., & McGonagle, K. (1999). The Effect of Incentives on Response Rates in Interviewer-Mediated Surveys. *Journal of Official Statistics*, 15(2), 217-230.
- Singer, E., & Ye, C. (2013). The Use and Effects of Incentives in Surveys. *The ANNALS of the American Academy of Political and Social Science*, 645, 112-141.
- Smith, T. W., Kim, J., Koch, A., & Park, A. (2006). Social-Science Research and the General Social Surveys. *Comparative Sociology*, 5(1), 33-43.
- Terwey, M., & Baltzer, S. (2011). *Variable Report ALLBUS / Allgemeine Bevölkerungsumfrage der Sozialwissenschaften 2010*. ZA-Nr. 4610. GESIS-Variable Reports 2011/02. Köln: GESIS – Leibniz Institut für Sozialwissenschaften. doi:10.4232/1.10760
- Wasmer, M., Scholz, E., & Blohm, M. (2010). *Konzeption und Durchführung der "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 2008*. GESIS-Technical Reports 2010|04. Mannheim: GESIS – Leibniz Institut für Sozialwissenschaften.
- Wasmer, M., Scholz, E., Blohm, M., Walter, J., & Jutz, R. (2012). *Konzeption und Durchführung der "Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften" (ALLBUS) 2010*. GESIS-Technical Reports 2012|12. Mannheim: GESIS – Leibniz Institut für Sozialwissenschaften.
- Wetzels, W., Schmeets, H., van den Brakel, J., & Feskens, R. (2008). Impact of Prepaid Incentives in Face-to-face Surveys: A Large-Scale Experiment with Postage Stamps. *International Journal of Public Opinion Research*, 20(4), 507-516.
- Wright, D., Bowman, K., Butler, D., & Eyerman, J. (2005). Non-response bias from the national household survey on drug abuse incentive experiment. *Journal of Economic and Social Measurement*, 30(2-3), 219-231.

Addresses of authors

Michael Blohm
GESIS – Leibniz Institute for the Social Sciences
P.O. Box 122155
68072 Mannheim
Germany
E-Mail: michael.blohm@gesis.org

Achim Koch
GESIS – Leibniz Institute for the Social Sciences
P.O. Box 122155
68072 Mannheim
Germany
E-Mail: achim.koch@gesis.org

Appendix 1: Definition of outcome rates

In household surveys like ALLBUS, survey participation is a sequential process, including several steps that have to be passed successfully in order to achieve a completed interview (Groves & Couper, 1998). First, the target persons have to be contacted. Second, the target persons' capability to respond has to be assessed. Third, target persons who have been successfully contacted and who have proven able to respond have to be motivated to actually provide the interview. Accordingly, we analyse four separate outcome variables: the contact rate, the capability rate, the cooperation rate and the response rate. We rely on the following procedures and assumptions when calculating the rates:

- All rates are respondent-level rates.
- There are no cases of unknown eligibility.
- "Partial" interviews are not accepted; they are classified as invalidated interviews.
- The outcome "Not able" also includes "cases not fully processed/invalidated interviews" in addition to the main codes "permanently ill/disabled" and "language difficulties".

Our outcome rates compare to the AAPOR Standard Definitions (2011) in the following way:

Contact rate: AAPOR CON 3

- Contacts excepting frame errors

$$(N_{\text{interviews}} + N_{\text{not able}} + N_{\text{refused}}) / (N_{\text{interviews}} + N_{\text{not contacted}} + N_{\text{not able}} + N_{\text{refused}})$$

Capability rate: no AAPOR definition available

- Capability excepting frame errors and noncontacts

$$(N_{\text{interviews}} + N_{\text{refused}}) / (N_{\text{interviews}} + N_{\text{not able}} + N_{\text{refused}})$$

Cooperation rate: AAPOR COOP 3

- Cooperation excepting frame errors, noncontacts and cases unable to respond

$$(N_{\text{interviews}}) / (N_{\text{interviews}} + N_{\text{refused}})$$

Response rate: AAPOR RR 5

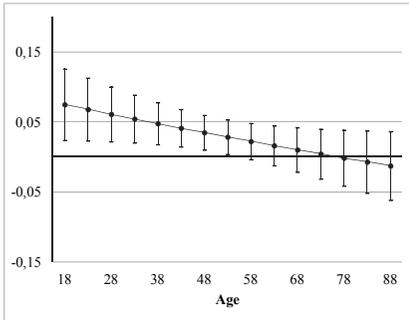
- Response excepting frame errors

$$(N_{\text{interviews}}) / (N_{\text{interviews}} + N_{\text{not contacted}} + N_{\text{not able}} + N_{\text{refused}})$$

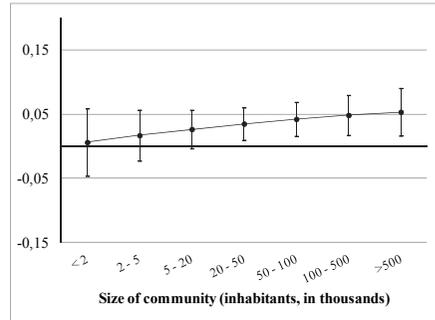
Appendix 2: Average Marginal Effects

Figure A2: Average Marginal Effects of incentives on the probability of 'response'

by age (with 95% CIs)



by size of community (with 95% CIs)



Appendix 3: Cross tabulation of ALLBUS 2010 variables by incentive

Notes: All cross tabulations are based on weighted counts. F (design based) is a variant of the second-order *Rao-Scott adjusted chi-square* statistic.

Table A3.1: Socio-demographic variables and use of incentives (column percent)

	No incentive	€10 incentive	F _{design based}	(df 1, df 2)	p-value
Gender			.044	(1, 147)	.835
Male	45.7	46.3			
Female	54.3	53.7			
N	827	628			
Age			2.174	(4.8, 699.1)	.059
18-29	15.9	18.2			
30-39	13.2	13.7			
40-49	18.3	22.2			
50-59	19.3	14.1			
60-69	16.1	16.4			
70 +	17.2	15.3			
N	826	627			
Size of community (inhabitants)			1.876	(5.4, 791.2)	.091
- 1,999	9.2	6.9			
2,000 - 4,999	15.1	13.5			
5,000 - 19,999	25.1	24.0			
20,000 - 49,999	14.9	17.3			
50,000 - 99,999	6.6	8.7			
100,000 - 499,999	18.9	17.3			
500,000 +	10.2	12.3			
N	827	628			
Citizenship			.150	(1, 147)	.699
German	95.4	94.9			
Non-German	4.6	5.1			
N	826	628			
Region			.380	(1, 147)	.538
Western Germany	82.7	81.7			
Eastern Germany	17.3	18.3			
N	827	628			

Table A3.2: Socio-economic variables and use of incentives (column percent)

	No incentive	€10 incentive	$F_{\text{design based}}$	(df 1, df 2)	p-value
Education			.369	(2.0, 288.4)	.688
Certificate of Secondary Education (<i>Hauptschule</i>)	37.8	36.9			
Advanced Technical College Entrance Qualification (<i>Mittlere Reife & Fachhochschulreife</i>)	33.3	32.2			
University Entrance Qualification/Adv. Tech. College Degree/University Degree (<i>Abitur/Fachhochschul-/Universitätsabschluss</i>)	28.8	30.9			
<i>N</i>	817	619			
HH equivalence income (quintiles, in €)			1.801	(3.8, 564.2)	.130
– 849	15.8	20.8			
850 – 1149	20.2	19.6			
1150 – 1459	23.1	17.7			
1460 – 1979	20.3	20.5			
1980 +	20.7	21.3			
<i>N</i>	653	509			
Personal income (quintiles, in €)			.408	(3.9, 575.3)	.799
None	20.4	22.6			
– 499	21.4	21.0			
500 – 999	22.6	20.1			
1000 – 1499	16.8	16.5			
1500 +	18.9	19.9			
<i>N</i>	716	548			
Tenant / owner			2.958	(2.9, 432.6)	.033
Tenant of flat	34.1	40.3			
Tenant of house	5.2	3.6			
Owner of flat	8.4	5.8			
Owner of house	52.3	50.3			
<i>N</i>	821	618			
Student			4.123	(1, 147)	.044
No	96.4	94.1			
Yes	3.6	5.9			
<i>N</i>	825	626			
Unemployed			1.269	(1, 147)	.262
No	94.9	96.0			
Yes	5.1	4.0			
<i>N</i>	825	626			

	No incentive	€10 incentive	$F_{\text{design based}}$	(df 1, df 2)	p-value
Economic situation (subj.)			1.348	(3.8, 564.0)	.252
Very good	3.2	4.9			
Good	49.8	50.8			
Partly good/partly bad	34.1	30.6			
Bad	10.3	9.9			
Very bad	2.7	3.9			
<i>N</i>	823	627			
Social class (subj.)			.691	(3.0, 435.1)	.556
Lower class	3.1	3.5			
Working class	26.9	24.3			
Middle class	59.3	59.4			
Upper middle class and upper class	10.6	12.7			
<i>N</i>	817	609			
Social class (Goldthorpe/Terwey)²³			.650	(9.6, 1417.6)	.766
Higher service class	14.1	14.0			
Lower service class	28.1	27.7			
Non-manual, clerical	7.9	8.7			
Proprietor, 2-49 employees	3.4	2.4			
Self-empl., 1/no employee	3.9	4.7			
Farmer, smallholder	1.9	1.4			
Foreman, technician	8.4	8.9			
Skilled worker	14.3	12.3			
Semi-/unskilled worker	11.4	13.8			
Agricultural worker	1.8	0.9			
Other non-manual worker	4.8	5.1			
Farmer, co-op farmer	0.1	0.1			
<i>N</i>	732	547			

23 For details see Terwey and Baltzer (2011).

Table A3.3: Involvement in community, society and politics and use of incentives (column percent)

