



# Melderegister, Mixed-Mode und Teilnahmeanreize

## Ein Modell zur Rekrutierung von Teilnehmern für sozialwissenschaftliche Panelbefragungen

Dissertation zur Erlangung der Doktorwürde  
durch den Promotionsausschuss Dr. rer. pol.  
der Universität Bremen

Fachbereich 8, Sozialwissenschaften

Gutachter: Prof. Dr. Uwe Engel und Prof. Dr. Frank Faulbaum

Vorgelegt von Björn Oliver Schmidt, M.A.

Bremen, den 28. August 2017

## Vorwort

Diese Arbeit ist im Rahmen des von der Deutschen Forschungsgemeinschaft geförderten Projekts „Access Panel and Mixed-Mode Internet Survey“ im Schwerpunktprogramm 1292 „Survey Methodology“ entstanden.

An erster Stelle möchte ich mich bei meinem Förderer und Gutachter Prof. Dr. Uwe Engel bedanken, ohne dessen Unterstützung das Dissertationsprojekt nicht hätte verwirklicht werden können. Des Weiteren möchte ich mich bei Prof. Dr. Frank Faulbaum für die Übernahme der Zweitbegutachtung bedanken.

Ich bedanke mich bei all denjenigen, die mir mit Rat und Tat zur Seite standen und mich durchweg motiviert haben. Ein besonderer Dank geht an: Julia Asseburg, Julia Christine Borowsky, Dr. Alexander Gattig, Dr. Andrea Gottlieb, Kai Heyse, Andreas Hoven, Britta Köster, Timo Peter, Florian Przybilla, Miriam Reußner, Dr. Julia Sievers, Sabine Sommer und das gesamte Team des Referat zwölf der Universität Bremen.

Diese Arbeit widme ich meinen Großeltern.

Björn Oliver Schmidt

# Inhaltsverzeichnis

Vorwort . . . . .	I
Tabellenverzeichnis . . . . .	VI
Abbildungsverzeichnis . . . . .	VIII
Abkürzungsverzeichnis . . . . .	IX
<b>1 Einleitung</b>	<b>1</b>
<b>2 Fehlerquellen wissenschaftlicher Umfragen</b>	<b>7</b>
2.1 Der Total Survey Error . . . . .	7
2.1.1 Noncoverage: Systematische Unterrepräsentation von Bevölkerungsgruppen . . . . .	11
2.1.2 Nonresponse: Niedrige Antwortraten als Quelle systematischer Ver- zerrungen . . . . .	22
2.2 Struktur fehlender Werte . . . . .	28
<b>3 Erklärungsmodelle der Umfrageteilnahme</b>	<b>32</b>
3.1 Kooperationsmodell nach Groves und Couper . . . . .	34
3.2 „Leverage-Saliency“ Theorie der Umfrageteilnahme . . . . .	39
<b>4 Methoden zur Reduzierung möglicher Fehler</b>	<b>42</b>
4.1 Mixed-Mode . . . . .	42
4.2 Teilnahmeanreize . . . . .	50
4.3 Melderegister . . . . .	60
4.4 Paradata . . . . .	65
4.5 Statistische Konzepte zur Kompensation von Verzerrungen . . . . .	70
<b>5 Ein integriertes Gesamtmodell</b>	<b>80</b>

<b>6</b>	<b>Datengrundlage und Operationalisierung</b>	<b>94</b>
6.1	Das PPSM Access Panel . . . . .	95
6.1.1	Generierung der Stichprobe, Rekrutierungs- und Initialisierungsphase	96
6.1.2	Nachrekrutierung - Bremen Stadt 2012 . . . . .	99
6.1.3	Befragung Winter 2014 . . . . .	104
6.1.4	Operationalisierung weiterer zentraler Konstrukte aus den Jahren 2012 und 2014 . . . . .	109
6.2	Statistische Analyseverfahren - Chronologie und Methoden . . . . .	112
<b>7</b>	<b>Ergebnisse</b>	<b>115</b>
7.1	Mixed-Mode . . . . .	116
7.1.1	Antwortraten und Moduspräferenzen . . . . .	116
7.1.2	Zur Bereitschaft eine Umfrage via Smartphone zu beantworten . . .	119
7.2	Wirkung von Teilnahmeanreizen . . . . .	129
7.2.1	Auswirkungen auf die Zusammensetzung der Teilnehmer und poten- tielle Verzerrungen . . . . .	129
7.2.2	Grenznutzen monetärer Teilnahmeanreize . . . . .	134
7.3	Das Melderegister als Auswahlrahmen - Analyse des Ausfallprozesses . . .	140
7.3.1	Teilnahme am Rekrutierungsinterview . . . . .	141
7.4	Unterschiede zwischen eingetragenen und nicht eingetragenen Personen . .	166
<b>8</b>	<b>Fazit und Handlungsempfehlungen</b>	<b>169</b>
	<b>Literaturverzeichnis</b>	<b>179</b>
	<b>Ehrenwörtliche Erklärung</b>	<b>201</b>
	<b>Anhang</b>	<b>202</b>
<b>A</b>	<b>Weitere Ergebnistabellen</b>	<b>203</b>
<b>B</b>	<b>Umfragevariablen Nachrekrutierung 2012</b>	<b>205</b>
<b>C</b>	<b>Soziodemografie</b>	<b>208</b>
<b>D</b>	<b>Einladungsschreiben 2012</b>	<b>209</b>

# Tabellenverzeichnis

2.1	Entwicklung der Internetnutzung in Deutschland . . . . .	12
2.2	Entwicklung der mobilen Internetnutzung in Deutschland von 2009 bis 2014 in Prozent . . . . .	15
4.1	Typen der Mixed-Mode Befragung, Gründe für die Anwendung und der Einfluss auf die Datenqualität . . . . .	44
4.2	Formen von Teilnahmeanreizen . . . . .	50
4.3	Arten von Paradata nach Befragungsmodus und Potential zur Fehlerredu- zierung . . . . .	66
5.1	Kontaktierungswege der zugrundeliegenden Erhebungen sowie zu analysie- rende Zielgrößen . . . . .	82
5.2	Forschungshypothesen zu Designaspekt A . . . . .	86
5.3	Forschungshypothesen zu Designaspekt B und C . . . . .	87
5.4	Forschungshypothesen zu Designaspekt D . . . . .	90
5.5	Forschungshypothesen zur Smartphonennutzung und zur Teilnahmebereit- schaft an einer mobilen Onlinebefragung . . . . .	92
6.1	Selektionsstufen der Rekrutierungsphase des Priority Programme on Survey Methodology (PPSM) Access Panels . . . . .	97
6.2	Die Initialisierungsphase des PPSM Access Panels . . . . .	98
6.3	Final Disposition Codes und übergeordnete Antwortrate - PPSM Access Panel Nachrekrutierung 2012 . . . . .	101
6.4	Studiendesign Nachrekrutierung 2012 . . . . .	102
6.5	Final Disposition Codes und Antwortrate - PPSM Access Panel Befragung Winter 2014 . . . . .	105

6.6	Soziodemografische Variablen und Parainformationen . . . . .	110
6.7	Zielgrößen und Analyseverfahren . . . . .	113
7.1	Antwortraten und Moduspräferenzen der einzelnen Teilgruppen in Prozent - Zielgröße: Wahl des Befragungsmodus . . . . .	117
7.2	Entscheidung über die Hypothesen - Zielgröße: Wahl des Befragungsmodus	119
7.3	Logistische Regression mit der Smartphonennutzung (0 = Nein; 1 = Ja ) als abhängige Variable . . . . .	121
7.4	Entscheidung über die Hypothesen - Zielgröße: Smartphonennutzung . . . .	124
7.5	Logistische Regression mit der Bereitschaft, eine Umfrage mobil zu beant- worten (0 = Nein; 1 = Via App/Via Link/Sowohl als auch) als abhängige Variable . . . . .	127
7.6	Entscheidung über die Hypothesen - Zielgröße: Teilnahme an mobiler Onli- neumfrage . . . . .	128
7.7	Vergleich der Bildungsabschlüsse in den experimentellen Untersuchungs- gruppen mit dem Mikrozensus 2013 in Prozent . . . . .	130
7.8	Auswirkung der vergebenen Anreize auf die Datenqualität: Lineare Re- gression mit der absoluten Häufigkeit von Item-Nonresponse als abhängige Variable . . . . .	133
7.9	Entscheidung über die Hypothesen - Zielgrößen: Kooperationswahrschein- lichkeit und Datenqualität . . . . .	134
7.10	Mittelwertvergleiche der experimentellen Bedingungen mit unterschiedlichen Anreizbeträgen . . . . .	137
7.11	Ergebnisse der qualitativen Inhaltsanalyse - „kleine Aufmerksamkeit“ vs. „Bezahlung“ . . . . .	138
7.12	Entscheidung über die Hypothesen - subjektive Wahrnehmung Anreizhöhe	139
7.13	Studiendesign Nachrekrutierung 2012 - Wiederholung von Tabelle 6.4 . . . .	142
7.14	Logistische Regression mit der Teilnahme am Rekrutierungsinterview (0 = Nein; 1 = Ja ) als abhängige Variable für die Bruttostichprobe . . . . .	143
7.15	Entscheidung über die Hypothesen - Zielgrößen: Teilnahmewahrscheinlich- keit und Wahl des Umfragemodi . . . . .	147

7.16	Logistische Regressionen mit der Teilnahme am Rekrutierungsinterview (0 = Nein; 1 = Ja) als abhängige Variable für jeweils unterschiedliche Designbedingungen . . . . .	150
7.17	Fortsetzung von Tabelle 7.16 . . . . .	151
7.18	Nonresponse Bias in ausgewählten Variablen nach dem RRM Model von Bethlehem (2002) . . . . .	157
7.19	Entscheidung über die Hypothesen - Zielgrößen: Erreichbarkeit, Kooperationswahrscheinlichkeit, Teilnahmewahrscheinlichkeit . . . . .	160
7.20	Logistische Regression mit der Teilnahme am Rekrutierungsinterview (0 = Nein; 1 = Ja ) als abhängige Variable und Parainformationen aus der CATI-Befragung . . . . .	162
7.21	Entscheidung über die Hypothesen - Zielgröße: Erreichbarkeit . . . . .	165
7.22	Logistische Regression mit dem Eintrag im Telefonbuch (0 = Nein; 1 = Ja ) als abhängige Variable für die Bruttostichprobe . . . . .	167
7.23	Entscheidung über die Hypothesen . . . . .	168
A.1	Logistische Regression mit der Teilnahme an der CAWI-Variante der Befragung (0 = Nein; 1 = Ja ) als abhängige Variable für die Bruttostichprobe .	203
A.2	Logistische Regression mit der Teilnahme an der CATI-Variante der Befragung (0 = Nein; 1 = Ja ) als abhängige Variable für die Bruttostichprobe .	204

# Abbildungsverzeichnis

2.1	Umfrageprozess und Total Survey Error nach Groves et al. (2011, 48) . . .	8
2.2	Selektionsstufen bei der Rekrutierung von Teilnehmern für eine Onlinebefragung in Anlehnung an Bandilla et al. (2009) . . . . .	27
2.3	Struktur fehlender Werte nach Groves und Peytcheva (2008) . . . . .	29
4.1	Zusammenhang zwischen Zielvariable, Teilnahmeentscheidung und Hintergrundvariable . . . . .	71
5.1	Ein integriertes Gesamtmodell zu Rekrutierung von Teilnehmern für Panelbefragungen . . . . .	80
7.1	Geschätzte Wahrscheinlichkeiten der Zielgröße Smartphonennutzung differenziert nach Geschlecht und Alter . . . . .	122
7.2	Relative Häufigkeiten der Bereitschaft eine Computer Aided Web Interviewing (CAWI) Umfrage mobil zu beantworten . . . . .	126
7.3	Subjektive Wahrnehmung der Incentivehöhe - feldexperimentelles Design mit 5, 10, 20 und 40 Euro . . . . .	135
7.4	Geschätzte Wahrscheinlichkeiten der Umfrageteilnahme differenziert nach Alter, Geschlecht und Modus . . . . .	144
7.5	Zusammenhang zwischen Zielvariable, Teilnahmeentscheidung und Hintergrundvariable - Wiederholung von Abbildung 4.1 . . . . .	149
7.6	Geschätzte Wahrscheinlichkeiten der Umfrageteilnahme für die feldexperimentellen Designbedingungen 5 Euro (vs. Kontrollgruppe)/Anruf angekündigt (vs. Anruf nicht angekündigt) . . . . .	152
7.7	Geschätzte Wahrscheinlichkeiten der Umfrageteilnahme differenziert nach Kontaktversuchen . . . . .	163



7.8	Geschätzte Wahrscheinlichkeiten der Umfrageteilnahme differenziert nach der Summe an Terminen mit der Kontakt- oder Zielperson . . . . .	164
7.9	Geschätzte Wahrscheinlichkeiten des Telefonbucheintrages differenziert Ge- schlecht und Alter . . . . .	166

## Abkürzungsverzeichnis

<b>AAPOR</b>	American Association for Public Opinion Research
<b>ALLBUS</b>	Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften
<b>ADM</b>	Arbeitskreis deutscher Markt- und Sozialforschungsinstitute e. V.
<b>CATI</b>	Computer Aided Telephone Interviewing
<b>CAPI</b>	Computer Aided Personal Interviewing
<b>CASI</b>	Computer Aided Self Interviewing
<b>CAWI</b>	Computer Aided Web Interviewing
<b>DFG</b>	Deutsche Forschungsgemeinschaft
<b>ESS</b>	European Social Survey
<b>LISS</b>	Longitudinal Internet Studies for the Social Sciences
<b>MCAR</b>	Missing Completely at Random
<b>MAR</b>	Missing at Random
<b>MNAR</b>	Missing Not at Random
<b>PPSM</b>	Priority Programme on Survey Methodology
<b>RLD</b>	Randomized Last Digit
<b>RRM</b>	Random Response Model
<b>RRF</b>	Structured Refusal Report Form
<b>SMS</b>	Short Message Service
<b>SOEP</b>	Sozio-oekonomisches Panel
<b>TSE</b>	Total Survey Error
<b>URL</b>	Uniform Resource Locator

Hinweis: Zugunsten einer besseren Lesbarkeit wird im gesamten Text das generische Maskulinum verwendet.

# Kapitel 1

## Einleitung

Wissenschaftliche Umfragen versuchen Meinungen, Werthaltungen, Ereignisse im Lebensverlauf einer Person oder die Verteilung soziodemografischer Merkmale der allgemeinen Bevölkerung möglichst genau abzubilden. Nicht selten wird die öffentliche Meinungsbildung oder gar die politische Entscheidungsfindung von Befragungsergebnissen beeinflusst. Durch gesellschaftlichen Wandel sowie technologischen Fortschritt muss sich die Umfrageforschung dabei stetig an neue Rahmenbedingungen anpassen.

Umfragen als Königsweg der quantitativen Sozialforschung leiden dabei heutzutage vor allem unter zwei Problemen: Die systematische Unterrepräsentation einzelner Bevölkerungsgruppen (Coverage Error) und niedrigen Antwortraten (Nonresponse Error). Abhängig vom Modus der Befragung variieren diese Entwicklungen und lassen sich im Kontext des „Total Survey Errors“ in ein umfassendes Konzept potentieller Fehlerquellen einordnen (Bartsch, 2012; Faulbaum, 2014). Wenn zur Grundgesamtheit gehörende Personen nicht im Auswahlrahmen enthalten sind, kann die Datenqualität unter Noncoverage leiden. Von Overcoverage spricht man, wenn Personen im Auswahlrahmen und schließlich in der Stichprobe enthalten sind, aber nicht zur Grundgesamtheit gehören. Seitdem die Verbreitung des Internets in den 1990er Jahren in Deutschland immer weiter zugenommen hat, sind Befragungen hierüber durch eine vergleichsweise kostengünstige Umsetzung immer populärer geworden. Online-Interviews haben 2016 ein Niveau von insgesamt 36 Prozent erreicht (Arbeitskreis deutscher Markt- und Sozialforschungsinstitute e.V., 2016). Daneben wird die Telefonbefragung mit einem gleichen Anteil heutzutage ebenfalls häufig genutzt.

Werden Befragungen über das Internet durchgeführt, geht hiermit, im Vergleich zur

telefonischen oder mündlich-persönlichen Befragung, ein besonderes Fehlerpotential einher. Denn von Verzerrungen durch Noncoverage sind vor allem Onlineumfragen betroffen, da es keinen Auswahlrahmen zur Befragung der allgemeinen Bevölkerung gibt (Baur und Florian, 2009). Aufgrund dessen wird häufig „offline“ per Telefon oder postalischem Anschreiben rekrutiert. Darüber hinaus sind unter anderem ältere Menschen, die über keinen Zugang zum Internet verfügen, systematisch unterrepräsentiert (Bartsch, 2012). Dieser Gruppe kann mit entsprechenden Befragungscomputern ausgestattet werden, was in seiner Umsetzung allerdings sehr teuer ist und deswegen nicht häufig zur Anwendung kommt. Um das Potential des Internets für die Umfrageforschung in Zukunft noch besser als bisher nutzbar zu machen, bedarf es weiterer Forschung.

Die Forschung im Bereich Coverage und Nonresponse gibt es nahezu schon so lange wie die Umfrageforschung selbst. Die Antwortrate galt hierbei über einen langen Zeitraum als das Qualitätskriterium. So wurden zunächst von der American Association for Public Opinion Research (AAPOR) sogenannte „Standard Definitions“ herausgegeben, um die Berechnungen von Antwortraten zu standardisieren und über Umfrageprojekte hinweg vergleichbar zu machen. Zusätzlich wurden verschiedene ausschöpfungssteigernde Designelemente, wie Teilnahmeanreize oder postalische Vorankündigungen eingesetzt und einer empirischen Überprüfung unterzogen, um dem Trend abnehmender Antwortraten entgegenzuwirken. Neuere Studien konnten allerdings nachweisen, dass eine hohe Antwortrate nicht notwendigerweise mit einem niedrigen Bias durch Nonresponse einhergeht und vice versa (Brick, 2013; Groves, 2006; Groves und Peytcheva, 2008; Kreuter, 2013a; Peytchev, 2013). Vor diesem Hintergrund beschäftigt sich die wissenschaftliche Methodenforschung im Bereich Nonresponse vor allem damit, die dahinterstehenden Mechanismen besser zu verstehen, um statistische Verfahren entwickeln zu können, mögliche Konsequenzen zu berücksichtigen.

Allerdings ist die Operationalisierung von Coverage beziehungsweise Nonresponse bedingter Fehler nicht wie bei Messeffekten oder anderen potentiellen Verzerrungsquellen über einen experimentellen Zugang möglich. Denn der Auswahlrahmen wird durch den zum Einsatz kommenden Umfragemodus vorgegeben und über Nonrespondenten liegen in der Regel nur wenig bis keine Informationen vor. Entsprechende Nachfassbemühungen, um zumindest einige Informationen über diejenigen zu bekommen, die nicht teilgenommen haben, sind häufig mit einem erheblichen Kostenaufwand verbunden. Solche Maßnahmen sind wiederum von Nonresponse betroffen und deswegen nur bedingt geeignet.

Das Grundprinzip zufallsbasierter Stichproben, dass jedes Element der Grundgesamtheit eine angebbare Wahrscheinlichkeit besitzt gezogen zu werden, ist abhängig vom Modus der Befragung in der Praxis häufig schwierig und zumeist nur mit erheblichem Kostenaufwand zu realisieren. Dadurch, dass kein allgemein gültiger Auswahlrahmen existiert oder der Zugang zum Internet fehlt, die Verbreitung von Festnetztelefonen rückläufig ist und es immer mehr Personen gibt, die nur noch über das Handy erreichbar sind (sogenannte „Mobile-Onlys“), werden bestimmte Gruppen von Personen systematisch von Befragungen ausgeschlossen. In Deutschland hat das Mobiltelefon bereits im Jahr 2014 in seiner Verbreitung das Festnetz überholt (Statistisches Bundesamt, 2015b). Zusätzlich besitzen heutzutage immer mehr Personen ein Smartphone und Nutzen dieses zur Beantwortung von Umfragen, was dieses Medium für die wissenschaftliche Forschung interessant macht (De Bruijne und Wijnant, 2014; Statistisches Bundesamt, 2015a). Eine aktuelle und zukünftige Herausforderung ist es daher das Potential von Mobilfunkbefragungen, vor allem über den Browser in Form von Mobile Web Befragungen, durch weitere Forschung nutzbar zu machen. Hinzukommen mehrheitlich geringe Antwortraten, wodurch zusammengekommen nicht nur die Wahrscheinlichkeit stichprobenbedingter, sondern gleichzeitig die ausfallbedingter Verzerrungen zunimmt. Der Grund hierfür besteht darin, dass es nicht zufällig ist, welche Personen an Umfragen teilnehmen. In der Forschung spricht man hier von Selektionseffekten. Gemeint ist damit, dass einerseits soziodemografische Hintergrundvariablen wie das Bildungsniveau oder das Einkommen einer Person auf die Teilnahmewahrscheinlichkeit wirken. Aufgrund niedriger Antwortraten wurden Designelemente entwickelt (Incentives, Art der Kontaktierung, Thema der Umfrage et cetera), um mehr Personen zu einer Teilnahme zu motivieren. Mit diesen Elementen können aber auch zusätzliche Verzerrungen einhergehen, da sich durch Selektionseffekte unterschiedliche Gruppen von Personen unterschiedlich stark von bestimmten ausschöpfungssteigernden Elementen angesprochen fühlen.

Ein besonderes Fehlerpotential entsteht dadurch, dass sich Respondenten systematisch von Nonrespondenten selektieren. Da es sich um systematische Einflüsse der Nichtteilnahme handelt, liegt der Schluss nahe, dass diese Systematik auch im Antwortverhalten zu finden ist. Diese Wahrscheinlichkeit ausfallbedingter Verzerrungen (Nonresponse Bias) war in der Vergangenheit und ist auch in der Zukunft Anlass wissenschaftlicher Untersuchungen. Denn das Problem, dass über die Personen, die sich aus unterschiedlichen Gründen nicht an wissenschaftlichen Umfragen beteiligen keinerlei Informationen vorliegen, besteht weiterhin.

Da es allerdings als nachgewiesen gilt, dass die Antwortrate nicht zwangsläufig mit der Höhe eines Bias durch Nonresponse zusammenhängt ist es nötig, alternative Vorgehensweisen und Indikatoren zu entwickeln (Brick, 2011; Kreuter, 2013a).

So sind die vergangenen Jahre vor allem durch den Begriff der Mixed-Mode Befragung geprägt worden (Couper, 2011; De Leeuw, 2005; De Leeuw et al., 2008; Dillman und Messer, 2010; Voogt und Saris, 2005). Das Besondere an diesem Ansatz ist, dass in Abhängigkeit von der jeweils zu befragenden Grundgesamtheit einzelne Umfragemodi so miteinander kombiniert werden, dass die Schwächen des einen durch die Stärken des anderen im Optimalfall nivelliert werden. Mixed-Mode Befragungen in Verbindung mit einer registerbezogenen Stichprobenziehung, Teilnahmeanreizen und einer postalischen Vorankündigung des Umfragegesuchs bieten für aktuelle und zukünftige Umfrageprojekte das Potential, obige Fehler zu kontrollieren und zu reduzieren. Eine Stichprobenziehung auf Basis des Einwohnermelderegisters bietet den Vorteil, dass auch über Personen, die sich letztlich nicht an der Umfrage beteiligen, einige Informationen vorhanden sind. Mit den Informationen des Auswahlrahmens in Kombination mit Paradata können Nonresponseanalysen durchgeführt und untersucht werden, inwiefern die Höhe der Antwortrate mit möglichen Verzerrungen zusammenhängt. Darüber hinaus steht die Postadresse als Kontaktmöglichkeit zur Verfügung, sodass eine schriftliche Umfrageeinladung erfolgen kann.

Der Einfluss von Teilnahmeanreizen ist bereits häufig untersucht worden (Becker et al., 2007; Church, 1993; Göritz, 2015; Hansen und Pedersen, 2012; Scherpenzeel und Toepoel, 2012; Singer und Ye, 2013; Toepoel, 2012). Ob es beim Einsatz monetärer Anreize, die unabhängig von der Teilnahme vergeben werden, einen Grenznutzen bezüglich der Höhe des Betrages gibt, wurde hingegen noch nicht ausreichend untersucht. Die hierzu vorliegenden Ergebnisse sind darüber hinaus uneindeutig, sodass es auch in diesem Bereich weiteren Forschungsbedarf gibt.

### **Ziel und Aufbau der Arbeit**

Im Zentrum des Interesses steht das Problem niedriger Antwortraten (Nonresponse) und der Ausschluss von Bevölkerungsgruppen bei der Durchführung wissenschaftlicher Umfragen (Noncoverage). Da diese beiden Phänomene in der Regel nicht zufällig sondern systematisch auftreten, kann dies negative Folgen für die Datenqualität haben. Das *Ziel* dieser Arbeit besteht darin, Effekte durch Noncoverage und selektiven Nonresponse bei

der Rekrutierung von Teilnehmern für sozialwissenschaftliche Panelbefragungen zu reduzieren und zu operationalisieren. Um dieses Ziel zu erreichen wurde ein heuristisches Gesamtmodell entwickelt, das aus vier Teilen besteht: (A) Registerstichprobe und Paradata, (B) Mixed-Mode Design, (C) Postalische Vorankündigung des Umfragegesuchs und (D) Teilnahmeanreize. Über die Auswirkung auf die Antwortrate hinaus soll mittels des Random Response Models nach Bethlehem (2002) untersucht werden, ob sich diese auf das Antwortverhalten der Teilnehmer auswirken und ob Verzerrungen aufgrund von selektivem Nonresponse vorliegt. Hierfür werden mittels multivariater Modelle Teilnahmewahrscheinlichkeiten geschätzt, die mit den vorliegenden Antworten in Beziehung gesetzt werden. Mit diesem Vorgehen soll eine Antwort auf die *übergeordnete Forschungsfrage* möglich werden: Lässt sich selektiver Nonresponse operationalisieren und auf die durch den Auswahlrahmen zur Verfügung stehenden Hintergrundvariablen sowie die einzelnen Designelemente einer Umfrage zurückführen?

An den Einsatz von Mixed-Mode schließt die Frage an, welche Personen sich für welchen Modus der Befragung entscheiden und die Fehlerbilanz hierdurch verringert werden kann. Denn durch den Registerbezug der Stichprobe kann Noncoverage auf der Stufe der Ziehung und Kontaktierung ausgeschlossen werden, da alle Personen in der Stichprobe durch eine postalische Vorankündigung kontaktiert werden können unabhängig davon, ob sie Internet oder ein Festnetztelefon besitzen. Relevant wird dieser Aspekt auf der Stufe der Teilnahme. Von Interesse ist hierbei, ob die Schwächen des einen Modus durch die Stärken des anderen ausgeglichen werden können.

Die Vergabe von Teilnahmeanreizen soll nicht nur hinsichtlich ihrer Wirkung auf die Antwortrate untersucht werden. Vielmehr soll eine Antwort auf die Frage gefunden werden, ob es einen möglichen Grenznutzen beim Einsatz monetärer Anreize in Abhängigkeit von dessen Höhe zu beobachten gibt, wenn diese unabhängig von einer Teilnahme vergeben werden. Darüber hinaus steht die Frage im Fokus, wie gesellschaftliche Gruppen auf monetäre Anreize reagieren und ob es hier Unterschiede zu beobachten gibt.

Zusätzlich soll ein Beitrag zur wissenschaftlichen Diskussion rund um den Einsatz innovativer Mobile Web Befragungen geleistet werden. Die Datengrundlage macht es möglich, einerseits Personen zu charakterisieren, die über ein Smartphone verfügen und dieses auch regelmäßig nutzen. Andererseits steht diesbezüglich die Frage im Fokus, ob dieser Kreis an Personen ihr mobiles Endgerät auch zur Beantwortung einer wissenschaftlichen Umfrage nutzen würde oder nicht.

Zunächst erfolgt ein Problemaufriss (Kapitel 2), gefolgt von einer theoretischen Herleitung der Fehler, die bei der Durchführung einer Umfrage entstehen können (Kapitel 3 und 4). Daraufhin werden mögliche Ansätze zur Fehlerreduktion vorgestellt (Kapitel 4) um hieraus ein Gesamtmodell zu entwickeln, woraus sich die empirisch zu überprüfenden Forschungshypothesen ableiten (Kapitel 5). Die Operationalisierung dieses Modells und die daran anknüpfende Datengrundlage sowie die zur Anwendung kommenden statistischen Analyseverfahren sind Teil darauffolgenden Kapitels (Kapitel 6). Im Anschluss werden die Ergebnisse differenziert bezüglich dem Einsatz von Mixed-Mode, Teilnahmeanreizen, der Analyse des Ausfallprozesses sowie möglicher Unterschiede zwischen Personen, die über einen Eintrag in das Telefonbuch verfügen und Personen, die nicht darin eingetragen sind, vorgestellt (Kapitel 7). Abschließend wird ein Fazit gezogen und daraus mögliche Handlungsempfehlungen abgeleitet (Kapitel 8).



# Kapitel 2

## Fehlerquellen wissenschaftlicher Umfragen

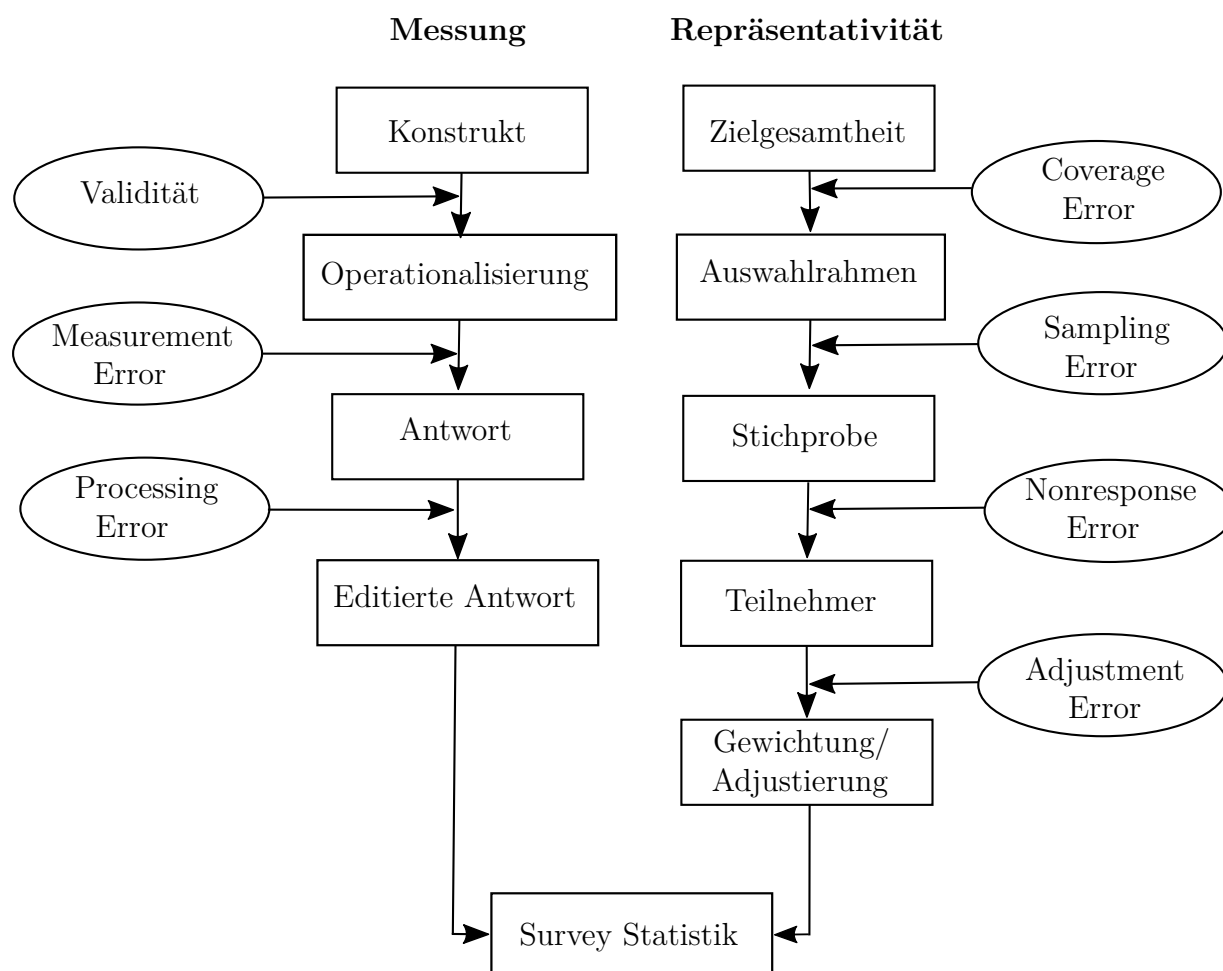
### 2.1 Der Total Survey Error

Bei der Durchführung einer Befragung kann die Datenqualität durch eine Reihe potentieller Fehlerquellen in Mitleidenschaft gezogen werden. Unter dem Begriff des Total Survey Error (TSE) werden in der Umfrageforschung sowohl zufällig als auch systematisch auftretende Fehler subsumiert. Problematisch für die Datenqualität sind insbesondere die systematischen, nicht zufälligen Fehlerquellen, die zu einer Abweichung zwischen wahren und geschätzten Werten führen können (Bartsch, 2012; Biemer und Lyberg, 2003; Bautista, 2012; Faulbaum, 2014; Groves und Lyberg, 2010; Groves et al., 2011; Weisberg, 2005). Unter wahren Werten kann die Verteilung einer Variable in der betreffenden Grundgesamtheit verstanden werden. Ein geschätzter Wert ergibt sich aus der durch eine Befragung erhobenen Datengrundlage. Führt eine systematische Fehlerquelle zu einer Differenz zwischen diesen Werten, wird in der Forschungspraxis häufig von einer Verzerrung oder einem Bias gesprochen. Wird etwa eine Frage nicht beantwortet, weil sie als zu persönlich empfunden wird oder ein Teilnehmer durch soziale Erwünschtheit nicht wahrheitsgemäß antwortet, ist dem Fehlen dieser Angaben eine Systematik zu unterstellen.

Historisch betrachtet stand anfangs vor allem die Frage danach im Fokus, wie man die Umfragequalität insgesamt sichern kann. Aufgrund dessen hat sich Deming (1944) zunächst darüber Gedanken gemacht, welche Fehler im Rahmen des Umfrageprozesses auftreten können. Aus diesen Überlegungen entstand eine erste Übersicht von insgesamt 13 möglichen

Fehlerquellen, die die „Brauchbarkeit“ einer Umfrage beeinflussen können. In darauf folgenden Arbeiten wie von Kish (1962) steht vor allem die Operationalisierung potentieller systematischer Fehler im Fokus, um einen möglichen Bias zu kontrollieren, respektive zu reduzieren (Faulbaum, 2014). Der Begriff TSE wurde schließlich durch die gleichnamige Monographie von Andersen et al. (1979) geprägt, in der die Ergebnisse verschiedener Studien in Bezug zu unterschiedlichen Fehlern im Rahmen von Gesundheitsumfragen diskutiert werden (Groves und Lyberg, 2010). Das Konzept ist in der Umfrageforschung mittlerweile weit verbreitet und wird häufig sowohl zur Operationalisierung, statistischen Kontrolle und Diskussion rund um einen möglichen Bias, als auch zur Planung und Qualitätssicherung von Umfragen genutzt. Groves et al. (2011) haben die Komponenten des TSE mit dem Ablauf des Umfrageprozesses, von der Generierung der Stichprobe über die Durchführung der Befragung bis hin zur Auswertung der Daten, in Verbindung gebracht. So soll mit von Abbildung 2.1 deutlich werden, *welche* Fehler im Befragungsprozess auftreten und *wann* diese zu einem Bias führen können.

Abbildung 2.1: Umfrageprozess und Total Survey Error nach Groves et al. (2011, 48)



In der Literatur unterscheiden sich die Konzepte in der Regel nach Umfang und Bezeichnung der jeweiligen Fehler, die unter dem Begriff des TSE insgesamt gefasst werden. Faulbaum (2014), Biemer und Lyberg (2003) oder auch Cobben und Bethlehem (2005) differenzieren zunächst allgemein zwischen einem *Sampling* und einem *Nonsampling Error*. Erstere entstehen aufgrund dessen, dass aus Kostengründen in der Praxis zumeist Stichproben gezogen werden und keine Vollerhebungen stattfinden können. Deshalb unterscheiden sich beispielsweise Stichprobenmittelwerte von Populationsmittelwerten, da sich bei jeder erneuten Ziehung von  $n$  Elementen ein anderer Mittelwert ergeben würde. Bei der fiktiven unendlichen Ziehung einer Stichprobe aus derselben Grundgesamtheit wird die sich daraus ergebende Varianz der Mittelwerte Sampling Error genannt (Diekmann, 2007). Die Mittelwerte der verschiedenen Stichproben sind im Optimalfall normalverteilt. Unter anderem wird deswegen in der Praxis etwa von erwartungstreuen Schätzern gesprochen.

Wird wie bei der Zensus-Erhebung aus dem Jahr 2011 oder einer Befragung von Mitarbeitern innerhalb von Betrieben eine Vollerhebung angestrebt, bedeutet dies nicht, dass Schätzungen hierdurch unverzerrt bleiben. Unabhängig davon, ob Stichproben gezogen werden oder nicht, kann der Nonsampling Error auch dann in unterschiedlichen Dimensionen auftreten (Cobben und Bethlehem, 2005, 2). Problematisch werden die hierunter gefassten potentiellen Fehlerquellen insbesondere dann, wenn sie nicht zufällig sind.