	No incentive	€10 incentive	$F_{\text{design based}}$	(df 1, df 2)	p-value
Interest in politics			.551	(3.9, 566.7)	.692
Very strong	10.4	9.7			
Strong	22.3	23.5			
Medium	41.6	40.7			
Weak	18.3	17.0			
Not at all	7.4	9.1			
<i>N</i>	827	627			
Voted last election			.819	(1, 147)	.367
No	18.7	20.6			
Yes	81.3	79.4			
<i>N</i>	825	626			
Size of personal network			.090	(2.9, 426.3)	.962
0	7.9	7.8			
1	13.4	13.4			
2	16.3	17.4			
3 or more persons	62.3	61.5			
<i>N</i>	792	590			
No. of association memberships			.053	(3.0, 435.3)	.983
0	43.3	43.8			
1	31.8	32.2			
2	14.8	14.1			
3 or more	10.1	9.9			
<i>N</i>	827	628			
Member of a party / a union			.174	(1, 147)	.678
No	84.0	83.2			
Yes	16.0	16.8			
<i>N</i>	827	628			

Empirische Forschungen an sozial- und politikwissenschaftlichen Lehrstühlen in Deutschland, der Schweiz und Österreich 2008 bis 2011

Empirical Research at Social and Political Science Departments in Germany, Austria and Switzerland from 2008 to 2011

Rolf Hackenbroch

Zusammenfassung

Die vorliegende Studie gibt einen Überblick über die in empirischen Forschungsprojekten an sozial- und politikwissenschaftlichen Lehrstühlen in den Jahren 2008 bis 2011 angewandten Methoden und deren Finanzierungsform. Anhand der Befragung von 41 Universitätsprofessoren wird aufgewiesen, dass Einstellungs- und Umfrageforschung – und hier vor allem repräsentative Bevölkerungsbefragungen – den weitaus größten Anteil der Forschungsprojekte stellen. Deren Finanzierung – dies im Unterschied zu nicht umfragebasierten Methoden – erfolgt vorwiegend über Drittmittel und die Durchführung wird zum überwiegenden Teil von externen Marktforschungsinstituten vorgenommen. Wie sich die Bedeutung von externen Instituten in der universitären Forschungslandschaft in Zukunft gestaltet, bleibt abzuwarten, wird doch die methodische Entwicklung in der empirischen Sozialforschung stark im Bereich der Online-Forschung gesehen.

Abstract

The present study provides an overview of the methods and the type of financing used in empirical research projects at social and political science departments in the years 2008 to 2011. Based on a survey of 41 chairs it can be shown that survey research – and especially representative surveys – provides the largest share of the research projects. Their funding – this in contrast to non-survey-based methods – primarily uses third-party funding and their implementation is done for the most part by external market research institutions. How the importance of such external institutions in the university research environment develops in the future remains to be seen, but the methodological development in empirical social research is greatly seen in the field of online research.



1 Ausgangspunkt und Zielsetzung¹

Von März bis Oktober 2011 führte die teleResearch GmbH aus Mannheim eine Erhebung zu Umfang, Finanzierung und Methodik empirischer Forschungsprojekte an sozial- und politikwissenschaftlichen Lehrstühlen durch. Die Untersuchung bezog sich auf Deutschland, Österreich und die Schweiz und hatte alle empirischen Primärerhebungen, die im Rahmen eines wissenschaftlichen Forschungsprojektes zwischen 2008 und 2011 durchgeführt wurden, zum Gegenstand. Erhebungen im Rahmen von Lehrveranstaltungen oder anderweitigen Projekten, die nicht in einen Forschungsbericht oder eine wissenschaftliche Publikation mündeten, waren nicht Gegenstand der Befragung.

Die Studie gibt einen Überblick über die in der universitären Forschung angewandten Methoden und Ansätze. Sie untersucht, welche Bedeutung Bevölkerungsumfragen im Vergleich zu Befragungen von Entscheidern in der Wirtschaft und der öffentlichen Verwaltung haben und ob es in Abhängigkeit von den Methoden verschiedene Finanzierungsformen und unterschiedliche Beteiligungen externer Meinungsforschungsinstitute gibt. Darüber hinaus werden die methodischen Schwerpunkte zukünftiger sozial- und politikwissenschaftlicher Forschungen aus Sicht der befragten Lehrstuhlinhaber dargestellt.

2 Methodik der Studie

Die Ausgangsstichprobe der Befragung bildeten 174 Adressen von Lehrstühlen für Sozialwissenschaften und Politikwissenschaften an deutschsprachigen Universitäten. Diese Adressen wurden in einem ersten Schritt danach qualifiziert, ob die jeweiligen Lehrstühle generell im Bereich empirischer Sozialforschung tätig sind. Es verblieben 93 Adressen, die das Ausgangssample für die vorliegende Erhebung darstellten. Von diesen Adressen verblieben nach Kontaktaufnahme und Qualifizierung der Zielperson 75 befragungsrelevante, verwertbare Adressen (= Grundgesamtheit). Von diesen konnten insgesamt 41 Universitätsprofessoren über ein ca. 12-15 minütiges telefonisches Interview zum Forschungsgegenstand befragt werden. Mit einer Quote von 55% haben wir in dieser schwer zu erreichenden Zielgruppe eine sehr gute Ausschöpfung erreicht.

1 Unser Dank für entscheidende Impulse bei der Durchführung der Befragung und der Fertigstellung der Auswertung gilt Herrn Gazmend Dika und Herrn Jens Schumm.

Abbildung 1 Ausschöpfungsübersicht

Ausgangsstichprobe (= Adressen insgesamt)	n=174	
<i>Ausgangssample Interviews (= verwendete Adressen von Lehrstühlen in D, A, CH</i>	n=93	100%
Davon nicht zur Grundgesamtheit gehörend:	n=18	19%
▪ falsche Zielgruppe	n=7	
▪ Befragungskriterien nicht erfüllt (keine wissenschaftlichen Forschungsprojekte in den letzten drei Jahren)	n=13	
<i>Grundgesamtheit (= abgeschlossene, verwertbare Adressen)</i>	n=75	100%
Davon Ausfälle gesamt:	n=34	45%
▪ verweigert / kein Interesse / keine Zeit	n=14	
▪ trotz bis zu 20-facher Kontaktversuche niemand erreichbar	n=20	
Realisierte Interviews	n=41	55%

* z.B. ZP an Uni nicht aufgeführt; falsches Institut; Hauptarbeitsplatz nicht am Lehrstuhl

3 Ergebnisse der Untersuchung

3.1 Art der Primärerhebungen an sozial-/ politikwissenschaftlichen Lehrstühlen 2008-2011

Fast alle Lehrstühle, die empirische Primärerhebungen in den Jahren 2008 bis 2011 durchführten, haben Einstellungs- und Umfragedaten erhoben (36 von 41). Bei zwei Dritteln dieser Lehrstühle wurden repräsentative Bevölkerungsbefragungen und bei ca. 40% spezifische Zielgruppen der Bevölkerung befragt. Auffallend ist insgesamt, dass Befragungen zu Zielgruppen in Wirtschaft und öffentlicher Verwaltung nur bei 4 der befragten 41 Lehrstühle stattfanden. Der Schwerpunkt der Erhebung von Einstellungs- und Umfragedaten liegt so eindeutig bei Bevölkerungsbefragungen und nicht in der Analyse von Entscheidern in Wirtschaft und öffentlicher Verwaltung.

Rund die Hälfte der befragten Universitätsprofessoren gab an, jenseits von Einstellungs- und Umfragedaten auch empirische Forschungsprojekte mittels anderer Methoden durchzuführen. Dabei wurde die Inhaltsanalyse als Erhebungsmethode vor Gruppendiskussion und Experiment (Feld-, Laborexperiment) genannt. Schaut man sich die absolute Anzahl der Studien an, so fällt die Vorrangstellung von Einstellungs- und Umfragedaten (insgesamt 149 Studien) vor anderen methodischen Ansätzen (56 Studien) sehr deutlich aus.

Besonders interessant ist, dass im Bereich von Einstellungs- und Umfragedaten telefonische, Online-, persönliche und auch schriftliche Befragungen gleichhäufig vertreten sind. Dominieren im Bereich der außeruniversitären Forschung eindeutig telefonische und Online-Befragungen, so spielen im universitären Bereich persönliche und schriftliche Befragungen eine große Rolle.

3.2 Finanzierung und Durchführung

Finanziert werden die Primärerhebungen von Einstellungs- und Umfragedaten größtenteils über Drittmittel von Forschungsförderungsinstitutionen (vier Fünftel aller Lehrstühle), nur ein geringerer Teil wird ausschließlich oder zusätzlich aus Eigenmitteln des Lehrstuhls/ des Instituts/ der Universität bzw. Drittmitteln anderer Auftraggeber gefördert.

Anders sieht das Bild bei der Finanzierung solcher Untersuchungen aus, die mittels Inhaltsanalyse, Gruppendiskussionen oder Experimenten durchgeführt werden. Diese erfolgt bei etwa der Hälfte der Lehrstühle über den eigenen Lehrstuhl/ Institut/ Universität und nur etwa zu 50% über Drittmittel von Forschungsförderungsinstitutionen bzw. Drittmittel anderer Auftraggeber. Die Finanzierung über eigene Mittel spielt also vor allem bei diesen Methoden eine herausragende Rolle; Einstellungs- und Umfragedaten werden hingegen schwerpunktmäßig über die Finanzierung durch Drittmittel von Forschungsförderungsinstitutionen sichergestellt.