### 1. *Measurement Error*

Dieser entsteht durch den eingesetzten Fragebogen, den Prozess der Antwortgenerierung, Interviewereffekte oder den Modus der Datenerhebung. Innerhalb eines Fragebogens hat sowohl die Reihenfolge der gestellten Fragen insgesamt als auch der mit dem Text einer Frage gesetzte kognitive Anker einen Einfluss auf das Antwortverhalten (Engel und Köster, 2015). Darüber hinaus wird die jeweils gegebene Antwort davon beeinflusst, inwiefern die einzelnen Punkte auf einer Likert-Skala benannt sind (Engel et al., 2012; Diekmann, 2007). Fragen werden durch den jeweils eingesetzten Umfragemodus unterschiedlich beantwortet (Christian et al., 2005). Effekte durch soziale Erwünschtheit entstehen hier vor allem durch den Einsatz von Interviewern. Besonders bei sensiblen Fragen, wie zum Konsum von Drogen oder abweichendem Verhalten, neigen Befragte durch die Anwesenheit eines Interviewers dazu, sich mit ihrer Antwort an sozialen Normen zu orientieren und weichen so häufig von der tatsächlich wahren Antwort ab (Glantz und Michael, 2014). Bei den sogenannten selbstadministrierten Befragungsformen - online und schriftlich - tritt dieser Effekt

ebenfalls auf, allerdings in deutlich abgeschwächter Form, da Antworten hier keinem fremden Interviewer mitgeteilt werden müssen.

## 2. *Processing Error*

Der Processing Error kann durch die Editierung oder Übertragung von Antworten in die Datenmatrix verursacht werden (Groves et al., 2011). Praktisch kann dies unter anderem auftreten, wenn die Antworten aus einem schriftlichen Fragebogen fehlerhaft digitalisiert werden.

## 3. *Adjustment Error*

Werden Umfragedaten durch Gewichtungs- oder Imputationsverfahren nachträglich adjustiert bzw. ergänzt, können zusätzliche Ungenauigkeiten entstehen. Das Ersetzen fehlender Werte durch Imputation basiert auf spezifischen Schätzverfahren (Spieß, 2010). Die jeweils hierdurch generierten Werte können der wahren Antwort zwar möglichst nahe kommen, diese allerdings nie zu hundert Prozent valide abbilden (Engel und Schmidt, 2014).<sup>1</sup>

## 4. *Coverage Error*

Allgemein wird zwischen Over- und Undercoverage unterschieden. Overcoverage liegt vor, wenn Personen im Auswahlrahmen enthalten sind, die nicht zur Grundgesamtheit gehören. Undercoverage oder auch Noncoverage liegt vor, wenn nicht alle Elemente der Grundgesamtheit im Auswahlrahmen enthalten sind und diese somit keine Wahrscheinlichkeit besitzen, in die Stichprobe gezogen zu werden. Unterscheiden sich für Personen in der Grundgesamtheit in Abhängigkeit vom Auswahlrahmen die Wahrscheinlichkeiten in die Stichprobe gezogen zu werden, spricht man von einem Coverage Error (Bartsch, 2012).

## 5. *Nonresponse Error*

Nehmen aufgrund von Verweigerung oder Nichterreichbarkeit nicht alle der in die Stichprobe gezogenen Personen an einer Befragung teil, kann ein Bias durch den sogenannten Nonresponse Error entstehen. Hier wird zusätzlich zwischen Unit- und Item-Nonresponse differenziert (Engel und Schmidt, 2014). Von Unit-Nonresponse wird gesprochen, wenn ein sogenannter Komplettausfall vorliegt. Item-Nonresponse

---

<sup>1</sup> Zu Imputationsverfahren siehe Allison (2000) oder Spieß (2010). Unterschiedliche Gewichtungszeduren werden in Gabler und Ganninger (2010) sowie in dieser Arbeit im Rahmen von Kapitel 4.5 beschrieben.

tritt auf, wenn sich Personen zwar an einer Umfrage beteiligen, allerdings nicht alle der gestellten Fragen beantworten. Die Verweigerung einer Antwort auf die Frage nach dem monatlichen Einkommen ist das klassische Beispiel für eine Frage, die in der Regel besonders von Item-Nonresponse betroffen ist (Diekmann, 2007).

Fehler durch Noncoverage und Nonresponse sind die für die Fragestellung dieser Arbeit zentralen Elemente des Total Survey Errors. Hiervon ist insbesondere die Datenqualität von Online- und Telefonbefragungen betroffen, sodass es mit diesen Modi schwierig und teilweise nur mit erheblichem Kostenaufwand möglich ist, Repräsentativität zu gewährleisten (Brick, 2011; Couper, 2011; Engel et al., 2004; Fan und Yan, 2010; Kaczmirek, 2015; Lavrakas, 2010). Die folgenden Kapitel greifen die Problematik um Noncoverage und Nonresponse auf und diskutieren differenzierter, wann diese zum Problem für die Datenqualität werden können.

### **2.1.1 Noncoverage: Systematische Unterrepräsentation von Bevölkerungsgruppen**

#### **Onlineumfragen**

Bei der Befragung über das Internet steht besonders die Frage im Zentrum, ob die definierte Grundgesamtheit über das Internet erreichbar ist. Vor allem bei der Realisierung bevölkerungsrepräsentativer Umfragen wird dies zum Problem. Denn wie Couper (2000) bereits festgestellt hat, leiden Onlineumfragen in besonderem Maße unter Noncoverage. Die Ziehung einer klassischen Zufallsstichprobe, die inferenzstatistische Analysen erlaubt, ist daher praktisch nicht umzusetzen (Baur und Florian, 2009). Einer der Hauptgründe dafür ist die auch heute noch vorherrschende „technische Barriere“ (Kaczmirek, 2015) oder auch „digitale Kluft“ (Wagner und Hering, 2014). Personen, die keinen Computer mit Zugang zum Internet besitzen und dieses privat nicht nutzen, werden hierdurch von Onlinebefragungen ausgeschlossen. Da es in Deutschland nicht zufällig ist, wer das Internet privat nutzt und wer nicht, entsteht eine Quelle systematischer Verzerrung. Folglich unterscheiden sich die Nutzer von den Nicht-Nutzern derart, dass die Datenqualität von Onlinebefragungen hierdurch beeinflusst wird (Bethlehem, 2015; Engel et al., 2004). Das Potential für Verzerrungen durch Noncoverage ist von zwei Faktoren abhängig:

1. Von dem Anteil an Personen, die nicht im Auswahlrahmen enthalten und somit unterrepräsentiert sind,

Tabelle 2.1: Entwicklung der Internetnutzung in Deutschland von 1997 bis 2014 in Prozent

	1997	2000	2003	2006	2009	2010	2011	2012	2013	2014
Gesamt	6,5	28,6	53,5	59,5	67,1	69,4	73,3	75,9	77,2	79,1
Zuwachs*		64,0	22,0	3,0	2,0	13,0	6,0	4,0	2,0	2,0
Männer	10,0	36,6	62,6	67,3	74,5	75,5	78,3	81,5	83,5	83,7
Frauen	3,3	21,3	45,2	52,4	60,1	63,5	68,5	70,5	71,1	74,6
14-19 Jahre	6,3	48,5	92,1	97,3	97,5	100,0	100,0	100,0	100,0	100,0
20-29 Jahre	13,0	54,6	81,9	87,3	95,2	98,4	98,2	98,6	97,5	99,4
30-39 Jahre	12,4	41,1	71,1	80,6	89,4	98,9	94,4	97,6	95,5	97,4
40-49 Jahre	7,7	32,2	67,4	7,2	80,2	81,9	90,7	89,4	88,9	93,9
50-59 Jahre	3,0	22,1	48,8	60,0	67,4	68,9	69,1	76,8	82,7	82,1
Ab 60 Jahre	0,2	4,4	13,3	20,3	27,1	28,2	34,5	39,2	42,9	45,4
In Ausbildung	15,1	58,5	91,6	98,6	98,0	100,0	100,0	100,0	100,0	100,0
Berufstätig	9,1	38,4	69,9	74,0	82,3	82,4	87,0	90,7	89,6	92,8
Rentner/n. berufst.	0,5	6,8	21,3	28,3	34,7	36,4	45,0	44,7	50,2	51,3

\*gegenüber dem Vorjahr in Prozent

Quelle: ARD-Onlinestudie 1997; ARD/ZDF-Onlinestudie 2000-2014

## 2. Von dem Grad der Unterschiedlichkeit zwischen Nutzern und Nicht-Nutzern (Bartsch, 2012; Dillman et al., 2014).<sup>2</sup>

Um deutlich zu machen, wie sich die Internetnutzung in Deutschland insgesamt entwickelt hat und sich die Nutzer von den Nicht-Nutzern unterscheiden, gibt Tabelle 2.1 einen Überblick hierüber.

Lag der Anteil an Internetnutzern laut ARD/ZDF-Onlinestudie im Jahr 1997 noch bei 6,5 Prozent, ist dieser über die Jahre hinweg stetig gewachsen, bis 2014 schließlich ein Niveau von 79,1 Prozent verzeichnet werden konnte. Die jährliche Zuwachsrate hat sich aufgrund des immer höheren Niveaus weiter abgeschwächt und lag in den Jahren 2013 und 2014 im Vergleich zum Vorjahr nur noch bei jeweils zwei Prozentpunkten (siehe hierzu ebenso Van Eimeren und Frees 2014). Der Ausstattungsgrad privater Haushalte ist laut Statistischem Bundesamt (2015a) im Jahr 2014 mit 78,8 Prozent etwas niedriger. Seit 2013

<sup>2</sup> Relativ zu vernachlässigen ist diese Problematik bei der Durchführung von face-to-face sowie postalischen Befragungen. Denn sind für die persönlich-mündliche oder schriftliche Beantwortung eines Fragebogens keine zusätzlichen Hilfsmittel wie ein Computer mit Internetzugang oder ein Festnetztelefon nötig. Andererseits besteht durch die gängigen Stichprobenverfahren wie „Random Route“ oder „Address Random“ eine relativ geringe Wahrscheinlichkeit für Noncoverage. Wenn Interviewer für eine face-to-face Befragung per „Random Route“ die jeweiligen Zielhaushalte auswählen, ist darüber hinaus kein Auswahlrahmen im klassischen Sinne, wie eine Liste aller zur Grundgesamtheit gehörenden Personen, nötig. Denn zur Bestimmung des jeweiligen Zielhaushaltes bekommt der Interviewer ab einem bestimmten Punkt einen zufälligen Laufweg zugewiesen, nach dem er den jeweiligen Haushalt auswählt. Werden für postalische Befragungen Adressen zufallsbasiert aus dem Melderegister gezogen („Address Random“), besteht ebenfalls eine nur geringe Wahrscheinlichkeit für Noncoverage. Die Vor- und Nachteile registerbezogener Stichprobenziehung werden ausführlich in Kapitel 4.3 (Melderegister) diskutiert.

ist zudem ein deutlicher Anstieg der mobilen Onlinenutzung zu beobachten (Van Eimeren und Frees, 2014).

Kosten- und Zeitersparnisse können bei der Durchführung von Onlinebefragungen allerdings nicht darüber hinwegtäuschen, dass dieser Modus von Noncoverage betroffen ist und repräsentative Befragungen der allgemeinen Bevölkerung hiermit nur schwer umzusetzen sind. Systematisch unterscheiden sich Nutzer und Nicht-Nutzer vor allem hinsichtlich Alter, Bildung, Geschlecht, Berufsstatus, Einkommen und Familienstatus (Bandilla et al., 2009; Bartsch, 2012; Baur und Florian, 2009; Dillman et al., 2014; Van Eimeren und Frees, 2014). Vor allem Menschen in den Altersgruppen von 14 bis 49 Jahren nutzen das Internet zumindest gelegentlich (Van Eimeren und Frees, 2014). Unter den 50-59 und den ab 60-Jährigen ist der Anteil an gelegentlichen Nutzern allerdings deutlich geringer. Wie Tabelle 2.1 zeigt, ist in diesen Altersgruppen ein Trend zunehmender Verbreitung sowie Nutzung zu beobachten, der sich auch in Zukunft fortsetzen sollte. Hierdurch wird das Potential für einen Bias durch Noncoverage in Bezug auf Altersunterschiede zukünftig sukzessive kleiner werden (Bandilla et al., 2009; Kaczmirek, 2015). Aus aktueller Perspektive bestehen hinsichtlich soziodemografischer Dimensionen allerdings auffällige Unterschiede, sodass bestimmte Bevölkerungsgruppen bei Befragungen über das Internet systematisch unterrepräsentiert sind (Van Eimeren und Frees, 2014). Dies führt in der Folge zu einer größeren Wahrscheinlichkeit für einen Bias durch Noncoverage.

In den Niederlanden wurde im Zuge der Longitudinal Internet Studies for the Social Sciences (LISS)<sup>3</sup> versucht, eine Lösung für diese Problematik zu finden. Um auch Haushalte ohne Internetanschluss in das Panel zu integrieren, wurden diese mit einem Zugang ausgestattet.<sup>4</sup> Wenn nötig, wurde darüber hinaus ein Computer für die Teilnahme zur Verfügung gestellt (Leenheer und Scherpenzeel, 2011). Im Ergebnis hat sich gezeigt, dass sich die Personen im Panel weiterhin von der niederländischen Population unterscheiden. Durch die Integration von Haushalten ohne Internetanschluss, die andernfalls nicht hätten teilnehmen können, sind diese allerdings kleiner geworden (Leenheer und Scherpenzeel, 2011, 17). Der offensichtliche Nachteil dieses Vorgehens liegt jedoch in den sehr hohen Kosten, die durch die Ausstattung der Haushalte mit dem notwendigen Equipment zur Beantwortung der Fragen entstehen. Hiermit fällt ein wesentlicher Vorteil der Onlinebefragung weg.

Ist man als Forscher daran interessiert, eine Befragung unter den Nutzern des Internets

---

<sup>3</sup> Für weitere Informationen über das LISS Panel siehe [www.lissdata.nl](http://www.lissdata.nl)

<sup>4</sup> Definitionen und unterschiedliche Formen von Panelbefragungen werden in Kapitel 6.1 differenzierter beschrieben.

durchzuführen, wird man vor eine weitere nicht unerhebliche Herausforderung gestellt. Ob für die Generierung einer Stichprobe hierfür ein Auswahlrahmen vorhanden ist, hängt davon ab, wie die jeweilige Grundgesamtheit definiert wird. Für repräsentative Befragungen der allgemeinen Bevölkerung über das Internet gibt es faktisch keinen geeigneten Auswahlrahmen (Brick, 2011). In der Praxis bedeutet dies, dass keine Liste von E-Mail Adressen vorhanden ist, aus der eine Zufallsstichprobe gezogen werden kann. Wird die Grundgesamtheit beispielsweise als wahlberechtigte Bevölkerung Deutschlands definiert, ist die Ziehung einer Stichprobe sowie die direkte Kontaktierung über E-Mail nicht möglich.

Werden Onlinebefragungen innerhalb von Universitäten oder Betrieben durchgeführt, existiert diese Problem nicht, da solche Listen hier in der Regel vorhanden sind. Häufig wird deswegen auf sogenannte „Opt-In-Befragungen“ zurückgegriffen. Hierbei wird keine Stichprobe aus einer entsprechenden Grundgesamtheit gezogen, sondern passiv durch Werbung oder Pop-Up's auf den entsprechenden Internetseiten versucht, Teilnehmer zu rekrutieren (Bartsch, 2012; Couper, 2000; Stoop, 2006; Vehre, 2011). Dieses Vorgehen ist für die Befragung von Nutzern einer Internetseite oder Kunden eines bestimmten Unternehmens geeignet. Da sich Leser spezifischer Internetseiten allerdings in der Regel systematisch von den Nicht-Lesern unterscheiden, ist diese Vorgehen für bevölkerungsrepräsentative Befragungen ungeeignet.

Aufgrund dieser Tatsache wird häufig „offline“ rekrutiert (Bartsch, 2012). Im Zuge dessen wird in der Regel zunächst eine Zufallsstichprobe nach den gängigen Verfahren generiert. Daraufhin wird per Telefon oder mündlich-persönlich versucht, im Rahmen einer ersten Umfrage nach Internetnutzern zu filtern, die benötigten Kontaktinformationen zu eruieren, um darauffolgend hierüber zu einer Onlinebefragung einladen zu können. Einem solchen Vorgehen sind Probleme durch Noncoverage in mehrfacher Hinsicht inhärent. Denn so muss nicht nur die Frage nach systematischen Unterschieden zwischen Nutzern und Nicht-Nutzern des Internets gestellt werden. Hieran knüpft ebenso die Frage danach an, welche sowohl Coverage- als auch Nonresponse-Probleme mit dem Modus der Rekrutierung selbst einhergehen.

### **Mobile Onlineumfragen**

Durch die zunehmende Verbreitung von Mobiltelefonen und Smartphones (Mobiltelefon mit Verbindung zum Internet) sind diese als Medium zur Durchführung von Umfragen seit einigen Jahren immer wichtiger geworden und in den Fokus wissenschaftlicher Forschung



gerückt (Busse und Fuchs, 2015; Mavletova, 2013). Für interviewervermittelte Befragungen sind durch diesen Trend etwa neue Stichprobenansätze entwickelt worden (Dual-Frame) (Häder und Häder, 2009). Hier ist bereits eine relativ breite Forschungsbasis hinsichtlich möglicher Fehlerquellen vorhanden.

Auch bei der mobilen Internetnutzung konnte ein deutlicher Zuwachs verzeichnet werden: Im ersten Quartal 2014 nutzen bereits insgesamt 50 Prozent das Internet in Deutschland, zumindest gelegentlich, auch mobil über Smartphones, Tablets oder Notebooks (ARD/ZDF-Onlinestudien 2009-2014; siehe Tabelle 2.2). Bereits 2016 waren insgesamt 95,1 Prozent aller deutschen Haushalte mit einem Handy oder Mobiltelefon ausgestattet, wovon wiederum 48,9 Prozent das Internet ebenso mobil nutzten (Statistisches Bundesamt, 2015a).

Tabelle 2.2: Entwicklung der mobilen Internetnutzung in Deutschland von 2009 bis 2014 in Prozent

	zumindest gelegentlich genutzt						täglich genutzt	
	2009	2010	2011	2012	2013	2014	2013	2014
Gesamt	11	13	20	23	41	50	21	22
Männer	15	16	26	27	46	52	25	22
Frauen	8	10	13	20	36	47	18	22
14-19 Jahre	12	21	28	46	64	77	45	46
20-29 Jahre	18	16	34	40	68	74	44	48
30-39 Jahre	11	15	23	28	46	66	23	31
40-49 Jahre	10	13	16	15	42	42	18	12
50-59 Jahre	8	9	10	12	24	32	10	8
Ab 60 Jahre	9	4	7	9	14	21	6	2

Quelle: ARD/ZDF-Onlinestudien 2009-2014

Die logische Konsequenz daraus ist, dass immer mehr Personen, die zu einer Onlinebefragung eingeladen werden, diese mit einem mobilen Endgerät beantworten. Vor diesem Hintergrund bezeichnen Toepoel und Lugtig (2015, 155) Onlineumfragen bereits als „Mixed-Device Surveys“, da von Laptops über Smartphones sowie Tablets oder auch mehrere Endgeräte quasi „hybrid“ zur Beantwortung von Fragebögen im Internet genutzt werden. Hierdurch ergibt sich eine Reihe methodischer Probleme, da sich diese Geräte hinsichtlich Displaygröße, dem Modus der Operationalisierung sowie den damit einhergehenden technischen Möglichkeiten voneinander unterscheiden (Toepoel und Lugtig, 2015, 155).

Diese Entwicklung zeigt, dass die Mehrheit an Befragten, auch wenn sie über ein mobiles Endgerät verfügen, sich zumeist von zu Hause aus an Umfragen beteiligen (Bruijne

und Wijnant, 2013). Stellt man den potentiellen Teilnehmern die Wahl, sodass je nach Präferenz entschieden werden kann, wie der Fragebogen beantwortet wird, haben Toepoel und Lugtig (2014, 544) in einer Pilotstudie herausgefunden, dass insgesamt 57 Prozent den Fragebogen mit ihrem Smartphone beantwortet haben. In einer Studie aus dem Jahr 2015, die diesbezüglich im Kontext des LISS Panels durchgeführt wurde, hat sich ein anderes Bild gezeigt. Nach Lugtig und Toepoel (2016, 15) nutzt immer noch eine Mehrheit von 90 Prozent einen stationären Computer zur Teilnahme an einer Onlinebefragung. Hierbei ist allerdings die Tatsache anzumerken, dass im Kontext dieses Panels ein Mode-Switch möglich gemacht wurde. Wenn die Teilnehmer an vorhergehenden Wellen einen stationären Computer genutzt haben, hatten sie die Möglichkeit den Fragebogen der Folgewelle mit einem mobilen Endgerät zu beantworten. Der Befund von Lugtig und Toepoel (2016) spricht dafür, dass die Wahrscheinlichkeit hoch ist, dass Personen über die Wellen eines Panels hinweg dem Modus der initialen Teilnahme treu bleiben.

Bereits Fuchs und Busse (2009, 25) haben in einer Untersuchung auf Basis des Eurobarometers aus dem Jahr 2007 geschlussfolgert, dass es aufgrund niedriger Nutzerzahlen von 20 Prozent in Deutschland (europäischer Durchschnitt: 31 Prozent) sowie systematischer Unterschiede zwischen Nutzern und Nicht-Nutzern noch zu früh für bevölkerungsrepräsentative Umfragen über das mobile Internet sei. In Zukunft wird der Anteil an Personen, die einen Fragebogen über ein mobiles Endgerät beantworten, sukzessive größer werden (Lugtig und Toepoel, 2016; Rivella et al., 2014). Dies zeigt ebenso die Entwicklung der Nutzerzahlen in Tabelle 2.2. Da sich die wissenschaftliche Diskussion hierzu allerdings noch am Anfang befindet ist es von zentraler Bedeutung, die Forschung hinsichtlich potentieller Fehlerquellen in diesem Bereich weiter voran zu treiben, um diesen Modus weiter zu etablieren und dessen Potential für die wissenschaftliche Umfrageforschung besser als bisher nutzbar zu machen. Denn selbst wenn eine Einladung nicht explizit auf eine mobile Variante des Fragebogens verweist, antwortet ein Teil der Befragten mit einem mobilen Endgerät (Rivella et al., 2014).<sup>5</sup> De Bruijne und Wijnant (2014) konnten einen deutlichen Anstieg dieser unintended mobile respondents nachweisen. Zur Beantwortung der Fragen wurden dabei insbesondere Tablets genutzt, die Nutzung von Smartphones bewegt sich in dieser Untersuchung auf einem niedrigen Niveau (De Bruijne und Wijnant, 2014). Ist ein Fragebogen nicht für die Darstellung auf Smartphones, Tablets oder

---

<sup>5</sup> In der englischsprachigen Literatur werden diese Personen häufig als „unintended mobile respondents“ bezeichnet (Rivella et al., 2014, 6).

Notebooks optimiert, ist durch diesen Trend mit einem Measurement Error zu rechnen. Softwarelösungen wie etwa LimeSurvey sind in der Lage zu dokumentieren, mit welchem Browser auf den Fragebogen zugegriffen wird. So kann im Nachhinein darauf reagiert werden, wenn unintended mobile response beobachtet wird (De Bruijne und Wijnant, 2014, 13). Vor diesem Hintergrund empfiehlt es sich bereits heute bei der Durchführung von Onlinebefragungen die Darstellung des Fragebogens ebenso für mobile Endgeräte zu optimieren. Erste Untersuchungen bezüglich eines möglichen Measurement Errors durch unterschiedliche Displaygrößen von mobilen Endgeräten und einer allgemein anderen Form der Eingabe (digitale Tastatur) haben ergeben, dass das jeweilige Endgerät, wenn die Programmierung und das Design der Fragen angepasst sind, keinen nennenswerten Einfluss auf das Antwortverhalten hat (Lugtig und Toepoel, 2016). Als einzigen Unterschied konnte Mavletova (2013) feststellen, dass die Antworten auf offene Fragen kürzer ausgefallen sind, wenn der Fragebogen per mobilem Endgerät im Vergleich zu einem stationären Computer ausgefüllt wurde. Insgesamt ist die Forschungslage bezüglich eines etwaigen Measurement Errors in Abhängigkeit des jeweiligen Endgerätes noch uneindeutig (Toepoel und Lugtig, 2015, 158).<sup>6</sup>

Auch hier wird der Fokus insbesondere auf zwei Komponenten des Total Survey Error gelegt: Noncoverage und Nonresponse. Die vorliegende Arbeit leistet einen Beitrag zu der Frage, wie die Nutzergruppen mobiler Endgeräte zu charakterisieren sind und darüber hinaus, welche Personen dazu bereit wären, sich hierüber an einer wissenschaftlichen Umfrage zu beteiligen. Die schon beschriebene Noncoverage-Problematik im Rahmen von Onlinebefragungen, die an einem stationären Computer ausgefüllt werden, ist, mit Blick auf die Verteilung von Smartphones und vorherrschende Nutzungsgewohnheiten des mobilen Internets in der Bevölkerung, gravierender (siehe Tabelle 2.2).

In Europa nutzen vor allem Männer, jüngere und höher gebildete Personen, die häufiger in Großstädten leben und seltener verheiratet beziehungsweise geschieden sind, das Internet mobil (Fuchs und Busse, 2009; Rivella et al., 2014). In Anbetracht dieser systematischen Unterschiede und dem damit einhergehenden Potential für einen Bias stellt sich die Frage danach, wie sich das Nutzungsverhalten seit der Untersuchung von Fuchs und Busse (2009) über die Jahre hinweg verändert hat und wie sich dieses vor dem Hintergrund aktuellerer Zahlen darstellt. Tabelle 2.2 gibt eine Übersicht über die diesbezüglichen

---

<sup>6</sup> Für weitere experimentelle Untersuchungen bezüglich Messeffekten siehe Peytchev und Hill (2010); Toepoel und Lugtig (2014).

Ergebnisse der im Kontext der ARD/ZDF-Onlinestudien stattfindenden Befragungen. Zu erkennen ist dabei zunächst, dass der Nutzeranteil seit 2009 stetig zugenommen hat und im Jahr 2014 zumindest 50 Prozent der Teilnehmer das Internet gelegentlich mobil nutzen, täglich 22 Prozent. Das Statistisches Bundesamt (2015c) gibt in seinem aktuellen Jahresbericht gar eine Nutzerzahl von insgesamt 62 Prozent an, allerdings sind hier Personen ab 10 Jahren beteiligt. Auch aktuell zeigen sich systematische Unterschiede im Nutzungsverhalten hinsichtlich zentraler soziodemografischer Dimensionen. So wird das Internet vor allem in den unteren Altersgruppen und eher von Männern mobil genutzt. Nach den Zahlen des statistischen Bundesamtes für das Jahr 2015 sind mit 69 Prozent unter allen Erwerbstätigen sowie mit 88 Prozent unter allen Schülern und Studierenden, hohe Nutzeranteile zu verzeichnen (Statistisches Bundesamt, 2015c). Es ist besonders vom Alter abhängig, ob das Internet auch über mobile Endgeräte genutzt wird oder nicht. Besonders in den unteren Altersgruppen können mit Blick auf Tabelle 2.2 relativ hohe Nutzerzahlen beobachtet werden. Mit steigendem Alter fallen diese allerdings deutlich ab. Während die Zahl der Nutzer des (stationären) Internets in der Altersgruppe der 50-59 Jährigen auf einem hohen Niveau von 82,1 Prozent liegt (siehe Tabelle 2.1), ist der Anteil von mobilen Nutzern mit 32 Prozent 2014 vergleichbar niedrig. Würde man eine Befragung lediglich über das mobile Internet durchführen, wäre vor dem Hintergrund dieser Zahlen mit erheblichen Verzerrungen zu rechnen. Eine mobile Variante kann deswegen aktuell in einem Mixed-Mode Design als Ergänzung dienen, damit die Personen, die es präferieren den Fragebogen mit einem mobilen Endgerät zu beantworten, auch die Möglichkeit dazu haben. Nichtsdestotrotz wird sich der Trend hin zu einer breiteren Ausstattung und Nutzung in der Bevölkerung fortsetzen, wodurch wissenschaftlich methodische Forschung bereits heute von Bedeutung ist, um mögliche verzerrende Effekte aufzudecken. So kann hierdurch versucht werden, die Antwortrate zu erhöhen und das Potential für einen Nonresponse Error dadurch zu verringern. Über zu erzielende Antwortraten im Rahmen von selbstadministrierten mobilen Onlinebefragungen liegen bisher sehr wenig Erkenntnisse vor. Lugtig und Toepoel (2016, 548) berichten in ihrer Untersuchung von einer „completion rate“ von 45 Prozent. Nutzen Personen das Internet mobil bedeutet dies allerdings nicht gleichzeitig, dass diese auch dazu bereit sind, einen Fragebogen hierüber zu beantworten. De Bruijne und Wijnant (2014) haben im Kontext des LISS Panels untersucht, inwiefern sich Personen in Abhängigkeit des jeweiligen Endgerätes, das zur Teilnahme genutzt wurde, voneinander unterscheiden. Dabei hat sich gezeigt, dass sich Nutzer stationärer

Computer von Smartphone-Nutzern und diese wiederum von Tablet-Nutzern unterscheiden. Smartphones werden zu einem geringeren Anteil unter den Jüngeren genutzt, während Tablets insbesondere in der arbeitenden Bevölkerung unter den 25 bis 54 Jährigen zur Beantwortung des Fragebogens zur Anwendung kommen (De Bruijne und Wijnant, 2014, 11). Beide Endgeräte werden eher von Frauen im Vergleich zu Männern genutzt. Werden Teilnehmer direkt nach ihrer Präferenz gefragt, lässt sich ein ähnliches Bild zeichnen: Mit den Kovariaten Alter, Geschlecht, Erwerbsstatus, Bildung (eher höher gebildete) sowie die Zusammensetzung des Haushaltes (eher in Mehrpersonenhaushalten) lässt sich die Präferenz für die Nutzung eines Tables schätzen, während lediglich Alter und Bildung die Präferenz für Smartphones beeinflussen (De Bruijne und Wijnant, 2014, 11).

### **Telefonumfragen**

Probleme durch Noncoverage sind bei der Durchführung telefonischer Befragungen zwar existent, allerdings in geringerem Umfang als dies bei Onlinebefragungen der Fall ist. Denn in Deutschland herrscht, anders als beispielsweise in den USA, eine unverändert hohe Abdeckung mit Festnetzanschlüssen (Lavrakas, 2010). Diese lag im Jahr 2014 bei 90,5 Prozent (Statistisches Bundesamt, 2015b). Zudem ist die Festnetzpopulation nach Meier et al. (2014) der Gesamtbevölkerung hinsichtlich soziodemografischer Merkmale relativ ähnlich. Die Unterschiede in der Geschlechter-, Alters-, Bildungs- oder Einkommensverteilung sind dementsprechend gering. Die Wahrscheinlichkeit eines Bias durch Noncoverage in Abhängigkeit vom Besitz eines Festnetzanschlusses existiert dennoch, da es in Deutschland einen immer größer werdenden Anteil an Personen gibt, die nur noch über den Mobilfunk erreichbar sind (Mobile-Onlys) (Busse und Fuchs, 2015).

Zur Generierung von Stichproben für Telefonumfragen konnte lange Zeit auf das Telefonbuch zurückgegriffen werden. Durch den Wegfall der gesetzlichen Eintragungspflicht Anfang der 1990er Jahre ist dieses als Auswahlrahmen für allgemeine Bevölkerungsumfragen jedoch nicht mehr geeignet (Deutschmann und Häder, 2002; Häder et al., 2009; Schnell, 2012). Zunächst konnten Personen Widerspruch gegen die Eintragung einlegen (Opt-Out-Regelung). Ab 1996 muss man für eine Eintragung wiederum selbst aktiv werden (Opt-in-Regelung) (Hüfken, 2014). Die Nicht-Eintragung bei der Vergabe von Neuanschlüssen ist mittlerweile allerdings die Regel, sodass Unterschiede zwischen Eingetragenen und Nicht-Eingetragenen über die Zeit immer größer geworden sind und letztere faktisch keine Auswahlwahrscheinlichkeit besitzen (Meier et al., 2014).

Bereits Heckel (2002, 30) ist zu dem Ergebnis gekommen, dass in Gemeinden mit 500.000 Einwohnern und mehr im Jahr 2000 bereits ein Anteil von 27,6 Prozent über keinen Telefonbucheintrag mehr verfügt. Der Trend zum Nicht-Eintrag hat sich über die Jahre hinweg deutlich verstärkt, sodass die Dichte sukzessive abgenommen hat. Folgt man der Übersicht von Häder et al. (2009, 33) wird dies besonders deutlich: Während im Jahr 1999 die Eintragungsdichte anteilig an der Gesamtzahl vorhandener Nummernblöcke<sup>7</sup> noch bei 47,3 Prozent lag, ist diese über die Jahre hinweg immer kleiner geworden und lag im Jahr 2008 bereits bei 18,8 Prozent. Diese Anteile sind allerdings nicht mit der Eintragungsdichte insgesamt zu verwechseln, da sich diese Zahl auf Nummernblöcke bezieht, in denen viele Rufnummern enthalten sind, hinter denen kein Anschluss steht. Schnell (2012, 272) schlussfolgert in Bezug zu Klingler und Müller (2001), dass der Anteil nicht eingetragener Anschlüsse in Abhängigkeit des Urbanisierungsgrades sowie zwischen den neuen und alten Bundesländern variiert und sich in einem Spektrum zwischen 18,9 Prozent bis 39,4 Prozent bewegt. Da die Befragung von Klingler und Müller (2001) allerdings auf kontaktierten Rufnummern basiert, weist Schnell (2012, 272) ebenso darauf hin, dass der Anteil nicht eingetragener Anschlüsse hierdurch tendenziell unterschätzt werden könnte. Mittlerweile ist vor diesem Hintergrund bei fortsetzendem Trend zum Nicht-Eintrag davon auszugehen, dass der Anteil aktuell bereits bei über 50 Prozent liegt.

Mit dem Gabler-Häder-Design, das auf dem Random-Digit-Dialing-Verfahren beruht, wurde hierfür in Deutschland eine Lösung gefunden (Gabler und Häder, 1997). Auf Basis des Telefonbuches werden zunächst Nummernblöcke gebildet, wobei die letzten zwei Ziffern abgeschitten und durch „00“ bis „99“ ersetzt werden. Mit diesen Hunderterblöcken wird ein Auswahlrahmen generiert, der nahezu alle vorhandenen Rufnummern unabhängig davon, ob diese eingetragen sind oder nicht, enthält (Meulemann und Hagenah, 2006, 23 f.). Folgend wird überprüft, ob in den einzelnen Blöcken eingetragene Nummern vorhanden sind. Ist dies der Fall, wird entsprechend versucht zu allen Nummern dieses Blocks Kontakt herzustellen. Somit wird es möglich, jegliche Haushalte eine gleiche und bekannte Auswahlwahrscheinlichkeit zuzuordnen unabhängig davon, ob diese im Telefonbuch verzeichnet sind oder nicht (Gabler und Häder, 1997).<sup>8</sup>

---

<sup>7</sup> Die Struktur von Festnetzzurufnummern in Blöcken muss man sich als 100 aufeinanderfolgende Ziffernfolgen vorstellen (Häder et al., 2009, 30). Innerhalb eines Ortsnetzbereiches gibt es jeweils eine konstante Vorwahl, die in Kombination mit Blöcken von 100-199, 200-299 etc. an die Provider vergeben werden. In diesen Blöcken sind wiederum nicht alle Kombinationen an Kunden vergeben, wodurch zusätzliche Lücken entstehen können (Häder et al., 2009, 30).

<sup>8</sup> Für eine differenziertere Darstellung des Gabler-Häder-Designs siehe Gabler und Häder (1997) oder Häder und Häder (2014). Für Random-Digit-Dialing siehe Waksberg (1978). Die Vor- und Nachteile dieses

Durch die zunehmende Verbreitung des Mobilfunks wurde die Umfrageforschung vor neue Herausforderungen gestellt. Mit 93,6 Prozent hat dieser das klassische Telefon mittlerweile überholt (Statistisches Bundesamt, 2015a). In Deutschland gehören nach Busse und Fuchs (2015) weniger als 20 Prozent zu der Gruppe der „Mobile-Onlys“. Häder et al. (2009) beziffern den Anteil auf ca. 10 Prozent. Diese unterscheiden sich systematisch von der Festnetzpopulation. „Mobile-Onlys“ sind eher jünger, allein lebend, leben eher in städtischen Mietwohnungen und weisen häufig niedrige Einkommen auf (Busse und Fuchs, 2015; Lynn und Kaminska, 2012; Mohorko et al., 2013). Somit sind klassische Telefonumfragen, in denen ausschließlich über das Festnetz befragt wird von einem Coverage-Error betroffen, da „Mobile Onlys“ faktisch keine Auswahlwahrscheinlichkeit besitzen. Der Anteil an „Festnetz-Onlys“ ist in Deutschland mit etwa drei bis fünf Prozent zwar gering, diese unterscheiden sich ebenso hinsichtlich Bildung, Einkommen und weiterer Dimensionen systematisch vom Rest der Population (Busse und Fuchs, 2015; Häder et al., 2009).<sup>9</sup> Telefonbefragungen können vor diesem Hintergrund heutzutage nicht mehr alleinig über das Festnetz stattfinden. Die 2008 ins Leben gerufene „Cell Phone Task Force“ der AAPOR schlägt deswegen in ihrem aktuellen Bericht eine Kombination aus Mobilfunk- und Festnetzauswahlrahmen in einem „Dual-Frame-Modell“ vor.<sup>10</sup> So ist gewährleistet, dass sowohl „Mobile-Onlys“, „Festnetz-Onlys“ als auch Personen, die über beide Anschlüsse verfügen, die gleiche Auswahlwahrscheinlichkeit besitzen. Hiermit werden Coverage-Probleme bezüglich der Gruppe der „Mobile-Onlys“ eliminiert (Brick et al., 2006). In Deutschland ist vor diesem Hintergrund die „Auswahlgrundlage Mobilfunk“ des Arbeitskreis deutscher Markt- und Sozialforschungsinstitute e. V. (ADM) entstanden. Auch die AAPOR gibt regelmäßig Empfehlungen heraus, inwiefern bei der Stichprobenziehung gewichtet werden sollte. Da die Generierung einer Stichprobe nach dem Dual-Frame-Modell ein kompliziertes statistisches Unterfangen darstellen würde, gibt es nach Häder et al. (2009) bis heute hierzu nur wenige valide Erkenntnisse.