3.3 Einsatz externer Marktforschungsinstitute

Eine der zentralen Fragen der vorliegenden Untersuchung war, in welchem Ausmaß externe Marktforschungsinstitute mit Primärerhebungen seitens der Lehrstühle betraut werden. Das Ergebnis ist überraschend eindeutig: Externe Marktforschungsinstitute werden fast ausschließlich bei der Erhebung von Einstellungs- und Umfragedaten beauftragt (23 von 36 Lehrstühlen beauftragten – teilweise neben der Erhebung über den eigenen Lehrstuhl – externe Institute), bei anderen Methoden wie Inhaltsanalyse, Experteninterviews und Gruppendiskussionen werden nur äußerst selten externe Institute beauftragt (3 von 22 Lehrstühlen). Als Gründe für die Beauftragung externer Markt- und Meinungsforschungsinstitute wird vor allem auf den Arbeitsaufwand und auf ökonomische Gründe verwiesen. Ein Viertel der befragten Lehrstühle nennen darüber hinaus die Kompetenz und Erfahrung externer Institute als Grund für das Outsourcing.

Die Motivation wiederum, Primärerhebungen am eigenen Lehrstuhl durchzuführen, liegt neben niedrigeren Kosten vor allem darin, dass Erhebungen am Lehrstuhl eine bessere Kontrolle der Erhebungen und eine bessere Handhabung der Gesamtstudie gewährleisten. Besonders bei Primärerhebungen über Inhaltsanalysen, Gruppendiskussionen und Experimente ist dies mit Abstand das am häufigsten genannte Argument.

Und nach welchen Kriterien fiel nun die Entscheidung für ein bestimmtes externes Marktforschungsinstitut? Hier steht die Reputation und Qualität des Instituts eindeutig an erster Stelle, gefolgt von Kostengründen und guten Erfahrungen. Dabei sind die Lehrstühle mit den Leistungen der Markt- bzw. Meinungsforschungsinstitute im Durchschnitt durchaus zufrieden (Mittelwert 2,1 auf einer Skala von 1 = sehr zufrieden bis 5 = überhaupt nicht zufrieden). Zur Zufriedenheit trug sowohl die Umsetzung der Studienanforderungen als auch die gute Zusammenarbeit bei. Kritik an den Leistungen der externen Institute wurde jedoch von rund der Hälfte der Lehrstühle formuliert. Diese bezog sich auf die Kommunikation, die Umsetzung der Methodik und auf die Ergebnisse an sich.

Im Vergleich zu den Erhebungen mit externen Instituten fällt die Zufriedenheit mit Erhebungen am eigenen Lehrstuhl bzw. der Universität jedoch deutlich höher aus (Mittelwert 1,5 bei Umfragedaten - 21 Lehrstühle; Mittelwert andere Methoden 1,6 - 18 Lehrstühle). Die geäußerte Zufriedenheit bezog sich hier vor allem auf die Ergebnisse selbst, aber auch auf die Kontrolle des Erhebungsprozesses und den Ablauf der Studie. Jedoch auch etwa die Hälfte der Lehrstühle äußerte Unzufriedenheit mit Erhebungen am eigenem Lehrstuhl bzw. der Universität. Diese bezog sich auf die Methodik und die Ergebnisse, aber auch auf die Qualität der Durchführung.

Wenn auch die Zufriedenheit mit den Erhebungen am eigenen Lehrstuhl – vor allem auf Grund der besseren Kontrolle über den Ablauf und den Prozess – größer ist, so zeigt sich doch auch hier, dass durchaus Mängel festgestellt wurden.

Trotzdem: Es erweist sich, dass bei der Forschungszusammenarbeit vor allem eine engere Verzahnung zwischen den Vorgaben und Vorstellungen der Lehrstühle und der Erhebung der Markt- und Meinungsforschungsinstitute notwendig ist.

4 Methodische Schwerpunkte zukünftiger Forschung

„Wo liegt in der empirischen Sozialforschung in Zukunft der Schwerpunkt bei der Anwendung der Methoden?“ Fast die Hälfte (18 Lehrstühle) nennen auf diese offen gestellte Frage die Online-Forschung als eindeutigen Schwerpunkt. Dazu gehören

Online-Panels, webbasierte Befragungen oder einfach nur Online-Befragungen. Weitere sechs Lehrstühle nennen die Kombination verschiedener Befragungs- und Forschungsmethoden als zukünftigen Schwerpunkt der Methodenausrichtung und weitere fünf Lehrstühle Panelstudien. Darüber hinaus gibt es noch eine Vielzahl an Einzelnennungen, jedoch neben den genannten Punkten keine dezidierten weiteren Schwerpunkte.

Das Resümee hieraus: Online-Forschung ist auch in der universitären Forschung als Methode bereits fest etabliert. Wie jedoch weiter oben gezeigt wurde, ist die derzeitige Situation bei Primärerhebungen in der wissenschaftlichen Forschung ganz stark durch das gleichzeitige Auftreten von schriftlichen, telefonischen, persönlich-mündlichen und Online-Befragungen gekennzeichnet. Dies findet seine Begründung nicht zuletzt darin, dass es sich bei einem Großteil der Studien um repräsentative Bevölkerungsbefragungen handelt.

Insofern ist es interessant zu sehen, dass der übergroße Teil (34 Lehrstühle) repräsentativen Bevölkerungsumfragen in Zukunft eine gleichbleibende oder sogar noch zunehmende Rolle zuspricht. Begründet wird dies mit der gesellschaftlichen Relevanz dieser Art von Querschnittserhebungen. Nur eine Minderheit von fünf Lehrstuhlinhabern geht hingegen von einer in Zukunft sinkenden Bedeutung aus. Deren Begründung liegt im zunehmenden Nonresponse-Problem und der damit verbundenen erschwerten Durchführbarkeit solcher Studien.

Trotz der starken Stellung, die der Online-Forschung perspektivisch zugeschrieben wird, werden repräsentative Bevölkerungsbefragungen - ob telefonisch, schriftlich oder face-to-face - weiterhin ihre starke Bedeutung behalten. Wie sich deren Stellenwert genau entwickelt, wird spannend bleiben zu beobachten.

Anschrift des Autors

Rolf Hackenbroch
teleResearch GmbH
D6, 5
68159 Mannheim
E-Mail: hackenbroch@tele-research.de

Rezensionen



UDO KUCKARTZ, 2012: Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. Weinheim und Basel: Beltz Juventa. ISBN 978-3-7799-2815-7 – 188 Seiten, EUR 14,95.

In der sozialwissenschaftlichen Literatur finden sich wenige praxisorientierte Anleitungen zur qualitativen Auswertung von Daten. Das Buch von Udo Kuckartz will diese Lücke schließen. Der Autor verspricht eine „anwendungsbezogene Anleitung zur inhaltsanalytischen Auswertung qualitativer Daten“ (S. 5). Sein Analyse-Paradigma ist dabei die „methodische Strenge“ – auch in der qualitativen Sozialforschung – d. h. eine möglichst genaue Beschreibung des Vorgehens und die Anerkennung der Existenz von Gütekriterien (S. 6). Die Publikation soll eine möglichst nachvollziehbare Beschreibung der Vorgehensweise bei qualitativen Inhaltsanalysen am Beispiel von leitfadenorientierten Interviews bereitstellen. Es will „Wege aufzeigen, wie man qualitative Daten methodisch kontrolliert und in systematischer Weise auswerten kann“ (S. 20). Kuckartz definiert in seinem Buch qualitative Daten als alle nicht numerische Daten: Textdaten, aber auch Bilder, Filme, Audio-Aufzeichnungen etc. gehören dazu. Schwerpunkt des Buches sind aber Texte und hier speziell transkribierte Interviews, obwohl der Autor betont, dass die Methode auch z. B. auf Audio-Material, Bilder etc. anwendbar wäre. Die Zielgruppe des Buches sind dabei alle, die das erste Mal eine qua-

litative Inhaltsanalyse durchführen wollen, z. B. DiplomantInnen, Masterstudierende oder DoktorantInnen.

Kapitel 1 beginnt mit einer generellen Diskussion darüber, was qualitative Daten sind und was deren Analyse bedeutet, da es in der Wissenschaft kein einheitliches Verständnis darüber gibt. Kuckartz beschreibt dabei Unterschiede und Gemeinsamkeiten der qualitativen sowie der quantitativen Datenanalyse. Desweiteren weist er darauf hin, dass es eine große Vielfalt von qualitativen Analyseansätzen gibt, die zudem sehr unterschiedlich gruppiert werden. Am Ende des Kapitels geht der Autor auf die aus seiner Sicht zentralen Punkte jeder (erfolgreichen) qualitativen Inhaltsanalyse ein: Zum einen hebt er hier die zentrale Stellung der Forschungsfrage hervor, die jedes weitere Vorgehen bestimmt. Zum anderen betont er die Notwendigkeit methodischer Strenge, z. B. wegen der besseren Nachvollziehbarkeit und des größeren Vertrauens in die Forschenden sowie deren Ergebnisse, wenn bestimmte Standards eingehalten werden.

Das zweite Kapitel befasst sich mit der Entwicklung der Inhaltsanalyse. Erste Ansätze der Inhaltsanalyse fanden sich schon sehr früh in der Literatur, beispielsweise 1910 bei Max Weber. Zunächst wurden in der klassischen Inhaltsanalyse manifeste Inhalte erfasst und die Resultate eher zur Quantifizierung genutzt. Schon 1952 führte aber Kracauer¹ den Begriff der „qualitative content analysis“ ein, die neben manifesten Inhalten auch latente Bedeutungen, auf die man sich intersubjektiv verständigen kann, erfassen sollte. Nach dieser Einführung zur Entwicklung der qualitativen Inhaltsanalyse wird im weiteren Verlauf des Kapitels

1 Kracauer, S. (1952). The challenge of qualitative content analysis. *Public Opinion Quarterly*, 16, 631-642.

erläutert, wie Textverstehen und -interpretieren im Sinne der klassischen Hermeneutik erfolgen kann. Danach leitet Kuckartz fünf grundlegende Handlungsregeln aus der Hermeneutik für die Durchführung einer qualitativen Inhaltsanalyse ab. Im Anschluss daran wird die Weiterentwicklung der qualitativen Inhaltsanalyse, die von Mayring² das erste Mal systematisch als Methode beschrieben wird, aufgezeigt, nicht ohne auf die verschiedenen Varianten der qualitativen Inhaltsanalyse zu verweisen. Im Schlusspunkt des Kapitels werden zur Unterscheidung von anderen qualitativen Analyseverfahren sechs Punkte hervorgehoben: die Zentralität der Kategorien, die systematische regelbasierte Vorgehensweise, die Klassifizierung des gesamten Materials, Einsatz von Techniken der Kategorienbildung am Material, die von der Hermeneutik inspirierte Reflexion des Materials und die Anwendung von Gütekriterien.