---

Vorgehens werden im folgenden Kapitel (2.1.2) diskutiert.

<sup>9</sup> Berücksichtigt werden muss zusätzlich, dass potentielle Teilnehmer heutzutage über mehrere Rufnummern (Festnetz, Mobilfunk) erreichbar sind oder ein Mobiltelefon von mehreren Personen genutzt wird („cell phone sharing“) (Busse und Fuchs, 2015). Somit besteht für diese Personen eine größere Auswahlwahrscheinlichkeit. In der Regel wird bei erfolgreicher Kontaktaufnahme nach der Anzahl an Rufnummern gefragt, über die eine Person erreichbar ist bzw. nach den Nutzern des Mobiltelefons. So kann im Nachhinein eine Gewichtung vorgenommen werden (Lavrakas, 2010).

<sup>10</sup> Eine registerbezogene Stichprobenziehung wäre eine weitere Möglichkeit. Für eine differenziertere Darstellung des Dual-Frame-Modells siehe Gabler und Ayhan (2007) oder auch Brick et al. (2006).

## 2.1.2 Nonresponse: Niedrige Antwortraten als Quelle systematischer Verzerrungen

### Telefonumfragen

Metastudien und international vergleichende Untersuchungen konnten wiederholt einen allgemeinen Trend abnehmender Antwortraten beobachten (Bethlehem et al., 2011; Brick, 2013; Engel und Schmidt, 2014; Engel et al., 2004; Groves und Peytcheva, 2008). Sowohl die kommerzielle Marktforschung, als auch wissenschaftliche Umfragen öffentlicher Institutionen sind hiervon betroffen (Aust und Schröder, 2009; Curtin et al., 2005; Kreuter, 2013a). So hat Baruch (1999) in einer Meta-Analyse von insgesamt 141 Studien festgestellt, dass Antwortraten wissenschaftlicher Umfragen über die Jahre hinweg stetig abgenommen haben. Differenzierter betrachtet ist die Allgemeinheit dieses Trends in Frage zu stellen, da die beobachtete Variation in den erzielten Antwortraten stark von der Feldarbeit, dem Design der jeweiligen Umfrage sowie dem Länderkontext abhängig ist (De Heer, 1999; De Leeuw und De Heer, 2002; Saßenroth, 2013). So kommen Engel et al. (2004) zu dem Ergebnis, dass unter Kontrolle relevanter Designmerkmale ein nur schwacher allgemeiner rückwärtiger Trend zu beobachten ist. Länderübergreifend variieren Antwortraten teilweise stark. In der dritten Welle des European Social Survey (ESS) bewegen sich diese in einem Spektrum von 46 bis 73 Prozent (Stoop et al., 2010). Zu bemerken ist hier, dass die Befragungen des ESS mündlich-persönlich durchgeführt werden. Hiermit sind immer noch die höchsten Antwortraten, gleichzeitig aber auch die höchsten Kosten, verbunden (Schnell, 2012).

Da es in der Praxis bei der Durchführung allgemeiner Bevölkerungsumfragen so gut wie unmöglich ist, Antwortraten von hundert Prozent zu erzielen, ist die Auseinandersetzung mit dem Thema Nonresponse schon so alt wie die Umfrageforschung selbst (Vehre, 2011). Unproblematisch ist Nonresponse, wenn es zufällig ist, welche Personen teilnehmen und welche nicht. Sobald hier allerdings eine Systematik auftritt und Teilnehmer hierdurch die gestellten Fragen anders beantworten, entsteht ein Bias durch Nonresponse. Wissenschaftliche Untersuchungen haben sich seit den 1980er Jahren nach Singer (2006) vor allem darauf fokussiert, welche Faktoren für Nonresponse verantwortlich sind und wie Antwortraten insgesamt erhöht werden können. Heute wird allerdings eher danach gefragt, wie die Höhe der Antwortrate mit einem möglichen Bias zusammenhängt (Groves und Peytcheva, 2008). Da jede Umfrage in unterschiedlichem Ausmaß betroffen ist und in der Regel nur wenige



bis keine Informationen über Personen vorliegen, die nicht teilgenommen haben, wird das Thema Nonresponse auch in Zukunft ein wichtiges Forschungsfeld bleiben.

Heutzutage sind selbst bei bester Planung und Durchführung mit einzelnen Umfragemodi wie Telefon oder Online lediglich Antwortraten von 20 Prozent oder geringer zu erzielen (Fan und Yan, 2010). Kritisch ist besonders die Entwicklung der Teilnahmebereitschaft im Rahmen telefonischer Befragungen zu sehen. Hierbei sind die Antwortraten nicht nur in Deutschland sondern auch in den USA und einigen europäischen Ländern auf ein niedriges Niveau gesunken (Couper, 2011; Curtin et al., 2005; De Leeuw et al., 2007; Häder et al., 2009; Engel et al., 2012; Engel und Schmidt, 2014; Lepkowski et al., 2007; Link et al., 2008; Tourangeau, 2004). Werden wissenschaftliche Umfragen über das Telefon durchgeführt, sind Antwortraten von 20 bis 30 Prozent zur Regel geworden. Im telefonischen Rekrutierungsinterview des PPSM Access Panels konnte zu insgesamt 55,7 Prozent der in die Stichprobe gezogenen Personen Kontakt hergestellt werden, von denen wiederum lediglich 24,5 Prozent an der ersten Befragung teilgenommen haben (Engel, 2013). In der CELLA-Hauptstudie haben sich in einer Befragung über das Festnetz lediglich 20 Prozent zu einer Teilnahme bereit erklärt (Häder et al., 2009). Auch Dillman et al. (2014, 259) berichten für regelmäßige telefonische Befragungen eines großen Umfrageinstituts ([www.pewresearch.org](http://www.pewresearch.org)), dass die Antwortraten im Jahr 1997 von 36 Prozent um insgesamt 27 Prozentpunkte auf schließlich 9 Prozent im Jahr 2012 gefallen sind.

Werden telefonische Befragungen von öffentlichen Institutionen durchgeführt, wie in Deutschland durch die statistischen Landes- oder Bundesämter, können in Verbindung mit vorheriger postalischer Ankündigung des Kontaktes noch die höchsten Antwortraten von über 50 Prozent erreicht werden (Vehovar et al., 2012). Für öffentliche Institutionen besteht in der Regel dadurch, dass sie bekannter sind als andere, eine höhere Legitimität für Befragungen und ein ausgeprägteres Vertrauen seitens der befragten Personen, was sich durch den sogenannten „Sponsorship-Effekt“ in einer höheren Teilnahmebereitschaft manifestiert (Dillman et al., 2014). So konnten Ladik et al. (2007) nachweisen, dass Antwortraten einer universitären Befragung in der Heimatstadt der Universität durch den hier vorherrschenden Bekanntheitsgrad höher liegen, als bei vergleichbaren Befragungen in „fremden“ Städten.

Die Hauptursachen niedriger Antwortraten sind vor allem in Verweigerung und Nichterreichbarkeit zu sehen (Engel und Schmidt, 2014; De Leeuw und De Heer, 2002). Wird ein Haushalt durch einen Interviewer per Telefon kontaktiert, besteht ein nur sehr kurzes

Zeitintervall, indem dieser den Befragten von der Seriosität des Umfragegesuchs überzeugen kann. In der Regel kann durch das Stichprobenverfahren (Gabler-Häder-Design) keine Vorankündigung des Anrufs stattfinden, da über die Rufnummer hinaus keine weiteren Kontaktinformationen vorliegen. Werden Personen ohne vorherige Ankündigung kontaktiert, spricht man von einem kalten Kontakt (De Leeuw et al., 2007, 414). Dadurch, dass Werbe- und Verkaufsanrufe zugenommen haben, hat sich das Umfrageklima in der Bevölkerung insgesamt verschlechtert (Engel et al., 2012; De Heer, 1999). In Deutschland melden sich trotz der seit 2009 verschärften gesetzlichen Regelungen zu unlauterer Telefonwerbung immer wieder Personen bei den zuständigen Verbraucherzentralen und beschweren sich über Werbe- und Verkaufsanrufe ohne vorherige Einwilligung (Verbraucherzentrale Nordrhein-Westfalen, 2012, 3). Eine im Jahr 2012 durch verschiedene Verbraucherzentralen durchgeführte Umfrage ist zu dem Ergebnis gekommen, dass 92 Prozent der betroffenen Verbraucher nicht mit dem erfolgten Werbeanruf einverstanden gewesen sind (Verbraucherzentrale Nordrhein-Westfalen, 2012, 8 f.). Im Durchschnitt wurden die befragten Personen mit 13 Werbeanrufen pro Woche belästigt und insgesamt 7 Prozent der Anrufenden hat sich dabei als Mitglied eines Markt- und Meinungsforschungsinstituts ausgegeben.

In Verbindung mit einer „kalten“ Kontaktierung hat dies zu einer Zunahme spontaner Verweigerungen geführt, sodass der Hörer häufig unmittelbar wieder aufgelegt wird. Curtin et al. (2005) sehen in der Zunahme von Werbe- und Verkaufsanrufen sowie häufigen Umfrageeinladungen den Hauptfaktor für zunehmende Nichterreichbarkeit und Verweigerungen in den letzten 25 Jahren. Die Zunahme von Nichterreichbarkeit kann in Kombination mit neueren Telefongeräten so interpretiert werden, dass der Hörer bei einer unbekannt anrufenden Telefonnummer gar nicht erst abgehoben wird. Der Anteil an spontanen Verweigerern lag im Rahmen des Rekrutierungsinterviews des PPSM Access Panels bei 44 Prozent von den insgesamt 56.848 kontaktierten Haushalten (Engel et al., 2012; Engel und Schmidt, 2014). Vorliegende Studien sind zu dem Ergebnis gekommen, dass Verweigerungsraten pro Jahr um 0,3 Prozentpunkte steigen (De Leeuw und De Heer, 2002, 22). Werden Stichproben nach dem Gabler-Häder-Design generiert, kommt es zusätzlich häufig zu Kontaktausfällen dadurch, dass viele Nummern, die zufällig generiert werden, nicht vergeben sind (Meier et al., 2014).

Nichterreichbarkeit wird weiterhin durch technologischen Fortschritt und gesellschaftlichen Wandel beeinflusst. Denn immer mehr Haushalte verfügen über eine digitale Rufnummernerkennung („Call Screening Technologie“) und Anrufbeantworter (Aust und Schröder,

2009; Brick, 2013; Link et al., 2008; Tourangeau, 2004). So muss man zum einen bei einer unbekannt anrufenden Nummer den Hörer entweder gar nicht erst abnehmen. Zum anderen kann die Privatsphäre zusätzlich durch einen Anrufbeantworter abgeschirmt werden. Durch die Zunahme von Singlehaushalten kommt es immer häufiger zu berufsbedingten Abwesenheitszeiten, was es heutzutage ebenso schwieriger macht, Personen über das Telefon zu erreichen (Engel und Schmidt, 2014).

Für Mobilfunkbefragungen werden im Vergleich zu Befragungen über das Festnetz häufig geringere Antwortraten berichtet (Vehovar et al., 2012). Eine niedrige Antwortrate ist allerdings nicht notwendigerweise mit einem hohen Bias durch Nonresponse verbunden und vice versa (Brick, 2013; Groves und Peytcheva, 2008; Klausch et al., 2015; Kreuter und Olson, 2013; Peytchev, 2013; Stoop, 2005; Wagner, 2012). Die Wahrscheinlichkeit ausfallbedingter Verzerrungen ist einerseits umso größer, je geringer die Antwortrate insgesamt ausfällt. Andererseits wird das Ausmaß vor allem durch den Grad der Unterschiedlichkeit zwischen Teilnehmern und Nicht-Teilnehmern bestimmt (Schnell, 2012).

Da es in der Regel nicht zufällig ist, wer teilnimmt und wer nicht, unterscheiden sich diese beiden Gruppen häufig systematisch voneinander. Hierfür sind sogenannte *Selektionseffekte* verantwortlich. Aus der Forschungspraxis ist bekannt, dass sich Teilnehmer und Nicht-Teilnehmer anhand spezifischer Merkmale selektieren (Dillman et al., 2002). Das bekannteste Beispiel ist wohl der „Bildungsbias“, denn an wissenschaftlichen Umfragen nehmen in der Regel höher gebildete Personen teil (Engel und Schmidt, 2014). Diese Teilnehmer rekrutieren sich darüber hinaus aus höheren Gesellschaftsschichten, sind tendenziell jünger, wohnen meist in urbanen Gebieten und gehören keiner ethnischen Minderheit an (Adua und Shap, 2010; Couper und Groves, 1996; Engel und Schmidt, 2014). Durch diese Unterschiede steigt die Wahrscheinlichkeit, dass Fragen anders beantwortet werden und somit auch ein höherer Bias verursacht wird. Dadurch, dass der Auswahlrahmen für telefonische Befragungen lediglich aus der Summe aller Festnetz und Mobilfunknummern besteht, ist die Operationalisierung eines möglichen Bias durch Nonresponse und somit auch die nachträgliche Gewichtung sehr beschränkt. Hier kann eine Charakterisierung der Personen, die nicht teilgenommen haben, lediglich über die Verknüpfung der jeweiligen Vorwahlen mit Gebietsdaten stattfinden.

Um trotzdem einige Informationen auch über diejenigen zu eruieren, die nicht teilgenommen haben, wird in der Praxis häufig versucht, die Nonrespondenten nach Ablauf der eigentlichen Feldphase nochmals zu kontaktieren (Bethlehem, 2009). Dieses Vorgehen ist

allerdings nur bedingt zielführend, da hier wiederum Nonresponse auftritt und darüber hinaus zusätzliche Kosten anfallen. Über ein flexibles Surveydesign oder auch den sogenannten „Basic Question Approach“ kann mit einer kürzeren Variante des eigentlichen Fragebogens versucht werden, zögerliche Personen von einer Teilnahme zu überzeugen (Cobben, 2007). Bereits Kersten und Bethlehem (1984) konnten diesbezüglich nachweisen, dass durch das Anbieten einer verkürzten Fragebogenvariante, die lediglich einige der zentralen Variablen beinhaltet, ein Teil der zögerlichen Personen zu einer Teilnahme motiviert werden können. Beim Aufbau des PPSM Access Panels wurde vor diesem Hintergrund in den Rekrutierungsinterviews eine Kern- und Exitvariante angeboten (Engel et al., 2012). Über die Exitvariante, die lediglich eine Frage enthielt, konnte dann festgestellt werden, dass die Umfrageerfahrung eine Rolle für die Teilnahmeentscheidung spielt: Diese wurde umso wahrscheinlicher, je mehr Umfrageerfahrung die potentiellen Befragten im Vorhinein bereits gesammelt hatten (Engel und Schmidt, 2014).

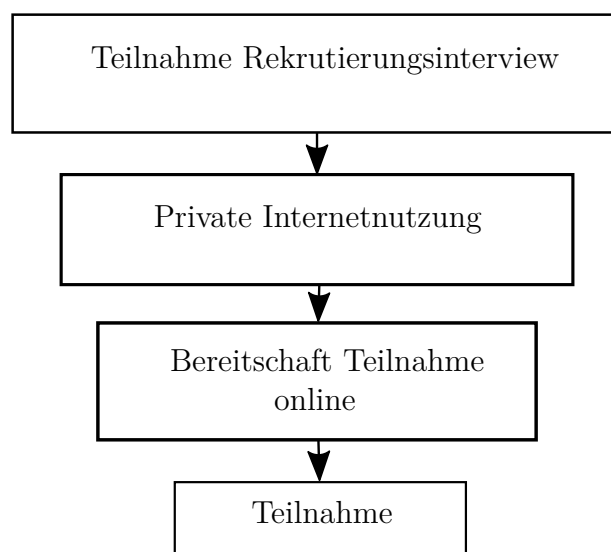
### **Onlineumfragen**

Onlineumfragen leiden ebenso wie Befragungen über das Telefon unter niedrigen Antwortraten (Fan und Yan, 2010). In Verknüpfung mit den Problemen durch Noncoverage ergibt sich ein besonderes Fehlerpotential, sodass inferenzstatistische Analysen auf Basis im Internet erhobener Daten kritisch zu hinterfragen sind. Allgemeine Bevölkerungsumfragen sind hierüber nach heutigem Stand lediglich mit erheblichem Kostenaufwand realisierbar (Baur und Florian, 2009; Leenheer und Scherpenzeel, 2011; Wagner und Hering, 2014).

Mit Onlinebefragungen lassen sich im Vergleich zu anderen Modi nachweislich nur geringere Antwortraten erzielen (Bech und Kristensen, 2009; Bethlehem, 2015; Millar und Dillman, 2011; Tourangeau, 2004). Shih und Fan (2008) haben in einer Meta-Analyse von 39 Studien festgestellt, dass die Antwortraten der untersuchten Onlinebefragungen im Vergleich zur postalischen Variante im Durchschnitt zehn Prozent niedriger ausgefallen sind. Aber auch hier ist es vom jeweiligen Design der Umfrage abhängig, wie erfolgreich Personen zu einer Teilnahme motiviert werden können. Bei einer Befragung unter Studierenden sind Kaplowitz et al. (2004) zu dem Ergebnis gekommen, dass bei feldexperimenteller Zuweisung und vorheriger E-Mail-Ankündigung keine Unterschiede im Teilnahmeverhalten zwischen Online- sowie postalischer Befragung beobachtet werden konnten. Dieses Vorgehen ist durch das Fehlen des Auswahlrahmens für allgemeine Bevölkerungsumfragen über das Internet allerdings nicht geeignet. So müssen potentielle Teilnehmer zunächst einmal

zwangsläufig die in Abbildung 2.2 dargestellten Selektionsstufen bis hin zur letztlichen Teilnahme durchlaufen. Hierbei fallen durch Nonresponse und Noncoverage Personen der ursprünglichen Ausgangsstichprobe systematisch aus, sodass zum Schluss lediglich ein kleiner Ausschnitt übrig bleibt. Im Rahmen eines Rekrutierungsinterviews wird zunächst danach gefragt, ob ein Internetanschluss verfügbar ist und auch privat genutzt wird. Bereits dieses erste Interview ist von Nonresponse betroffen, was überwiegend mit systematischen Verzerrungen einhergeht. Coverage-Effekte beim Übergang zur nächsten Stufe „private Internetnutzung“ lassen den Kreis an Personen weiter kleiner werden. In der Folge kann die Frage nach der Teilnahmebereitschaft an einer späteren Onlineumfrage lediglich einem Bruchteil derer gestellt werden, die am Rekrutierungsinterview teilgenommen hat. Wird seitens der Befragten Bereitschaft signalisiert, bedeutet dies allerdings nicht, dass sich letztendlich alle Personen beteiligen. In der Regel wird auch hier Nonresponse auftreten.

Abbildung 2.2: Selektionsstufen bei der Rekrutierung von Teilnehmern für eine Onlinebefragung in Anlehnung an Bandilla et al. (2009)



Jüngere Personen mit einem hohen Bildungsabschluss sind eher zur Teilnahme bereit und zeichnen sich durch eine höhere Wahrscheinlichkeit aus, die einzelnen Rekrutierungsstufen bis zur letztlichen Beantwortung des Onlinefragebogens zu durchlaufen (Bandilla et al., 2009, 137). Auf Basis der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) aus dem Jahr 2008 versuchen Bandilla et al. (2009), den Fehler durch Coverage und Noresponse auf den einzelnen Selektionsstufen zu quantifizieren. Im Ergebnis hat sich gezeigt, dass der Fehler durch Nonresponse durch die fehlende Teilnahmebereitschaft größer ist als der, der durch die letztliche Nicht-Teilnahme derer entsteht, die sich zuvor befragungsbereit erklärt hatten (Bandilla et al., 2009). In Abhängigkeit

des jeweiligen Rekrutierungsmodus kann versucht werden, zögerliche Personen durch Überzeugungsversuche seitens eines Interviewers zu einer Teilnahme zu bewegen.

Festzuhalten bleibt an dieser Stelle, dass das Potential für einen Bias durch Nonresponse bei der Rekrutierung von Teilnehmern für eine Onlinebefragung vor allem in der Frage nach der Teilnahmebereitschaft und hiermit einhergehenden Selektionseffekten begründet liegt. Dieses Ergebnis konnte bereits durch weitere Studien bestätigt werden (Bartsch, 2012). Um die Potentiale der Onlinebefragung in Zukunft noch besser als bisher nutzbar zu machen, besteht vor allem im Bereich der Rekrutierungsmechanismen Forschungsbedarf. Auch wenn die Verbreitung des Internets in Zukunft immer weiter zunimmt, wird das Problem des fehlenden Auswahlrahmens für bevölkerungsrepräsentative Umfragen weiterhin bestehen bleiben und somit daran anknüpfend auch die Nonresponse-Problematik bei der Frage nach der Teilnahmebereitschaft.

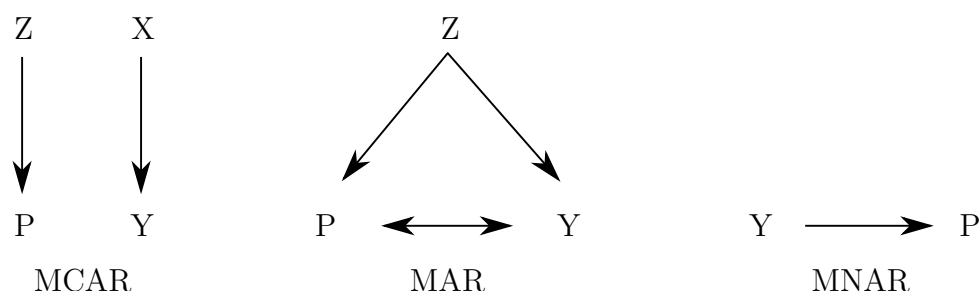
## 2.2 Struktur fehlender Werte

Im Kontext des Umfrageprozesses können fehlende Werte in unterschiedlicher Form auftreten. Hierbei ist nicht nur zwischen Item und Unit-Nonresponse zu unterscheiden. Untersucht man den Zusammenhang zwischen Nonresponse und einem möglichen Bias muss zwangsläufig die Frage danach gestellt werden, welche Mechanismen dazu führen, dass Personen einzelne Fragen nicht beantworten oder sich erst gar nicht an einer Umfrage beteiligen. Wichtig wird dies im Besonderen, um zunächst einmal theoretisch herzuleiten, wann fehlende Werte zum Problem für die Qualität der Daten und daraus entstehende Analysen werden können. Abbildung 2.3 stellt den Zusammenhang zwischen der Hilfs- oder auch Hintergrundvariable ( $Z$ ), der Antwortwahrscheinlichkeit (Item-Nonresponse) oder Teilnahmewahrscheinlichkeit (Unit-Nonresponse) ( $P$ ) und den gegebenen Antworten ( $Y$ ) grafisch dar. Der Zusammenhang zwischen diesen drei Komponenten bestimmt das potentielle Ausmaß für einen Bias durch Nonresponse und somit auch die Notwendigkeit einer nachträglichen Operationalisierung beziehungsweise Adjustierung (Groves, 2006; Groves und Peytcheva, 2008).

Hierbei kommt es vor allem darauf an, ob einzelne Werte (Item-Nonresponse) oder komplette Beobachtungen (Unit-Nonresponse) zufällig fehlen oder dem dahinterstehenden (Ausfall-)Mechanismus eine bestimmte Systematik unterstellt werden muss. Ist dieser systematisch und korreliert darüber hinaus mit den Umfragevariablen ( $Y$ ), ist mit einem

Bias zu rechnen und es sind Maßnahmen zur Korrektur zu ergreifen (Groves, 2006). In der wissenschaftlichen Umfrageforschung hat sich hierbei eine Unterscheidung zwischen Missing Completely at Random (MCAR), Missing at Random (MAR) und Missing Not at Random (MNAR) etabliert (De Leeuw et al., 2003; Groves und Peytcheva, 2008; Spieß, 2010). Diese Klassifizierung geht ursprünglich auf Rubin (1976) zurück. In der Regel wird diese zur Beschreibung von Item-Nonresponse herangezogen, ist allerdings ebenso dazu geeignet, um Unit-Nonresponse und die dahinterstehenden Mechanismen herzuleiten.

Abbildung 2.3: Struktur fehlender Werte nach Groves und Peytcheva (2008)



Fehlen Werte vollkommen zufällig spricht man von MCAR. Der Ausfallprozess und die dahinter stehende Hilfsvariable ( $Z$ ) sowie auch jegliche unbeobachtete Variablen haben dabei keinerlei Einfluss auf die Ausprägung der Zielvariablen ( $Y$ ). Beide Prozesse finden unabhängig voneinander statt und werden deswegen in Abbildung 2.3 nebeneinander abgebildet (Bartsch, 2012, 33). De Leeuw et al. (2003) nennen hierzu das Beispiel, wenn ein Interviewer oder ein Teilnehmer beim Ausfüllen eines Papierfragebogens eine Frage übersieht und keine Antwort einträgt. Denkbar wäre ebenso, wenn die Teilnahme an einer Umfrage dadurch verhindert wird, dass die hierzu benötigte Technik wie ein Computer oder ein Festnetztelefon aktuell nicht funktionsfähig ist. Für die Datenanalyse ist diese Form unproblematisch und kann ignoriert werden (Schnell, 2012, 172).

MAR liegt vor, wenn das Fehlen einzelner Werte oder kompletter Fälle im Datensatz mit der Hilfsvariable ( $Z$ ) korreliert ist und die Variation der Zielvariablen ( $Y$ ) hierdurch erklärt werden kann. Die Einkommensfrage produziert üblicherweise häufig viele fehlende Werte. De Leeuw et al. (2003) beschreiben für MAR die Situation, wenn ältere Personen bestimmte Ereignisse nicht erinnern können. Dann ist das Fehlen des Wertes nicht mit dem Ereignis an sich sondern mit dem Alter der teilnehmenden Person korreliert. MAR liegt auch dann vor, wenn die Angabe des Einkommens ( $Y$ ) in einem Fragebogen mit steigendem Einkommen häufiger verweigert wird ( $P$ ) und dies durch das jeweilige (beobachtete) Bildungsniveau

(Z) erklärt werden kann (Schnell, 2012, 172).

Das Problem besteht hierbei häufig darin, dass die Hilfsvariable (Z) in der Regel nicht beobachtbar ist. Bezogen auf Unit-Nonresponse liegen zumeist keine Informationen über die Personen vor, die sich nicht beteiligt haben. Zur Veranschaulichung von systematischen Komplettausfällen (Unit-Nonresponse) kann auf das schon bekannte Beispiel des Bildungsbias zurückgegriffen werden. Bekannt ist, dass höher gebildete Personen in wissenschaftlichen Umfragen überrepräsentiert sind (Engel und Schmidt, 2014). Höher gebildete Personen beteiligen sich eher an Wahlen (Aarts und Weßels, 2005). Stellt man nun die Frage nach der Wahlbeteiligung (Y) werden die Antworten darauf nach oben hin verzerrt sein. Durch den Zusammenhang zwischen dem Bildungsniveau einer Person (Z) und der Teilnahmewahrscheinlichkeit (P) und dessen Einfluss auf die Zielvariable (Y) liegt MAR vor, solange das Bildungsniveau beobachtet wurde. Hier muss eine nachträgliche Adjustierung unter Rückgriff auf die Hilfsvariable (Z) stattfinden. Ob der systematische Ausfallprozess im Falle von MAR beleuchtet werden kann, der zu Item- oder Unit-Nonresponse führt und das Potential für einen Bias vergrößert, hängt von der Verfügbarkeit relevanter Hilfsvariablen ab.

Treten fehlende Werte auf, die der Gruppe MNAR zuzuordnen sind, besteht ein Zusammenhang zwischen Zielvariable (Y) und dem fehlenden Wert, der lediglich durch die Zielvariable (Y) selbst erklärt werden kann. In der Praxis wird dieser Fall häufig beobachtet, wenn Fragen gestellt werden, die nicht gerne beantwortet werden, weil der Teilnehmer mit seiner Antwort etwa deviantes Verhalten preisgeben müsste. Empfindet dieser eine Frage als zu privat und verweigert eine Antwort darauf, ist das Fehlen der Antwort ausschließlich durch die Frage selbst zu erklären.

Theoretisch ist die Unterscheidung zwischen den verschiedenen Mustern, die zu fehlenden Werten führen, eindeutig möglich. Die Differenzierung zwischen MAR und MNAR ist in der Praxis allerdings nicht möglich und auch nicht überprüfbar, da der dahinterstehende Mechanismus mit den unbekanntem fehlenden Werten zusammenhängt (Allison, 2001). Vor diesem Hintergrund kann lediglich eine Aussage darüber getroffen werden, ob fehlende Werte MCAR sind oder nicht.

Da gültige Werte in der Zielvariable (Y), wie Bartsch (2012) richtig festgestellt hat, lediglich für Respondenten vorliegen, sind diesbezügliche Nonresponse Analysen entsprechend beschränkt. Hier kann lediglich über logistische Regressionsmodelle versucht werden, die Teilnahmewahrscheinlichkeit möglichst genau vorherzusagen, um diese daraufhin in Zu-



sammenhang mit der Zielvariable und einem möglichen Bias durch systematische Ausfälle zu untersuchen. Das Potential zur Verringerung eines möglichen Bias ist vor allem von den zur Verfügung stehenden Hilfsvariablen abhängig.<sup>11</sup> Genau an dieser Stelle setzt die vorliegende Arbeit an. Das Design der durchgeführten Umfrage wurde so angelegt, dass möglichst viele Hilfsvariablen beobachtet werden können, um den Ausfallprozess auf seine Systematik hin überprüfen zu können. Hierfür kommen vor allem Variablen infrage, die für alle Personen in der Stichprobe zur Verfügung stehen unabhängig davon, ob diese auch teilgenommen haben oder nicht. Der diesbezügliche Stand der Forschung und die in der Praxis zur Anwendung kommenden Methoden werden in Kapitel 4.3 und 4.4 ausführlich beschrieben und diskutiert.

Der folgende Abschnitt beschäftigt sich mit theoretischen Erklärungsmodellen der Umfrageteilnahme und beschreibt diese näher.

---

<sup>11</sup> Im Kontext internationaler Forschung wird hier von „auxiliary variables“ gesprochen. Die statistischen Konzepte zur Adjustierung möglicher Verzerrungen werden ausführlich in Kapitel 4.5 diskutiert.

## Kapitel 3

# Erklärungsmodelle der Umfrageteilnahme

Die Entscheidung für oder gegen die Teilnahme an einer Umfrage wird in der sozialwissenschaftlichen Methodenforschung zumeist handlungsorientiert im Kontext von Rational-Choice-Theorien erklärt. Dabei steht vor allem der potentielle Befragte als Subjekt und die jeweiligen Faktoren, die die Entscheidungsfindung der Teilnahme positiv oder negativ beeinflussen, im Fokus.

Esser (1986) beschreibt diesen Prozess allgemein als Abwägung zwischen den Alternativen „Teilnahme“ und „Verweigerung“. Übersteigt der Nutzen- den Kostenfaktor ist die Wahrscheinlichkeit entsprechend höher, dass ein potentieller Befragter einem Umfragegesuch zustimmt (Esser, 1990). Unter einem konkreten Nutzen können Aspekte wie das Gefühl, einen Beitrag zu Wissenschaft und Forschung zu leisten oder die Aussicht auf einen Teilnahmeanreiz verstanden werden (Engel und Schmidt, 2014). Kosten entstehen für den Befragten unter anderem durch die notwendige Investition an Zeit.

Der Begriff der Entscheidungsfindung darf allerdings nicht darüber hinwegtäuschen, dass die Entscheidung für oder gegen die Beteiligung an einer Umfrage mitunter relativ schnell im Rahmen eines kurzen Zeitintervalls getroffen wird beziehungsweise werden muss (Dillman et al., 2014). Wird per Telefon oder mündlich-persönlich befragt, müssen sich potentielle Teilnehmer, sofern kein weiterer Interviewtermin vereinbart wird, innerhalb weniger Sekunden festlegen. Häufig findet so kein langwieriges Abwegen vieler Aspekte statt, um zu einer Bilanz zwischen Kosten und Nutzen zu gelangen, sondern es muss spontan entschieden werden. Bei postalischen oder Onlinebefragungen ist das Zeitfenster

entsprechend größer, da die Entscheidung nicht unmittelbar getroffen werden muss. Werden Umfrageeinladungen entsprechend per E-Mail oder Post verschickt, kann etwa das Anschreiben erst einmal zur Seite gelegt und länger über die möglichen Vor- und Nachteile einer Teilnahme abgewogen werden.

Da deshalb auch allgemeinere Faktoren wie Reziprozität, Vertrauen und Altruismus eine Rolle spielen, definieren Dillman et al. (2014) den Prozess der Entscheidungsfindung als eine Situation des sozialen Austauschs. Reziprok wird diese durch Erwartungen seitens der Befragten in der Form, dass mit einer Teilnahme etwas geleistet wird und man dafür etwas zurück bekommt. Dabei kann es sich etwa um Informationen durch Ergebnisse oder Broschüren, Teilnahmeanreize, Anerkennung oder das Gefühl einen Beitrag zu Wissenschaft und Forschung geleistet zu haben, handeln (De Leeuw et al., 2007). Vertrauen manifestiert sich derart, dass Institutionen des öffentlichen Sektors wie Universitäten oder statistische Landesämter allgemein bekannter in der Bevölkerung sind als etwa Marktforschungsinstitute. Darüber hinaus wird universitären Institutionen durch ihren öffentlichen Charakter mehr Seriosität seitens der Befragten zugeschrieben (Sponsorship Effekt) (Ladik et al., 2007). Hierdurch bringen die Befragten einem Umfragegesuch mehr Vertrauen entgegen und zeigen damit eine höhere Bereitschaft teilzunehmen. Altruismus kann als weitere latente Dimension für die Teilnahmeentscheidung eine Rolle spielen, wenn Personen die Beteiligung an einer Umfrage als Art Bürgerpflicht ansehen und dies etwa über ihre individuellen Motive stellen und somit „selbstloser“ entscheiden.

Das Verhältnis zwischen Kosten und Nutzen kann teilweise direkt vom jeweiligen Design der Umfrage und somit vom Forscher selbst beeinflusst werden. Hierzu haben Couper und Groves (1996) in ihrem Kooperationsmodell (Kapitel 3.1) die Faktoren, die die Kooperationsbereitschaft der Befragten beeinflussen danach differenziert, welche direkt durch den Forscher beeinflussbar sind oder sich jeglicher Kontrolle entziehen. In diesem Kontext wird zunächst einmal deutlich, welche Determinanten für den Entscheidungsprozess von Bedeutung und vor allem durch den Forscher strategisch einsetzbar sind, um eine Teilnahme wahrscheinlicher werden zu lassen.

Die „Leverage-Saliency Theory of Survey Participation“ nach Groves et al. (2000) erweitert das Kooperationsmodell um einen wichtigen Aspekt. Mit diesem Ansatz lässt sich theoretisch herleiten, wieso ausschöpfungssteigernde Elemente eines Umfragedesings, wie Incentives oder Vorankündigungen, den Bias durch Nonresponse nicht nur verringern sondern ebenso vergrößern können.

### 3.1 Kooperationsmodell nach Groves und Couper

Um die Mechanismen aus theoretischer Sicht nachvollziehen zu können, die hinter der Entscheidung für oder gegen eine Teilnahme stehen, ist es zunächst einmal notwendig zu hinterfragen, welche Determinanten bzw. Gruppen von Determinanten hierfür von Bedeutung sind. Couper und Groves (1996) diskutieren diese in ihrer Arbeit über die „Household-Level Determinants of Survey Nonresponse“ und beziehen sich mit ihrem Modell auf die Interaktion zwischen Interviewer und dem Mitglied eines Haushaltes sowie der letztlichen Teilnahme oder Verweigerung. Die dabei diskutierten *Determinanten der Teilnahmebereitschaft* sind ebenso auf selbstadministrierte Befragungsformen (postalisch, online) zu übertragen. Hierbei fällt der entsprechende Schritt der Interaktion zwischen Interviewer und Haushalt weg. Allgemein differenzieren Couper und Groves (1996) zunächst zwischen den Einflussgrößen Umfragedesign, Interviewer, sozialer Umwelt und den Eigenschaften des Haushaltes selbst. Zentral ist dabei, dass die Auswirkungen der sozialen Umwelt und der Eigenschaften des Haushaltes auf die Interaktion zwischen Interviewer und Haushalt sowie der letztlichen Teilnahmeentscheidung nicht zu kontrollieren sind. Die von Couper und Groves (1996, 65) benannten Einflussgrößen differenzieren sich in weitere Unterdimensionen, deren Wirkmechanismen zusätzlicher Klärung bedürfen.

#### Umfragedesign (kontrollierbar)

Durch die einzelnen Aspekte des Umfragedesigns lässt sich die Entscheidung für oder gegen eine Teilnahme teilweise direkt beeinflussen. Deshalb sind Antwortraten, auch wenn diese standardisiert nach den Richtlinien der AAPOR berechnet werden, über Studien hinweg häufig nur schwer miteinander vergleichbar (Engel et al., 2004). So sind dem Umfragedesign weitere Dimensionen inhärent, die die Teilnahmewahrscheinlichkeit positiv oder negativ beeinflussen können.