Nach dieser eher theoretischen Einordnung der Methode in den ersten beiden Kapiteln beginnt in Kapitel 3 die Beschreibung des praktischen Vorgehens bei einer qualitativen Inhaltsanalyse. Zunächst werden zentrale Begriffe einer Inhaltsanalyse wie „Kategorien“, „Einheiten“ und „Codierer“ erläutert. Der Autor unterscheidet dabei verschiedene Arten von Kategorien wie Faktenkategorien, inhaltliche, analytische, natürliche, evaluative und formale Kategorien. Er geht dabei auch auf Konzepte, Codes und Variablen ein. Danach wird der Ablauf einer quantitativen und einer qualitativen Inhaltsanalyse gegenübergestellt. Der entscheidende Teil des Kapitels sind allerdings die ersten Schritte einer Inhaltsanalyse, die der Autor als initiiierende Textarbeit beschreibt, bei der das Textmaterial gesichtet, Memos geschrieben und erste Fallzusammenfassungen erstellt werden. Daran anschließend folgt ein für alle For-

men der qualitativen Inhaltsanalyse sehr bedeutender Absatz zur deduktiven Kategorienentwicklung ohne empirisches Material aus der Theorie oder Hypothese heraus und zur induktiven Kategorienbildung am Material. Es werden Mischformen genauso diskutiert wie die Frage, welche Art der Kategorienbildung sich wann eignet.

Nach der Einführung in die Kategorienentwicklung unterscheidet der Autor in Kapitel 4 drei grundlegende Methoden der Inhaltsanalyse: 1) die inhaltlich strukturierende qualitative Inhaltsanalyse, die schwerpunktmäßig mit thematischen Kategorien arbeitet; 2) die evaluative qualitative Inhaltsanalyse, die Bewertungskategorien verwendet und 3) die typenbildende Inhaltsanalyse, in deren Mittelpunkt die Erstellung einer Typologie steht. Zunächst beginnt das Kapitel mit Gemeinsamkeiten und Unterschieden der drei Ansätze. Folgend wird für jede der drei Methoden zunächst in einem Schaubild der Ablauf einer Auswertung in verschiedenen Phasen dargestellt. Danach wird jede Phase des jeweiligen Verfahrens detailliert beschrieben – jeweils belegt mit Beispielen, um das Herangehen an die Texte und deren Analyse nachvollziehbar zu machen. Bei der Beschreibung der typenbildenden Inhaltsanalyse wird vor Beschreibung der Phasen grundsätzlich erläutert, was Typenbildung in der Sozialforschung allgemein bedeutet und wie eine Typenbildung abläuft (von der Bestimmung des Merkmalsraums über die Bildung einer Typologie, der Beschreibung der Typen, der Zuordnung der Fälle zu den Typen bis hin zu einer Zusammenhangsanalyse). Zudem legt Kuckartz bei allen drei Formen der Inhaltsanalyse Wert auf eine ausführliche Beschreibung der letzten Phase im Ablauf einer Inhaltsanalyse – die Ergebnisdarstellung, die häufig in der Literatur zu kurz kommt. Weiterhin schlägt er vor, wie die verschiedenen Ansätze sinnvoll kombiniert werden können bzw. wie sie sich gegenseitig ergänzen.

Im Anschluss an die detaillierte Beschreibung, wie eine Inhaltsanalyse durchge-

2 Mayring, P. (2010). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (11. Aufl.). Weinheim: Beltz.

führt werden sollte, folgt in Kapitel 5 eine Schilderung, wie bei einer qualitativen Inhaltsanalyse Softwareunterstützung eingesetzt werden kann. Das Kapitel gliedert sich in drei unterschiedliche Teile. Zunächst beschreibt der Autor das Vorgehen bei der Transkription von Texten. Er zeigt auf, was bei dieser zu beachten ist, um erfolgreich auszuwerten. Hierbei befasst er sich z. B. damit, was überhaupt abgeschrieben werden muss oder wie abgeschrieben werden sollte. Andere Fragen betreffen das Formulieren von Transkriptionsregeln und die Dringlichkeit das Textmaterial zu anonymisieren. Im zweiten Teil geht es um die Vorbereitung der eigentlichen Codierung in einem Softwareprogramm für die qualitative Datenanalyse. Dabei werden das Strukturieren des Textmaterials zur Codierung, z. B. durch das Festlegen von Codier-einheiten oder die Möglichkeiten des Arbeitens im Team diskutiert. Die dritte Passage beschäftigt sich dann mit der Kategorienbildung, dem Codieren und dem Durchführen einer Inhaltsanalyse mit Softwareunterstützung. Zusätzliche Möglichkeiten wie Multimedia-Unterstützung, Visualisierung, Text- und Hyperlinks werden genauso angesprochen wie Worthäufigkeitslisten, Keyword-in-Context-Listen, dictionärbasierte Wortsuche und Codiertechniken aus der computerunterstützten Inhaltsanalyse. Hier wird zwar im Wesentlichen auf Funktionen des Programms MAXQDA eingegangen, aber das Vorgehen lässt sich problemlos auf andere Programme übertragen. Einzig die Erstellung des Worthäufigkeits-, Keyword-in-Context-Listen, die dictionärbasierte Wortsuche und Codiertechniken aus der computerunterstützten Inhaltsanalyse sind MAXQDA-spezifisch oder finden sich sonst eher in Programmen zur quantitativen Inhaltsanalyse.

Das sechste, eher kurze Kapitel widmet sich den in der qualitativen Forschung durchaus kontrovers diskutierten Gütekriterien. Kuckartz unterscheidet dabei zwei Arten von Gütekriterien: die interne Studiengüte und

externe Gütekriterien. Zu ersterer gehören Zuverlässigkeit, Glaubwürdigkeit und Verlässlichkeit. Eine kurze Checkliste soll helfen, dies sicherzustellen. Zu den externen Gütekriterien gehören Übertragbarkeit und Verallgemeinerbarkeit der Ergebnisse (geeignetes Sampling und Validierungsmöglichkeiten wie Peer Briefing, Member checking, längerer Aufenthalt im Feld, Triangulation). Einen zweiten Schwerpunkt des Kapitels bildet die Erstellung von Forschungsberichten und die Dokumentation der Arbeit. Hier gibt der Autor sehr allgemeine Hinweise, in welcher Form das Vorgehen und die Ergebnisse berichtet werden sollten. Diese Hinweise sind aber so gehalten, dass sie im Grunde für jede Art der Forschung – qualitativ oder quantitativ gelten können.

Kapitel 7 umfasst nur etwas mehr als eine Seite und stellt die Stärken der qualitativen Inhaltsanalyse dar. Den Abschluss des Buches bilden eine Übersicht über einige regelmäßig stattfindende (Methoden-) Konferenzen und Workshops zur qualitativen Inhaltsanalyse sowie einige Links zum Thema.

Zu Beginn des Buches weist der Autor darauf hin, dass die Kapitel aufeinander aufbauen und bewusst keine Hyperlink-Struktur haben, d. h. sie sollten nacheinander gelesen werden. Diese Empfehlung kann für das Buch in der Tat gegeben werden: Es ist als Nachschlagewerk, um schnell etwas zu finden, eher nicht geeignet, dafür aber besonders gut als Leitfaden für das Vorgehen bei der eigenen Analyse. Jedes Kapitel beginnt mit einer Übersicht zum Inhalt des jeweiligen Textabschnitts. Diese Gliederung verhilft zu einem guten Überblick, was im jeweiligen Kapitel zu erwarten ist.

„Qualitative Inhaltsanalyse. Methoden, Praxis und Computerunterstützung“ hält, was der Titel verspricht. Es ist ein empfehlenswertes Lehrbuch für alle, die zum ersten Mal qualitativ auswerten wollen und bisher wenig oder keine Erfahrung mit der Methode haben. Um es vorweg zu neh-

men: Obwohl das Buch im Titel die Computerunterstützung betont, ist es auch für alle diejenigen geeignet, die keine spezielle Software einsetzen wollen. Computerunterstützung findet nur in einem Kapitel des Buches ihren Niederschlag. Alle anderen Beschreibungen sind völlig unabhängig vom Softwareeinsatz. Zudem findet sich im Teil zur Computerunterstützung ein Absatz zur Transkription, der auch für diejenigen von Interesse sein dürfte, die auf den Einsatz einer speziellen qualitativen Analyse-Software verzichten, aber die Texte zur weiteren Analyse transkribieren wollen. Das Buch ist klar gegliedert, leicht nachvollziehbar geschrieben und viele Beispiele veranschaulichen das Vorgehen.

In den anwendungsbezogenen Kapiteln 3 und 4 werden das Vorgehen bei einer qualitativen Inhaltsanalyse und die verschiedenen Phasen der Auswertung gut verständlich dargestellt. Alle Schritte werden mit Beispielen aus einer Studie zur Wahrnehmung des Klimawandels verdeutlicht. Diese Beispiele eignen sich gut, um die Anwendung einer Inhaltsanalyse zu verstehen und die Umsetzung für die eigenen Texte zu erkennen. Auch die Einordnungen der drei verschiedenen Ansätze der Inhaltsanalyse und die Beschreibung ihrer Kombinationsmöglichkeiten sind eine gute Hilfestellung.