Diese variiert zunächst mit dem *Thema der Umfrage*. Besteht besonderes Wissen über einen bestimmten Sachverhalt oder Interesse am thematischen Schwerpunkt der Befragung, wird eine Teilnahme wahrscheinlicher (Adua und Shap, 2010; Engel et al., 2012; Esser, 1986; Geuzinge et al., 2000; Groves et al., 2004; Groves und Peytcheva, 2008; Roose et al., 2007).

Werden *Teilnahmeanreize* zur Erhöhung der Antwortrate eingesetzt, besteht die Frage zunächst darin, worin die Wirkung begründet liegt. Wird diese eher über den moralischen

Aspekt im Sinne einer „kleinen Aufmerksamkeit“ oder den ökonomischen Gesichtspunkten im Sinne einer Bezahlung vermittelt (Engel, 2006)?

Nach Dillman et al. (2014) soll der Einsatz von Teilnahmeanreizen zu einer Situation des sozialen Austausches führen und somit über den moralischen Aspekt vermittelt auf die Teilnahmewahrscheinlichkeit wirken. Auch wenn diese in monetärer Form und unabhängig von einer Teilnahme vergeben werden, soll die Situation seitens der Befragten nicht als ökonomischer Tausch wahrgenommen werden. Denn findet eine solche Interpretation der Situation statt und der Anreiz wird somit nicht als „kleine Aufmerksamkeit“ im moralischen Sinne sondern als Bezahlung wahrgenommen, kann aus dem positiven Effekt ein negativer werden. Nämlich genau dann, wenn die Höhe des Anreizes für die zu investierende Zeit und Mühe als nicht ausreichend bewertet wird (Dillman et al., 2014; Engel, 2006; Engel und Schmidt, 2014). Eine Ablehnung des Umfragegesuchs würde entsprechend erleichtert.<sup>1</sup> Die Wahrnehmung eines Anreizes als sozialen Austausch im moralischen Sinne entsteht praktisch etwa dadurch, dass einem postalischen Einladungsschreiben eine Geldnote beiliegt. So entsteht nach dem Prinzip der Gegenseitigkeit (Adloff und Mau, 2005) ein reziprokes Verhältnis zwischen anfragender Institution und den potentiellen Teilnehmern. Diese fühlen sich durch die Geldnote quasi dazu verpflichtet etwas zurück zu geben, womit eine Teilnahme wahrscheinlicher wird (Dillman et al., 2014). Werden Teilnahmeanreize in Form von Gutscheinen oder Verlosungen von einer Beteiligung abhängig gemacht (konditional), entsteht diese Situation der Reziprozität nicht, da der Befragte mit der Beantwortung der Fragen erst einmal in „Vorleistung“ treten muss. Vor diesem Hintergrund ist der Einsatz monetärer nicht konditionaler Teilnahmeanreize zur Erhöhung der Antwortrate in der Praxis insgesamt erfolgreicher (Görritz, 2010; Millar und Dillman, 2011; Scherpenzeel und Toepoel, 2012; Singer und Ye, 2013).

Abhängig von der Präferenz für einen bestimmten *Umfragemodus* wird die Teilnahmewahrscheinlichkeit davon beeinflusst, in welcher Form die Umfrage durchgeführt wird. Eine Person füllt etwa einen Fragebogen bevorzugt im Internet aus, weil die Beantwortung hier relativ schnell vonstattengeht sowie mit wenig Aufwand verbunden ist und zu einem späteren Zeitpunkt fortgesetzt werden kann. Wenn zu einer Onlinebefragung eingeladen wird, ist diese Person in der Folge auch eher zur Partizipation bereit.

Die *Dauer des Interviews* und die damit einhergehende *Länge des Fragebogens* sind

---

<sup>1</sup> Die Forschungslage bezüglich der Wahrnehmung der Situation ist nicht immer eindeutig und wird ausführlich in Kapitel 4.2 diskutiert.

wichtige Aspekte des Umfragedesigns, die das Abwägen zwischen Teilnahme und Verweigerung maßgeblich mitbestimmen (Diekmann, 2007; Dillman et al., 2014; Schnell, 2012). Nehmen Befragte die zu investierende Zeit subjektiv als zu lange wahr, wird eine Verweigerung wahrscheinlicher. Die Investition des Faktors Zeit ist für die Teilnahmeentscheidung wesentlich, da hierdurch die Kosten einer Beteiligung bedeutend zunehmen können. Deswegen sollte ein optimales Gleichgewicht zwischen der Länge des Fragebogens und dessen Informationsvielfalt sowie der Dauer des Interviews gefunden werden (Engel und Schmidt, 2014).

Eine *Vorankündigung* des Umfragegesuchs kann in Abhängigkeit vom Umfragemodus und dem damit einhergehenden Auswahlrahmen postalisch erfolgen. Dazu sollte ein entsprechendes Anschreiben personalisiert sein und alle wichtigen Informationen enthalten, wie das Ziel der Studie, die voraussichtliche Dauer der Befragung sowie über die ausführende Institution aufklären (De Leeuw et al., 2007). So können sich die potentiellen Teilnehmer schon im Vorhinein über die durchführende Institution und deren Grad an Seriosität überzeugen. Wird im Nachhinein ein Interviewer zum jeweiligen Haushalt geschickt oder telefonisch Kontakt aufgenommen, sind die Personen eher dazu bereit, an der Umfrage teilzunehmen (Baur und Florian, 2009; Brehm, 1994; Hansen und Pedersen, 2012; Lavrakas, 2010; Rao et al., 2010; Scherpenzeel und Toepoel, 2012). Denn so haben sie bereits Kenntnis über die bevorstehende Kontaktaufnahme und sind über die Studie, deren Ziele und die dahinterstehende Institution informiert.

Vor diesem Hintergrund bestimmt die *Stichprobenziehung* die Wahrscheinlichkeit einer Teilnahme indirekt mit. Denn vom jeweiligen Auswahlrahmen und den hierüber zur Verfügung stehenden Kontaktinformationen ist es abhängig, ob Befragte im Vorhinein über das Umfragegesuch informiert werden und Teilnahmeanreize im Zuge dessen vergeben werden können.

Ähnlich verhält es sich, wenn per Erinnerungsschreiben (englisch „Reminder“) nach einer bestimmten Zeit an eine noch ausstehende Teilnahme erinnert wird. Häufig wird hierzu per Post oder E-Mail ein entsprechendes Schreiben verschickt. Die Teilnahmewahrscheinlichkeit kann durch den Einsatz von Erinnerungsschreiben noch einmal erhöht werden, sodass dies in der Regel zu einer höheren Antwortrate führt (Atkeson und Tafoya, 2008; Dillman et al., 2014; Link et al., 2008; Schonlau et al., 2003; Toepoel und Lugtig, 2014).

Die *Länge der Feldphase* kann dafür verantwortlich sein, ob eine Person an einer Befragung teilnimmt oder nicht. Ist diese etwa zu kurz und somit das Zeitfenster für die

Vergabe von Terminen entsprechend klein kann es vorkommen, dass potentielle Teilnehmer während der angestrebten Feldphase zeitlich verhindert sind oder ein Kontakt erst gar nicht zustande kommt.

Die *Administrierung und Pflege der Kontaktdaten* ist vor allem im Kontext von Panelbefragungen wesentlich, damit die Wahrscheinlichkeit möglichst groß ist, die Personen bei der Durchführung einer weiteren Welle erneut zu erreichen. Wird die Pflege der Adressdatenbank nicht kontinuierlich verfolgt, steigt die Anzahl an Ausfällen aufgrund von Nichterreichbarkeit. Den Teilnehmern kann etwa über eine kostenlose Hotline oder allgemeine E-Mail-Adresse die Möglichkeit gegeben werden im Falle eines Umzuges die aktuellen Kontaktinformationen mitzuteilen.

### **Interviewer (kontrollierbar)**

Der Einsatz von Interviewern bei face-to-face und telefonischen Befragungen bestimmt die Wahrscheinlichkeit der Teilnahme mit. Diese ist maßgeblich von dessen Erfahrung, soziodemografischen Charakteristika sowie der Einstellung, inwiefern zögerlicher Personen zu einer Teilnahme überzeugt werden können, abhängig (De Leeuw et al., 2008; Engel et al., 2012; Hox und De Leeuw, 2002; Watson und Wooden, 2009). Als relevant haben sich dabei vor allem die gesammelte Erfahrung sowie die Einstellung gegenüber Personen, die zunächst von einer Teilnahme überzeugt werden müssen, erwiesen. Die soziale oder auch kommunikative Kompetenz des Interviewers sowie eine positive Einstellung derart, dass potentiell jeder von einer Teilnahme überzeugt werden kann, wirken sich positiv auf die Antwortrate insgesamt aus. Hat ein Interviewer zusätzlich eine freundliche und offene Art, fällt die Interaktion mit dem potentiellen Teilnehmer leichter, sodass dieser in der Folge eher dazu bereit ist, die Fragen zu beantworten. Ist darüber hinaus die nötige Schulung und Erfahrung vorhanden, können Interviewer auf Situationen reagieren, in denen Personen zögerlich sind und das jeweils passende Überzeugungsargument anbringen oder die Einleitung in das Gespräch an unterschiedliche Gegebenheiten anpassen („Tailoring“). Soziodemografische Charakteristika wie Alter oder Geschlecht konnten bisher nur ein marginaler Effekt auf die Teilnahmewahrscheinlichkeit nachgewiesen werden (Kane und Macaulay, 1993).

### **Soziale Umwelt (nicht kontrollierbar)**

Die soziale Umwelt und deren Einfluss auf die Teilnahmeentscheidung manifestiert sich

über das allgemeine Umfrageklima in der Bevölkerung, den Urbanitätsgrad oder den Charakteristiken der Nachbarschaft. Häufen sich Verkaufsanrufe unter dem Deckmantel von Befragungen (Engel et al., 2004) oder findet eine „Überbefragung“ innerhalb der Bevölkerung statt, kann sich dies direkt negativ auf die Kooperations- und schließlich auch auf die Teilnahmebereitschaft auswirken (Toepoel, 2012). Ende der 1990er Jahre hat De Heer (1999, 139) das Umfrageklima und dessen Wirkung als „[...] largely unknown factor“ bezeichnet. Curtin et al. (2005, 25) sehen einige Jahre später im deutlichen Zuwachs von Verkaufs- sowie Anrufen zu Umfragezwecken eine Erklärung für die Zunahme an Verweigerungen und Kontaktausfällen für telefonische Befragungen. Werden vermehrt Verkaufsanrufe unter dem Deckmantel von Befragungen durchgeführt, steigt auch gegenüber seriösen Umfrageinstituten die Skepsis seitens der Befragten, da der Anrufer zunächst unbekannt ist. Bekannt ist darüber hinaus, dass sich Personen eher an Umfragen beteiligen, wenn sie in urbanen Gebieten leben (Adua und Shap, 2010; Couper und Groves, 1996). Ein möglicher Erklärungsansatz hierfür ist, dass das Leben in Städten im allgemeinen anonymere im Vergleich zum Leben auf dem Land ist. Insofern ist es für Menschen in urbaneren Gebieten nicht außergewöhnlich, von zunächst fremden Personen kontaktiert und zur Teilnahme an einer Umfrage aufgefordert zu werden.

### **Eigenschaften des Haushalts (nicht kontrollierbar)**

Vor allem die soziodemografischen Charakteristika oder auch die psychologischen Prädispositionen sowie die Struktur eines Haushaltes an sich beeinflussen das Teilnahmeverhalten. So nehmen etwa eher höher gebildete Personen an Umfragen teil (Goyder, 1987, 84). Hier spricht man vom „Bildungsbias“. Je mehr Personen in einem Haushalt leben, umso wahrscheinlicher ist es, dass zum Zeitpunkt der Kontaktierung jemand erreichbar ist. Durch bessere Erreichbarkeit ist eine höhere Teilnahmewahrscheinlichkeit insgesamt gegeben (Engel et al., 2012). Die psychologischen Prädispositionen einer Person wirken über die Persönlichkeitseigenschaften auf die Teilnahmewahrscheinlichkeit. Ist eine Person eher „verträglich und extravertiert“ und gleichzeitig weniger neurotisch und gewissenhaft, ist diese auch eher dazu bereit, sich an einer Umfrage zu beteiligen (Saßenroth, 2012, 2013).



## 3.2 „Leverage-Saliency“ Theorie der Umfrageteilnahme

Die „Leverage-Saliency Theory of Survey Participation“ geht auf Groves et al. (2000) zurück. Im Zentrum dieses Ansatzes stehen die Designelemente einer Umfrage und deren Wirkung auf die Antwortrate sowie einen möglichen Nonresponse Bias. Denn durch den Einsatz von Incentives oder in Abhängigkeit vom jeweiligen Umfragethema treten spezifische Selektionseffekte in Kraft, die letztendlich dafür mit verantwortlich sind, dass es nicht zufällig ist, welche Personen sich an einer Umfrage beteiligen und welche nicht (Tourangeau, 2004). Dieses Modell lässt sich der Familie der Rational-Choice Ansätze zuordnen und fügt diesem zentrale, darüber hinausgehende Aspekte hinzu (Roose et al., 2007).

In diesem Kontext ist theoretisch herleitbar, inwiefern Personen die Designelemente einer Umfrage unterschiedlich gewichten und diese eine unterschiedliche „Zugkraft“ auf die potentiellen Teilnehmer ausüben. Fühlt sich eine Person etwa von einem monetären Anreiz zu einer Teilnahme motiviert, findet das Thema der Umfrage im Gegensatz dazu allerdings uninteressant, kann die vermeintlich positive Wirkung des Anreizes hierdurch nivelliert werden.

Die Entscheidung für oder gegen eine Teilnahme wird von Couper (2000) dabei als eine Art Waage dargestellt. Die einzelnen Elemente des Umfragedesings werden durch Gewichte repräsentiert, die auf der einen Seite eine negative Wirkung auf die Teilnahmewahrscheinlichkeit ausüben und auf der linken Hälfte der Achse platziert werden. Auf der rechten Seite finden die Gegengewichte ihren Platz, die einen allgemein positiven Einfluss in Richtung Teilnahme ausüben. Die jeweils relative Position auf der Achse der Waage, die Hebelwirkung (leverage), sowie die Größe des Gewichts an sich (saliency), bestimmen letztendlich, ob sich der Winkel in Richtung Verweigerung oder Teilnahme neigt. Die Hebelwirkung ist dabei von den spezifischen Eigenschaften der Person selbst abhängig und macht es vor diesem Hintergrund erklärbar, warum ein einzelner Aspekt des Umfragedesings unterschiedlich wirkt. So kann das Umfragethema mit dem Alter in dem Sinne korrelieren, dass ältere Personen sich eher an Umfragen beteiligen, die auf Inhalte dieser betreffenden Altersgruppe rekurrieren (Groves et al., 2000). Ein weiteres Beispiel hierfür ist, dass Personen, die über Umfrageerfahrung verfügen eher dazu bereit sind, sich befragen zu lassen (Engel et al., 2012; Engel und Schmidt, 2014). Hiermit ist ebenso das Ergebnis von Vehre et al. (2013) zu erklären, dass sich Personen mit einem niedri-

gen Bildungsniveau eher von Teilnahmeanreizen angesprochen fühlen. Daran anknüpfend entspricht die Hebelwirkung der Interaktion zwischen den Eigenschaften des Befragten selbst und den einzelnen Designelementen einer Umfrage. Dies mündet schließlich in der Auswirkung dieses Zusammenspiels in einer individuellen Teilnahmewahrscheinlichkeit.

Die Größe des Gewichts (salience) repräsentiert letztlich, wie „prominent“ die einzelnen Designelemente einer Umfrage von der jeweils durchführenden Institution gemacht werden. So ist die Zugkraft in eine bestimmte Richtung des Kontinuums „Teilnahme - Verweigerung“ ebenso davon abhängig, wie eindrücklich auf das Umfragethema hingewiesen wurde oder welchen Stellenwert Teilnahmeanreize im Rahmen einer Umfrageeinladung oder des Erstkontaktes durch einen Interviewer einnehmen. Dieser Effekt kann durch feldexperimentelle Designbedingungen operationalisiert und so eingeschätzt bzw. kontrolliert werden. Folglich interagieren die einzelnen Designelemente miteinander, können sich in ihrer Wirkung gegenseitig nivellieren und Richtung sowie Stärke des Effekts sind letztlich von den Eigenschaften des Befragten selbst sowie der „Prominenz“ der einzelnen Aspekte abhängig.

In der Praxis hat sich gezeigt, dass diese theoretischen Annahmen empirisch haltbar sind. Groves et al. (2000) sind zu dem Ergebnis gekommen, dass der Effekt nicht konditionaler monetärer Teilnahmeanreize bei Personen schwächer ausgeprägt ist, die besser in die Gesellschaft integriert sind. Der Grad an Integration wird durch einen Index abgebildet, der sich dadurch ergibt, ob eine Person etwa in Vereinen oder Institutionen ehrenamtlich aktiv ist, sich lokalpolitisch engagiert oder gegenseitige Hilfestellung unter Nachbarn geleistet wird (Groves et al., 2000, 202 f.). Roose et al. (2007) schlussfolgern aus ihren Befunden, dass der Effekt von Remindern und Ersatzfragebögen, die versandt werden, wenn der ursprüngliche Fragebogen verloren gegangen sein sollte, auf die Antwortrate kleiner wird, wenn das Interesse am Thema der Umfrage zunimmt.

Die zentrale, an diese theoretische Herangehensweise anknüpfende Frage besteht darin, wie in diesem Kontext ein Bias durch Nonresponse erklärbar ist. Dieser kann in diesem Zusammenhang entstehen, wenn durch das jeweilige Umfragedesign spezifische Selektionseffekte entstehen. Hierdurch ist es folglich nicht zufällig, welche Personen sich an einer Umfrage beteiligen. Damit steigt die Wahrscheinlichkeit, dass sich Respondenten von Nonrespondenten systematisch unterscheiden und die gestellten Fragen anders beantwortet werden. Außerdem kann hierdurch nicht nur Unit- sondern ebenfalls Item-Nonresponse auftreten und die Datenqualität zusätzlich in Mitleidenschaft ziehen. Adua und Shap

(2010) sind diesbezüglich zu dem Ergebnis gekommen, dass Wissen und Interesse zu einem Umfragethema negativ mit der Häufigkeit an nicht beantworteten Fragen korreliert. Anders formuliert, die Eigenschaften des Befragten selbst sind nicht nur Determinanten für Unit- sondern ebenso für Item-Nonresponse.

Um das Fehlerpotential bei der Durchführung einer Umfrage zu verkleinern, etwa durch die Erhöhung der Antwortrate, gibt es bestimmte Elemente, die strategisch eingesetzt werden können. Diese sind Teil des nächsten Kapitels.

## Kapitel 4

# Methoden zur Reduzierung möglicher Fehler

Die Umfrageforschung steht auch heute noch vor der Herausforderung eine Lösung für die Stichprobenproblematik im Kontext von Onlinebefragungen zu finden und die Antwortraten telefonischer sowohl als auch webbasierter Umfragen wieder zu erhöhen. Zentral ist es dabei, Ansätze und Indikatoren (weiter) zu entwickeln, um das Ausmaß eines möglichen Bias durch Nonresponse messbar zu machen und somit den Zusammenhang zwischen Antwortrate und Nonresponse Bias besser als bisher beschreiben zu können (Kreuter, 2013a).

In Zukunft wird es notwendig sein, jegliche im Rahmen des TSE diskutierten Fehlerkomponenten durch methodische Begleitforschung zu operationalisieren, um eine bestmögliche Datenqualität und Kosteneffektivität gewährleisten zu können (Fuchs, 2010). In der Umfrageforschung werden diverse Ansätze diskutiert, um eine Lösung für die beschriebene Stichprobenproblematik sowie niedrige Antwortraten und der Operationalisierung eines möglichen Bias durch Nonresponse zu finden. Ein vielversprechendes Modell stellt die Verknüpfung eines Mixed-Mode Designs mit Teilnahmeanreizen und einer Stichprobenziehung auf Basis des Melderegisters dar.

### 4.1 Mixed-Mode

Bei Mixed-Mode Befragungen handelt es sich um den Versuch, mit der Kombination verschiedener Modi das Ausmaß des TSE zu verringern. Die Kombination von mehreren Erhebungsmodi innerhalb einer Befragung ist in den letzten Jahren deutlich populärer

geworden, da es aufgrund der schon beschriebenen Problematik schwierig sein kann, eine akzeptable Datenqualität mit einem Single-Mode Design zu gewährleisten. So hat Couper (2011, 892) das Phänomen treffend beschrieben, indem er ein weitergehendes Verständnis für den Begriff „Modus“ anregt: „I believe that ‚mode‘ should no longer be thought of as a categorical descriptor, but rather as a multidimensional construct.“ Die Forschungslage ist vor diesem Hintergrund recht breit gefächert, der Vergleich der einzelnen Studien wird allerdings durch große Variation in der zur Anwendung kommenden Umfragedesings sowie der jeweils betreffenden Grundgesamtheiten erschwert.

Einen Überblick über die verschiedenen Typen von Mixed-Mode Befragungen, die Gründe für deren Anwendung sowie den möglichen Einfluss auf die Qualität der Daten gibt Tabelle 4.1. Der Einsatz eines solchen Designs zeichnet sich durch diverse Kombinationsmöglichkeiten aus, um mit einer guten Untersuchungsplanung und Verknüpfung der einzelnen Modi potentielle Fehlerquellen im Idealfall zu minimieren.

Mixed-Mode Befragungen sind Single-Mode Befragungen allerdings nicht notwendigerweise überlegen. Der Erfolg des Einsatzes für die Fehlerbilanz einer Umfrage ist vor allem davon abhängig, ob Selektionseffekte zu erwarten sind oder nicht (Vannieuwenhuyze, 2014). Treten diese auf und mit ihnen eine mögliche Verzerrung durch Noncoverage und Nonresponse kann eine Mixed-Mode Befragung effektiver als eine Single-Mode Befragung sein. Hierbei muss allerdings genau zwischen den möglichen Fehlern, die durch Selektions- und Moduseffekte hervorgerufen werden können, abgewogen werden. Letztere treten in der Regel in der Form auf, dass die gleiche Frage abhängig vom jeweils eingesetzten Umfragemodus anders beantwortet werden kann (Dillman et al., 2009). Diese beiden potentiell verzerrenden Elemente sind in einem Mixed-Mode Design üblicherweise konfundiert (Buelens und van den Brakel, 2014). Denn häufig kann kein experimentelles Design in der Form angewendet werden, dass die Befragungspersonen zufällig zu einem Modus zugeordnet werden, sodass Modus- und Verzerrungen durch Selektionseffekte statistisch sauber voneinander getrennt werden können. Dies liegt etwa daran, dass mit dem Argument der Fehlerreduzierung den Befragten im Vorhinein zur Wahl gestellt wird, in welchem Modus sie partizipieren wollen, um die Antwortrate zu erhöhen. Oder es liegen durch den Auswahlrahmen nur begrenzte Kontaktinformationen vor, die es notwendig machen zunächst in einem anderen Modus, im Vergleich zur eigentlichen darauffolgenden Befragung, Kontakt zu den potentiellen Teilnehmern herzustellen. Häufig kommt dieses Vorgehen aufgrund des fehlenden Auswahlrahmens bei der Durchführung von Onlinebefragungen

zum Einsatz, wenn zunächst per Telefon nach Internetnutzern und befragungsbereiten Personen gefiltert wird (Bandilla et al., 2009). Vor diesem Hintergrund werden mögliche Moduseffekte in der Praxis eher in Kauf genommen, da der TSE durch den Einsatz eines Mixed-Mode Designs verringert werden kann (Voogt und Saris, 2005).

Tabelle 4.1: Typen der Mixed-Mode Befragung, Gründe für die Anwendung und der Einfluss auf die Datenqualität

Mixed-Mode Typ	Gründe für Implementierung	Einfluss auf die Datenqualität
Moduswechsel in der <i>Kontaktphase</i> für...		
(1)...Vorankündigung und nachfolgende Datenerhebung	Korrektur des Auswahlrahmens Erhöhung der Antwortrate Erhöhung des Vertrauens	Reduzierung des Coverage und Nonresponse Error
(2)...Rekrutierung/ Screening/Auswahl und nachfolgende Datenerhebung	Reduzierung der Kosten Erhöhung der Effizienz Erweiterung Kontaktinformationen für Haupterhebungsmodus	Gefahr von Moduseffekten bei Screening mit anschließender Befragung
Moduswechsel in der <i>Feldphase</i> für...		
(3)...dieselbe Stichprobe mit gleicher Feldzeit und Fragebogen	Reduzierung der Kosten Reduzierung von Noncoverage Reduzierung von Nonresponse	Reduzierung des Coverage und Nonresponse Error Moduseffekte innerhalb der Subgruppen konfundiert
(4)...Teile eines Fragebogens bei gleicher Stichprobe	Mehr Privatsphäre Reduzierung des Effekts sozialer Erwünschtheit	Erhöhung der Datenqualität im Rahmen besonders sensibler Fragen
(5)...unterschiedliche Zeitpunkte bei gleicher Stichprobe	Reduzierung der Kosten	Messeffekte, da Zeit- und Moduseffekte konfundiert sind
(6)...unterschiedliche Stichproben, teilweise mit unterschiedlichen Zeitpunkten und Fragebögen	Vergleichende Untersuchungen Verschiedene Forschungstraditionen Verschiedene Coverage-Effekte	Reduzierung Coverage und Nonresponse Error Moduseffekte innerhalb der Subgruppen konfundiert
Moduswechsel in der <i>Follow-up Phase</i> für...		
(7)...Erinnerungsschreiben und Erhebungsmodus	Reduzierung von Nonresponse Verschiedene Forschungstraditionen Verschiedene Coverage-Effekte	Reduzierung des Nonresponse Error Moduseffekte

Übersetzte Darstellung in Anlehnung an De Leeuw et al. (2008, 303 f.).

Die zur Verfügung stehenden Umfragemodi schriftlich, face-to-face, online und telefonisch können dabei im Rahmen der Zeitfenster Kontakt-, Feld- sowie Follow-up Phase einer Umfrage zu einem Mixed-Mode Design miteinander kombiniert werden. Häufig findet bereits im Vorfeld der eigentlichen Befragung eine Vorankündigung in der Kontaktphase in der Form statt, dass ein schriftliches Ankündigungsschreiben per Post versandt wird,

um zu einer Onlinebefragung einzuladen oder einen Telefonanruf anzukündigen (Typ 1, siehe Tabelle 4.1). Hiermit gehen gleichzeitig mehrere Vorteile und somit positive Effekte für die Qualität der Daten einher. Denn mit einer Vorankündigung kann nicht nur der Auswahlrahmen insgesamt korrigiert, sondern ebenso die Antwortrate erhöht und somit der Coverage und Nonresponse Error verringert werden (Dillman et al., 2009; De Leeuw, 2005; De Leeuw et al., 2008; Fuchs, 2010; Groves und Peytcheva, 2008). Eine Korrektur des Auswahlrahmens kann in der Form stattfinden, dass zunächst alle Personen in der Stichprobe postalisch kontaktiert werden unabhängig davon, ob sie über einen zu einer Teilnahme notwendigen Anschluss zum Internet oder ein Festnetztelefon verfügen. Wird ein Mixed-Mode Design etwa im Rahmen von besonders durch Noncoverage und Nonresponse beeinträchtigen Modi wie der Online- oder der telefonischen Befragung eingesetzt, besteht trotz möglicher Moduseffekte ein hinreichendes Potential für die Verringerung der Fehlerbilanz.

Eine weitere dem Typ 1 zugehörige Kombinationsmöglichkeit besteht darin, mit einer postalischen Vorankündigung während der Kontaktphase darauf zu verweisen, dass eine Teilnahme auf mehrere Arten möglich ist. Können sich diese je nach individueller Präferenz für einen Modus entscheiden, hat dies vor dem Hintergrund vorliegender Studien allerdings keinen nennenswerten Effekt auf die Antwortrate insgesamt (Couper, 2011).

Wird eine postalische Vorankündigung dadurch erweitert, dass nach Personen mit einem Internetanschluss gefiltert wird, um daraufhin per E-Mail zu einer Onlinebefragung einzuladen, können damit Kosten reduziert werden bei gleichzeitiger Erhöhung der Effizienz (Typ 2). Denn hiermit können zusätzliche Kontaktinformationen wie E-Mail-Adressen oder Rufnummern für die Haupterhebung gesammelt werden. Schließt die eigentliche Befragung, oder Teile dieser, allerdings direkt an das „Screening“ an, besteht die Gefahr, dass sich der Erhebungsmodus auf das Antwortverhalten auswirkt (De Leeuw et al., 2008).

Geht die Befragung von der Kontakt- in die Feldphase und somit zur eigentlichen Erhebung über, gibt es innerhalb dieser weitere Möglichkeiten für den Einsatz von Mixed-Mode Befragungen. Voogt und Saris (2005) differenzieren im Rahmen der zugrundeliegenden Stichprobe, für welche Personen eine Telefonnummer eruiert beziehungsweise nicht eruiert werden kann. Bei bekannter Rufnummer werden die Personen entsprechend angerufen und per Telefon befragt. Ist diese allerdings unbekannt, wird ein schriftlicher Fragebogen versandt (Typ 3). So werden im ersten Schritt die zur Verfügung stehenden Kontaktinformationen genutzt und hierüber der entsprechende Kontakt hergestellt und befragt, sofern

sich die potentiellen Teilnehmer dazu bereit erklären. Dabei hat sich gezeigt, dass die Antwortrate durch die Kombination der beiden Modi insgesamt erhöht werden kann. Die Personen, die im Rahmen dieses ersten Schrittes nicht erreicht werden können, werden in Form eines Follow-ups erneut kontaktiert (Typ 7). Dieses Vorgehen hat einen substantiell positiven Effekt auf die Antwortrate. Voogt und Saris (2005) konnten zeigen, dass einerseits der Versandt eines zusätzlichen postalischen Fragebogens einen positiven Effekt hat, wenn zuvor telefonische Kontaktversuche stattgefunden haben. Andererseits konnte die Anzahl an Teilnehmern durch den Besuch eines Interviewers im Anschluss an die postalische Variante, wenn noch nicht teilgenommen wurde, erhöht werden. Zusätzlich konnten die Selektionseffekte durch dieses Vorgehen verringert werden, ohne negative Konsequenzen bezüglich der Vergleichbarkeit der Antworten über die einzelnen Modi hinweg. Schließlich wurde in die Haushalte, die zunächst telefonisch und im Nachgang postalisch kontaktiert wurden, ein Interviewer geschickt, wodurch die Antwortrate in dieser Gruppe nochmals, wenn auch nur leicht, erhöht werden konnte. Als Fazit halten Voogt und Saris (2005, 385) fest: „All in all, we conclude that a mixed mode design is an efficient way of fighting bias in survey research“.

Dieses sequentielle Vorgehen (Typ 3) bietet zur Reduzierung des Coverage und Non-response Error ein größeres Potential, als den Befragten selbst zwischen verschiedenen Teilnahmealternativen wählen zu lassen. Eine vielversprechende Möglichkeit um die Stichprobenproblematik zu überwinden besteht darin, zunächst postalisch zu einer Onlineumfrage einzuladen, um daraufhin in einem zweiten oder dritten Kontaktierungsschritt all diejenigen, die bis dahin noch nicht teilgenommen haben, die Möglichkeit zu geben, einen Papierfragebogen auszufüllen. So wird Personen ohne Zugang zum Internet trotzdem die Möglichkeit gegeben, sich an der Umfrage zu beteiligen. Messer und Dillman (2010) sind diesbezüglich zu dem Ergebnis gekommen, dass mit einer sequentiellen Kontaktierung im Rahmen eines solchen Mixed-Mode Designs einerseits die Antwortrate von Onlinebefragungen erhöht werden kann. Andererseits kann der Fehler durch Noncoverage und Nonresponse verringert werden. Hierbei treten zwar Selektionseffekte derart auf, dass sich die Teilnehmer der Online- von denen der postalischen Befragung unterscheiden. Messer und Dillman (2010, 19) schlussfolgern allerdings, dass der Bias bei Kombination der beiden Modi insgesamt geringer ausfällt. Zusätzlich können die Kosten derart gesenkt werden, dass in einem ersten sequentiellen Kontaktierungsschritt zunächst zu einem kostengünstigeren Modus eingeladen wird, wie etwa online oder postalisch. In einem zweiten Schritt



wird denjenigen, die bis dahin noch nicht teilgenommen haben, die Teilnahme in einem kostenintensiveren Modus angeboten. Schonlau et al. (2007) haben daran anknüpfend mit dem Argument der Kosteneffektivität zunächst per Post zu einer Onlineumfrage eingeladen. Dabei hat sich gezeigt, dass sich an der Onlineumfrage insgesamt 35,2 Prozent beteiligt haben. In einem zweiten Kontaktierungsschritt wurden die Personen, die noch nicht teilgenommen haben, per Telefon daran erinnert. Dies hat die Antwortrate um 12,5 Prozentpunkte erhöht (Schonlau et al., 2003, 219).

Zur Erhöhung der Privatsphäre und der Verringerung sozialer Erwünschtheitseffekte können Teile eines Fragebogens in unterschiedlichen Modi beantwortet werden. Im Beziehungs- und Familienpanel (pairfam)<sup>1</sup> wird die Haupterhebung durch einen Interviewer geführt (Computer Aided Personal Interviewing (CAPI)) und die sensitiven Fragen, wie etwa zum Sexualleben, werden ohne Hilfe des Interviewers beantwortet (Computer Aided Self Interviewing (CASI)) (Typ 4) (Huinink et al., 2011, 94). So müssen diese teils sehr privaten und sensiblen Informationen keinem fremden Interviewer mitgeteilt werden, wodurch sich die Befragten nicht in dem Maße an sozialen Normen orientieren und ehrlichere Antworten geben (Diekmann, 2007). In der Folge sind Mess- und Moduseffekte allerdings konfundiert, da hier in der Regel kein experimentelles Design zur Anwendung kommt. So sind mögliche Auswirkungen der unterschiedlichen Befragungsmodi auf das Antwortverhalten im Nachhinein nicht mehr nachzuvollziehen.

Werden die selben Personen in regelmäßigen Abständen über mehrere Zeitpunkte hinweg befragt, spricht man von einer Panelstudie (Lynn, 2009). Innerhalb eines Panels kann es möglich sein, dass die Teilnehmer den Fragebogen zum ersten Zeitpunkt telefonisch und zum zweiten online beantworten (Typ 5) oder sich bereits nach dem Erstkontakt, je nach Präferenz, für einen Umfragemodus entscheiden können. In sogenannten Access Panels kann ein Mixed-Mode Design in allen drei Umfragephasen zur Anwendung kommen. Diese zeichnen sich dadurch aus, dass zunächst in einen ausgewählten Modus Kontakt zu den potentiellen Teilnehmern hergestellt wird. Im Anschluss an eine erste Befragung in diesem Modus werden die Teilnehmer danach gefragt, ob sie bereit sind, sich in regelmäßigen Abständen erneut befragen zu lassen. Zur Rekrutierung von Teilnehmern für das PPSM<sup>2</sup> Access Panel wurden die Teilnehmer telefonisch kontaktiert (Engel et al., 2012; Engel, 2013; Vehre, 2011; Bartsch, 2012). Im Anschluss an das telefonische Rekrutierungsinterview wurde

---

<sup>1</sup> Für weiterführende Informationen zum Familien- und Beziehungspanel siehe Huinink et al. (2011) oder [www.pairfam.de](http://www.pairfam.de)

<sup>2</sup> Das PPSM Access Panel wird ausführlich in Kapitel 6.1 (Das PPSM Access Panel) beschrieben.

danach gefragt, ob die Befragten bereit sind, an einem weiteren Interview teilzunehmen und wenn ja, auf welche Weise der Kontakt wieder aufgenommen werden darf (telefonisch oder online). Haben die Befragten mehr als eine Kontaktinformation angegeben (Festnetz, mobil oder online), wurden diese zufällig zu einem dieser Modi zugeordnet, um in dieser Gruppe eine statistische Trennung zwischen Selektions- und Moduseffekten vornehmen zu können. Das nach wenigen Wochen auf das Rekrutierungs- folgende Initialisierungsinterview wurde dann bereits in dem jeweils präferierten Umfragemodus durchgeführt. Im Anschluss daran wurden die Teilnehmer danach gefragt, ob sie bereit sind, sich erneut befragen zu lassen. Über die darauf folgenden Wellen hinweg besteht im PPSM Access Panel jederzeit die Möglichkeit den Teilnahmemodus, je nach Präferenz, zu wechseln (Typ 5) (Engel et al., 2012; Engel, 2013; Vehre, 2011; Bartsch, 2012).

Typ 6 wird in der Regel bei der Durchführung länderübergreifender Befragungen angewandt. Dies hat häufig pragmatische Gründe wie etwa, dass in einem Land über das Melderegister Kontaktinformationen eruiert werden können, die in einem anderen nicht verfügbar sind (De Leeuw et al., 2008). Ein weiterer Grund kann sein, dass eine face-to-face Befragung für ein bestimmtes Land durch eine geringe Bevölkerungsdichte in einzelnen Regionen nicht infrage kommt. So wurden die Erhebungen in den ersten drei Runden des European Social Surveys als PAPI-Variante<sup>3</sup> oder CAPI-Variante in den jeweiligen Ländern durchgeführt (Stoop et al., 2010, 77). PAPI wurde dabei vor allem in den zentraleuropäischen Ländern eingesetzt, da CAPI Befragungen seinerzeit nicht sehr verbreitet gewesen waren.