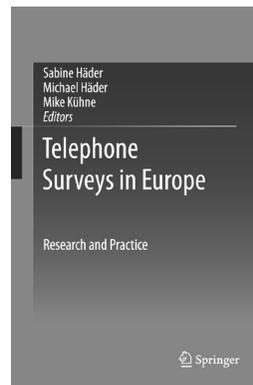
Die Kapitel 1 und 2 sind aus meiner Sicht zwar nicht unwichtig, aber für ein Praxisbuch zu ausführlich ausgefallen. Dagegen könnte Kapitel 6 zu den Gütekriterien und der Forschungsdokumentation durchaus detaillierter sein, denn gerade bei der Validierung und Darstellung der Ergebnisse zeigt sich bei der Zielgruppe des Buchs häufig eine große Unsicherheit. Die auf einer der letzten Seiten unter Ressourcen aufgeführte Liste mit Links ist nicht besonders hilfreich, da sie nur relativ wenige Links aufgreift und einige davon schon seit längerem nicht mehr aktualisiert werden, wie der Autor selbst schreibt. Dagegen fehlen Links zu den drei am häufigsten genutzten

Programmen Atlas.ti, MaxQda und Nvivo, die ich mir an dieser Stelle wünschen würde.

Trotz dieser kleinen Schwächen ist „Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung“ von Udo Kuckartz ein sehr empfehlenswertes Einführungsbuch, das sich sowohl zum Selbststudium als auch zum Einsatz in der Methodenausbildung sehr gut eignet und hoffentlich dazu beiträgt, dass zukünftig viele Unsicherheiten bei der Anwendung der qualitativen Inhaltsanalyse vermieden werden.

CORNELIA ZÜLL

DOI: 10.12758/mda.2013.006



SABINE HÄDER,
MICHAEL HÄDER, MIKE
KÜHNE (Eds.), 2012:
Telephone Surveys
in Europe: Research
and Practice.
Berlin, Heidelberg:
Springer. ISBN 978-
3-642-25410-9.
324 Seiten. 139€

Mit dem Buch „Telephone Surveys in Europe. Research and Practice“ haben dessen Herausgeber, Sabine Häder (GESIS, Mannheim), Michael Häder (TU Dresden) und Mike Kühne (TU Dresden) ein Thema aufgegriffen, das angesichts der gesellschaftlichen und technologischen Entwicklung schon längst einer Behandlung bedurft hätte. Der vorliegende Sammelband, der sich in fünf Kapitel untergliedert, widmet sich aus unterschiedlichen Perspektiven – aus Forschung und Praxis – den aktuellen Herausforderungen für die Telefonumfrageforschung in Europa. Die Autoren, welche aus der akademischen

und kommerziellen Sozialforschung sowie aus der Sicht der amtlichen Institutionen berichten, behandeln folgende zentrale Bereiche, welche sich zugleich in den Kapiteln des Buches widerspiegeln: Auswahlrahmen, Gewichtung und Nicht-Teilnahme, sowie Datenqualität.

Im ersten Kapitel des Buches „Spotlights on Selected European Countries“ wird aus Sicht der jeweiligen Länder die dortige Situation exemplarisch dargestellt. Eröffnet wird die Länderreihe von Anna Andreenkova mit einem Beitrag über Russland. In der einleitenden historischen Skizze weist die Autorin zunächst auf die Bedeutung der politischen „Gorbachev era“ und auf den kulturelle Kontext sowie deren Einfluss auf die Umfrageforschung allgemein und auf die Praxis zu Telefonbefragungen speziell hin. Nationale Umfragen werden in Russland vorwiegend als persönlich-mündliche Befragungen durchgeführt, da die Ausstattung der Privathaushalte mit Telefon in der Fläche nicht ausreichend ist. Der Versorgungsgrad (einschließlich Mobilfunkanschlüssen) ist u. a. durch ein deutliches Stadt-Land-Gefälle gekennzeichnet. Gleichwohl gehören Telefonumfragen in Russland zum Alltag. Beschränkt ist dies allerdings meist auf kommunale bzw. städtische Telefonumfragen. Bei der Auswahl von Zielpersonen im Haushalt und auch bei der Teilnahme an Telefonumfragen (Selection- und Nonresponse Effect) werden Besonderheiten aufgezeigt, wie sie auch für andere postkommunistische Übergangsgesellschaften nicht unüblich sind.

Der Bericht über die Situation in den Niederlanden von Dirkjan Beukenhorst stellt aus der Sicht des Amtes für Statistik die wesentlichen Nachteile der Telefonumfrage, anhand ausgewählter niederländischer Arbeiten, dar. Im Mittelpunkt der Diskussion steht dabei die Stichprobenziehung. Wengleich in den Niederlanden nahezu alle Privathaushalte über einen Telefonanschluss (incl. der sogenannten Mobile-only-Haushalte) verfügen, kommt

eine Stichprobenziehung auf der Basis von Telefonnummern nicht in Betracht. Auch das nationale Adressregister wird mangels Nummernabdeckung als Ausgangsbasis für eine anschließende Zuweisung von Telefonnummern verworfen, da lediglich für zwei Drittel der Adressen eine Telefonnummer ermittelt werden kann. Abschließend werden auch die Teilnahmequote und die Datenqualität (insbesondere bei Mobilfunkbefragungen) kritisch bewertet – u. a. auf Grund von 'on the move'. Im Ergebnis hält der Autor den Einsatz von Telefonumfragen, zumindest gegenwärtig, für nicht praktikabel.

Ganz anders wird das Instrument der Telefonumfrage aus Sicht der Amtlichen Statistik in der Schweiz gesehen. Dort werden Telefonumfragen, wie Michèle Ernst Stähli berichtet, seit den 80er Jahren durchgeführt. Der hohe Versorgungsgrad, die lückenlose Eintragung ins Telefonbuch und die hohe Teilnahmebereitschaft ließ die Telefonumfrage für das Statistische Bundesamt zum wichtigen Erhebungsverfahren werden. Gesellschaftliche Veränderungen wie z. B. steigende Nichteintragung ins Telefonbuch und die Verbreitung von Mobilfunkanschlüssen sowie politische Entwicklungen (Liberalisierung des Telefonmarktes) führten jedoch zu methodisch unbefriedigenden Ergebnissen. So wirkte der zunehmende Einsatz der RDD-Auswahlverfahren dem sog. Coverage Error entgegen. Nicht gelöst ist das Problem einer zunehmenden Teilnahmemüdigkeit. Sie fällt aber in der Gruppe derjenigen Personen deutlich geringer aus, die zuvor mit einem Schreiben über die anstehende Untersuchung informiert wurde. Das im Aufbau befindliche Zentralregister liefert zudem eine sinnvolle Ergänzung zur Reduzierung des Coverage Error und zur solideren Analyse des Nonresponse Errors. Inwieweit dieses Zentralregister nicht nur dem Statistischen Bundesamt zur Nutzung vorbehalten ist, sondern auch der akademischen Forschung zur Verfügung stehen wird, bleibt abzuwarten.

Vesa Kuusela und Matti Simpanen zeichnen für Finland die Entwicklung und Verbreitung der Telefonanschlüsse für die Zeit von 1998 bis 2009 nach. Insbesondere heben sie dabei die rasante Entwicklung der Mobile-only-Haushalte hervor. Grundlage der Stichprobenziehung ist das nationale Adressregister mit anschließender Ergänzung der Telefonnummern u. a. aus dem öffentlich zugänglichen Telefonregister. Durch ergänzende Recherchen können für über 95% aller Adressen die relevanten Telefonnummern ermittelt werden. Mögliche Störfaktoren, wie sie durch außerhäusliche Bedingungen bei Mobilfunkinterviews zu erwarten sind, scheinen weniger relevant, da nahezu 80% aller Telefoninterviews im häuslichen Kontext durchgeführt werden. Ergänzend ist noch auf die hohe Teilnahmequote von 60-70% zu verweisen. Sie wird erreicht, obwohl mehr als 90 % der Telefoninterviews via Mobiltelefon durchgeführt werden.

Den Überblick über Großbritannien gibt Elizabeth Nelson aus der Perspektive der kommerziellen Umfrageforschung. Sie berichtet zunächst über die allgemeine Entwicklung der Umfrageforschung. Dabei spannt sie perspektivisch den Bogen von der klassischen Face-to-face-Umfrage bis hin zur zukünftigen Entwicklung (z. B. Mobile Internet Research). Implikationen und Veränderungen, die sich bei der Telefonumfrage ergeben haben, werden anhand ausgewählter Arbeiten vorgestellt. Folgende Aspekte stehen dabei im Vordergrund: Auswahlrahmen (Telefonverzeichnis vs. RDD), Auswahl der Zielperson im Haushalt und Teilnehmerate. Der Beitrag schließt mit dem Ausblick auf die neueren Entwicklungen der Informations- und Kommunikationstechnik und deren Auswirkungen auf die Marktforschung.

Obwohl sich auch in Italien mit der Liberalisierung des Telefonmarktes und mit der Entwicklung im Bereich des Mobilfunks deutliche Veränderungen vollzogen haben (u. a. bezüglich der Abdeckung und Vollständigkeit der Angaben im Telefonbuch),

dient das Telefonbuch dort als Grundlage für die Stichprobenziehung, wie Teresio Poggio und Mario Callegaro berichten. Die Durchsicht einer Datenbank (einschließlich akademischer Studien) und die Auskunft der zentralen Vereinigung der Markt- und Sozialforscher hatten ergeben, dass nach wie vor überwiegend das Telefonbuch oder aber das Quotenverfahren zur Anwendung kommen. Zufällige Auswahlverfahren auf der Basis der Festnetz- und/oder Mobilfunknummern gibt es kaum. Die Autoren sehen drei Möglichkeiten, zukünftig mit dieser Situation umzugehen: Adjustierung durch Gewichtung, der Einsatz eines RDD-Verfahrens als Dual Frame und ein Mixed-Mode-Ansatz wobei in Letzterem ein breiter Konsens (unter der akademischen und kommerziellen Sozialforschung sowie den statistischen Ämtern) gesehen wird.