Wenn eine Umfrage zunächst einmal abgeschlossen ist, gibt es in der sogenannte Follow-up Phase weiterhin die Möglichkeit all diejenigen, die der ersten Umfrageeinladung nicht gefolgt sind, an die Teilnahme zu erinnern und ggf. einen weiteren Modus zur Teilnahme anzubieten. Klausch et al. (2015) haben auf diese Weise versucht, einerseits die Kosten möglichst gering zu halten und andererseits im Rahmen eines sequentiellen Mixed-Mode Designs Selektions- von Moduseffekten statistisch zu trennen und deren Wirkung auf die Qualität der Daten zu beurteilen. Dabei wurden die potentiellen Befragten in den Niederlanden auf Basis einer Melderegisterstichprobe zufällig zu den Modi face-to-face, telefonisch, postalisch oder online zugeordnet und zunächst nur zur Teilnahme in einem dieser Modi eingeladen. In einem zweiten Schritt wird dieses Single- zu einem Mixed-Mode Design, indem eine Nachfassaktion (Follow up - Typ 7) unter den Nonrespondenten durchgeführt

---

<sup>3</sup> Paper and Panel Interview

wurde und diese die Möglichkeit hatten, sich mündlich-persönlich befragen zu lassen. Dabei konnten keine nennenswerten Unterschiede hinsichtlich der durch die einzelnen Umfragemodi verursachten Selektionseffekte in den jeweiligen Single-Mode Designs beobachtet werden (Klausch et al., 2015). Die mündlich-persönliche Befragung in der Follow-up Phase hat die Antwortraten deutlich erhöht, besonders wenn zuvor zu einer Onlinebefragung eingeladen wurde. Dabei hat sich gezeigt, dass sich die Selektionseffekte zwischen Single- und Mixed-Mode Design nicht substantiell unterscheiden und die Repräsentativität trotz teilweise deutlich höherer Antwortraten nicht nennenswert erhöht werden konnte. Da in der Studie von Klausch et al. (2015) neben der Erhöhung der Antwortraten kein positiver Effekt auf die Datenqualität durch den Einsatz eines Mixed-Mode Designs nachgewiesen werden konnte, spricht dies eher für den Einsatz von Single-Mode Befragungen. Was hierbei allerdings nicht außer Acht gelassen werden darf ist die Tatsache, dass diese Studie in den Niederlanden durchgeführt wurde und hier andere Voraussetzungen bestehen, als sie etwa für Deutschland gelten. Hier wäre unter anderem die hohe Abdeckung mit Internetanschlüssen zu nennen, die 2011 bereits bei über 90 Prozent lag oder die nach wie vor hohe Anzahl an im Telefonregister verzeichneten Rufnummern (Klausch et al., 2015, 15). Für Deutschland wären aufgrund vorliegender Studien andere Ergebnisse zu erwarten, insbesondere was Selektionseffekte und ein möglicher Bias durch Noncoverage und Nonresponse angeht.

Auch Link et al. (2008) konnten zeigen, dass mehr Personen teilnehmen, wenn eine Erinnerungskarte im Anschluss an eine telefonische Befragung versandt wird. Wenn per Telefon an die Teilnahme an einer Onlinebefragung erinnert wird, lassen sich hiermit mehr Teilnehmer gewinnen und die Antwortrate von 17,8 Prozent auf 30,3 Prozent erhöhen (Typ 7) (Schonlau et al., 2003, 219).

Festzuhalten bleibt nach der Durchsicht vorliegender Studien, dass Mixed-Mode Befragungen ein vielversprechendes Mittel zur Verringerung des TSE darstellen. Sind bei der Durchführung einer Umfrage die in Kapitel 2.1 beschriebenen Selektionseffekte in Form von Noncoverage und Nonresponse zu erwarten, ist die Kombination der Modi ein probates Mittel, um die Datenqualität zu erhöhen. So haben Dillman et al. (2009, 1) das Design und die Implementierung von Mixed-Mode Befragungen zurecht als „[...] major survey trends in the 21st century [...]“ bezeichnet.

In Deutschland steht die Forschung in Bezug auf die optimale Kombination und den sequentiellen Einsatz von Mixed-Mode Befragungen noch relativ am Anfang.

## 4.2 Teilnahmeanreize

Teilnahmeanreize (Incentives) werden primär dazu eingesetzt, um die Antwortrate von Umfragen zu erhöhen (Singer und Ye, 2013). Gleichzeitig wird hierdurch der Spielraum ausfallbedingter Verzerrungen (Nonresponse Error) zwar kleiner, bleibt allerdings trotzdem bestehen. Seitdem Antwortraten, abhängig vom Modus der Umfrage, ein teilweise bedenklich niedriges Niveau erreicht haben, ist die Frage nach einem möglichen Nonresponse Error, auch in Verbindung mit Teilnahmeanreizen, in den Interessensfokus gerückt. Denn wenn diese die Antwortrate erhöhen, kann durch den Einsatz ein zusätzlicher Bias entstehen, wenn sich etwa bestimmte Bevölkerungsgruppen hiervon stärker angesprochen fühlen als andere. Zunächst einmal ist eine Unterscheidung und Einordnung der in der Praxis zur Anwendung kommenden Anreizformen notwendig, um diesen Gedanken weiterführen zu können.

### Formen von Teilnahmeanreizen

Im Allgemeinen werden diese hinsichtlich drei dichotomer Dimensionen differenziert (Göritz, 2015, 340). Tabelle 4.2 informiert über die sich ergebenden Kombinationsmöglichkeiten. Die erste Unterscheidung findet in Bezug zur Dimension konditional und nicht konditional statt.<sup>4</sup>

Tabelle 4.2: Formen von Teilnahmeanreizen

	monetär	nicht monetär
konditional - Pro-Kopf	KPM	KP
nicht konditional - Pro-Kopf	NKPM	NKP
konditional - Lotterie	KLM	KL
nicht konditional - Lotterie	NKLM	NKL

Von Bedeutung ist hierbei, ob die Vergabe von Anreizen von einer Teilnahme abhängig gemacht wird (konditional - Variante **KPM**, **KP**, **KLM**, **KL**) oder generell unabhängig davon an alle Personen in der Stichprobe vergeben werden (nicht konditional - Variante **NKPM**, **NKP**, **NKLM**, **NKL**). Diese können wiederum „Pro-Kopf“ (Variante **KPM**, **KP**, **NKPM**, **NKP**) oder an einen zufällig ausgewählten Anteil an Personen in Form einer Lotterie beziehungsweise Verlosung vergeben werden (Variante **KLM**, **KL**, **NKLM**, **NKL**), abhängig davon, ob dies mit einer Teilnahme verknüpft wird oder nicht. Alle in Tabelle 4.2

<sup>4</sup> In der englischsprachigen Literatur wird hier auch häufig von „prepaid“ und „postpaid“ vergebenen Teilnahmeanreizen gesprochen.

abgebildeten Varianten können zuletzt in monetärer oder nicht monetärer Form eingesetzt werden (Zusatz **M**, wenn Anreiz monetär).

Monetär werden diese häufig nicht konditional in Form eines kleinen Geldbetrages (NKPM) oder konditional im Rahmen einer Verlosung mit höheren Beträgen eingesetzt (KPM). In der Praxis findet eine „Pro-Kopf“-Vergabe häufig experimentell statt, sodass auch hier nur ein zufällig ausgewählter Teil der Befragten einen Anreiz erhält. Hiermit ist im Nachhinein festzustellen, ob der Einsatz zu einer Erhöhung der Antwortrate geführt hat oder nicht. Andernfalls würden sehr hohe Kosten anfallen, wenn etwa an die komplette Stichprobe ein kleiner Geldbetrag vergeben werden würde. Deswegen ist es in der wissenschaftlichen Umfrageforschung unüblich, allen Personen in der Stichprobe oder jedem Teilnehmer einer Umfrage einen monetären Anreiz zukommen zu lassen (KPM, NKPM). Kugelschreiber, Tassen oder andere kleinere Sachanreize werden zumeist unabhängig davon vergeben, ob eine Person teilnimmt oder nicht (NKP). Hierbei kann es durchaus vorkommen, dass diese kleinen Gegenstände alle Personen in der Stichprobe erhalten (NKP), da die hiermit verbundenen Kosten nicht so hoch ausfallen, als wenn etwa an alle fünf Euro versendet werden. Die Verlosung höherpreisiger Anreize wie Smartphones, Gutscheine oder Bargeld ist üblicherweise mit einer Beteiligung an der Umfrage verknüpft (KLM, KL).

Die Auswahl der jeweiligen Form hängt in der Praxis einerseits von dem zur Verfügung stehenden Auswahlrahmen und dem sich daraus ergebenden Umfragemodus ab. Wird etwa per Telefon befragt, stehen aufgrund des Stichprobendesigns (beispielsweise Gabler-Häder-Design, siehe Kapitel 2.1.1) in der Regel über die Rufnummern der Personen hinaus keine weiteren Kontaktinformationen zur Verfügung. Deswegen ist es nicht möglich, vor der telefonischen Kontaktaufnahme einen nicht konditionalen Teilnahmeanreiz zu versenden, sodass die Personen diesen bei einem Anruf quasi „schon in den Händen hält“. Andererseits richtet sich die Auswahl der Form des Anreizes nach der zugrundeliegenden Grundgesamtheit und somit nach der Erwartung an eine potentielle Erhöhung der Antwortrate in Abhängigkeit bisheriger Forschungsergebnisse.

### **Kooperationsbereitschaft und Antwortrate**

Diesbezüglich gibt es zahlreiche Metastudien, die Teilnahmeanreizen einen positiven Effekt auf die Antwortrate bescheinigen. Church (1993) hat insgesamt 38 Studien und deren experimentellen beziehungsweise quasi-experimentellen Einsatz von Teilnahmeanreizen in unterschiedlicher Form bei postalischen Befragungen untersucht. Dabei hat sich

eindeutig gezeigt, dass nicht jede Art von Anreiz notwendigerweise zu einer höheren Antwortrate führt. Insbesondere der Versand in Verbindung mit einer postalischen initialen Umfrageeinladung, zunächst einmal unabhängig davon ob monetärer Natur oder nicht, hat zu einer signifikanten Erhöhung der Teilnehmerzahlen geführt (NKPM, NKP, NKLM, NKL). Die Antwortrate der monetären Variante (NKPM) fällt mit einer durchschnittlichen Erhöhung von 19,1 Prozent im Vergleich zur nicht monetären Variante (NKP) mit 7,9 Prozent mehr als doppelt so hoch aus (Church, 1993, 62). Vergleicht man diese Befunde etwa mit der Vergabe von Kugelschreibern, Schlüsselanhängern oder anderen nicht von der Teilnahme abhängigen materiellen Anreizen, zeigt die empirische Evidenz stringent in eine Richtung. So haben etwa Hansen und Pedersen (2012, 244) bei einer Rekrutierung von Teilnehmern für ein Onlinepanel einen nicht konditionalen materiellen Anreiz (NKP - Markenschokolade) vergeben. Im Ergebnis hat sich gezeigt, dass dies nicht zur Erhöhung der Rekrutierungsrate beigetragen hat.

Engel (2006) ist bei der Untersuchung von insgesamt 68 Studien unter Kontrolle von 177 Designbedingungen zu dem Ergebnis gekommen, dass durch monetäre Anreize mehr Personen für eine Umfrage gewonnen werden können. Dabei kommt es auf den Zeitpunkt an, zu dem diese eingesetzt werden. Am effektivsten hat sich dabei die Variante NKPM (nicht konditional - pro Kopf - monetär) gezeigt. So kann die Antwortrate vor allem dann mit einem Anreiz erhöht werden, wenn dieser nicht konditional und im Voraus an die potentiellen Teilnehmer in Form eines kleinen Geldbetrages gesendet wird (Engel, 2006).

Eine weitere Metastudie von Göritz (2006) untersucht den Effekt von Teilnahmeanreizen bei der Durchführung von Onlinebefragungen auf die Datenqualität in Form von Item- und Unit-Nonresponse. Die 32 untersuchten experimentellen Designbedingungen hinsichtlich der Wirkung auf die Antwortrate weisen in die erwartete Richtung. Einerseits besteht eine um 19 Prozent höhere Wahrscheinlichkeit in den jeweiligen experimentellen Anreizgruppen dafür, dass sich eine Person beteiligt im Vergleich zu den jeweiligen Kontrollgruppen (kein Anreiz) (Göritz, 2006, 65). Andererseits steigt in den Untersuchungsgruppen ebenso die Wahrscheinlichkeit um 27 Prozent im Vergleich zur jeweiligen Kontrollgruppe, dass der Fragebogen komplett beantwortet wird und somit die Qualität der Daten steigt (ebd.). Dass Anreize generell positiv auf die Antwortrate wirken, wenn sie monetärer Natur und nicht an eine Teilnahme geknüpft sind, gilt als gesicherte empirische Evidenz, da dieses Ergebnis bereits durch eine Vielzahl weiterer Studien repliziert werden konnte.

Werden Anreize allerdings in Aussicht gestellt und somit konditional vergeben, wie

in Form einer Verlosung (KLM, KL), sind hiermit nur geringe bis keine Effekte auf die Antwortrate verbunden (Church, 1993; Toepoel, 2012). In der Analyse von Becker et al. (2007) lag die Antwortrate in der Kontrollgruppe (kein Incentive) sogar 5 Prozentpunkte höher im Vergleich zur Gruppe derer, die einen Anreiz in Aussicht gestellt bekommen haben. Dabei ist es vor allem auch davon abhängig, in welchem Modus die Befragung durchgeführt wird und welche Kontaktinformationen durch den zugrundeliegenden Auswahlrahmen zur Verfügung stehen. Werden Befragungen der allgemeinen Bevölkerung online durchgeführt, liegen, wie schon in Kapitel 2.1.1 erwähnt, in der Regel keine E-Mail Adressen vor. Selbst wenn diese zur Verfügung stehen ist es nicht möglich, dass der Befragte in Verknüpfung mit der Umfrageeinladung einen Teilnahmeanreiz erhält. Hier werden häufig in Verbindung mit einer E-Mail-Einladung Gutscheinodes verschickt oder eine elektronische Bezahlung via PayPal oder ähnlichem angeboten. Göritz et al. (2008) haben im Kontext eines Onlinepanels den Teilnehmern eine elektronische Bezahlung von 1,50 Euro über PayPal in Aussicht gestellt (KPM). In der ersten Welle ist kein positiver Effekt auf die Antwortrate hierdurch zu beobachten gewesen. In der zweiten Welle konnte diese allerdings erhöht werden, wenn zuvor eine Bezahlung für die Teilnahme an der ersten Befragung stattgefunden hat, während in der dritten Welle sogar ein negativer Effekt beobachtet werden konnte (Göritz et al., 2008).

Becker et al. (2007), Nederhof (1983) oder auch Singer und Ye (2013) schlussfolgern, dass im Voraus gezahlte monetäre Anreize am besten wirken und Antwortraten hiermit um bis zu 30 Prozent erhöht werden können. James und Bolstein (1990, 374) bezeichnen das Versenden eines Teilnahmeanreizes in Verbindung mit einer postalischen Umfrageeinladung gar als eine von zwei Methoden, die zur Erhöhung der Antwortrate am effektivsten ist. Nicht konditionale monetäre Anreize sind folglich den konditionalen nicht monetären überlegen. Millar und Dillman (2011) sowie Messer und Dillman (2011) betonen dabei, dass ein Vergleich zwischen Online- und postalischer Befragung gezeigt hat, dass durch monetäre Anreize mehr Teilnehmer für die Onlinevariante gewonnen werden können. Werden diese zusätzlich nicht konditional eingesetzt verstärkt sich der Effekt durch Interaktion dieser beiden Vorgehensweisen nochmals (Berger, 2006; Millar und Dillman, 2011; Scherpenzeel und Toepoel, 2012; Toepoel, 2012). Für intervieweradministrierte Befragungsformen kommen Singer et al. (1999) sowie Singer et al. (2000) zu ähnlichen Ergebnissen. In einer telefonischen Befragung konnte die Antwortrate mit einem monetären Anreiz innerhalb eines Vorausbriefes um 10 Prozent erhöht werden (Singer et al., 2000, 186). Auch Jobber

et al. (2004, 23) konnten mit einer Metastudie über Anreizeffekte bei einer postalischen Befragungen nachweisen, dass mit dem Einsatz nicht konditionaler Geldbeträge (NKPM) die Antwortrate über alle Studien hinweg um 15 Prozent erhöht werden kann.

### **Anreizhöhe und Grenznutzen**

Das erfolgreiche Einsetzen von monetären Anreizen wirft zwangsläufig die Frage auf, ob mit der Höhe auch die positive Wirkung auf die Antwortrate verstärkt wird und ob es möglicherweise einen Grenznutzen gibt. Dieser soll so verstanden werden, dass bis zu einer bestimmten Höhe eines monetären Anreizes die Teilnahmewahrscheinlichkeit linear zunimmt und dieser positive Effekt lediglich bis zu einem Grenzwert zu beobachten ist, sich danach abschwächt oder nicht mehr nachzuweisen ist.

Wenn im Sinne eines ökonomischen Tausches argumentiert wird, dass die Befragten für die Investition von Zeit für die Beantwortung der Fragen als Gegenleistung bezahlt werden, wäre ein linearer Zusammenhang zwischen Anreizhöhe und Wirkung auf die Antwortrate zu erwarten. Wie Befragte die Situation subjektiv wahrnehmen, ist bis heute relativ wenig untersucht worden. Zahlreich empirisch belegt ist zwar, wie einzelne Designelemente einer Umfrage auf die Antwortrate wirken. Nach Singer und Ye (2013) gibt es allerdings keine Studien dazu, inwiefern Verbindlichkeit zur Teilnahme seitens der Befragten in der konkreten Kontaktsituation überhaupt entsteht. Empirische Evidenz zur Wirkung von Anreizen gibt es dahingehend, dass die Vergabe von monetären nicht konditionalen Anreizen (NKPM) die Haltung begünstigt, dies im Rahmen des nächsten Umfragegesuchs wieder zu erwarten. Singer et al. (1998, 163) haben diesbezüglich nachgewiesen, dass Personen, die einen Anreiz erhalten haben, signifikant eher der Aussage „people should get paid for doing surveys like this“ zustimmen. Der positive Effekt des Anreizes, der zum ersten Erhebungszeitpunkt beobachtet werden konnte, lässt sich bei folgenden Erhebungswellen wieder beobachten: Sechs Monate später ist eine Teilnahme für die Experimentalgruppe in Abhängigkeit dieses Sachverhaltes wahrscheinlicher im Vergleich zur Kontrollgruppe, die zuvor keinen Anreiz erhalten hatte (Singer et al., 1998).

Die Forschungslage bezüglich der Anreizhöhe ist keinesfalls eindeutig. Einige Studien begünstigen den ökonomischen Aspekt und kommen zu dem Ergebnis, dass die Teilnahmewahrscheinlichkeit mit der Höhe zunimmt. Andere konnten unter Kontrolle relevanter Merkmale keine Zunahme des positiven Effekts mit der Höhe des Anreizes feststellen. Engel (2006) resümiert, dass der Anreizeffekt potentiell durch die Interview- und Fragebogenlänge



kompensiert wird, was gegen eine ökonomische Interpretation spricht. Werden die Teilnehmer direkt danach gefragt, „[...] unter welchen Bedingungen es für die befragte Person eher unwahrscheinlich oder eher wahrscheinlich sei, an einer Befragung teilzunehmen, wenn sie dazu eingeladen werde“ (Engel, 2006, 7), deuten die Antworten eher in Richtung einer ökonomischen Interpretation. Kein Anreiz und eine Verlosung würde aus Sicht der Befragten einen negativen, eine kleine Aufmerksamkeit, vor allem konditional in Form von fünf Euro, einen positiven Effekt haben. Würden mehr als fünf Euro angeboten, erhöht dies entsprechend die positive Wirkung. Engel (2006) weist klar auf die Grenzen dieses Ergebnisses hin: Einerseits antizipieren die Teilnehmer eine hypothetische Situation, in der ein Anreiz vergeben werden würde. Andererseits bezieht sich dieses Ergebnis naturgemäß lediglich auf die Teilnehmer, wodurch in Frage zu stellen ist, ob dieses Ergebnis replizierbar wäre, wenn auch die Nonrespondenten an der Befragung teilgenommen hätten. Nichtsdestotrotz schlussfolgert Engel (2006, 7), „[...] dass es sinnvoll ist, die Wirkung einer Anreizbedingung über das Gegenseitigkeitsprinzip zu erklären und [...] dass dieses Prinzip eher in seiner ökonomischen als sozialen Variante zu wirken scheint“. Werden Teilnehmer nicht nach ihrer subjektiven Meinung befragt und der Anreizeffekt innerhalb experimenteller Designbedingungen untersucht, sind die Befunde eindeutiger.

Scherpenzeel und Toepoel (2012) haben jeweils 10, 20 sowie 50 Euro nicht konditional an die potentiellen Teilnehmer vergeben (NKPM). Dabei konnten keine Unterschiede hinsichtlich der erzielten Antwortrate in den einzelnen Gruppen festgestellt werden. James und Bolstein (1990) haben hingegen herausgefunden, dass die beobachtete Antwortrate mit der Höhe des monetären Anreizes steigt. Allerdings wird in der Studie mit sehr geringen Beträgen von 0,25 bis 1 Dollar experimentiert. Wird die Anreizhöhe schrittweise um 0,25 Dollar erhöht, ist hiermit im Ergebnis eine höhere Antwortrate verbunden (James und Bolstein, 1990, 352). Trussell und Lavrakas (2004) kommen in einer groß angelegten Studie in den Vereinigten Staaten mit einer Bruttostichprobe von insgesamt 49.657 Adressen zu einem ähnlichen Ergebnis. Es werden neun experimentelle Gruppen von 0 bis 10 Dollar untersucht, wobei es keine Gruppe gibt, die 9 Dollar erhalten hat. Dabei wurde zunächst telefonisch Kontakt zu den Personen in der Stichprobe aufgenommen und im Nachhinein ein schriftlicher Fragebogen versendet. Dieser wurde unabhängig davon versandt, ob die Personen bei der initialen Kontaktaufnahme erreicht wurden, einer Beteiligung an der schriftlichen Befragung zugestimmt haben oder nicht. Im Ergebnis zeigt sich, dass der Effekt auf die Antwortrate sowie Kooperationsbereitschaft generell positiv ist und mit der

Anreizhöhe zunimmt (Trussell und Lavrakas, 2004, 362). Konsistent mit der „Leverage-Saliency“ Theorie der Umfrageteilnahme (siehe Kapitel 3.2) ist bei lokaler Betrachtung der Untersuchungsgruppen, dass die Ausprägung des Anreizeffektes vom Ergebnis des initialen telefonischen Kontaktes abhängig ist. Ist dieser erfolgreich gewesen, ist der Effekt auf die Antwortrate am größten im Vergleich zwischen der Null- und Ein-Dollar-Gruppe, steigt bis zu 6 Dollar weiter moderat an und verliert danach bis hin zur Obergrenze von 10 Dollar seine Signifikanz. Dies ist nach Trussell und Lavrakas (2004) ebenfalls eine Evidenz für die Gültigkeit der Theorie des sozialen Austausches nach Dillman et al. (2014), denn es sind vor allem die kleinen Beträge unter 6 Dollar, die dafür sorgen, dass mehr Personen den Fragebogen zurück schicken. Wurden die Haushalte nicht erreicht oder haben diese eine Teilnahme an der postalischen Befragung verweigert, zeigt sich ein anderes Bild. Hier ist zwar ebenso ein positiver Effekt zu beobachten, dieser nimmt aber nur marginal sukzessive mit der Höhe des Anreizes zu. Diese Befundlage erhärtet sich weiter durch die bereits erwähnte experimentelle Untersuchung von Jobber et al. (2004). Diese konnten zeigen, dass der Effekt auf die Antwortrate mit jedem weiteren Viertel Dollar in dem beobachteten Intervall von 0,25 bis hin zu 1 Dollar linear um jeweils 2 Prozent zunimmt (Jobber et al., 2004, 23).

Diese Befundlage spricht eher gegen eine ökonomische Interpretation, da man bei solch niedrigen Beträgen nicht von einer Bezahlung für die zu investierende Zeit im Sinne eines (Aus-)Tausches sprechen kann. Hier kommt in der Praxis offensichtlich eher der austauschtheoretische Aspekt nach Dillman et al. (2014) zum Tragen. Widmet man sich dem zweiten Teil der Ausgangsfrage dieses Abschnitts, ist nach Durchsicht relevanter Untersuchungen eine recht eindeutige Antwort möglich. Darüber, dass zwischen der Anreizhöhe und dessen Effekt auf die Antwortrate ein linearer Zusammenhang besteht, liegt in der Forschungspraxis ein Konsens vor. Denn es konnte im Bereich von bis zu 5 Euro/Dollar ein Anstieg der positiven Auswirkung auf die Antwortrate nachgewiesen werden. Ob ein bestimmter Grenzwert existiert und sich der Zusammenhang somit ab einem bestimmten Wert anders darstellt, kann an dieser Stelle nicht abschließend beantwortet werden. Wird die Höhe eines Anreizes wesentlich auf 20, 30 oder 40 Euro erhöht sind die empirischen Evidenzen uneindeutig, wenn hiermit nicht gar ein negativer Effekt verknüpft sein kann (Berger, 2006).

## Panelteilnahme und -mortalität

An den Einsatz von Teilnahmeanreizen im Kontext eines Panels knüpft die Frage an, inwiefern die Teilnehmer hierdurch etwa langfristig motiviert werden können an möglichst vielen der einzelnen Befragungswellen teilzunehmen. Die Forschungslage offenbart diesbezüglich eine teilweise differierende Befundlage im Vergleich zu reinen Querschnittserhebungen.

In einer Untersuchung von Göritz und Wolff (2007, 104) mittels eines Onlinepanels hat sich ergeben, dass der Hinweis auf eine Verlosung von Gutscheinen im Wert von jeweils 20 Euro in der Einladungsmail (KL) zu einer Erhöhung der Antwortrate der ersten Welle um 12,7 Prozent führt. Dies hat allerdings nicht mit einer Rekrutierung von Teilnehmern für eine Panelbefragung stattgefunden, sondern die befragten Personen sind zum Zeitpunkt der Kontaktierung bereits Mitglied darin gewesen. Die initiale Rekrutierungsbefragung hat hier schon stattgefunden. Diese haben zunächst den Charakter einer Querschnittsbefragung, da erst anschließend an ein erstes Interview erhoben wird, ob die Teilnehmer dazu bereit sind, sich in regelmäßigen Abständen befragen zu lassen. Dieser positive Einfluss wird über die einzelnen Panelwellen hinweg mitgetragen und konnte in vorliegenden Studien zumindest für die jeweils zweite Welle, wenn nur in abgeschwächter Form, beobachtet werden (Göritz et al., 2008). Arzheimer und Klein (1998) konnten in der zweiten Welle keine Unterschiede im Teilnahmeverhalten zwischen Personen feststellen, die in der ersten Welle einen Anreiz oder keinen Anreiz erhalten haben. Aus der Forschung ist diesbezüglich bekannt, dass für Personen, die an der ersten Welle teilgenommen haben, die Wahrscheinlichkeit entsprechend höher ist, sich auch an den Folgewellen zu beteiligen (Engel et al., 2012; Engel und Schmidt, 2014). Dieser Effekt wird in der Praxis als „Markov Prozess“ (Göritz und Wolff, 2007, 100) oder „Markov Kette“ (englisch Markov Chain) (Toepoel, 2012, 217) bezeichnet. Eine mögliche Konklusion hieraus kann sein, dass durch die Einwilligung seitens der Teilnehmer, sich regelmäßig befragen zu lassen, ein gewisses Vertrauensverhältnis entsteht und eine Verlosung somit insgesamt seriöser wahrgenommen wird, etwa durch die Veröffentlichung von Gewinnern vorhergehender Verlosungen auf der Internetseite der Panelstudie (Toepoel, 2012).

In der Marktforschung wird häufig mit Punktesystemen gearbeitet. Mit jeder Beteiligung an einer weiteren Welle sammeln die Personen Punkte, die sie sich nach bestimmter Zeit in Form von Gutscheinen oder bar auszahlen lassen können. Göritz (2004) hat diesbezüglich innerhalb eines Onlinepanels herausgefunden, dass die Antwortrate durch Bonuspunkte

im Vergleich zu einer Verlosung insgesamt höher ausfällt. Die Höhe der vergebenen Bonuspunkte hängt dabei linear mit der Höhe der Antwortrate zusammen.

### **Zusammensetzung der Stichprobe und Nonresponse Bias**

Treten durch den Einsatz von Anreizen Selektionseffekte derart in Kraft, dass trotz einer höheren Antwortrate ein größeres Maß an Verzerrung zu beobachten ist? Diese Frage ist zumindest in Bezug zu Onlinebefragungen bis heute relativ wenig untersucht worden (Singer und Ye, 2013, 129).

Einige Studien konnten in Abhängigkeit des eingesetzten Anreizes und des jeweiligen Befragungsmodus einen Effekt auf die Zusammensetzung der Stichprobe feststellen, andere konnten dies nicht bestätigen. Groves et al. (2000) sind daran anknüpfend zu dem Ergebnis gekommen, dass Personen mit einer niedrigen gesellschaftlichen Einbindung signifikant eher dazu bereit sind, sich an einer Umfrage zu beteiligen, wenn ein monetärer Anreiz vergeben wird. Zudem gibt es weitere empirische Evidenz derart, dass sich eher niedriger gebildete Personen von Teilnahmeanreizen angesprochen fühlen, insbesondere wenn diese in monetärer nicht konditionaler Form vergeben werden (Martin und Winters, 2001; Nederhof, 1983; Petrolia und Bhattacharjee, 2009; Singer und Ye, 2013; Vehre et al., 2013). Da aus der Forschungspraxis hinlänglich bekannt ist, dass es die Personen aus den unteren gesellschaftlichen Schichten mit niedrigem Bildungsniveau sind, die in wissenschaftlichen Umfragen unterrepräsentiert sind, spiegelt diese Befundlage das Potential von Anreizen wider, den Spielraum für mögliche Verzerrungen zu senken. Esser (1973) hat hier einst von einem sogenannten Mittelschichts-Bias gesprochen, womit gemeint ist, dass mehrheitlich Personen aus gehobeneren gesellschaftlichen Milieus an Umfragen teilnehmen, da hier entsprechende Einstellungen und Wertesysteme prominenter sind. Aktuellere Studien bezeichnen dieses Phänomen als Bildungsbias (Engel und Schmidt, 2014; Schnell, 1991; Vehre et al., 2013). Berger (2006, 94) hat eine zusammenfassende Erklärung für den Effekt von Anreizen auf die Zusammensetzung der Teilnehmer und damit einhergehende mögliche Verzerrungen geliefert: „Materielle Anreize können aber je nach kulturellem Hintergrund unterschiedliche Wertesysteme aktivieren und bei differentieller Anwendung möglicherweise ein Mittel zur Reduktion systematischer Ausfälle darstellen.“ Das in Kapitel 3.1 beschriebene Prinzip der Reziprozität scheint bei Personen unterer Gesellschaftsschichten nicht in der Form ausgeprägt zu sein, dass die Teilnahme an Befragungen etwa als eine Art Bürgerpflicht angesehen wird. Vor diesem Hintergrund entsteht durch den Einsatz von Anreizen in

dieser Gruppe von Personen ein subjektiv wahrnehmbarer Nutzen, sodass der Effekt tendenziell stärker ausgeprägt ist. Brehm (1994) konnte allerdings keine Unterschiede im Antwortverhalten in Abhängigkeit davon feststellen, ob die Befragten einen Dollar oder einen Stift als nicht konditionalen Anreiz mit dem Einladungsschreiben erhalten haben.

Anreize können nicht zuletzt als Maßnahme zur Konvertierung von Verweigerern eingesetzt werden. Diese sogenannten „refusal conversation incentives“ können etwa dann Anwendung finden, wenn potentielle Befragte weder mit Argumenten noch mit anderen Mitteln von einer Teilnahme zu überzeugen sind. Currivan (2005) haben Personen, die zunächst verweigert haben, 20 Dollar für eine Teilnahme in Aussicht gestellt. In der Folge hat sich ein größerer Anteil an Personen beteiligt, die älter als 55 Jahre sind, keinen Universitätsabschluss besitzen und aktuell keiner bezahlten Arbeit nachgehen (Currivan, 2005, 16). Diese Selektionseffekte bringen das Potential mit, die Repräsentativität der Daten zu verbessern, da dieser Kreis an Personen üblicherweise unterrepräsentiert ist.

Weiterhin kann das Thema der Umfrage und ein mögliches Desinteresse seitens der Befragten eine eher untergeordnete Rolle spielen, wenn diese im Vorhinein und unabhängig von einer Teilnahme einen kleinen Geldbetrag zugesandt bekommen (Groves et al., 2000, 2004; Marcus et al., 2007; Singer und Ye, 2013). Anreize und Umfragethema interagieren wie im Rahmen der „Leverage-Saliency“ Theorie der Umfrageteilnahme beschrieben: Besteht ein eher geringes Interesse am Thema der Umfrage kann dieser potentiell negative Effekt durch einen Anreiz kompensiert werden. Einer geringen intrinsischen Motivation seitens der Befragten kann offensichtlich mit dem Einsatz von Anreizen entgegengewirkt werden (Vehre et al., 2013).

Dieser Befund birgt allerdings die Gefahr einer weiteren potentiellen Verzerrungsquelle, vermittelt über ein geringeres Interesse am Thema der Umfrage. Denn aus der Forschungspraxis ist bekannt, dass Personen, die etwa wegen mangelndem Interesse erst zu einer Teilnahme überzeugt werden müssen, in der Folge weniger kognitive Mühen in die Beantwortung der Fragen investieren (Engel und Köster, 2014, 48). Hier spricht man von Satisficing-Effekten (Krosnick, 1991, 1999). Denkbar wäre ebenso, dass durch eine geringere intrinsische Motivation zur Teilnahme die Wahrscheinlichkeit steigt, dass einzelne Fragen nicht beantwortet werden und die Qualität der Daten negativ beeinflusst wird. Die Datenqualität wird häufig über das Maß an Item-Nonresponse sowie den Umfang von Antworten auf offene Fragen operationalisiert. Singer und Ye (2013) resümieren diesbezüglich, dass es einerseits relativ wenig Studien gibt, die sich dieser Problematik annehmen.

Andererseits konnte in der Mehrzahl der Untersuchungen bislang kein Effekt von Anreizen auf die Qualität der Daten in Form von Item-Nonresponse nachgewiesen werden (Singer und Ye, 2013, 125).

Göritz (2006) ist dieser Frage nachgegangen und hat in einer Metaanalyse insgesamt 26 Feldexperimente hinsichtlich der Wirkung von Anreizen auf die sogenannte „Retention-Rate“ untersucht. Hiermit soll operationalisiert werden, inwiefern Personen, die ein Interview begonnen haben, die Fragen bis zum Schluss beantworten. Materielle Anreize erhöhen nach Göritz (2006, 66) nicht nur die Antwortrate um 2,7 Prozent, sondern ebenso den Anteil an Personen um 5,07 Prozent, die den Fragebogen komplett beantworten. Insgesamt gibt es empirische Evidenz dahingehend, dass Anreize nicht nur mehr Personen zu einer Teilnahme motivieren, sondern auch der Anteil an Item-Nonresponse leicht verringert werden kann.

### 4.3 Melderegister

Stichproben aus den allgemeinen Einwohnermelderegistern gelten rechtlich als Gruppenauskunft (Koch, 1997) und haben aufgrund einiger Vorteile gegenüber anderen Verfahren einen guten methodischen Ruf. So basierte etwa die letzte Zensus-Befragung (früher sog. Volkszählung) aus dem Jahr 2011 auf einer deutschlandweiten registergestützten Stichprobe.<sup>5</sup>

Liegt eine Studie „im öffentlichen Interesse“, wird das Melderegister als Auswahlrahmen zu Forschungszwecken zur Verfügung gestellt. Im öffentlichen Interesse liegt eine Studie, wenn „[...] ein wohlverstandenes Interesse der Allgemeinheit, dem ein soziologisch/gesellschaftlicher Bezug immanent ist oder die Rede ist von einem Interesse der Allgemeinheit, was zum Beispiel wissenschaftliche Forschungsvorhaben sind“ (Albers, 1997, 118). Hat man das öffentliche Interesse der geplanten Studie gegenüber der zuständigen Meldebehörde nachgewiesen, wird von dieser die Stichprobe generiert und der anfragenden Institution zur Verfügung gestellt. Strebt man allerdings eine deutschlandweite Befragung an, ergibt sich daraus eines der zentralen Probleme. Da das Meldewesen in Deutschland Ländersache ist, hat dies einerseits zu unterschiedlichen gesetzlichen Rahmenbedingungen geführt, was sich darin äußern kann, dass es zu Abweichungen hinsichtlich der Informationsvielfalt der gelieferten Stichprobendatenbanken kommen kann. Andererseits unterscheiden

---

<sup>5</sup> [www.zensus2011.de](http://www.zensus2011.de)

sich die einzelnen zuständigen Behörden in Anbetracht der benötigten Zeit zur Bearbeitung der Anfragen. Durch teilweise große zeitliche Variation unterscheidet sich der Aktualitätsgrad der Stichproben. So hat Von der Heyde (2014, 192) nicht umsonst darauf verwiesen, dass „es unter Umständen sehr lange dauern kann, bis die ausgewählten Gemeinden mit den Daten rausrücken“. Praktisch wird dies dadurch relevant, dass in älteren Stichproben der Anteil an Personen größer ist, die verzogen sind und deren Anschrift nicht mehr korrekt ist. In der Folge kann dies zu mehr Kontaktausfällen führen. Lokal betrachtet wird die Aktualität der jeweiligen Registerdaten dadurch in Mitleidenschaft gezogen, dass Zu- und Fortzüge, Sterbefälle und Geburten erst nach einiger Zeit registriert werden (Von der Heyde, 2014). Da regional beschränkte Befragungen allerdings relativ zeitnah nach der Stichprobenziehung erfolgen, kann dies nach Von der Heyde (2014) vernachlässigt werden.