Auch in Portugal, so berichten Paula Vicente und Elizabeth Reis, werden Telefonumfragen seit den 80er Jahren durchgeführt. Sie beschränken sich jedoch überwiegend auf die Marktforschung und deren Kundenbefragungen. Bevölkerungsumfragen werden seit Mitte der 90er Jahre auf Grund einer Versorgungsquote der Privathaushalte mit Telefon in Höhe von ca. 80% und auf Grund der einsetzenden CATI-Technologie zum dominierenden Erhebungsverfahren der Markt- und Sozialforschung. Die Entwicklungen in der Informations- und Kommunikationstechnik wie auch die allgemeinen gesellschaftlichen Veränderungen werfen auch für Portugal Probleme auf. Es handelt sich um vergleichbare Probleme, wie diese zuvor aus den anderen Ländern berichtet wurden: Auswahlrahmen, Teilnehmerate und Messung. Auch in Portugal gibt es einen Rückgang der Festnetzanschlüsse und eine Zunahme der Mobile-only-Haushalte. Dementsprechend wurden die Telefonverzeichnisse in den vergangenen Jahren zunehmend unvollständiger. Klassische Telefonumfragen weisen daher eine entsprechende Verschiebung in der sozialen Zusammensetzung auf. Entsprechende

Zufallsauswahlverfahren, welche die Festnetz- und die Mobilfunknummern berücksichtigen, kommen in Portugal nicht zur Anwendung bzw. werden nicht berichtet. Ergänzend zur Auswahlproblematik werden Aspekte des Erhebungsmodus thematisiert. So zeigen sich keine Unterschiede zwischen Festnetz- und Mobilfunkinterviews im Hinblick auf die Kooperationsrate, jedoch scheint es geringe Unterschiede bezüglich der Interviewdauer und der Vollständigkeit von Interviews zu geben. Unklar bleibt, welche Faktoren für die genannten Unterschiede bestimmt werden können.

Im zweiten Kapitel des Sammelbandes „Sampling for Telephone Surveys“ werden aktuelle Forschungsarbeiten präsentiert, die sich mit dem Problem der Abdeckung (der Untererfassung) und aktuellen bzw. neueren Konzepten zu Auswahlrahmen beschäftigen. Dominique Joye, Alexandre Pollien, Marlène Sapin und Michèle Ernst Stähli thematisieren die Problematik Abdeckung (Coverage) und Ausfälle (Nonresponse) im Rahmen von Sekundäranalysen des ESS08 und des EVS08. Veranschaulicht wird zunächst, welche Konsequenzen es hätte, wenn auf der Basis des Telefonbuches eine Stichprobe gezogen würde. Es wird dargestellt wie groß die Unterversorgung ist bzw. wie viele Haushalte keinen Telefonanschluss haben. Im Anschluss daran wird gezeigt, wie groß der Anteil der Personen (Haushalte) ist, die sich nicht mehr im Telefonbuch verzeichnen lassen. In bivariaten Analysen wird ergänzend verdeutlicht, dass sich die Telefonbesitzer in ihren politischen Einstellungen von den Nicht-Besitzern unterscheiden. Differenzen, wie sie auch anhand der Analysen mit dem EVS08 für die gelisteten und nicht-gelisteten („Non-Pubs“) Haushalte berichtet werden. Unklar ist bisher, inwiefern Merkmale oder Merkmalsmuster des o.g. Noncoverage Ähnlichkeiten mit Personen aufweisen die an Umfragen nicht teilnehmen. Wenn dem so wäre, könnten sich die Effekte verstärken. Die berichteten Ergebnisse, die im Rahmen der

Non-Response-Studie zum EVS08 durchgeführt wurde, zeigen jedoch keine klaren Beziehungen. Abschließend behandeln die Autoren Gewichtungspeduren zur Adjustierung o.g. Effekte. Keines der eingesetzten Gewichtungsverfahren vermochte es allerdings, die Unterschiede, wie sie auf der Ebene der sozialen Zusammensetzung und auf der Ebene der politischen Einstellungen berichtet wurden, zu egalisieren.

Christiane Heckel und Kathrin Wiese behandeln Auswahlrahmen von Festnetz- und Mobilfunknummern sowie die Möglichkeit, diese auf europäischer Ebene zu realisieren. Den Ausgangspunkt der Betrachtung bzw. die Grundlage bildet für die Autoren das sog. Häder-Gabler-Modell. Analysen mit dem Eurobarometer und Angaben von Eurostat verdeutlichen zunächst die unterschiedlichen Entwicklungen in den jeweiligen europäischen Ländern und der damit verbundenen Problematik. Anhand ausgewählter Länder (Deutschland, Frankreich, Großbritannien, Italien und Spanien) – von den Autoren als „Top 5“ bezeichnet – wird die Segmentierung der Telefonmärkte hervorgehoben und zugleich ein proportionales Zufallsverfahren auf der Basis der Netzbetreiber als mögliches Auswahlverfahren aufgezeigt.

Götz Schneiderat und Tino Schlinzig stellen die Realisierung von Telefonumfragen auf der Grundlage von Mobilfunk- und Festnetznummern (Dual-Frame) in den Vordergrund ihres Beitrags. Die Grundlage bildet das Projekt CELLA2 (CEL=CELL phone, LA=Landline). Die Autoren behandeln in diesem Zusammenhang Fragen der Ausschöpfungsquote und Aspekte der realisierten Stichprobe – insbesondere die der sozialen Zusammensetzung (einschließlich auf der Ebene relevanter Subgruppen). Auch wird aufgezeigt, dass sich die „Mobile only's“ nicht nur durch demographische und sozioökonomische Merkmale, sondern auch bezüglich einiger Persönlichkeitseigenschaften (Big Five) unterscheiden. Abschließend widmen sich die Autoren dem Stel-

lenwert der Kontakthäufigkeit und deren Einfluss auf die Datenqualität. Insbesondere vor dem Hintergrund der mode-bedingten (Mobil- vs. Festnetz) unterschiedlichen Erreichbarkeit und der unterschiedlichen Kontaktsituationen (Vorabinformation via SMS) stellt dies ein Novum dar.

Unter „Weighting and Nonresponse“, dem dritten Kapitel des Bandes, werden Arbeiten vorgestellt, die sich der Ausfall- wie auch der Auswahlproblematik widmen. So behandeln die Arbeiten die Möglichkeiten der Adjustierung von Ausfällen wie auch die Berücksichtigung von Inklusionswahrscheinlichkeiten und deren Gewichtung, im Rahmen des dual-frame Ansatzes. Siegfried Gabler, Sabine Häder, Iris Lehnhoff und Elisabeth Mardian stellen in ihrem Beitrag Überlegungen und Vorgehensweisen zur Adjustierung von ungleichen Inklusionswahrscheinlichkeiten und Ausfällen, im Zusammenhang einer Festnetz- und Mobilfunkbefragung, vor. Zunächst werden einige ausgewählte Parameter zum Telefon behandelt, welche anschließend in modifizierte Schätzer – Horvitz-Thompson (HT) und der allgemeine Regressionschätzer (GREG) – überführt werden. Bezüglich der Ausfallproblematik thematisieren die Autoren ein Konzept zur Messung von Prokrastination. Anhand von Interviewabbrüchen – Grundlage der Analysen bilden die Daten aus dem „Cella 2“ Projekt – wird dargestellt, inwiefern Prokrastination ein brauchbarer Proxy für die Kooperation darstellt. Abschließend wird gezeigt, welche Adjustierungsverfahren vergleichbare Verteilungen wie die im Mikrozensus für Geschlecht, Schulabschluss, Haushaltsgröße und Familienstand erzielen.

In dem Beitrag von Fannie Cobben, Barry Schouten und Jelke Bethlehem werden ein Modewechsel von CAPI auf CATI-Befragungen sowie dessen Konsequenzen auf die Datenqualität thematisiert. Grundlage der Arbeit bildet eine repräsentative computerbasierte mündliche Befragung (CAPI), welche auf einer Zufallsauswahl

aus dem Melderegister basiert. In ihrer Untersuchung behandeln die Autoren insbesondere die Problematik Noncoverage und Non-response. Anhand verschiedener Gewichtungsverfahren, deren Verwendung die genannten Verzerrungen (coverage und non-response) adjustieren sollen, werden deren Leistungsfähigkeiten dargestellt. Dies geschieht anhand ausgewählter relevanter Merkmale (Schulbildung, Religionszugehörigkeit und Erwerbstätigkeit). Die besten Resultate werden von einem kombinierten Verfahren, einem allgemeinen Regressionschätzers (GREG) und einer coverage propensity stratification-Gewichtung, berichtet. Entsprechend dem berichteten o. g. positiven Ergebnis sehen die Autoren Potential für Mixed mode designs.

In der Schweiz bildet seit einigen Jahren ein nationales Personenregister die Auswahlgrundlage für Bevölkerungsumfragen. Oliver Lipps und Kathrin Kissau thematisieren in ihrem Beitrag, ob sich dieser Auswahlrahmen auch für Telefonumfragen fruchtbar anwenden lässt. Zur Vorbereitung einer kantonalen Telefonumfrage (Kanton Luzern), werden unterschiedliche Quellen und Strategien dargestellt, anhand derer Telefonnummern ermittelt und den Personenangaben zugewiesen werden können. Der Einsatz der verschiedenen Quellen und Strategien erbrachte eine Abdeckung von ca. 85%. Welche Personengruppen mit welchen Quellen und Strategien zur sukzessiven Verbesserung der sozialen Zusammensetzung des Gesamtsamples führen, verdeutlicht eine Untergliederung nach sozialen Merkmalen (Alter, Geschlecht, Familienstand, Nationalität und Ortsgröße).

Im vierten Kapitel „Data Quality“, dem vorletzten Teil des Bandes, wird der Frage nachgegangen, inwiefern unterschiedliche Erhebungsmodi (Mobil vs. Festnetz) einen Einfluss auf die Datenqualität haben. Die Angaben der Befragten bzw. die Datenqualität wird durch eine Vielzahl von Faktoren beeinflusst. Peter Lynn und Olena Kaminska stellen in einer konzeptionellen Arbeit

die Faktoren dar, die sich insbesondere im Rahmen der Befragung via Mobilfunk bzw. Handy auf die Angaben der Befragten auswirken können. Zunächst werden – basierend auf der allgemeinen Literatur – die jeweiligen Ebenen des Interviews dargestellt. Die Autoren differenzieren nach Interviewsituation, Eigenschaften des Befragten und Eigenschaften der Umfrage (z. B. Design der Studie). Des Weiteren thematisieren sie, welche Faktoren sich wie bzw. in welcher Richtung und mit welcher Stärke auf die Datenqualität auswirken können. Bisher wurden die Eigenschaften, die sich im Rahmen einer Mobilfunkbefragung von denen der herkömmlichen Festnetzbefragung unterscheiden, kaum näher untersucht. So zeichnet sich eine Befragung via Handy durch spezifische situationale, technische und psychologische Eigenschaften aus. Diese Eigenschaften und deren Einfluss auf die Datenqualität näher zu bestimmen, heben die Autoren hervor und formulieren dies als Aufgabe für die künftige Forschung.

Im Vordergrund des Beitrags von Mike Kühne und Michael Häder steht die Frage, ob unterschiedliche Bedingungen der Datenerhebung (Festnetz vs. Mobilfunk) einen Einfluss auf die Qualität der Angaben haben. Folgende Indikatoren zur Messung der Antwortqualität werden bei der Analyse berücksichtigt: Item-Nonresponse, Response Stability und Social Desirability. Unter Kontrolle der Interviewsituation (Anwesenheit Dritter und Befragungsort) zeigen sich beim Vergleich beider Erhebungsformen keine nennenswerten Unterschiede.

Mit einem Beitrag von Michael Häder schließt das Kapitel zur Datenqualität. Der Autor geht der Frage nach, ob und ggf. welche Unterschiede je nach Befragungssituation (Festnetz vs. Mobilfunk) bestehen und inwiefern diese sich auf die Qualität der Angaben auswirken können. Die Datengrundlage bilden die Ergebnisse aus den beiden Projekten CELLA1 und CELLA2. Als theoretischen Ausgangspunkt werden sowohl kognitionspsychologische Ansätze als auch

Aspekte der Persönlichkeitspsychologie behandelt. Skalen zur sozialen Erwünschtheit, die Anzahl von Item-Nonresponse und ein Experiment (Replikation „forbid/not allow“ Rugg 1941) werden als Kriterien für die Antwortqualität in die Studie eingeführt. Hinsichtlich der Befragungssituation zeigen sich insgesamt nicht sehr deutliche Unterschiede. Gleichwohl konnten je nach Setting (Ort des Interviews, Anwesenheit Dritter, Aktivitäten kurz vor dem Interview) Variationen zwischen den Erhebungsmodi ermittelt werden.

Im abschließenden fünften Kapitel „Recommendations for Telephone Surveys“ sollen Arbeiten vorgestellt werden, welche die Befunde vorheriger Arbeiten aufgreifen und Empfehlungen zur Verbesserung und Optimierung darstellen bzw. formulieren. Kennzeichnend für die Arbeit von Gerd Meier ist eine experimentelle Untersuchung zur Steigerung der Teilnehmerate. Bezugnehmend auf die Persuasionsforschung und die Reziprozitätsnorm werden verschiedene Einleitungs- bzw. Gesprächseröffnungstexte hergeleitet und eingeführt. Die Untersuchungsergebnisse lassen erkennen, dass die Kooperationsrate in Abhängigkeit von der Komplexität der Gesprächseröffnung bzw. Einstiegsfrage variiert.

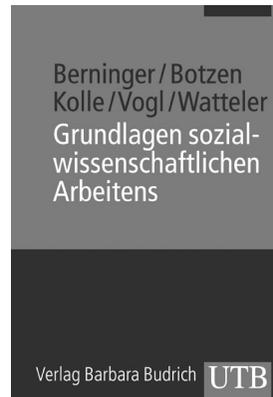
Sollten bei telefonischen Bevölkerungsbefragungen Mobilfunknummern im Auswahlrahmen enthalten sein? Eine Frage, der die Autoren Vasja Vehovar und Ana Slavec in einer slovenischen Telefonumfrage (einschließlich verschiedener Simulationsberechnungen) nachgehen. Da sich die Mobile-only's von den Personen mit Festnetzanschluss nicht nur hinsichtlich soziodemographischer Merkmale unterscheiden, berücksichtigen die Autoren zudem einige ausgewählte Themenbereiche. In Anlehnung an die in der Literatur verschiedentlich berichteten und bereits seit den 60er Jahren angewendeten dual-frame Designs, wird zudem das Kosten-Nutzen Verhältnis thematisiert.

Klassische Aspekte der Feldarbeit und deren Einfluss auf die soziale Zusammensetzung stehen im neunzehnten und somit letzten Beitrag des Sammelbandes im Mittelpunkt der Arbeit von Patrick Schlich und Franziska Jentsch. Die Grundlage bildet der regelmäßige telefonische Gesundheitssurvey des Robert Koch Institutes, der ausschließlich Haushalte in der Erhebung berücksichtigt, die über einen Festnetzanschluss verfügen. Folgende feldspezifische Maßnahmen werden dargestellt: Anzahl der Kontakte, Uhrzeit bzw. Tageszeit der Kontaktierung, Wochentage der Kontakte und erfolgreicher Zweitkontakt nach ablehnendem Erstkontakt (z. B. „keine Zeit“). Welchen Einfluss die genannten Maßnahmen auf der Ebene ausgewählter soziodemographischer Merkmale haben, wird mittels bivariater Analysen untersucht.

Selbst bei cursorischer Sichtung wird dem Leser schnell klar, dass der vorliegende Band – nicht nur wegen seiner Länderperspektive – seines Gleichen sucht. So wird der Gegenstand, Telefonumfragen in Zeiten gesellschaftlichen und technologischen Wandels aus einer breit angelegten Perspektive betrachtet. Mit einem Einblick in die jeweiligen nationalen Herausforderungen und Herangehensweisen werden zudem Akzente aus der akademischen wie aus der nicht-akademischen Umfrageforschung gesetzt. Im Mittelpunkt der aktuellen Diskussion, so auch in dem vorliegenden Band, stehen folgende Themen: Auswahlrahmen, Gewichtung, Non-response und Datenqualität. Summarisch lässt sich festhalten, dass die Autoren der neunzehn Beiträge versierte Kenner der behandelten Materie sind. Wenngleich, was beim Umfang der Beiträge üblich ist, die Qualität einzelner Arbeiten variiert. Auch zeigen sich Differenzen bei den Formalia, die beim Redigieren hätten auffallen sollen. Alles in allem wird ein Forschungsstand präsentiert der die gegenwärtigen relevanten Forschungsfragen aufgreift und die zukünftigen skizziert.

VOLKER HÜFKEN

DOI: 10.12758/mda.2013.007



BERNINGER, INA et al. (2012): Grundlagen sozialwissenschaftlichen Arbeitens. Eine anwendungsorientierte Einführung. Opladen: Budrich

Zwei Autorinnen und drei Autoren legen gemeinsam ein neues Werk zum „Handwerkszeug“ der wissenschaftlichen Arbeit vor. Ist dies ein weiteres Grundlagenbuch unter vielen oder unterscheidet sich diese Neuerscheinung von dem großen, bereits existierenden Angebot für Studierende und WissenschaftlerInnen? Der Untertitel lässt es zumindest vermuten: „Eine anwendungsorientierte Einführung“ wird den Lesern versprochen. Im Klappentext wird damit geworben, dass die Schritte des wissenschaftlichen Arbeitens verständlich und umfassend erklärt werden. Als Zielgruppe werden Studierende aller sozialwissenschaftlicher Fächer genannt.

Das 200 Seiten umfassende Buch ist in 11 Kapitel unterteilt. Anstelle einer Einleitung gibt Ina Berninger im ersten Kapitel „Einführung in das wissenschaftliche Arbeiten“ einen Überblick darüber, was Wissenschaft ist und daraus folgend, was wissenschaftliches Arbeiten bedeutet und beinhaltet. Anhand eines auch grafisch dargestellten Ablaufs eines Forschungsprozesses wird ein Überblick über die Inhalte des vorliegenden Buches gegeben.

Katrin Botzen geht im zweiten Kapitel „Konzeption einer wissenschaftlichen Arbeit“ auf die Entwicklung einer bearbeitbaren Fragestellung ein. Vom Brainstorming, der Themenstrukturierung und Eingrenzung, der Gliederung bis hin zur Erstellung eines

Exposés wird alles Wichtige dazu auf zehn (!) Seiten demonstriert. An dieser Stelle zeigt sich bereits eine Stärke, vielleicht sogar *die* Stärke des Buches: Die kurze und dennoch genügende Erklärung der wichtigsten Techniken zu den einzelnen Arbeitsfeldern. Anschließend führt Dominikus Vogl noch in das Thema Zeitplanung ein und gibt Tipps zur Arbeitsmotivation.