Wenn man als Forscher bereit ist, die benötigte Zeit und Kosten für eine deutschlandweite Registerstichprobe aufzubringen, gelten diese als der „Rolls Royce der Umfrageforschung“ (Von der Heyde, 2014, 192), denn diese Vorgehensweise gilt als methodisch sauber (Albers, 1997). Kosten entstehen dadurch, dass pro Stichprobe eine Aufwandsentschädigung an die zuständige Behörde gezahlt werden muss, die in der Regel zwischen 100 und 150 Euro liegt. Für deutschlandweite Befragungen können hier schnell hohe Summen anfallen, denn die meisten Meldeämter sind auf Gemeindeebene organisiert.

Mit dem jeweils zugrundeliegenden Stichprobenansatz, der im Wesentlichen von dem zum Einsatz kommenden Erhebungsverfahren und der betreffenden Grundgesamtheit abhängig ist, gehen praktische Limitationen einher. Diese sind insbesondere davon abhängig, welche Informationen über die Personen in der Bruttostichprobe durch den Auswahlrahmen zur Verfügung stehen.

Wie bereits in Kapitel 2.1.2 beschrieben, kann die gängige Stichprobenpraxis zur Durchführung telefonischer Befragungen benannt werden. Dadurch, dass der Auswahlrahmen lediglich die Rufnummern der Haushalte enthält, ist der Einsatz nicht konditionaler Teilnahmeanreize oder postalischer Vorankündigungen nicht möglich. Diese unzweifelhaft ausschöpfungssteigernden Designelemente können somit nicht strategisch eingesetzt werden, um die Antwortrate zu erhöhen. Ähnlich verhält es sich bei Onlinebefragungen. Da kein Auswahlrahmen für allgemeine Bevölkerungsumfragen zur Verfügung steht wird, wie in Kapitel 2.1.2 bereits diskutiert, auf andere Kommunikationskanäle, wie das Telefon, zurückgegriffen. Dieses Vorgehen wird häufig für die Rekrutierung von Access Panels oder Internet Panels im letzten Teil eines telefonischen oder postalischen Interviews eingesetzt

(Bandilla et al., 2013; Engel und Schmidt, 2015). Ein vielversprechender Ansatz um einerseits der Stichprobenproblematik der telefonischen sowie Onlineumfrage entgegen zu wirken sowie andererseits etwas mehr Licht in die „Black-Box“ der Nonrespondenten zu bringen und das Ausmaß eines möglichen Bias einschätzen zu können, ist die Generierung einer Stichprobe auf Basis des Einwohnermelderegisters (Albers, 1997; Von der Heyde, 2014; Koch, 1997; Messer und Dillman, 2010). Messer und Dillman (2010, 3) schlussfolgern diesbezüglich folgerichtig: „A potential way to overcome coverage and sampling problems of both the Internet and RDD is to use ABS (address based sampling).“

Denn wird eine Stichprobe auf Basis des Melderegisters gezogen, besteht der wesentliche Vorteil darin, dass Anschriften und darüber hinausgehende Informationen von allem Personen, unabhängig davon ob diese letztlich teilnehmen oder nicht, zur Verfügung stehen. Hierdurch ist es möglich, die Befragung durch einen postalischen Vorausbrief in Verknüpfung mit einem Teilnahmeanreiz anzukündigen. Damit geht, nach vorliegenden Studien zu urteilen, nicht nur eine Erhöhung der Antwortrate einher. Ebenso ist eine Verringerung des potentiellen Bias aufgrund von Nonresponse möglich, da Anreize offenbar besonders Personen ansprechen, die üblicherweise unterrepräsentiert sind (siehe Kapitel 4.2). Weiterhin sind im Auswahlrahmen Informationen wie Alter, Geschlecht und Nationalität enthalten. Über die Anschriften kann zusätzlich der jeweilige Stadtteil, in dem eine Person lebt, charakterisiert werden, etwa durch die jeweils vorherrschende Arbeitslosenquote oder den Anteil an Personen mit Migrationshintergrund. Dieses Set an Informationen macht einen Vergleich zwischen Respondenten und Nonrespondenten hinsichtlich potentieller Selektionseffekte möglich. Somit kann festgestellt werden, ob in Abhängigkeit der Informationen, die für die gesamte Stichprobe zur Verfügung stehen, ein Bias durch Nonresponse in den Daten vorliegt (Koch, 1997). Um dieses Potential zur Adjustierung von Nonresponse ausschöpfen zu können, müssen die Hilfsvariablen allerdings sowohl mit der Teilnahmewahrscheinlichkeit (Response Propensity) als auch mit den interessierenden Umfragevariablen korrelieren (Kreuter, 2013b; Little und Vartivarian, 2005).<sup>6</sup> Engel (2006) hat festgestellt, dass die Erklärungskraft der über das Melderegister verfügbaren Informationen bezüglich der Teilnahmewahrscheinlichkeit mit 1,1 bis 3,2 Prozent nur gering ausfällt. Weiterhin wird die Erklärungskraft dieser Informationen für selektiven Nonresponse nach Engel und Schmidt (2015, 208) sowie Lynn (2003, 239) dadurch beeinträchtigt, dass die Informationen

---

<sup>6</sup> Die dahinterstehenden statistischen Konzepte zur Kompensation von Nonresponse werden ausführlich in Kapitel 4.5 beschrieben und diskutiert.



über den Auswahlrahmen im besten Falle nur moderat mit den Umfragevariablen korreliert sind. Das Set an Hintergrundinformationen über die Bruttostichprobe kann somit allenfalls erweitert werden. Darüber hinaus lässt sich in der Forschungspraxis häufig beobachten, dass nicht jede Dimension eines Fragebogens mit den Variablen aus dem Auswahlrahmen korrelieren. Da bei vielen Methoden zur Adjustierung eines selektiven Nonresponse das Problem besteht, aussagekräftige Hintergrundvariablen zu eruieren (Schouten, 2004, 2), muss eine lokale Untersuchung im Kontext der jeweiligen Studie stattfinden. Die Stärken des Melderegisteransatzes scheinen insbesondere darin zu liegen, dass die potentiellen Teilnehmer vorab kontaktiert werden, mehrere Befragungsmodi zu einem Mixed-Mode Design kombiniert und Teilnahmeanreize eingesetzt werden können.

Ein Mixed-Mode Design kann dazu beitragen, den Total Survey Error zu verringern. Das Potential zur Verringerung möglicher Fehlerquellen kann noch besser genutzt werden, wenn sequentiell vorgegangen wird. Dies bedeutet, dass in einem ersten Kontaktierungsschritt ausschließlich zu einer Onlinebefragung eingeladen wird und in weiteren entweder zunächst eine Erinnerung stattfindet oder ein anderer Modus zur Teilnahme angeboten wird. So kann durch die Adressinformationen eine postalische Einladung zu einer Onlinebefragung versandt werden. Daraufhin kann an die Teilnahme an der Umfrage durch einen Reminder erinnert werden. Wer nach dem zweiten oder dritten sequentiellen Kontaktierungsschritt noch nicht teilgenommen hat, beispielsweise da das Internet nicht regelmäßig genutzt wird oder kein Internetanschluss vorhanden ist, kann der Versand eines schriftlichen Fragebogens erfolgen. So gibt man den Personen, die bei internetbasierten Umfragen unterrepräsentiert sind die Möglichkeit zur Teilnahme, sodass der Effekt durch Noncoverage verringert oder gar beseitigt werden kann. Messer und Dillman (2010, 2011) haben in ihren Feldexperimenten hierzu nachgewiesen, dass mit einem solchen sequentiellen Mixed-Mode Design auf Basis einer Melderegisterstichprobe die Antwortrate in einer Onlinebefragung erhöht werden und gleichzeitig der Fehler aufgrund von Coverage-Effekten verringert werden kann (wie bereits in Kapitel 4.1 erläutert). Die Kombination Online- und postalischer Befragungen mit Vorankündigungen und nicht konditionalen Anreizen führt zu einer ähnlichen Zusammensetzung der Teilnehmer, als wenn ausschließlich postalisch befragt wird (Messer und Dillman, 2011, 449).

Obwohl keine Rufnummern in einer Registerstichprobe vorhanden sind, ist eine Befragung über das Telefon möglich. So kann über die Adressen der Personen in der Stichprobe ein Abgleich mit dem Telefonregister erfolgen, um so die Rufnummern derer zu recherchie-

ren, die darin verzeichnet sind. Dies muss durch einen weiteren Befragungsmodus flankiert werden, da das Telefonregister allein nicht mehr als Auswahlrahmen geeignet ist (siehe Kapitel 2.1.1). Andernfalls muss mit Coverage-Effekten und einem Bias gerechnet werden, da sich Eingetragene von Nicht-Eingetragenen systematisch unterscheiden (Schnell, 2012). Personen, deren Rufnummern feststellbar sind, können dann per Post auf den Anruf vorbereitet und daraufhin telefonisch befragt werden. So wird ein kalter Kontakt vermieden und eine Teilnahme wahrscheinlicher. Ferner kann ein Anruf als eine Art Reminder eingesetzt werden. Die Personen, die sich nach einer gewissen Zeit nicht an einer postalischen oder internetbasierten Umfrage beteiligt haben, werden per Telefon daran erinnert, um die Antwortrate zu erhöhen. Link et al. (2008) haben nachgewiesen, dass Personen, die über öffentliche Datenbanken wie das Telefonregister identifizierbar sind, eher an Umfragen teilnehmen. Kann die Rufnummer recherchiert werden, ist in dieser Gruppe von Personen eine höhere Teilnahmewahrscheinlichkeit zu erwarten.

An dieser kann Stelle festgehalten werden, dass die Vorteile einer Registerstichprobe die Nachteile überwiegen, vor allem wenn Befragungen lokal beschränkt durchgeführt werden. Für deutschlandweite Studien müssen teils lange Wartezeiten sowie relativ hohe Kosten in Kauf genommen werden. Liegt nicht allzu viel Zeit zwischen der Auslieferung der einzelnen Stichprobendateien, bekommt man allerdings einen methodisch fast einwandfreien Auswahlrahmen zur Verfügung gestellt, der den Möglichkeitsspielraum zur Verringerung des Total Survey Errors über die Informationsvielfalt gegenüber anderen gängigen Verfahren deutlich vergrößert. So können etwa Mixed-Mode Befragungen strategisch eingesetzt werden, wobei zwischen potentiellen Fehlerquellen wie Noncoverage und Nonresponse sowie Moduseffekten abgewogen werden muss.

In der Forschungspraxis wurde bereits darauf verwiesen, dass aufgrund der Gefahr, dass Variablen aus dem Melderegister nicht mit den Umfragevariablen korrelieren, darüber hinausgehende Informationen insbesondere über die Gruppe der Nonrespondenten routinemäßig bei jeder Befragung gesammelt werden sollten (Wagner, 2010, 239). Diese sogenannten Paradata sind in diesem Kontext unverzichtbar für die Umfrageforschung geworden.

## 4.4 Parاداتen

Computer sind aus der Umfrageforschung nicht mehr wegzudenken. Seitdem diese zur Durchführung von Befragungen immer häufiger in unterschiedlicher Form verwendet werden, sind Parاداتen, beispielsweise zur Kontrolle und Verbesserung des Feldprozesses, immer wichtiger geworden. Selbst face-to-face Befragungen (CAPI, CASI) werden bereits seit längerer Zeit durch Befragungscomputer unterstützt. So können die Antworten unmittelbar digitalisiert werden, wodurch die Daten schneller für nachfolgende Analysen zur Verfügung stehen. Parاداتen sind sprichwörtlich formuliert „Daten neben den Daten“, die computergestützt automatisiert mit erhoben oder durch Interviewer in Kontakt- oder Beobachtungsprotokollen festgehalten werden (Engel und Schmidt, 2014; Felderer et al., 2014; Kreuter und Carolina, 2010; Kreuter und Olson, 2013). Offiziell zugängliche Sekundärdaten, wie etwa das Sozio-oekonomisches Panel (SOEP) oder der ALLBUS, enthalten solche Informationen in der Regel nicht, da der Fokus eher auf den Inhalten der jeweiligen Themenschwerpunkte und nicht auf der Methodenforschung liegt (Vehre et al., 2013).

Da jede Umfrage von Nonresponse betroffen ist, bedarf es auf der einen Seite weiterer Forschung, um Teilnahmewahrscheinlichkeiten möglichst zu maximieren. Auf der anderen Seiten ist es von Bedeutung, einen möglichen Bias durch Nonresponse durch ein breites Set an Hintergrundvariablen zu operationalisieren und wenn notwendig zu adjustieren. Das Problem wissenschaftlicher Umfragen bleibt hierbei stets das gleiche: Über die Personen, die nicht teilgenommen haben, liegen keine oder nur sehr begrenzte Informationen vor. Deswegen geht die Umfrageforschung seit einigen Jahren neue Wege, um durch methodische Begleitforschung möglichst genau abschätzen zu können, ob eine niedrige Antwortrate mit einem Bias einhergeht und vice versa.

Parاداتen haben, vergleichbar mit den Informationen, die über den Auswahlrahmen in einer Registerstichprobe verfügbar sind, einen wesentlichen Vorteil: Einige stehen sowohl für Respondenten als auch für Nonrespondenten zur Verfügung (Kreuter und Olson, 2013). Sind diese Hilfsvariablen, wie schon in Kapitel 4.3 erwähnt, sowohl mit der Teilnahmewahrscheinlichkeit als auch mit zentralen Umfragevariablen korreliert, bieten diese neben Informationen aus dem Auswahlrahmen das Potential zur Operationalisierung und Verringerung des Nonresponse Error (Maitland et al., 2009). Groves und Heeringa (2006, 448) bezeichnen Parاداتen bezüglich ihrer Erklärungskraft für die Wahrscheinlichkeit der Umfrageteilnahme als endogene Zielgröße als „[...] good predictors of the propensity of

a sample unit to participate in the survey.“ Vor diesem Hintergrund werden Paradata in der Praxis genutzt, um einen möglichen Bias durch Nonresponse und Messeffekte zu untersuchen (Couper und Kreuter, 2013; Kreuter et al., 2010; Smith, 2011). Tabelle 4.3 gibt eine Übersicht der verfügbaren Informationen, die automatisiert oder durch den Interviewer dokumentiert werden können und welche Fehlerquelle hiermit adressiert werden kann.

Tabelle 4.3: Arten von Paradata nach Befragungsmodus und Potential zur Fehlerreduzierung

Modus	Art der Hilfsinformation (nicht erschöpfend)	Gewichtungs-/Adjustierungspotential
<b>Computer Aided Telephone Interviewing (CATI)</b>	<i>Kontaktversuche</i> <i>Kontaktresultate</i> <i>Anrufzeiten</i> <i>Anrufbeantworter besprochen</i> <i>Interviewer pro Kontakt</i> <i>Regionalkennziffern</i> Antwortrate	Verbesserung/Anpassung der Feldphase/des Studiendesigns Reduzierung Nonresponse Error Reduzierung Measurement Error
<b>Computer Aided Personal/Self Interviewing (CAPI/CASI)</b>	Verweildauer bei einer Frage GPS-Koordinaten <i>Kontaktresultate</i> <i>Interviewer pro Kontakt</i> <i>Regionalkennziffern</i> Antwortrate  Über Kontaktprotokoll: <i>Charakteristika der Nachbarschaft</i> <i>Charakteristika des Haushaltes</i> <i>Charakteristika der Mitglieder eines Haushaltes</i>	Verbesserung/Anpassung der Feldphase/des Studiendesigns Reduzierung Nonresponse Error Reduzierung Measurement Error Reduzierung Coverage Error
<b>Computer Aided Web Interviewing (CAWI)</b>	Verweildauer bei einer Frage Start- und Endzeitpunkt Typ des Browsers Unterbrechungen Antwortrate Wortanzahl (offene Fragen)	Verbesserung/Anpassung der Feldphase/des Studiendesigns Reduzierung Nonresponse Error Reduzierung Measurement Error

Anmerkung: *Kursiv* - Information für Respondenten sowohl als auch Nonrespondenten vorhanden.

Da es sich bei schriftlichen Befragungen um eine selbstadministrierte Befragungsform handelt, bei der weder ein Interviewer eingesetzt wird noch durch Softwareunterstützung Parainformationen erhoben werden, sind diese nicht in Tabelle 4.3 aufgeführt (Smith, 2011). Sogenannte Kontextdaten können im Nachhinein an einen bestehenden Datensatz

angefügt werden (Vehre et al., 2013). Hierbei handelt es sich um Regionalkennziffern, über die Unterschiede der jeweiligen Region in der eine Person lebt, kontrolliert werden können. Telefonische (Computer Aided Telephone Interviewing (CATI)), face-to-face (CAPI; CASI) und Onlinebefragungen (CAWI) bieten durch Softwareunterstützung sowie den Einsatz von Interviewern (CATI; CAPI/CASI), im Vergleich zu schriftlichen Befragungen, ein wesentlich umfangreicheres Set an Paradata. Über Kontaktversuche, Kontaktergebnisse sowie Kontaktierungszeiten, die etwa von einer CATI-Software automatisiert dokumentiert werden, können nicht zuletzt Feldprozesse optimiert werden. In einem Access Panel können diese dazu genutzt werden, um in einer zweiten oder dritten Welle ein Zeitintervall einzugrenzen, in dem die betreffende Person am besten zu erreichen ist. Dieses Vorgehen ist einem „Responsive Design“ relativ ähnlich. Groves und Heeringa (2006) unterteilen den Feldprozess in drei Phasen und nutzen Parainformationen, um vor dem Übergang in die nächste Phase die Teilnahmewahrscheinlichkeit derer zu schätzen, die bis dahin noch nicht teilgenommen haben. Dabei greifen Groves und Heeringa (2006) auf Interviewereinschätzungen über die allgemeine Teilnahmebereitschaft, Gebietsdaten oder die Tageszeit und den Wochentag des Kontaktes sowie auf das Ergebnis des vorausgegangenen Kontaktversuches zurück. Mit diesen Informationen können über logistische Regressionsmodelle schwierig zu rekrutierende Teilgruppen identifiziert werden, um in der nächsten Phase des Feldprozesses bestimmte Designelemente, wie etwa Anreize oder einen anderen Befragungsmodus, einzusetzen und so die Wahrscheinlichkeit einer Teilnahme zu erhöhen. Groves und Heeringa (2006) sind in ihrer Untersuchung zu dem Ergebnis gekommen, dass mit einem derartigen „Responsive Design“ nicht nur Kosten gesenkt, sondern ebenso der Nonresponse Error verringert werden kann.

Mit den finalen Kontaktergebnissen werden im Anschluss an eine Feldphase Antwortraten nach den gängigen AAPOR-Standards berechnet (AAPOR, 2015). Die finalen Kontaktergebnisse können ebenso dazu genutzt werden, um den Kontaktverlauf als Determinante der Umfrageteilnahme zu untersuchen. Engel et al. (2012) haben in der Aufbauphase des PPSM Access Panels die Teilnahmewahrscheinlichkeit in Abhängigkeit des Kontaktverlaufs geschätzt. Dabei hat sich gezeigt, dass die Teilnahmewahrscheinlichkeit durch das Vereinbaren von Terminen mit der Zielperson erhöht werden kann (Engel et al., 2012).

Software zur Durchführung von Onlinebefragungen (CAWI), wie etwa LimeSurvey oder Globalpark,<sup>7</sup> dokumentieren die Verweildauer bei einer Frage, Start- und Endzeitpunkt

---

<sup>7</sup> Für detaillierte Informationen siehe [www.limesurvey.org](http://www.limesurvey.org) oder [www.questback.com](http://www.questback.com).

der Teilnahme, Unterbrechungen oder die Wortanzahl bei offenen Fragen (Callegaro, 2013; Couper und Bosnjak, 2010). Weiterhin kann festgestellt werden mit welchem Webbrowser eine Person den Fragebogen beantwortet hat und hierüber, welches Betriebssystem auf dem jeweiligen Endgerät installiert ist. Diese „session-level“ Paradata werden meist auf der ersten Fragebogenseite mit dem Anklicken des Uniform Resource Locator (URL) erhoben (Callegaro, 2013). Hiermit können mögliche Probleme einzelner Fragen lokalisiert und bereinigt werden, sodass die Wahrscheinlichkeit für das Auftreten von Item-Nonresponse und eines Measurement Errors verringert werden kann. Mit CAWI-Befragungen lassen sich nur sehr begrenzt Parainformationen erheben, die für die komplette Stichprobe zur Verfügung stehen. Lediglich der Zeitpunkt des Versendens von Einladungs- und Erinnerungsmails kann festgehalten werden.

Insgesamt erweitern Paradata bei der Durchführung interviewervermittelter Befragungsformen den Spielraum für Analysen über Umfang und Bedingungsfaktoren eines möglichen Bias durch Nonresponse (Lavrakas, 2010). So kann mit der Schätzung von Kontakt- und Teilnahmewahrscheinlichkeiten versucht werden, auf Variationen im Antwortverhalten zu schließen (Felderer et al., 2014; Olson, 2013). Hierzu können jegliche Parainformationen, wie Unterschiede hinsichtlich der Erreichbarkeit, herangezogen werden, die sowohl für Respondenten als auch Nonrespondenten zur Verfügung stehen. In Tabelle 4.3 sind das diejenigen, die kursiv gesetzt sind.

Der Einsatz von Interviewern vergrößert die Vielfalt möglicher Paradata erheblich (Bethlehem et al., 2011, 124). Über Kontaktprotokolle, die im Anschluss an ein Interview oder eine Verweigerung ausgefüllt werden, können weitere Informationen erhoben und im Nachhinein zur Verbesserung der Datenqualität genutzt werden. Allerdings in etwas eingeschränkterer Form, da zumindest für CATI-Befragungen zwischen zweierlei Formen von Protokollen zu unterscheiden ist: dem Teilnahme- und dem Verweigerungsprotokoll.<sup>8</sup> Aus der jeweiligen Bezeichnung geht bereits hervor, dass diese Parainformationen entweder für Respondenten *oder* Nonrespondenten zur Verfügung stehen, was zumindest eine Einschätzung der Auswirkung dieser Variablen auf einen möglichen Nonresponse Error beschränkt. Nichtsdestotrotz können über die Interviewatmosphäre, angebrachte Überzeugungsargumente, die Verweildauer bei einer bestimmten Frage oder eine allgemeine Einschätzung über das Antwortverhalten (beispielsweise: Hat der Befragte sich

---

<sup>8</sup> In der englischsprachigen Literatur wird von einer „Structured Refusal Report Form (RRF)“ gesprochen (Lavrakas, 2010, 485).

Mühe bei der Beantwortung der Fragen gegeben?) auf Probleme im Feldprozess und einen etwaigen Measurement Error geschlossen werden. Denn es konnte bereits nachgewiesen werden, dass Personen, die durch einen Interviewer zu einer Teilnahme überzeugt werden müssen, weniger kognitive Mühen in die Beantwortung der Fragen investieren und der Measurement Error hierdurch entsprechend höher ausfallen kann (Engel und Köster, 2014). Couper und Kreuter (2013) untersuchen daran anknüpfend, welche Parainformationen die Zeit beeinflusst, die ein Teilnehmer benötigt, um eine Frage zu beantworten. Für CAPI-Befragungen gilt diese Beschränkung in dieser Form nicht, da Interviewer durch Ortsanwesenheit in der Lage sind, Beobachtungsprotokolle zu führen. Beobachtungen über das soziale Umfeld, das Stadtviertel, in dem eine Person lebt oder die Anzahl an Kindern in einem Haushalt sind face-to-face Befragungen vorbehalten.<sup>9</sup> Werden Interviewer in ausgewählte Haushalte zur Durchführung einer Befragung geschickt, können Informationen über Charakteristika der Nachbarschaft, der Wohngegend (Urbanisierungsgrad, Häufigkeit von Graffiti usw.), des Haushalts im Allgemeinen und der einzelnen Mitglieder eines Haushaltes dokumentiert werden (Kreuter und Olson, 2013). Diese können im Nachhinein, ähnlich wie die Kontaktinformationen in der telefonischen Befragung, für Nonresponse Analysen genutzt werden, da zumindest einige der Informationen unabhängig von einer Teilnahme zur Verfügung stehen.

Vorliegende Studien konnten belegen, dass Paradata zur Erklärung der Teilnahmewahrscheinlichkeit herangezogen werden können. Stoop et al. (2010) sind auf Basis des ESS zu dem Ergebnis gekommen, dass Personen, die von einer Nachbarschaft umgeben sind, deren Gebäude in einem generell schlechten Zustand sind, eher dazu neigen, nicht zu kooperieren und ein Umfragegesuch abzulehnen. Vergleichbare Untersuchungen sind zu ähnlichen Befunden gekommen: So konnte nachgewiesen werden, dass die subjektiv wahrgenommene Sicherheit in einem Stadtviertel negativ mit der Erreichbarkeit sowie der Kooperationsbereitschaft korreliert ist (Durrant und Steele, 2009; Olson, 2013). Wenn die Interviewer etwa über „Random Route“ in CAPI/CASI-Befragungen in die Stichprobenziehung eingebunden sind, können GPS-Koordinaten die Möglichkeit einer zusätzlichen Kontrolle bieten. Denn so kann über den mitgeführten Befragungscomputer überprüft werden, ob diese die vorgegebene Strecke eingehalten haben oder aber davon abgewichen sind. So kann ein potentiell verursachter Coverage Error verringert werden.

---

<sup>9</sup> Diese Informationen können in Befragungssituationen, in denen der Interviewer nicht vor Ort anwesend ist, lediglich erfragt und nicht beobachtet werden.

Darüber hinaus sind die Interviewer in der Lage, zentrale soziodemografische Charakteristika wie Geschlecht, Alter und Herkunft von einem Großteil der Befragten erfolgreich zu bestimmen (Bauman et al., 1992; Lavrakas, 2010). Es liegt im jeweiligen Modus begründet, dass dies im Rahmen von CAPI-Befragungen einfacher als am Telefon zu realisieren ist. Diese Informationen müssen in der Folge nicht abgefragt werden, sodass die „Last“ einer Teilnahme durch das Verkürzen der Befragungsdauer verringert werden kann.

Die durch Teilnahme- oder Verweigerungsprotokolle erhobenen Informationen können durch den subjektiven Charakter der Antworten fehlerbehaftet sein (Kreuter und Carolina, 2010; Kreuter, 2013b). Ein Bias kann verursacht werden, wenn Alter oder Herkunft falsch beurteilt wird, die benötigte Zeit und Aufmerksamkeit fehlt und so Ungenauigkeiten beim Ausfüllen auftreten oder bestimmte Vorkommnisse nicht dokumentiert werden.

Werden Mobiltelefone zur Durchführung von Befragungen eingesetzt, kann nach einer Einwilligung der Teilnehmer die Aufzeichnung von GPS-Koordinaten erfolgen. Hierdurch kann festgestellt werden, wo sich die Befragten zur Zeit der Teilnahme befinden. Dies ist für innovative Mobile Web Befragungen von Bedeutung. Denn es kann einen Unterschied ausmachen, ob die Person sich bei der Beantwortung fortbewegt und den Fragebogen auf dem morgendlichen Weg zur Arbeit ausfüllt oder sich an einem festen Punkt, beispielsweise zu Hause, aufhält (Toepoel und Lugtig, 2014).

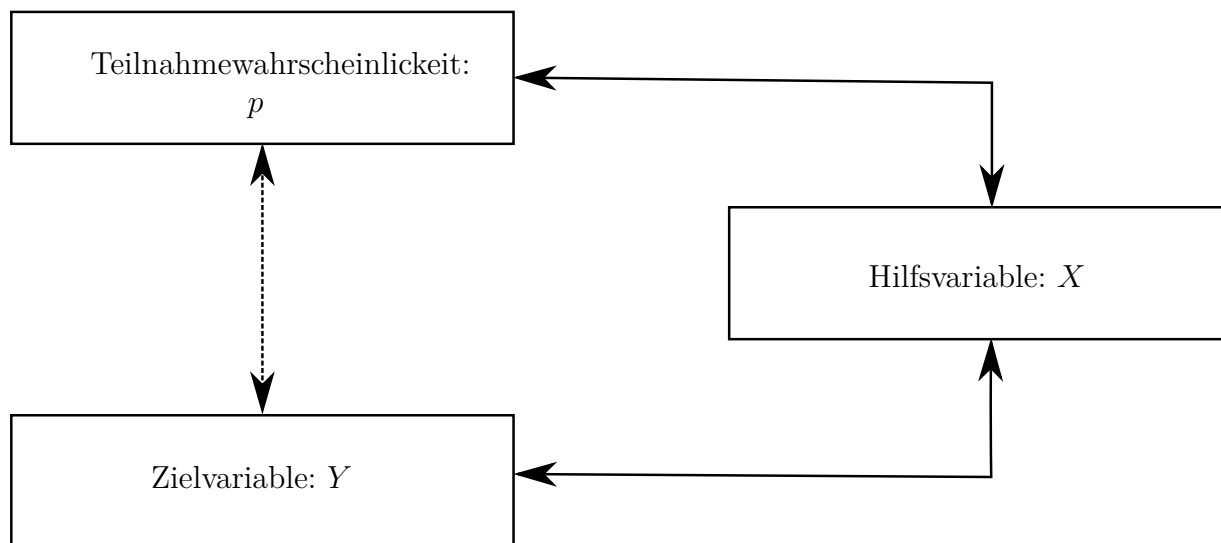
## 4.5 Statistische Konzepte zur Kompensation von Verzerrungen

Es gibt eine Reihe statistischer Verfahren, um einen Bias durch Nonresponse nachträglich zu adjustieren und so die Qualität statistischer Analysen zu erhöhen. Ob eine Gewichtung durch bekannte Randverteilungen oder Teilnahmewahrscheinlichkeiten (Propensity Scores) oder gar das Ersetzen fehlender Werte durch Imputation angewendet werden können, hängt von dem Anteil an fehlenden Werten selbst und deren Muster ab. In Kapitel 2.2 (Struktur fehlender Werte) wurden die unterschiedlichen Typen bereits beschrieben. Abbildung 4.1 greift diese Problematik erneut auf.

Eine zentrale Annahme zur Anwendung verschiedener Verfahren zur Kompensation von Verzerrungen besteht darin, dass fehlende Werte dem Mechanismus MAR folgen und somit Hilfsvariablen ( $X$ ) zur Verfügung stehen, durch die der Ausfallmechanismus möglichst



Abbildung 4.1: Zusammenhang zwischen Zielvariable, Teilnahmeentscheidung und Hintergrundvariable



Quelle: Bethlehem et al. (2011, 123)

genau erklärt werden kann. Wie bereits deutlich geworden ist, ist dies in der Praxis zumeist nicht der Fall. Es ist nicht überprüfbar, ob MCAR oder MAR vorliegt, sodass aufgrund eines Mangels an Alternativen MAR lediglich angenommen wird (Spieß, 2010; Schnell, 2012, 178). In diesem Fall besteht eine Beziehung zwischen der Hilfsvariable ( $X$ ) und der Teilnahmeentscheidung ( $p$ ). Die Hilfsvariablen können somit zur Gewichtung eines etwaigen selektiven Bias durch Nonresponse herangezogen werden (Bethlehem et al., 2011, 123). Es ist zwar ebenso möglich, Modelle für fehlende Werte zu schätzen, die dem Muster MNAR folgen, diese sind allerdings komplex und nicht zu testen (Allison, 2000, 302).

Vor diesem Hintergrund benötigt man für die Analyse von Unit-Nonresponse Bruttodatensätze, die ein möglichst breites Set an Informationen für sowohl Respondenten als auch Nonrespondenten enthalten. Wie bereits Schnell (2012, 183) hierzu folgerichtig festgestellt hat sind ohne diese „[...] kaum Nonresponse-Analysen und nur sehr begrenzt eventuelle Korrekturen möglich.“ Eine Anpassung von Umfragevariablen an bekannte Randverteilungen oder die Schätzung fehlender Werte über (multiple) Imputationsverfahren sind für die Adjustierung eines Bias durch Unit-Nonresponse nur bedingt geeignet. In der Umfrageforschung wird nicht selten auf die Anpassung zentraler Merkmale, wie etwa soziodemografischer Charakteristika, an bekannte Randverteilungen öffentlich verfügbarer Datenquellen zurückgegriffen. Unter Rückgriff auf Zensusdaten werden so Gewichte berechnet, unter deren Anwendung die Verteilung zentraler soziodemografischer Variablen

(Alter, Bildung etc.) in einem Datensatz exakt an die tatsächlichen Werte in der Bevölkerung angepasst werden.<sup>10</sup> Hierbei wird allerdings ignoriert, dass der Mechanismus, der zu Unit-Nonresponse führt, in der Regel systematisch ist und dem durch dieses Vorgehen in keiner Weise Rechnung getragen werden kann. Ein möglicher Bias würde somit unentdeckt bleiben. Imputationsverfahren schätzen unter bestimmten Vorannahmen fehlende Werte und füllen den Datensatz mit inhaltlich möglichst plausiblen Werten auf. Es gibt verschiedene Verfahren, wie die multiple Imputation, bei der auf Basis von Regressionsmodellen mehrere imputierte Datensätze geschätzt werden (in der Regel fünf) und mit der hierdurch zusätzlich entstehenden Variation eine Abschätzung über die mit der Imputation einhergehende Unsicherheit stattfinden kann (Spieß, 2010).<sup>11</sup> Diese Schätzungen können lediglich auf den beobachteten Informationen basieren und sind anzuwenden, „[...] wenn die imputierten Werte mit Sicherheit identisch mit den unbeobachteten Werten sind“ (Spieß, 2010, 124). Ob diese Annahme im Bezug auf Unit-Nonresponse und möglichen dahinterstehenden systematischen Ausfällen in Verbindung mit niedrigen Antwortraten zutrifft, muss bezweifelt werden. Werden in einem solchen Falle fehlende Werte für Komplettausfälle geschätzt, sind diese mit einer großen Unsicherheit behaftet. Denn nach Spieß (2010, 128) sind multiple Imputationsverfahren nicht für Umfragen mit niedrigen Antwortraten konzipiert (maximaler Anteil fehlender Werte für multiple Imputation: 30 Prozent), wie sie heutzutage in der Praxis häufig beobachtet werden (siehe Kapitel 2.1.2 (Nonresponse)).

Da die Wahrscheinlichkeit der Teilnahme ( $p$ ) ein latentes Konstrukt ist, dass nicht direkt beobachtet werden kann, ist es notwendig, diese über statistische Modelle zu schätzen (Kreuter und Olson, 2013; Little und Rubin, 2002; Olson, 2013). Dies betrifft die gestrichelte Beziehung zwischen  $p$  und  $Y$  in Abbildung 4.1. Nimmt man nun an, dass fehlende Werte dem Mechanismus MAR folgen, kann über Hilfsvariablen ( $X$ ) versucht werden, ein Modell für die Teilnahmewahrscheinlichkeit ( $p$ ) zu modellieren. Dieses Modell kann genutzt werden, um festzustellen, ob mit Selektionseffekten unterschiedliche Teilnahmewahrscheinlichkeiten einhergehen und diese die beobachteten Werte in ( $Y$ ) beeinflussen (einen Bias verursachen).

Daran anknüpfend wurde zur Aufdeckung und Adjustierung eines möglichen Bias durch Unit-Nonresponse die Schätzung von Propensity Scores entwickelt. Hiermit wird es möglich, die gestrichelte Beziehung zwischen  $p$  und  $Y$  unter Rückgriff auf ein Set von

---

<sup>10</sup> Für einen anwendungsorientierten Überblick zum Thema Gewichtung siehe Gabler und Ganninger (2010) oder Valliant et al. (2013).

<sup>11</sup> Für einen Überblick zum Thema Imputation siehe Allison (2001), Little und Rubin (2002) oder Spieß (2010).

Hintergrundvariablen ( $X$ ) zu schätzen.

### **Schätzen von Teilnahmewahrscheinlichkeiten und die Adjustierung über Propensity Scores**

Die Idee und das Konzept der Propensity Scores geht auf Rosenbaum und Rubin (1983) zurück. Die Grundidee besteht darin, den kausalen Effekt einer nicht experimentellen Zuteilung zu treatments (Untersuchungsgruppen) auf beobachtete Variablen zu schätzen (Brick, 2011, 881). So lassen sich Rosenbaum und Rubin (1983) von der Annahme leiten, dass die Zugehörigkeit zu einer der Gruppen und die Disparität in der Zusammensetzung dieser, durch eine nicht experimentelle Zuteilung, einen Einfluss auf die unter diesen verschiedenen Bedingungen gegebene Antwort ( $Y$ ) hat. Wenn sich Respondenten und Nonrespondenten bei der Durchführung einer Umfrage systematisch voneinander differenzieren, kann man von einer nicht-experimentellen Zuteilung sprechen.

Da lediglich Antworten in der Gruppe der Respondenten vorliegen, kann über Propensity Scores die unterschiedliche Gruppenzusammensetzung mittels Hintergrundvariablen ( $X$ ) „balanciert“ werden, die zu einer Varianz der Teilnahmewahrscheinlichkeit ( $p$ ) führt, um so möglichst einen Effekt ohne Bias schätzen zu können (Bethlehem et al., 2011). Im Kern geht es bei der Adjustierung von Unit-Nonresponse über Propensity Scores darum, den Selektionseffekt und dessen Dimensionen zu operationalisieren, an denen entlang sich die treatments Respondent/Nonrespondent differenzieren. In Abhängigkeit eines Sets von Hintergrundvariablen werden konditionale Wahrscheinlichkeiten geschätzt, zu einer der beiden Untersuchungsgruppen zu gehören (Little und Rubin, 2002). Schafft man es als Forscher auf diese Art eine „Balance“ zwischen den Gruppen herzustellen, ist es bei korrekter Spezifikation des Modells möglich, den kausalen Effekt der Zugehörigkeit unverzerrt zu schätzen (Brick, 2011). Dies erfolgt über logistische Regressionsmodelle, wobei die unterschiedlichen treatments die abhängige Variable bilden, die durch die Hintergrundvariablen möglichst genau vorhergesagt werden soll. Besitzt ein Element in der Stichprobe die Eigenschaft (Response), wird ihr der Wert 1 zugewiesen, bei Nonresponse wird die Indikatorvariable entsprechend auf 0 gesetzt. Diese geschätzten Wahrscheinlichkeiten können auf unterschiedliche Weise verwendet werden. So kann über Propensity Scores oder auch Teilnahmewahrscheinlichkeiten ( $p$ ) jeweils eine neue Variable generiert und als inverse zur Gewichtung herangezogen werden (Little und Rubin, 2002; Olson, 2013). Darüber hinaus ist es möglich einen Betrag zu schätzen, mit dem sich eine Aussage über die Höhe eines

potentiellen Bias durch Nonresponse treffen lässt. In die Schätzgleichung dieses Modells können ausschließlich Variablen einbezogen werden, die für die gesamte Bruttostichprobe zur Verfügung stehen. Werden experimentelle Designelemente in einer Umfrage platziert, etwa ein Experiment zur Wirkung von Teilnahmeanreizen, kann hiermit überprüft werden, wie sich dies auf Selektionseffekte und einen damit einhergehenden Bias auswirkt.