Im dritten Kapitel „Wissenschaftliche Literaturrecherche“ erklärt Christian Kolle die unterschiedlichen Literatortypen sowie deren Beschaffung und Verwaltung. Dabei geht er auch auf die Online-Literaturrecherche in unterschiedlichen Fach-Datenbanken ein und zeigt mit Hilfe von Abbildungen sowohl Suchstrategien als auch die richtige Bedienung von Suchmasken am Beispiel von sowiport (GESIS).

Der Inhalt des vierten Kapitels „Wissenschaftliche Texte lesen und verstehen“ wird wieder von Ina Berninger dargelegt. Die Autorin verdeutlicht anhand unterschiedlicher Methoden, wie Studierende einen wissenschaftlichen Text richtig und schnell erfassen und zusammenfassen können.

Auch das fünfte Kapitel „Theorien, Modelle und Hypothesen“ präsentiert Frau Berninger. Hier geht es darum, ein Theorieverständnis zu erlangen, welches man zur Erstellung des theoretischen Bezugs einer wissenschaftlichen Arbeit benötigt. Nach der Darstellung der unterschiedlichen Arten von Hypothesen und eines soziologischen Erklärungsmodells, der „Colemanschen Badewanne“, wird an praktischen Beispielen die Anwendung von Theorienmodellen dargelegt.

„Aufbau, Inhalt und Formalia einer wissenschaftlichen Arbeit“ werden im sechsten Kapitel von Katrin Botzen beschrieben. Neben der strukturellen Gliederung einer Arbeit (Deckblatt, Inhaltsverzeichnis, Einleitung etc.) ist hier besonders die verständliche Erklärung der beiden gängigen Zitierweisen zu empfehlen. Eine durchaus auch für die Lehre verwendbare Aufführung der gebräuchlichsten Literaturarten zum Nach-

weis in einem Literaturverzeichnis und Kriterien für eine mögliche Bewertungsgrundlage einer wissenschaftlichen Arbeit runden diesen wichtigen Absatz des Grundlagenbuchs ab.

Abermals Ina Berninger gibt im siebten Kapitel wertvolle Anregungen zum wissenschaftlichen Schreiben, der Schreibmotivation, dem Schreibstil und zur Überarbeitung eines Textes.

Im achten Kapitel „Recherche nach sozialwissenschaftlichen Forschungsdaten“ vermittelt Oliver Watteler einen durchaus interessanten Überblick über Informationsgewinnung und Arten von Daten in den Sozialwissenschaften. Außer Suchstrategien und einem Überblick zu Daten generierenden Einrichtungen finden potentielle Leser auch eine kleine Zuordnung von Forschungsinhalten (Themen) zu passenden Daten unterschiedlicher Surveys.

In aller Kürze geht nochmal Katrin Botzen im neunten Kapitel auf die „Darstellung von Tabellen, Grafiken und Diagrammen“ ein. Dies gewährt nicht nur einen kleinen Einblick in unterschiedliche Modelle und Analysemethoden, sondern es werden die grundlegende Regeln zum Gestalten und Beschriften sowie Vor- und Nachteile von unterschiedlichen Diagrammen und Tabellen vermittelt.

Im zehnten Kapitel widmen sich Ina Berninger und Christian Kolle der wissenschaftlichen Präsentation. Hierbei wird nicht nur auf die Visualisierung des Erarbeiteten eingegangen, sondern auch auf den Vortrag und den oder die Vortragenden.

Dominikus Vogl widmet sich zum Ende des Buchs im Kapitel elf dem „Arbeiten mit Software“. Anfangs geht er auf sehr grundlegendes, aber im Alltag häufig vernachlässigtes Wissen, wie das logische Anlegen einer Ordnerstruktur und der Datensicherung ein. Besonders praktisch für Anfänger ist das folgende Unterkapitel: Das Kapitel 11.3 Textverarbeitung mit Word kann Schritt für

Schritt im Selbststudium - oder leicht abgewandelt sogar als E-Learnigeinheit denkbar - nicht nur für Studierende zur Erstellung einer Dokumentenvorlage für wissenschaftliche Arbeiten mit Word verwendet werden. Unter anderem werden in diesem Kapitel das Anpassen der Formatvorlagen, das richtige automatische Anzeigenlassen der Seitenzahlen im Dokument und das Erstellen automatischer Verzeichnisse anschaulich beschrieben. Diese hier behandelten grundlegenden EDV-Kenntnisse zusätzlich zu vermitteln, halte ich für Einführungsbuch in das wissenschaftliche Arbeiten für unverzichtbar, da nach meinen Erfahrungen nicht nur Studienanfänger über diese Basics nicht oder nur teilweise verfügen. Für „Profis“ bietet dieses Kapitel abschließend eine kleine Einführung in LaTeX.

Halten die Autorinnen und Autoren die Versprechen auf dem Schutzumschlag? Durchaus. Die Ankündigung der Anwendungsorientiertheit wird durch die anschaulichen Erklärungen anhand von vielen Beispielen und Vorlagen, wie auch den wie Lehreinheiten aufgebauten Teilen zur Datenbankrecherche, den Darstellungen zur Literaturverwaltungssoftware und zu Textverarbeitungsprogrammen, eingehalten. Das Buch bietet eine begleitende Anleitung von der noch relativ unkonkreten Idee bis zur Fertigstellung einer Seminar- oder Abschlussarbeit. Besonders die Beispiele zur Themeneingrenzung (Seite 22) können Studierende bei der Konkretisierung ihrer

anfänglich oft viel zu komplexen Themenvorschläge für wissenschaftliche Arbeiten nutzen.

Werden die einzelnen Schritte der sozialwissenschaftlichen Forschung verständlich und umfassend erklärt? Auch diese Frage ist positiv zu beantworten. Die Verständlichkeit zeigt sich vor allem darin, dass die einzelnen Themengebiete auf das Wichtigste reduziert sind. Trotzdem werden für ein Grundlagenbuch die einzelnen Bereiche umfassend genug behandelt. Vielleicht ist gerade diese Pointiertheit besonders positiv zu erwähnen, wenn man einen Vergleich zu anderen Büchern in diesem Bereich zieht.

Aufgrund der zum Teil unterschiedlichen Struktur und Darbietung der jeweiligen Kapitel durch die verschiedenen AutorInnen leidet ein wenig die Einheitlichkeit dieser ersten Ausgabe, was jedoch einer leichten Rezeption nicht im Wege steht. Durchaus hilfreich zum schnellen Nachschlagen wäre ein kleines Schlagwortregister mit den wichtigsten Begriffen zur Thematik.

Ich werde dieses Buch wegen der kurzen, aber dennoch ausreichend präzisen Darstellungen, welche einen schnellen Zugang zu den Fertigkeiten des wissenschaftlichen Arbeitens ermöglichen, meinen Studierenden (Bachelor und Lehramt) empfehlen.

RALF SCHULTHEISS

DOI: 10.12758/mda.2013.008

Information for Authors

Methods, data, analyses (mda) publishes research on all questions important to quantitative methods, with a special emphasis on survey methodology. In spite of this focus we welcome contributions on other methodological aspects.

Manuscripts that have already been published elsewhere or are simultaneously submitted to other journals will not be considered. As a rule we do not restrict authors' rights. All rights remain with the author, and articles in mda are published under the CC-BY open-access license.

Mda aims for a quick peer-review process. All papers submitted to mda will first be screened by the editors for general suitability and then double-blindly reviewed by at least two reviewers. The decision on publication is made by the editors based on the reviews. The editorial team will contact the authors by email with the result at the latest eight weeks after submission; if the reviews have not been received by then, we provide a status update with a new target date.

When preparing a paper for submission, please consider the following guidelines:

- Please submit your manuscript by e-mail to [mda\(at\)GESIS\(dot\)org](mailto:mda(at)GESIS(dot)org).
- The total length of the manuscript shall not exceed 10.000 words.
- Manuscripts should...
 - be written in English, using American English spelling. Please use correct grammar and punctuation. Non-native English speakers should consider a professional language editing prior to publication.
 - be typed in a 12 pt Roman font, double-spaced throughout.
 - start with a cover page containing the title of the paper and contact details / affiliations of the authors, but be anonymized for review otherwise.
- Please also send us an abstract of your paper (approx. 300 words), a brief biographical note (no longer than 250 words), and a list of 5-7 keywords for your paper.
- Acceptable formats for Graphics are
 - Tiff
 - Jpeg (uncompressed, high quality)
 - EMF/WMF
- Please ensure a resolution of at least 300 dpi and take care to send high-quality graphics. Line art images should have a resolution of 500-1000 dpi. Please note that we cannot print color images.
- The type area of our journal is 11.5 cm (width) x 18.5 cm (height). Please consider this when producing tables or graphics.
- Footnotes should be used sparingly.
- By submitting a paper to mda the authors agree to make data and program routines available for purposes of replication.

Please follow the APA guidelines when preparing in-text references and the list of references.

Entire Book:

Groves, R. M., & Couper, M. P. (1998). *Nonresponse in household interview surveys*. New York: John Wiley & Sons.

Journal Article (with DOI):

Klimoski, R., & Palmer, S. (1993). The ADA and the hiring process in organizations. *Consulting Psychology Journal: Practice and Research*, 45(2), 10-36. doi:10.1037/1061-4087.45.2.10

Journal Article (without DOI):

Abraham, K. G., Helms, S., & Presser, S. (2009). How social processes distort measurement: The impact of survey nonresponse on estimates of volunteer work in the United States. *American Journal of Sociology*, 114(4), 1129-1165.

Chapter in an Edited Book:

Dixon, J., & Tucker, C. (2010). Survey nonresponse. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of Survey Research*. Second Edition (pp. 593-630). Bingley: Emerald.

Internet Source (without DOI):

Lewis, O., & Redish, L. (2011). Native American tribes of Wisconsin. Retrieved April 19, 2012, from the Native Languages of the Americas website: www.native-languages.org/wisconsin.htm

For more information, please consult the *Publication Manual of the American Psychological Association* (Sixth ed.).