Daran anknüpfend hat Bethlehem (2002) die Idee der Propensity Scores aufgegriffen und im Kontext des Random Response Model (RRM) eine Vorgehensweise zur Operationalisierung von Selektionseffekten im Kontext von Unit-Nonresponse und eines potentiellen Bias in den beobachteten Variablen ( $Y$ ) vorgeschlagen. Anhand der Herleitung dieses Modells soll die Idee und das dahinterstehende Prinzip näher erläutert werden.<sup>12</sup>

Die Grundannahme des RRM besteht zunächst darin, dass jedes Element  $k$  der zugrundeliegenden Grundgesamtheit eine unbekannte Teilnahmewahrscheinlichkeit  $p_k$  besitzt. Wird Element  $k$  in eine Stichprobe aus dieser Grundgesamtheit gezogen, wird ein zufälliger Prozess ausgelöst, der sich in der Wahrscheinlichkeit  $p_k$  für eine Teilnahme sowie  $1 - p_k$  für eine Nicht-Teilnahme manifestiert. In der Folge liegen die Beobachtungseinheiten  $R_1, R_2, \dots, R_n$ , vor, wobei  $R_k = 1$  den Fall repräsentiert, wenn Element  $k$  teilnimmt und antwortet; andernfalls ist  $R_k = 0$ . Zusammenfassend kann die Wahrscheinlichkeit von Response als  $P(R_k = 1) = p_k$  und die von Nonresponse als  $P(R_k = 0) = 1 - p_k$  beschrieben werden. Es wird angenommen, dass aus dieser Population eine Stichprobe mit der Größe  $n$  ohne Zurücklegen gezogen wird, woraus sich eine Gruppe von Beobachtungseinheiten  $a_1, a_2, \dots, a_n$  ergibt. Die Teilnehmer bestehen aus den Elementen  $k$ , wofür  $a_k = 1$  sowie  $R_k = 1$  gilt, sodass die Summe der beobachteten Fälle Gleichung 4.1 sowie die Summe der nicht beobachteten Fälle Gleichung 4.2 entspricht:

$$n_R = \sum_{k=1}^N a_k R_k, \quad (4.1)$$

$$n_{NR} = \sum_{k=1}^N a_k (1 - R_k), \quad (4.2)$$

wobei sich die Bruttostichprobe  $n$  aus der Summe aus  $n_R$  und  $n_{NR}$  zusammensetzt. Beobachtete Werte liegen lediglich vor, wenn sich eine Person beteiligt hat ( $n_r$ ). Der

---

<sup>12</sup> Die mathematischen Gleichungen sind aus Bethlehem et al. (2011, 43 f.) übernommen, woran sich ebenso die textliche Herleitung orientiert.

Durchschnitt der beobachteten Werte lässt sich folglich bezeichnen mit

$$\bar{y}_R = \frac{1}{n_R} \sum_{k=1}^N a_k R_k Y_k. \quad (4.3)$$

Der Erwartungswert des Durchschnitts der Antworten der Teilnehmer ist nahezu gleichzusetzen mit

$$E(\bar{Y}_R) \approx \tilde{Y}, \quad (4.4)$$

wobei sich die Schätzgleichung für den Durchschnitt in der Population aus

$$\tilde{Y} = \frac{1}{N} \sum_{k=1}^N \frac{p_k Y_k}{\bar{p}}, \quad (4.5)$$

ergibt und

$$\bar{p} = \frac{1}{N} \sum_{k=1}^N p_k, \quad (4.6)$$

die durchschnittliche Teilnahmewahrscheinlichkeit, sozusagen die durchschnittliche „Response Propensity“, aller Elemente der Stichprobe darstellt. Gleichung 4.5 verdeutlicht, dass der erwartete Wert des Durchschnitts einer Antwort durch Selektionseffekte nicht dem in der Grundgesamtheit entspricht und somit verzerrt ist. Der Bias eines Mittelwerts in den beobachteten Werten lässt sich mathematisch beschreiben als

$$Bias(\bar{y}_R) = \tilde{Y} - \bar{Y} = \frac{R_{pY} S_p S_Y}{\bar{p}}, \quad (4.7)$$

wobei  $R_{pY}$  den Zusammenhang zwischen Zielvariable ( $Y$ ) (gestrichelte Linie in Abbildung 4.1) und der Teilnahmewahrscheinlichkeit ( $p$ ) repräsentiert.  $S_Y$  und  $S_p$  sind die jeweiligen Standardabweichungen von  $p$  und  $Y$ . Die Teilnahmewahrscheinlichkeit ( $p$ ) ergibt sich, wie schon erwähnt, aus der Schätzung eines logistischen Regressionsmodells mit

$$Pr(p = 1 | X'_k) = \log \left( \frac{p(X_k)}{1 - p(X_k)} \right) = \text{logit}(p(X_k)) = X'_k \beta. \quad (4.8)$$

Hierfür wird über Gleichung 4.8 die Wahrscheinlichkeit für  $p = 1$  (Response) in Abhängigkeit eines Vektors von beobachteten Hintergrundvariablen ( $X'$ ) für *jedes* Element  $k$  in der Stichprobe geschätzt. Die sich daraus ergebende Variation in Form der Teilnahmewahrscheinlichkeit ( $p$ ) fließt schließlich in die Schätzung für den Betrag des Bias

durch selektiven Nonresponse in Gleichung 4.7 des RRM ein. Der Betrag wird dabei umso größer, je stärker der Zusammenhang zwischen der Teilnahmewahrscheinlichkeit ( $p$ ) und der beobachteten Zielvariablen ( $Y$ ) ist (Ausdruck  $R_{pY}$  in Gleichung 4.7). Liegt daran anknüpfend die Ursache für die Variation in der Zielvariable ( $Y$ ) in  $p$  begründet, kann von einem „nonignorable nonresponse“ (Olson, 2013, 144) oder auch MNAR gesprochen werden. Liegen sowohl  $p$  als auch  $Y$  in den Hintergrundvariablen ( $X$ ) begründet und kann deren Variation hierdurch möglichst genau aufgeklärt werden, liegt „ignorable nonresponse“ (ebd., 144) beziehungsweise MAR vor. Der Begriff „ignorable nonresponse“ ist dabei etwas irreführend, da auch hier Verzerrungen auftreten können, die es aufzudecken und zu adjustierten gilt.

Das Potential für die Operationalisierung und Adjustierung eines selektiven Bias durch Nonresponse ist maßgeblich von den zur Verfügung stehenden Hintergrundvariablen abhängig (Bethlehem et al., 2011; Kreuter und Olson, 2011; Little und Rubin, 2002). Eine ideale Hintergrundvariable weist nach Olson (2013, 144 f.) insgesamt vier Eigenschaften auf:

1. Gültige Werte müssen für Respondenten und Nonrespondenten vorliegen.
2. Es sollte kein Item-Nonresponse sowie keine Verzerrungen, wie etwa durch einen Measurement Error, vorliegen.
3. Die Variable sollte in einem starken Zusammenhang ( $R^2 \geq .48$ ) mit den interessierenden Zielvariablen ( $Y$ ) stehen.
4. Um einen Bias zu reduzieren, sollte die Variable letztlich ein ebenso starker Prädiktor der Teilnahmewahrscheinlichkeit ( $p$ ) sein (siehe zu den Auswahlkriterien für Hintergrundvariablen auch Kreuter und Olson, 2011, 2013; Little und Vartivarian, 2005).

In der Praxis steht der Forscher einer komplizierten Situation gegenüber: Häufig ist es sehr schwierig, Hintergrundvariablen auszumachen, die diesen Eigenschaften entsprechen (Kreuter und Olson, 2011). Denn in der Regel steht kein umfangreiches Set an Variablen über den Bruttodatensatz zur Verfügung, das sich zur Adjustierung eines Bias durch selektiven Nonresponse heranziehen lässt. Selbst wenn diese vorhanden sind, ist die Gewichtung eines Bias, wie Bethlehem et al. (2011, 287) bereits festgestellt haben, ein mitunter kompliziertes Unterfangen: „In practice, it is very difficult to assess the possible

negative effect of nonresponse“. Verfügbare Variablen über den Auswahlrahmen sind ebenso bei Melderegisterstichproben und der Erhebung möglicher Paradata relativ beschränkt. Wobei bei dieser Art der Stichprobenziehung mehr Informationen für die Bruttostichprobe zur Verfügung stehen als dies etwa bei CATI-Befragungen der Fall ist (siehe Kapitel Melderegister (4.3)). Außerdem können zusätzlich Parainformationen herangezogen werden, die für sowohl Respondenten als auch Nonrespondenten vorliegen (siehe Tabelle 4.3 in Kapitel Paradata (4.4)). Die Auswahl an Hintergrundvariablen bleibt trotzdem beschränkt. Deswegen ist es in der Regel nicht möglich, eine Auswahl an Variablen zu treffen, auf die die von Olson (2013) formulierten Eigenschaften am ehesten zutreffen. Trivialerweise wird aus diesem Grund zumeist auf all die Variablen zurück gegriffen, die zur Verfügung stehen.

Trotzdem sollte in der Praxis eine Beurteilung der Qualität verfügbarer Hintergrundvariablen und dem damit einhergehenden Gewichtungspotential stattfinden. Denn Little und Vartivarian (2005) haben bereits betont, dass speziell die Eigenschaften drei und vier nach Olson (2013) von Relevanz sind. Besteht kein Zusammenhang zwischen Hintergrund- und beobachteter Variable sowie der Teilnahmewahrscheinlichkeit oder lässt sich dieser lediglich zwischen zwei der Variablen beobachten, kann ein potentieller Bias zwar operationalisiert und bei korrekter Modellspezifikation auch verringert werden. Hiermit geht allerdings die Gefahr einher, dass die Varianz der Schätzer in Form von Mittelwerten der beobachteten Umfragevariablen vergrößert wird und Schätzungen mit einer größeren Unsicherheit belastet werden können (Little und Rubin, 2002; Little und Vartivarian, 2005).

Simulationsstudien haben sich unter anderem der Frage angenommen, wie sich die Ausgestaltung der Zusammenhänge zwischen Teilnahmewahrscheinlichkeit  $p$ , Umfragevariablen  $Y$  und Hintergrundvariablen  $X$  im Kontext unterschiedlicher erzielter Antwortraten auf das Gewichtungspotential und die Varianz der Schätzer auswirkt. Kreuter und Olson (2011) haben sich mit den bis dahin unbekanntem Folgen einer Variation in den Zusammenhängen dieser Variablen in Bezug zu unterschiedlichen Antwortraten beschäftigt. Ausgangspunkt ist dabei, dass in der Praxis häufig auf Basis der Differenz zwischen gewichteten und nicht gewichteten Schätzungen eine Aussage über die potentielle Reduzierung eines Bias getroffen wird, was nach Kreuter und Olson (2011) nicht notwendigerweise informativ ist. Vielmehr muss zusätzlich die Beziehung zwischen den zur Gewichtung verwendeten Hintergrundvariablen und den Umfragevariablen sowie der Teilnahmewahrscheinlichkeit untersucht werden. Denn als zentrales Ergebnis hat sich gezeigt, dass sich der Bias sowie der Standardfehler adjustierter Mittelwerte der beobachteten Umfragevariablen in

Abhängigkeit der Beziehung zwischen  $p$ ,  $Y$  und  $X$  substantiell unterscheiden können.

Sind die Hintergrundvariablen  $X$  ein starker positiver Prädiktor sowohl für die Umfragevariable  $Y$  als auch für die Teilnahmewahrscheinlichkeit  $p$ , können Standardfehler und Bias mittlerer Werte im Rahmen eines korrekt spezifizierten Gewichtungsmodells reduziert werden (Kreuter und Olson, 2011, 17). Korreliert  $X$  lediglich schwach positiv mit  $Y$ , führt dies in der Folge zu höheren Standardfehlern sowie einer größeren Unsicherheit in den Signifikanztests.

Werden Hintergrundvariablen zur Adjustierung herangezogen, die sowohl die Kontaktwahrscheinlichkeit als auch die Kooperationswahrscheinlichkeit in unterschiedlicher Weise konditional beeinflussen, konnten Kreuter und Olson (2011) in ihrer Simulationsstudie keinen Bias in den beobachteten Umfragevariablen feststellen. Denkbar wäre, dass auf der einen Seite ein negativer Zusammenhang zwischen Alter  $X$  und der Teilnahmewahrscheinlichkeit  $p$  beobachtet werden kann, sowie, dass ältere Personen ein höheres Einkommen  $Y$  erzielen. Auf der anderen Seite ist das Bildungsniveau  $X$ , als weitere Hintergrundvariable, ein starker positiver Prädiktor sowohl für die Wahrscheinlichkeit der Teilnahme als auch für die Höhe des Einkommens (Kreuter und Olson, 2011, 17). Ältere Menschen sind einfacher zu erreichen, häufig aufgrund unterschiedlicher Faktoren aber nicht zu einer Teilnahme bereit (Engel und Schmidt, 2014). Da nun das Alter in *negativen* und das Bildungsniveau in *positiven* Zusammenhang mit der Teilnahmewahrscheinlichkeit steht, und beide Prädiktoren darüber hinaus einen positiven Einfluss auf die Umfragevariable Einkommen ( $Y$ ) ausüben, können diese konkurrierenden Einflüsse zu einer verzerrungsfreien Schätzung nicht gewichteter Mittelwerte führen (Kreuter und Olson, 2011, 17).<sup>13</sup> Als Konklusion kann festgehalten werden, dass jegliche Zusammenhänge zwischen den relevanten Hintergrundvariablen (sofern vorhanden) sowohl mit der Teilnahmewahrscheinlichkeit als auch den Umfragevariablen zur Gewichtung eines selektiven Nonresponse überprüft werden sollten. Hierüber sind erste Rückschlüsse möglich, inwiefern die Variablen zur Gewichtung herangezogen werden können und wie etwaige adjustierte Ergebnisse zu interpretieren sind. Werden Propensity Scores als Inverse zur Gewichtung genutzt, sollte diese Prüfung ebenfalls stattfinden, da sich gezeigt hat, dass die Varianz vergrößert und der Bias nicht verkleinert werden kann, wenn die Hintergrundvariablen nicht mit den Umfragevariablen korreliert sind (Alanya et al., 2015). Daran anknüpfend haben Alanya et al. (2015) in einer

---

<sup>13</sup> Hierbei ist die starke Annahme Voraussetzung, dass keine weiteren Variablen die Teilnahmewahrscheinlichkeit beeinflussen. In der Praxis ist diese wahrscheinlich in den wenigsten Fällen haltbar.



neueren Simulationsstudie die Gewichtung durch Propensity Scores mit dem Schätzen fehlender Werte durch multiple Imputation unter verschiedenen Bedingungen miteinander verglichen. Im Allgemeinen hat sich gezeigt, dass das Potential eines Bias durch Nonresponse generell niedrig ist, wenn die Hintergrundvariablen  $X$  ein nur schwacher Prädiktor für sowohl die Teilnahmewahrscheinlichkeit  $p$  als auch die Umfragevariablen  $Y$  sind. Mit Anwendung multipler Imputation im Falle von Unit-Nonresponse geht im Vergleich zur Schätzung einzelner fehlender Werte durch Item-Nonresponse eine größere Unsicherheit in den Schätzungen einher. Deswegen haben Alanya et al. (2015) nicht wie üblicherweise lediglich fünf imputierte Datensätze geschätzt, sondern 100. Ob nun das inverse Gewichten durch Propensity Scores oder die multiple Imputation besser zur Adjustierung von selektiven Nonresponse geeignet ist, konnte nicht hinreichend beantwortet werden. Es ist besonders von der Antwortrate, der Modellspezifikation sowie den zur Anwendung kommenden Hintergrundvariablen und deren Beziehung zu Teilnahmewahrscheinlichkeit und Umfragevariablen abhängig, welches Verfahren validere Ergebnisse liefert.

Im folgenden Kapitel 5 werden die in den vorherigen Abschnitten vorgestellten Methoden und deren Potential zur Reduzierung eines möglichen Bias durch Nonresponse und Noncoverage in einem empirisch zu testenden Gesamtmodell vereint. Dieses dient als Grundlage für die abzuleitenden Forschungshypothesen.

## Kapitel 5

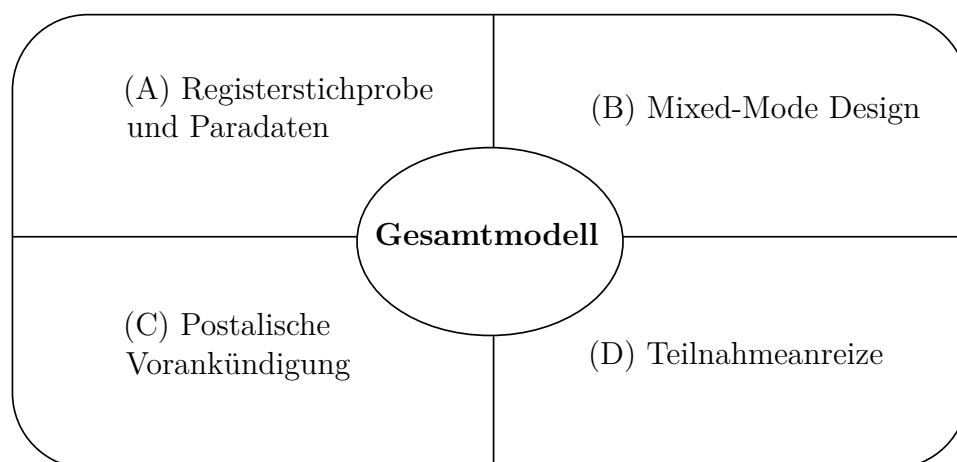
# Ein integriertes Gesamtmodell

Zur Reduzierung von Nonresponse und Noncoverage sowie zur Operationalisierung eines möglichen Bias wurden die in Kapitel 4 (Ansätze zur Fehlerreduktion) beschriebenen Designelemente und die daran anknüpfende Rekrutierung von Teilnehmern in ein empirisch zu testendes Gesamtmodell integriert. Abbildung 5.1 gibt eine entsprechende Übersicht.

Das Rekrutierungsmodell besteht aus vier Bestandteilen: (A) Registerstichprobe und Paradata, (B) Mixed-Mode Design, (C) postalische Vorankündigung und (D) dem Einsatz von Teilnahmeanreizen. Die Vor- und Nachteile sowie eine mögliche fehlerreduzierende Wirkung wurden bereits diskutiert. In diesem Kapitel geht es darum, die aus diesem Gesamtmodell abzuleitenden und empirisch zu testenden Forschungshypothesen herzuleiten und zu beschreiben. Die folgenden Ausführungen orientieren sich an der Chronologie des in Abbildung 5.1 dargestellten Gesamtmodells.

Darüber hinaus muss zwischen Brutto- und Nettostichprobe differenziert werden. Denn

Abbildung 5.1: Ein integriertes Gesamtmodell zu Rekrutierung von Teilnehmern für Panelbefragungen



die Bruttostichprobe, die alle gezogenen Personen repräsentiert, teilt sich bei der Durchführung einer Umfrage naturgemäß in zwei Gruppen auf: Respondenten und Nonrespondenten. Für letztere stehen in der Folge deutlich weniger Informationen zur Verfügung, als für all diejenigen, die sich an der Befragung beteiligt haben (Nettostichprobe).

Daran anknüpfend gehen mit der Rekrutierung von Teilnehmern für Panelbefragungen über das Rekrutierungsinterview und der Unterscheidung zwischen Brutto- und Nettostichprobe weitere Selektionsstufen einher, entlang derer sich Respondenten von Nonrespondenten differenzieren. Dies liegt im notwendigen Vorgehen begründet, da die Teilnehmer im Anschluss an ein erstes Rekrutierungsinterview, das zunächst die gleichen methodischen Eigenschaften hat wie eine Querschnittbefragung, um Einverständnis gebeten werden müssen, sich in regelmäßigen Abständen befragen zu lassen. Nur bei Einverständnis ist es nach dem Bundesdatenschutzgesetz §14 erlaubt, personenbezogene Daten zum Zweck der erneuten Kontaktierung über einen längeren Zeitraum zu speichern (BMJV, 2015, 15). Ferner hängt es vom jeweiligen Rekrutierungsdesign selbst ab, inwiefern mögliche verzerrende Effekte durch Nonresponse oder Noncoverage zu erwarten und zu operationalisieren sind. In Kapitel 2 wurde dies bereits für die relevanten Modi, CATI und CAWI, in Bezug zum Rekrutierungsinterview diskutiert. Eine Unterscheidung muss ferner hinsichtlich der Linie Teilnahme am Rekrutierungsinterview, Wiederbefragungsbereitschaft sowie Teilnahme am Panel getroffen werden. Denn einerseits wirken auf diesen einzelnen Selektionsstufen unterschiedliche Prädiktoren (Bartsch, 2012; Vehre, 2011; Watson und Wooden, 2009). Andererseits variiert der Umfang an zur Verfügung stehenden Informationen über die einzelnen Stufen hinweg teilweise deutlich (Lynn, 2009). Der Fokus der vorliegenden Arbeit liegt auf der initialen Teilnahme an der ersten Befragung - dem Rekrutierungsinterview. Die nachfolgenden Stufen werden in den Analysen nicht berücksichtigt. Dies liegt darin begründet, dass die Wiederbefragungsbereitschaft sowie die tatsächliche Teilnahme am PPSM Access Panel bereits umfangreich untersucht wurde (Bartsch, 2012; Engel et al., 2012; Engel, 2013; Vehre, 2011). Darüber hinaus ist die Forschungslage zu Attrition beziehungsweise Panelmortalität breit aufgestellt, denn zur Nonresponsekontrolle durch die chronologisch zuvor stattgefundenen Befragungswellen stehen deutlich mehr Informationen zur Verfügung (Lepkowski und Couper, 2002; Lynn, 2009). Am schwierigsten ist es die Teilnahmewahrscheinlichkeit für das Rekrutierungsinterviews zu schätzen. Diese Problematik besteht auf den folgenden Selektionsstufen nur bedingt (Watson und Wooden, 2009). Nehmen Personen an einer ersten Befragung teil und selektieren sich in der Folge

Tabelle 5.1: Kontaktierungswege der zugrundeliegenden Erhebungen sowie zu analysierende Zielgrößen

Erhebungszeitpunkt	E-Mail	Post	Telefon	Zielgrößen
Nachrekrutierung 2012		x	x	<ul style="list-style-type: none"> <li>• Erreichbarkeit (<math>p_e</math>),</li> <li>• Kooperationswahrscheinlichkeit (<math>p_k</math>)</li> <li>• <math>p_e + p_k =</math> Teilnahmewahrscheinlichkeit (<math>p_{tn}</math>)</li> </ul>
Panelwelle 2014	x		x	<ul style="list-style-type: none"> <li>• Wahl des Befragungsmodus (<math>\delta_{modus}</math>)</li> <li>• Smartphonennutzung (<math>\delta_{smartp}</math>),</li> <li>• Teilnahme an mobiler Online-/ Webbefragung (<math>\delta_{mw}</math>),</li> <li>• subjektive Wahrnehmung der Anreizhöhe</li> </ul>

entlang der Linie wiederbefragungsbereit oder nicht, liegt durch die zuvor gestellten Fragen ein umfangreiches Set an Informationen vor, dass zum Zweck der Nonresponsekontrolle herangezogen werden kann. Ebenso verhält es sich, wenn das eigentliche Panel mit seinen sukzessiven Erhebungswellen startet.

Daran anknüpfend ergeben sich für die vorliegende Arbeit in Abhängigkeit des jeweiligen Untersuchungsdesigns ausgewählte Zielgrößen (abhängige Variablen). Tabelle 5.1 gibt eine entsprechende Übersicht.

Soll eine statistische Modellierung der Kooperationswahrscheinlichkeit ( $p_k$ ) und die Auswirkung der Variation dieser auf die Beantwortung von Fragen (Nonresponse Bias) stattfinden, muss die Erreichbarkeit der Zielpersonen ( $p_e$ ) mit einbezogen werden. Denn Merkmale wie die Anzahl an Kontaktversuchen oder das Alter der Zielperson wirken vermittelt über die Erreichbarkeit ( $p_e$ ) auf die Kooperations- ( $p_k$ ) und letztlich Wahrscheinlichkeit der Teilnahme insgesamt ( $p_{tn}$ ). Da der Auswahlrahmen aus dem zentralen Einwohnermelderegister besteht und alle Personen in der Bruttostichprobe postalisch kontaktiert worden sind, trifft dies im Zusammenhang mit den zugrundeliegenden Daten aus den Jahren 2009 und 2012 lediglich auf die Gruppe der Personen zu, die zusätzlich angerufen worden sind. Denn im chronologisch zuerst stattgefundenen Kontaktierungsschritt wurden alle Personen in der Bruttostichprobe postalisch kontaktiert. Hier gab es aufgrund nicht mehr aktueller Adressinformationen ebenso Ausfälle, die von ihrer Anzahl her allerdings gering sind und somit nicht gesondert geschätzt und als Prädiktor einbezogen werden. Im Rahmen der Erhebungswelle aus dem Jahr 2014 ist diese Unterscheidung nicht relevant, da lediglich inhaltliche Variablen, wie die Smartphonennutzung, die Bereitschaft zur Teilnahme an einer mobilen Onlineumfrage oder die subjektive Wahrnehmung der

Anreizhöhe, zur Analyse herangezogen werden.<sup>1</sup> Über die Teilnahmewahrscheinlichkeit hinaus soll im Rahmen des Mixed-Mode Desings der Nachrekrutierung aus dem Jahr 2012 untersucht werden, inwiefern sich die Präferenzen für unterschiedliche Befragungsmodi ( $\delta_{modus}$ ) verteilen. Anschließend werden Hypothesen hergeleitet, die sich mit der subjektiven Wahrnehmung seitens der Befragten bezüglich der Höhe eines Anreizes beschäftigen, die erwartbar selektive Nutzungswahrscheinlichkeit von Smartphones mit einbeziehen ( $\delta_{smartp}$ ) sowie die Bereitschaft abbilden, sich über ein mobiles Endgerät an Umfragen zu beteiligen ( $\delta_{mw}$ ).

### Hypothesen zum Rekrutierungsinterview

Als „globale“, forschungsleitende Hypothese wird die Annahme formuliert, dass die Teilnahmewahrscheinlichkeit durch selektiven Nonresponse mit den gegebenen Antworten in Umfragen korreliert und ein Bias entsteht. In der empirischen Analyse wird der Ausfallprozess bezüglich der in Bremen stattgefundenen Nachrekrutierung in den Fokus gerückt und statistisch modelliert. Das Formulieren von gerichteten Hypothesen wird stellenweise durch die Tatsache verhindert, dass erst einmal offen bleiben muss, wie genau die in die Schätzgleichung eingehenden Variablen schließlich mit den gegebenen Antworten zusammenhängen. Diese Problematik wurde bereits in Kapitel 4.5 (Statistische Konzepte zur Kompensation von Verzerrungen durch Unit-Nonresponse) diskutiert. Somit hat dieser Teil der empirischen Analyse notwendigerweise den Charakter einer explorativen Untersuchung.

Es kann zwar aufgrund der Forschungslage angenommen werden, dass selektiver Nonresponse die Beantwortung von Fragen beeinflusst und ein Bias entstehen kann. Einerseits werden von Studie zu Studie unterschiedliche inhaltliche Fragen gestellt. Andererseits ist es speziell vom Umfang und der Qualität der zur Verfügung stehenden Hintergrundvariablen abhängig, ob ein Bias in den gegebenen Antworten überhaupt operationalisiert werden kann. Daran anknüpfend ist die Aufstellung einer solchen forschungsleitenden, ungerichteten Annahme möglich.

Die Formulierung von gerichteten Hypothesen ist somit ausschließlich für den Zusammenhang zwischen der Teilnahmewahrscheinlichkeit  $p$  und den Hilfsvariablen  $X$  möglich (siehe Abbildung 4.1).<sup>2</sup>

---

<sup>1</sup> Eine differenzierte Beschreibung des jeweils zugrundeliegenden Untersuchungsdesigns erfolgt in Kapitel 6 (Datengrundlage und Operationalisierung).

<sup>2</sup> Die Operationalisierung der in die Analysen eingehenden Variablen wird in Kapitel 6 (Daten und Operationalisierung) beschrieben.

Der zugrundeliegende Auswahlrahmen sowie relevante Paradata, wie etwa die Anzahl an Kontaktversuchen, haben den Vorteil, dass diese für alle Personen in der Bruttostichprobe zur Verfügung stehen und somit für Nonresponseanalysen herangezogen werden können. Bei der Durchführung einer CATI-Umfrage muss in einem ersten Schritt erfolgreich Kontakt zur Zielperson hergestellt werden, bevor eine Befragung stattfinden kann. Deswegen muss eine Differenzierung der unterschiedlichen Zielgrößen Erreichbarkeit und Kooperationsbereitschaft erfolgen. Denn einige Prädiktoren wirken vermittelt über die Wahrscheinlichkeit der Erreichbarkeit ( $p_e$ ) auf die der Kooperationswahrscheinlichkeit ( $p_k$ ). Die Summe dieser beiden Faktoren ergibt die letztliche Teilnahmewahrscheinlichkeit. Tabelle 5.2 gibt eine Übersicht über die Annahmen des Zusammenhangs der unterschiedlichen Zielgrößen mit den zur Verfügung stehenden Prädiktoren.

Soziodemografische Determinanten wie Alter, Geschlecht oder Bildung haben sich in der Vergangenheit für die Teilnahmewahrscheinlichkeit an Befragungen insgesamt als nicht erklärungsrelevant erwiesen (Engel und Schmidt, 2014; Stoop et al., 2010). Ein allgemeiner Effekt des Alters auf die Wahrscheinlichkeit der Teilnahme am Rekrutierungsinterview kann somit nicht vermutet werden (Haunberger, 2011, 86). Vielmehr muss die Frage gestellt werden, über welche Drittvariablen ein möglicher Alterseffekt transportiert wird. In der Umfrageforschung werden Ergebnisse für ältere Menschen berichtet, die darauf hindeuten, dass diese aus unterschiedlichen Gründen Umfragen eher skeptisch gegenüber stehen und sich deswegen in geringerem Maße an solchen beteiligen (Hypothese  $H_{A5}$ ). Als Grund dafür werden Angst vor Kriminalität oder gesundheitliche Probleme genannt, die eine Teilnahme verhindern (Couper und Groves, 1996; Goyder, 1987; Haunberger, 2011). Stoop et al. (2010) weisen ebenso darauf hin, dass ältere Menschen über schlechtere Bildungsabschlüsse verfügen oder Jüngere einen Lebensstil verfolgen, der als eine Ursache für geringe Erreichbarkeit (Zielgröße:  $p_e$ ) angesehen werden kann (Hypothese  $H_{A1}$ ). Der sogenannte Bildungsbias, der meint, dass insbesondere Personen mit höheren Bildungsabschlüssen an Umfragen teilnehmen, gilt als gesicherte empirische Evidenz (Adua und Shap, 2010; Couper und Groves, 1996; Engel und Schmidt, 2014). Personen im jüngeren und mittleren Alter sind eher seltener zu Hause anzutreffen als Rentner im höheren Alter (Hypothese  $H_{A1}$ ). Hinzukommen, vermittelt über das Alter, häufigere Abwesenheitszeiten durch Erwerbstätigkeit, was sich in Ein-Personen-Haushalten bemerkbar macht. Denn lebt eine Person in einem Haushalt, in dem mehr als zwei Personen wohnen, steigt die Wahrscheinlichkeit eines Kontaktes (Zielgröße:  $p_e$ ) (Engel et al., 2012; Lynn, 2009). Wenn die Zielperson nicht

anwesend sein sollte, kann die Kontaktperson, wie etwa der Partner oder die Kinder, bei Anwesenheit einen Termin für einen erneuten Kontakt vereinbaren, zu der die Zielperson besser erreichbar ist (Hypothese  $H_{A4}$ ). Personen in Singlehaushalten sind in der Folge schlechter zu erreichen und haben somit, vermittelt über die Kontaktwahrscheinlichkeit, eine niedrigere Teilnahmewahrscheinlichkeit.

Die Altersvariable wird speziell in Abhängigkeit von dem jeweiligen Umfragemodus in Bezug zu Noncoverage relevant. Im Rahmen von Kapitel 2.1.1 (Noncoverage) ist bereits deutlich geworden, dass auch heute noch speziell die Altersgruppe der ab 60 Jährigen bei Befragungen über das Internet unterrepräsentiert sind (ARD/ZDF-Onlinestudie, 2015; Baur und Florian, 2009; TNS Infratest, 2014). Daran anknüpfend und in Bezug auf das in der Befragung aus dem Jahr 2012 angewandte Mixed-Mode Design kann angenommen werden, dass sich dieser Kreis an Personen eher für die CATI- im Vergleich zur CAWI-Befragung entscheidet (Hypothese  $H_{A7}$ ). Das Alter wirkt sich demnach vermittelt über die Häufigkeit der Internetnutzung auf die Präferenz aus, sich für einen bestimmten Modus zu entscheiden.

Unterschiede zwischen den Geschlechtern gibt es in Bezug zur Internetnutzung und daran anknüpfend zu der Wahrscheinlichkeit an einer Umfrage hierüber teilzunehmen. Wie bereits in Kapitel 2.1.1 (Noncoverage) beschrieben und aus Tabelle 2.1 (Seite 12) ersichtlich, nutzen Männer das Internet häufiger als Frauen (TNS Infratest, 2014). 2014 lag der Unterschied laut TNS Infratest (2014) bei zehn und 2015 liegt dieser nach der ARD/ZDF-Onlinestudie (2015) bei sieben Prozentpunkten. Demnach ist vermittelt über die Internetnutzung anzunehmen, dass sich Männer im Kontext der angebotenen Teilnahmemöglichkeiten eher für den Online- im Gegensatz zum Telefonmodus entscheiden (Hypothese  $H_{A8}$ ). Ein Effekt auf die Teilnahmewahrscheinlichkeit ist aufgrund von Geschlechterunterschieden nicht zu vermuten (Hypothese  $H_{A6}$ ).

Die durch Softwareunterstützung eruierte Anzahl an Kontaktversuchen hat sich für die Kontaktwahrscheinlichkeit ( $p_k$ ) in der Vergangenheit ebenfalls als erklärungsrelevant erwiesen.<sup>3</sup> Denn die Anzahl an Kontaktversuchen sollte auf ein notwendiges Maß reduziert werden, um Personen nicht zu verärgern. Es konnte nachgewiesen werden, dass die Kontaktwahrscheinlichkeit im Bereich von eins bis fünf Versuchen am höchsten ist (Engel, 2013) (Hypothese  $H_{A2}$ ). Dabei konnte für das Rekrutierungsinterview des PPSM Access

---

<sup>3</sup> Das Methodenzentrum der Universität Bremen verwendet die CATI-Software VOXCO. Für ausführliche Informationen siehe [www.voxco.com](http://www.voxco.com).

Tabelle 5.2: Forschungshypothesen zu Designaspekt A

<b>Zielgröße: Erreichbarkeit (<math>p_e</math>)</b>	
$H_{A1}$ :	Je älter die Zielperson, desto höher ist die Wahrscheinlichkeit eines erfolgreichen telefonischen Kontaktes.
$H_{A2}$ :	Je mehr Kontaktversuche benötigt werden, um eine Person zu erreichen, desto niedriger ist die Wahrscheinlichkeit eines erfolgreichen telefonischen Kontaktes.
$H_{A3}$ :	Ist ein Anrufbeantworter in den Kontaktverlauf eingebunden, ist die Wahrscheinlichkeit eines erfolgreichen telefonischen Kontaktes niedriger im Vergleich dazu, wenn kein Anrufbeantworter eingebunden ist.
<b>Zielgröße: Kooperationswahrscheinlichkeit (<math>p_k</math>)</b>	
$H_{A4}$ :	Wird ein Termin mit der Ziel- oder Kontaktperson vereinbart, erhöht dies die Wahrscheinlichkeit für eine Teilnahme am Rekrutierungsinterview.
$H_{A5}$ :	Je älter die Zielperson ist, desto niedriger ist die Wahrscheinlichkeit für eine Teilnahme am Rekrutierungsinterview.
$H_{A6}$ :	Das Geschlecht hat keinen Einfluss auf die Wahrscheinlichkeit der Teilnahme am Rekrutierungsinterview.
<b>Zielgröße: Wahl der Befragungsmodi (<math>\delta_{modus}</math>)</b>	
$H_{A7}$ :	Je höher das Alter einer Person, desto eher entscheiden sich die Teilnehmer für die CATI im Vergleich zur CAWI-Befragung.
$H_{A8}$ :	Männer beteiligen sich im Vergleich zu Frauen eher an der CAWI-Befragung.

Panels nachgewiesen werden, dass die Teilnahmewahrscheinlichkeit mit zunehmenden Kontaktversuchen geringer wird (Engel, 2013, 26).

Wie in Kapitel 2 diskutiert, haben technische Innovationen, wie das „Call Screening“ und Anrufbeantworter dazu geführt, dass die Erreichbarkeit abgenommen hat (Aust und Schröder, 2009; Brick, 2013; Link et al., 2008; Tourangeau, 2004) (Hypothese  $H_{A3}$ ). Denn erscheint eine unbekannte Rufnummer auf dem Display eines Telefons, kann der Hörer entweder gar nicht erst abgehoben werden oder eine Person lässt den Anrufbeantworter in der Leitung, sodass ein direkter Kontakt nicht zustande kommen kann.

Im noch zu beschreibenden initialen Rekrutierungsdesign des PPSM Access Panels haben die zugrundeliegende Stichprobenziehung und die damit einhergehenden Kontaktinformationen eine kalte telefonische Kontaktierung zugelassen. Durch die Zusammenführung von Stichproben- und Telefonregisterdatenbank konnte im Rahmen der Nachrekrutierung von 2012 zu einigen Adressen erfolgreich eine Rufnummer zugeordnet werden. Besteht daran anknüpfend die Möglichkeit aufgrund des Vorhandenseins notwendiger Kontaktinformationen einen Telefonanruf im Vorhinein postalisch anzukündigen, konnten zahlreiche Studien einen positiven Effekt auf die Kooperations- und Teilnahmewahrscheinlichkeit



Tabelle 5.3: Forschungshypothesen zu Designaspekt B und C

<b>Zielgröße: Kooperationswahrscheinlichkeit (<math>p_k</math>)</b>	
$H_{B1}$ :	Verfügt eine Person über einen Telefonbucheintrag, erhöht dies die Wahrscheinlichkeit der Teilnahme am Rekrutierungsinterview.
$H_{B2}$ :	Personen, die über einen Telefonbucheintrag verfügen, unterscheiden sich hinsichtlich soziodemografischer Merkmale systematisch von Personen, die nicht darin eingetragen sind.
$H_{B3}$ :	Ist ein Telefonanruf in den Kontaktverlauf eingebunden, erhöht dies die Wahrscheinlichkeit der Teilnahme am Rekrutierungsinterview.
$H_{C1}$ :	Die postalische Vorankündigung der telefonischen Kontaktaufnahme wirkt sich positiv auf die Wahrscheinlichkeit der Teilnahme am Rekrutierungsinterview aus.
<b>Zielgröße: Wahl des Befragungsmodi (<math>\delta_{modus}</math>)</b>	
$H_{C2}$ :	Wird ein Anruf durch ein Anschreiben angekündigt, entscheidet sich ein größerer Teil an Zielpersonen für die Teilnahme an der CATI-Befragung.
$H_{C3}$ :	Wird kein Anruf durch ein Anschreiben angekündigt, entscheidet sich ein größerer Teil an Zielpersonen für die Teilnahme an der CAWI-Befragung.

nachweisen (Hypothese  $H_{C1}$ ) (Baur und Florian, 2009; Brehm, 1994; Hansen und Pedersen, 2012; Lavrakas, 2010; Rao et al., 2010; Scherpenzeel und Toepoel, 2012). So muss kein kalter Kontakt stattfinden und die Befragten können sich zuvor von der Seriosität der anfragenden Institution überzeugen und verfügen zum Zeitpunkt der persönlichen Kontaktaufnahme bereits über zentrale Informationen zum Befragungsprojekt.

Der Eintrag in das Telefonbuch ist heutzutage optional, sodass Personen selbst aktiv werden müssen, um darin verzeichnet zu sein (siehe Kapitel 2.1.1 (Noncoverage)). Diese Tatsache lässt die Vermutung zu, dass sich Personen mit und ohne Eintrag hinsichtlich der Umfragebereitschaft voneinander unterscheiden. Denn werden Personen selbst aktiv und lassen ihren Anschluss verzeichnen, sollte damit eine höhere Bereitschaft einhergehen, Anrufe von zunächst unbekanntem Personen entgegen zu nehmen. Lässt man seine Rufnummer nicht verzeichnen, kann darüber hinaus angenommen werden, dass hiermit eine größere Abneigung gegen unbekannte Anrufer vorhanden ist und diese ihre Privatsphäre gegen solche versuchen abzuschirmen (Hypothese  $H_{B1}$ ). Deutschmann und Häder (2002) haben darauf verwiesen, dass sich Personen mit und ohne Telefonbucheintrag hinsichtlich soziodemografischer Merkmale teilweise deutlich voneinander unterscheiden (Hypothese  $H_{B2}$ ). Systematische Unterschiede konnten hinsichtlich Alter und Bildung nachgewiesen werden: Personen, die über einen Eintrag verfügen, sind älter (im Mittel 41,8 gegenüber 35,1 Jahre) und verfügen eher über niedrige Bildungsabschlüsse (Deutschmann und Häder, 2002, 77). Begründet liegen diese Unterschiede ebenso darin, dass der Anteil von

Mobile-Onlys in den letzten Jahren größer geworden ist (Busse und Fuchs, 2015). Auch hier gibt es ausreichend empirische Evidenz für systematische Unterschiede zwischen dieser Gruppe und Personen, die zusätzlich über Festnetz oder ausschließlich über das Festnetz (Landline-Onlys) zu erreichen sind. Mobile-Onlys sind eher jünger, allein lebend, leben eher in städtischen Mietwohnungen und weisen häufig niedrige Einkommen auf (Busse und Fuchs, 2015; Lynn und Kaminska, 2012; Mohorko et al., 2013). Diese Eigenschaften gehen wiederum mit einer niedrigeren Wahrscheinlichkeit für einen Eintrag in das Telefonbuch einher (Hypothese  $H_{B2}$ ).

Bei der im Jahr 2012 in Bremen durchgeführten Nachrekrutierung von Teilnehmern für das PPSM Access Panel wurde in einer der Untersuchungsgruppen ein Telefonanruf angekündigt.<sup>4</sup> Erhalten potentielle Teilnehmer eine schriftliche Ankündigung und vergessen diese, kann der eine Woche später erfolgende Telefonanruf in seiner Funktion ähnlich wie ein klassischer Reminder betrachtet werden. Man wird durch häufige telefonische Kontaktversuche seitens der Interviewer (maximal bis zu zehn) mehrmals an die Umfrage erinnert. Die positive Wirkung von Remindern auf die Rücklaufquote insgesamt gilt als gesicherter empirischer Befund (Hypothese  $H_{B3}$ ) (Atkeson und Tafoya, 2008; Dillman et al., 2014; Link et al., 2008; Schonlau et al., 2003; Toepoel und Lugtig, 2014).

Einen inhaltlich relativ trivialen Effekt vermutet Hypothese  $H_{C2}$  sowie  $H_{C3}$  in Bezug zur Präferenz der Teilnehmer für eine der angebotenen Teilnahmealternativen. Mit dem postalischen Anschreiben wurde in einer der Untersuchungsgruppen nicht nur ein Telefonanruf angekündigt, sondern die potentiellen Teilnehmer haben darüber hinaus einen Link und ein Zugangspasswort zu einer inhaltlich homogenen CAWI-Umfrage zugesandt bekommen. Mit einer Teilnahme an dieser ist ein höheres Maß an Einsatzbereitschaft verbunden, denn URL und Passwort müssen abgetippt werden. Zur Teilnahme an der CATI-Befragung ist es lediglich notwendig, den Anruf abzuwarten. Die Antworten werden zusätzlich von einem Interviewer in der Befragungssoftware festgehalten, bei der CAWI-Variante müssen diese von den Teilnehmern selbst eingetragen werden. Durch die Unterschiede hinsichtlich des zu investierendem Aufwand in die Beantwortung der Fragen in den beiden Modi ist zu erwarten, dass sich Personen in dieser Untersuchungsgruppe eher für eine Teilnahme an der CAWI-Befragung entscheiden (Hypothese  $H_{C3}$ ). Ähnlich verhält sich die Variation des Anschreibens zwischen den jeweiligen Untersuchungsgruppen in

---

<sup>4</sup> Eine ausführliche Beschreibung der Datengrundlage sowie Untersuchungsgruppen findet im folgenden Kapitel statt. Der genaue Wortlaut der postalischen Anschreiben kann dem Anhang entnommen werden.

Abhängigkeit des Sachverhaltes, ob ein Anruf nach erfolgreicher Eruiierung der Rufnummer angekündigt werden kann oder nicht. Konnte dieser nicht angekündigt werden hatten die potentiellen Teilnehmer die Möglichkeit, ihre Rufnummer über eine Postkarte sowie einen vorfrankierten Briefumschlag zurück zu senden. Da hiermit ein größerer Aufwand mit der Teilnahme an der CATI- im Vergleich zur CAWI-Befragung in dieser Untersuchungsgruppe notwendig ist, werden sich die Personen eher für die Teilnahme an letzterer entschieden haben (Hypothese  $H_{C3}$ ). Wurde ein Anruf angekündigt ist es schließlich wahrscheinlicher, dass die eingeladenen Personen den Anruf abwarten und dann eher in diesem Modus partizipieren (Hypothese  $H_{C2}$ )

### **Teilnahmeanreize**

Dass sich Anreize in unterschiedlichen Formen positiv auf die Teilnahmebereitschaft und somit auf die Antwortrate auswirken, gilt als gesicherte empirische Erkenntnis. Wie bereits in Kapitel 4.2 (Teilnahmeanreize) beschrieben, ist dieser Effekt bei den nicht konditionalen Teilnahmeanreizen, die zumeist in Form kleinerer Geldbeträge (Form: NKPM - vgl. Kapitel 4.2 Tabelle 4.2) an alle Befragten in der jeweiligen Untersuchungsgruppe vergeben werden, zu beobachten (Hypothese  $H_{D1}$ ) (Becker et al., 2007; Berger, 2006; James und Bolstein, 1990; Millar und Dillman, 2011; Messer und Dillman, 2011; Nederhof, 1983; Scherpenzeel und Toepoel, 2012; Singer et al., 1999; Singer und Ye, 2013; Toepoel, 2012).

Im Gegensatz dazu hat sich in der Vergangenheit gezeigt, dass konditionale nicht monetäre Anreize, wie die Aussicht auf einen Gewinn in Abhängigkeit einer Teilnahme, keine oder nur marginale ausschöpfungssteigernde Wirkung im Kontext wissenschaftlicher Umfragen nachzuweisen ist (Hypothese  $H_{D2}$ ) (Hansen und Pedersen, 2012; Church, 1993).

Die Hypothesen  $H_{D4}$  und  $H_{D5}$  beziehen sich auf die subjektive Wahrnehmung der Befragten in Abhängigkeit der Anreizhöhe (kleine Aufmerksamkeit vs. Bezahlung). Wie dies bereits ausführlich in Kapitel 4.2 (Teilnahmeanreize) beschrieben wurde, ist die Forschungslage bei niedrigen Beträgen von bis zu zehn Euro/Dollar relativ eindeutig. Werden kleinere Anreize von einem Dollar vergeben und diese nur geringfügig erhöht, konnte ein linear positiver Effekt im Sinne einer ökonomischen Interpretation auf die Antwortrate nachgewiesen werden (Berger, 2006; James und Bolstein, 1990; Trussell und Lavrakas, 2004). Hypothese  $H_{D4}$  vermutet entsprechend, dass ein von der Teilnahme unabhängig vergebene Anreiz in Höhe von bis zu zehn Euro von den potentiellen Teilnehmern noch als „kleine Aufmerksamkeit“ wahrgenommen wird.

Tabelle 5.4: Forschungshypothesen zu Designaspekt D

<b>Zielgröße: Kooperationswahrscheinlichkeit (<math>p_k</math>)</b>	
$H_{D1}$ :	Der Einsatz monetärer un konditionaler Teilnahmeanreize (fünf Euro, unabhängig von der Teilnahme; Form: NKPM) hat einen positiven Effekt auf die Wahrscheinlichkeit der Teilnahme am Rekrutierungsinterview.
$H_{D2}$ :	Der Einsatz nicht monetärer konditionaler Teilnahmeanreize (Verlosung in Abhängigkeit einer Teilnahme; Form: NKP) hat keinen positiven Effekt auf die Wahrscheinlichkeit der Teilnahme am Rekrutierungsinterview.
$H_{D3}$ :	Der Einsatz monetärer nicht konditionaler Teilnahmeanreize (fünf Euro, unabhängig von der Teilnahme; Form: NKPM) erhöht insbesondere die Wahrscheinlichkeit der Teilnahme am Rekrutierungsinterview für schlechter gebildete Zielpersonen.
<b>Zielgröße: Subjektive Wahrnehmung der Anreizhöhe</b>	
$H_{D4}$ :	Beträge von bis zu zehn Euro werden von Befragten subjektiv als kleine Aufmerksamkeit wahrgenommen.
$H_{D5}$ :	Beträge von mehr als zehn Euro werden von Befragten subjektiv nicht mehr als kleine Aufmerksamkeit wahrgenommen.
<b>Zielgröße: Datenqualität</b>	
$H_{D6}$ :	Der Einsatz nicht konditionaler monetärer Anreize führt zu keiner Erhöhung von Item-Nonresponse

Werden größere Beträge vergeben, konnte kein Effekt auf die Antwortrate nachgewiesen werden (Berger, 2006; Scherpenzeel und Toepoel, 2012). Stellt man einen Zusammenhang zwischen den vorliegenden Studien her, zeigt sich, dass die Erhöhung niedriger Beträge zu einer höheren Teilnahmewahrscheinlichkeit führt. Da dies für größere Beträge ab zehn Euro nicht gilt scheint eher eine austauschtheoretische Interpretation nach Dillman et al. (2014) sinnvoll zu sein, die den sozialen Aspekt betont. Denn kleinere Beträge erhöhen den positiven Effekt auf die Antwortrate, sind aufgrund ihrer Höhe aber nicht als Bezahlung im Sinne eines ökonomischen Tausches zu betrachten. Werden mehr als zehn Euro als Anreiz vergeben ist eine Interpretation im ökonomischen Sinne naheliegend. Um dies zu überprüfen wurden die Teilnehmer der PPSM Access Panel Welle aus dem Jahr 2014 mittels eines offenen Antwortfeldes danach gefragt, welchen Anreiz sie als „Kleine Aufmerksamkeit“ und welchen sie als „Bezahlung“ wahrnehmen würden. Engel (2006) hat eine ähnliche Untersuchung durchgeführt, indem er Teilnehmer einer Umfrage danach gefragt hat, unter welchen Bedingungen eine Teilnahme eher wahrscheinlich beziehungsweise eher unwahrscheinlich ist. Die subjektive Einschätzung der Befragten zeigt dabei, dass der Einsatz von fünf Euro unabhängig von einer Teilnahme eher zu einer Beteiligung führt im Vergleich dazu, wenn kein Anreiz angeboten wird. Anreize von mehr als fünf Euro verschieben die Verteilung der Antworten in die Richtung, dass eine

Teilnahme hierdurch eher wahrscheinlich wird (Engel, 2006). Werden Feldexperimente durchgeführt und dabei tatsächlich Anreize vergeben, sind die vorliegenden Studien zu ähnlichen Ergebnissen gekommen. Werden größere Beträge vergeben, kann nicht mehr von eindeutigen Ergebnissen in eine bestimmte Richtung gesprochen werden. Daran knüpft sich die Frage, inwiefern Befragte bestimmte Anreize subjektiv wahrnehmen und ob sie diese als „kleine Aufmerksamkeit“ oder „Bezahlung“ beurteilen. Denn Dillman et al. (2014) haben drauf verwiesen, dass als zu gering empfundene Anreize, die als Bezahlung wahrgenommen werden, zu einer Ablehnung der Teilnahme führen können. Dies wird umso wahrscheinlicher, je mehr die potentiellen Teilnehmer die Höhe des Anreizes als unzureichend für die zu investierende Zeit empfinden.

Interpretiert man den Effekt eines monetären nicht konditionalen Anreizes von über zehn Euro und höher austauschtheoretisch und im ökonomischen Sinne, kann angenommen werden, dass sich die Wahrnehmung in Richtung „Bezahlung“ verschiebt. Diese Vermutung wird in Form von Hypothese  $H_{D5}$  und  $H_{D6}$  einer empirischen Überprüfung unterzogen.

### **Smartphonennutzung und mobile Onlineumfragen**

Der technologische Fortschritt unserer Gesellschaft ist ein Grund dafür, warum die Umfrageforschung regelmäßig vor neue Herausforderungen gestellt wird. Diese Herausforderungen sind gleichzeitig als Möglichkeit zu sehen, neue Technologien für wissenschaftliche Umfragen durch die Erforschung der damit einhergehenden methodischen Sachverhalte nutzbar zu machen. Eine dieser Entwicklungen ist in der zunehmenden Verbreitung des Internets in der Bevölkerung zu sehen, insbesondere im mobilen Bereich. Die Methodenforschung im Bereich mobiler Endgeräte (Smartphone, Tablet, Laptop) als Medium zur Beantwortung wissenschaftlicher Umfragen, steht durch diese neuere Entwicklung in den letzten Jahren noch am Anfang.

Wie bereits in Kapitel 2.1.1 (Noncoverage) diskutiert, unterscheiden sich Nutzer und Nicht-Nutzer des mobilen Internets trotz der zunehmenden Verbreitung in den letzten Jahren, weiterhin systematisch voneinander (ARD/ZDF-Onlinestudie, 2015). Diese sind besser gebildet, jünger und eher männlich (Hypothesen  $H_{S1}$ ,  $H_{S2}$ ,  $H_{S3}$ ). So sind auch heute noch vergleichbare Effekte durch Noncoverage bei der Durchführung von Onlineumfragen in Abhängigkeit davon zu erwarten, ob die Beantwortung der Fragen durch ein stationäres und/oder mobiles Endgerät möglich ist. Die Forschungslage hierzu ist dadurch, dass Onlineumfragen über mobile Endgeräte durch eben diese Probleme für allgemeine

Tabelle 5.5: Forschungshypothesen zur Smartphonennutzung und zur Teilnahmebereitschaft an einer mobilen Onlinebefragung

<b>Zielgröße: Smartphonennutzung</b> ( $\delta_{smartp}$ )	
$H_{S1}$ :	Smartphonennutzer verfügen über höhere Bildungsabschlüsse im Vergleich zu Personen, die diese nicht nutzen.
$H_{S2}$ :	Smartphonennutzer sind jünger im Vergleich zu Personen, die diese nicht nutzen.
$H_{S3}$ :	Männer nutzen Smartphones häufiger als Frauen.
<b>Zielgröße: Teilnahme an mobiler Onlinebefragung</b> ( $\delta_{mw}$ )	
$H_{MO1}$ :	Personen, die bereit sind an einer mobilen Onlinebefragung teilzunehmen, verfügen über höhere Bildungsabschlüsse.
$H_{MO2}$ :	Personen, die bereit sind an einer mobilen Onlinebefragung teilzunehmen, sind eher jünger.
$H_{MO3}$ :	Mehr Männer als Frauen erklären sich zu einer Teilnahme an einer mobilen Onlinebefragung bereit.

Bevölkerungsumfragen streng genommen noch nicht geeignet sind, bislang nicht breit aufgestellt. Trotzdem wird dieses Forschungsfeld in Zukunft an Bedeutung gewinnen, denn in der Vergangenheit hat sich gezeigt, dass Befragungsteilnehmer Onlinefragebögen häufig mobil beantworten und davon auszugehen ist, dass sich dieser Trend fortsetzt (Rivella et al., 2014; Toepoel und Lugtig, 2014). Bruijne und Wijnant (2013) sind zu dem Ergebnis gekommen, dass 57 Prozent der Teilnehmer den Fragebogen über ein mobiles Endgerät beantwortet haben. Für Access Panels werden Zahlen berichtet, die deutlich unter diesem Niveau liegen. Poggio et al. (2013) berichten zwar von einem nicht zu ignorierenden Anteil an Personen, der den Fragebogen per mobilem Endgerät beantwortet hat. Dieser liegt über die einzelnen Wellen hinweg lediglich bei acht Prozent. Wird versucht, sowohl Personen zu charakterisieren, die ein mobiles Endgerät besitzen und dies regelmäßig nutzen, als auch bereit sind, sich hierüber an Befragungen zu beteiligen, liegen bis heute keine eindeutigen Erkenntnisse vor. So berichten Bruijne und Wijnant (2013, 486), dass jüngere Mitglieder des Panels sowie mehr Männer und eher höher gebildete zu den Nutzern gehören und die Wahrscheinlichkeit für diese Personengruppe entsprechend höher ausfällt, einen Fragebogen mobil zu beantworten (Hypothesen  $H_{MO1}$ ,  $H_{MO2}$ ,  $H_{MO3}$ ). Poggio et al. (2013) konnten lediglich nachweisen, dass die Nutzung mobiler Endgeräte abnimmt, je älter ein Panelmitglied ist. Signifikante Effekte für Bildung und den Grad der Urbanisierung konnten nicht gefunden werden. Vermittelt über die Verfügbarkeit sowie Nutzungsintensität von Smartphones und Tablets variiert offensichtlich die Bereitschaft, diese für die Beantwortung von Umfragen zu nutzen. Die Entwicklung der Verbreitung und Nutzung mobiler Endgeräte

scheint dem Trend der allgemeinen Internetnutzung zu ähneln.

## Kapitel 6

# Datengrundlage und Operationalisierung

Die folgenden Unterkapitel beschreiben die empirischen Grundlagen der vorliegenden Arbeit. Zunächst wird das PPSM Access Panel und die damit einhergehende Datenstruktur erläutert. Erhoben wurden die Daten dieser Arbeit innerhalb von Lehrforschungsprojekten an der Universität Bremen. Die Studierenden wurden als Interviewer geschult und darin unterrichtet, was theoretisch sowie praktisch zur Planung, Durchführung und Berichterstattung einer Erhebungswelle innerhalb eines Access Panels gehört. Durchgeführt wurden die Erhebungen im Befragungslabor des Sozialwissenschaftlichen Methodenzentrums unter der Leitung von Prof. Dr. Uwe Engel. Das Befragungslabor ist ausgestattet mit sechs Computerarbeitsplätzen sowie der Befragungssoftware VOXCO.<sup>1</sup> Der Aufbau des Panels mit Rekrutierung und Initialisierung wurde im Jahr 2009 vom Sozialwissenschaftlichen Umfragezentrum in Duisburg durchgeführt.<sup>2</sup>

Die Operationalisierung zentraler Konstrukte zur Überprüfung der in Kapitel 5 hergeleiteten Forschungshypothesen und die damit einhergehenden feldexperimentellen Designbedingungen werden in den jeweiligen Abschnitten der betreffenden Erhebungen des PPSM Access Panels beschrieben. Mit einer Chronologie der zur Anwendung kommenden Analyseverfahren schließt dieses Kapitel ab.

---

<sup>1</sup> Informationen zur Befragungssoftware VOXCO sind unter <http://www.voxco.com/de/> zu finden.

<sup>2</sup> <http://www.suz-umfragen.de/>



## 6.1 Das PPSM Access Panel

Panelbefragungen zeichnen sich dadurch aus, dass dieselben Personen zu mehr als einem Zeitpunkt in regelmäßigen Abständen befragt werden (Schupp, 2014). So wird es möglich, intraindividuelle Veränderungen zu untersuchen und mittels komplexerer Analyseverfahren den Einfluss unbeobachteter Heterogenität besser als mit Querschnittsdaten zu kontrollieren (Allison, 2009; Giesselmann und Windzio, 2013). Eine der bekanntesten und umfangreichsten Panelbefragungen in Deutschland, die bereits seit 1984 durchgeführt wird, ist das SOEP (Schnell, 2012, 331 f.).

Das PPSM Access Panel ist eine spezielle Form der Längsschnittbefragung und Teil des von der Deutsche Forschungsgemeinschaft (DFG) geförderten Schwerpunktprogrammes 1292 „Survey Methodology - Access Panel and Mixed-Mode Internet Survey“ (Engel et al., 2012, 38). Der Begriff „Access Panel“ definiert sich durch Zustimmung seitens der Teilnehmer im Anschluss an eine erste Befragung, die benötigten Kontaktinformationen zu speichern sowie sich in regelmäßigen Abständen befragen zu lassen (Vehre, 2011, 37).<sup>3</sup> Die „Panelisten“ wurden aktiv durch vorhergehende zufällige Stichprobenziehung in Verbindung mit einer anschließenden telefonischen Kontaktierung inklusive der Nachfrage nach regelmäßiger Mitwirkung rekrutiert. Von einem passiv rekrutierten Access Panel spricht man im Gegensatz dazu, wenn Personen sich selbst in Adressdatenbanken von Umfrageinstituten eintragen und dem keine Stichprobenziehung vorausgegangen ist (Engel et al., 2004; Engel, 2015, 147). Dieses Verfahren wird in der kommerziellen Marktforschung eingesetzt. Dabei wird üblicherweise auf eine bestimmte Art von Anreizsystem zurückgegriffen, wie das Sammeln von Punkten für die Teilnahme an Befragungen, die man sich kumuliert in Form von Gutscheinen oder Bargeld auszahlen lassen kann (Schupp, 2014, 936). Für die wissenschaftliche Umfrageforschung ist dieses Vorgehen weniger geeignet. Denn für inferenzstatistische Analysen und einer möglichst fehlerfreien Hochrechnung auf zu bestimmende Grundgesamtheiten ist eine Stichprobenziehung nach dem klassischen Vorgehen (jedes Element der Grundgesamtheit hat die gleiche angebbare Wahrscheinlichkeit, in die Stichprobe gezogen zu werden) notwendig ist.

Das PPSM Access Panel wurde im Jahr 2009 mit einem Telefon- (Festnetz und Mobil) sowie einem Onlinezweig aufgebaut. Mit den regelmäßig stattfindenden Befragungen der

---

<sup>3</sup> Für eine Übersicht zur internationalen Verbreitung von Access Panels differenziert nach Rekrutierungsform (aktiv/passiv) sowie Modus der Befragung siehe Bartsch (2012, 27) sowie Vehre (2011, 39).

gleichen Personen wurden unterschiedliche Ziele an mehreren Projektstandorten verfolgt.<sup>4</sup> Das Hauptziel bestand darin, die Selektionseffekte zu operationalisieren, die bei der Rekrutierung von Teilnehmern eines solchen Access Panels auftreten (Engel et al., 2012, 38). Vor diesem Hintergrund wurde eine praktische Unterscheidung zwischen Rekrutierungs- und Initialisierungsphase vorgenommen, um die einzelnen selektiven Schritte bis zur letztlichen Teilnahme am Access Panel untersuchen zu können. Zudem wurden eine Reihe feldexperimenteller Designelemente implementiert. Zur Generierung zusätzlicher Paradata wurde ein flexibles Umfragedesign eingesetzt, womit über Personen, die eine Teilnahme zunächst verweigert haben, einige wenige Informationen erhoben werden konnten. Darüber hinaus sollten Modus- und Selektionseffekte statistisch sauber getrennt voneinander untersucht werden können, weshalb die Teilnehmer in Abhängigkeit davon, welche Kontaktinformationen zur Verfügung standen, zufällig auf den Telefon- (Festnetz und Mobil) oder Onlinezweig verteilt worden sind (Engel et al., 2012, 39).

Der thematische Schwerpunkt der Befragungen im PPSM Access Panel wurde sowohl in der Aufbauphase als auch im Rahmen der einzelnen Wellen mit dem Titel „Deutschland im Trend der Zeit“ umschrieben. Inhaltlich wurde versucht, tagesaktuelle Themen aufzugreifen, die gesellschaftlich diskutiert werden. Für diese Themen sollte sich ein größerer Teil der in die Stichprobe gezogenen Personen interessieren. Eine inhaltliche Auswertung der einzelnen Fragen der jeweiligen Wellen hat ebenso stattgefunden.<sup>5</sup>

In den nächsten Abschnitten werden die Stichprobengenerierung sowie Rekrutierungs- und Initialisierungsphase näher erläutert. Das PPSM Access Panel bildet dabei den Rahmen der zur Anwendung kommenden Datengrundlage. Daraufhin wird auf die Panelstruktur sowie Nachrekrutierung und die Operationalisierung zentraler Konstrukte eingegangen.

### **6.1.1 Generierung der Stichprobe, Rekrutierungs- und Initialisierungsphase**

Die Stichprobe für das PPSM Access Panel wurde mittels eines auf die Grundgesamtheit an Rufnummern in Deutschland abgestimmtes Randomized Last Digit (RLD) Verfahren

---

<sup>4</sup> Eine Übersicht über Projektstandorte und Themenschwerpunkte ist auf [www.survey-methodology.de](http://www.survey-methodology.de) zu finden.

<sup>5</sup> Eine inhaltliche Auswertung diverser Wellen ist in Engel (2014) zu finden. Eine Übersicht über die Erhebungsinstrumente der Aufbauphase des PPSM Panels liefert der Downloadbereich der Webseite des Methoden zentrums unter <http://www.methodenzentrum-bremen.de/>. Die Erhebungsinstrumente sowie relevante Inhalte und Fragen der Wellen, die dieser Arbeit als Datengrundlage dienen, werden in den folgenden Kapiteln und im Anhang ausführlicher beschrieben.

Tabelle 6.1: Selektionsstufen der Rekrutierungsphase des PPSM Access Panels

	Schritt 1.1 Davon gültig	Schritt 1.2 Kontakt zu HH	Schritt 1 Befragung	Schritt 2 Bereit beizutreten
<b>Bruttostichprobe (170.349)</b>	59,9% (102.138)	55,7% (56.848)	24,5% (14.784)	70,7% (10.452)

Anmerkungen: Quelle: Engel (2013, 11); absolute Fallzahlen (N) in Klammern; die geführten Interviews im Rahmen von Schritt 1 setzen sich aus 13.910 Voll-, 347 Kern- sowie 527 Exit-Versionen des Fragebogens zusammen.

(häufig in Bezug zu Deutschland mit Gabler/Häder-Design bezeichnet, siehe Kapitel 2.1.2 (Nonresponse)) generiert. Das Verfahren wurde seinerzeit entwickelt, um der abnehmenden Eintragungsdichte von Anschlüssen im Telefonverzeichnis entgegen zu wirken und so auch Haushalte in den Auswahlrahmen zu integrieren, die nicht darin eingetragen sind (Gabler und Häder, 2002). Tabelle 6.1 gibt eine Übersicht über die Selektionsstufen der Rekrutierungsphase. Als Grundgesamtheit wurden alle in Deutschland lebenden und wahlberechtigten Personen definiert (Vehre, 2011, 58).

Die Bruttostichprobe bestand aus der Gesamtanzahl von  $n = 170.349$  zufällig generierten Rufnummern. 59,9 Prozent ( $n = 102.138$ ) der generierten Rufnummern konnten als gültig identifiziert werden (Schritt 1.1), wovon zu insgesamt 55,7 Prozent ( $n = 56.848$ ) der Haushalte in dem darauffolgenden Schritt 1.2 Kontakt hergestellt werden konnte. Bei ungültigen Rufnummern handelt es sich um Faxanschlüsse, Rufnummern von Unternehmen oder nicht vergebene Anschlüsse.

Zur Teilnahme an einer ersten Befragung im Rahmen von Schritt 1 haben sich 24,5 Prozent ( $n = 14.784$ ) der per Last Birthday Methode ausgewählten Personen in den Haushalten bereit erklärt.<sup>6</sup> Schritt 1.2 sowie Schritt 1 haben zeitlich voneinander getrennt stattgefunden. Bei erfolgreichem Kontakt wurde zu der in darauffolgenden Schritt 1 erfolgten Befragung eingeladen. Dies hatte den Vorteil, dass es eine geringere Hürde darstellt, zunächst einmal eine Teilnahme hieran zuzusagen. So müssen die Befragten nicht beim ersten Kontakt einem für sie zunächst unbekanntem Umfrageinstitut der Speicherung von Kontaktdaten zur Realisierung weiterer Befragungswellen zustimmen. Darüber hinaus

<sup>6</sup> Die Next-/Last-Birthday-Methode ist eine Vorgehensweise zur randomisierten Auswahl von Personen in Haushalten. So wird sichergestellt, dass bei der Auswahl keine Selbstselektionseffekte in Kraft treten (Binson und Catania, 2000; Reuband, 2001).

Tabelle 6.2: Die Initialisierungsphase des PPSM Access Panels

Schritt 3 Willkommensbefragung	Schritt 4.1	Schritt 4.2			
	$S_{11}$	sukzessive Panelwellen $S_{12}/S_{21}$			
		1	2	3	n...
71,9% (7.021)	69,9% (1.804)	63,2% (4.128)	68,1% (2.811)	82,6% (2.323)	...

Anmerkungen: Quelle: Engel (2013, 11); absolute Fallzahlen (N) in Klammern.

kann durch eine zuvor stattfindende Befragung ein höheres Maß an Vertrauen hergestellt werden, damit sich in Schritt 2 möglichst viele Teilnehmer dazu bereit erklären, dem Panel beizutreten (Engel et al., 2012, 38 f.). Dies äußert sich in der Folge mit 70,7 Prozent in einer hohen Bereitschaft, sich regelmäßig befragen zu lassen. Auf diese Weise können Informationen über Personen gesammelt werden, die in Schritt 1 aus unterschiedlichen Gründen kein Interview gegeben haben. Darüber hinaus hatte das Rekrutierungsinterview (Schritt 1) die Aufgabe, weitere Kontaktmöglichkeiten wie Mobilfunknummern oder E-Mail Adressen zu ermitteln.

Der Blick auf Tabelle 6.2 zeigt, dass nicht alle Personen, die zum Ende des Rekrutierungsinterviews ihre Zustimmung zur Teilnahme am Access Panel gegeben haben, auch bei Schritt 3 an der Willkommensbefragung teilgenommen haben. 71,9 Prozent ( $n = 7.021$  Personen) haben schließlich an der Willkommensbefragung der Initialisierungsphase teilgenommen. Die zuvor definierte Zielgröße von  $n = 6.000$  Panelisten konnte somit erreicht werden (Engel et al., 2012, 38). Schritt 4.1 sowie 4.2 repräsentieren die einzelnen sukzessiven Panelwellen. Die Panellinien ( $S_{11}$ )  $S_{12}/S_{21}$  unterscheiden sich dadurch, dass die Linie  $S_{11}$  von einer Projektgruppe in Bremen und  $S_{12}/S_{21}$  durch das sozialwissenschaftliche Umfragezentrum (SUZ) in Duisburg, unter der Leitung von Prof. Dr. Faulbaum, realisiert worden sind (Engel, 2013, 9).

Die letzte von neun Kurzumfragen innerhalb des Panels hat im Winter 2014 stattgefunden. Es wurde simultan sowohl im Telefon- als auch im Onlinezweig mit ähnlichen Feldzeiten befragt. Von Welle zu Welle wurden unterschiedliche Methodenexperimente implementiert und getestet. Die Anzahl an wiederbefragungsbereiten Personen hat dabei,

wie bei einer Panelbefragung üblich, über die Wellen hinweg abgenommen. In der Forschungspraxis wird hier von Panelattrition oder Panelmortalität gesprochen (Lynn, 2009; Schnell, 2012, 331). Die Ausfallquoten haben sich in einem Bereich von circa 15 bis teilweise fast 40 Prozent bewegt. Wobei zu den Personen, die in einer Welle nicht erreicht werden konnten, in der jeweiligen Folgewelle nochmals versucht wurde, Kontakt herzustellen. Vor dem Hintergrund des relativ hohen Niveaus an Panelmortalität haben in den Jahren 2010 und 2012 drei Nachrekrutierungen stattgefunden, um das Panel wieder aufzufüllen. Methodisch haben sich diese von der eigentlichen Rekrutierung dadurch unterschieden, dass im Anschluss an das erste Interview direkt danach gefragt wurde, ob man bereit ist, sich in regelmäßigen Abständen befragen zu lassen.

Das Gerüst der zu untersuchenden Daten für die vorliegende Arbeit bildet das PPSM Access Panel. Die explizite Grundlage ergibt sich sowohl durch eine im Jahr 2012 im Land Bremen stattgefundene Nachrekrutierung von Teilnehmern, als auch durch die Daten der Panelwelle aus dem Jahr 2014. Die Daten der Nachrekrutierung lassen sich zur Analyse des Stichprobenansatzes und dessen Wirkung auf einen etwaigen Bias durch Nonresponse heranziehen. Als Auswahlrahmen wurde das allgemeine zentrale Melderegister genutzt. Darüber hinaus lassen die in den folgenden Kapiteln zu beschreibenden experimentellen Designbedingungen eine empirische Untersuchung der Wirkung unterschiedlicher Teilnahmeanreize, Vorausbriefer oder des Vorhandenseins eines Telefonbucheintrages zu.

Eine Aussage über die Charakteristika von Personen, die Smartphones und mobiles Internet nutzen, sowie die subjektive Wahrnehmung seitens der Befragten bezüglich der Wirkung unterschiedlicher Teilnahmeanreize ist mit den Daten der zuletzt stattgefundenen Panelwelle aus dem Winter 2014 möglich.

### **6.1.2 Nachrekrutierung - Bremen Stadt 2012**

Nachrekrutierungen finden im Kontext von Panelbefragungen statt, um diese wieder aufzufüllen. Sind nach einer bestimmten Anzahl an Befragungswellen durch Panelmortalität nicht mehr ausreichend befragungsbereite Personen vorhanden, müssen Teilnehmer nachrekrutiert werden. Ausfälle über die Befragungswellen eines Panels hinweg sind selbst mit höchstem Aufwand bei der Panelpflege nicht zu vermeiden (Schnell, 2012, 331). Unter dem Begriff der Panelpflege werden Vorgänge gefasst, um möglichst viele Personen dazu zu motivieren, an der nächsten Welle teilzunehmen. Praktisch handelt es sich dabei um die

Pflege der Datenbank mit den Kontaktinformationen oder das Zusenden von Ergebnissen einzelner Panelwellen an Teilnehmer auf Anfrage.

Die Art und Weise der Rekrutierung sowie die Stichprobenziehung unterscheiden sich nicht von dem der Aufbauphase eines Panels. Bei der Nachrekrutierung im Jahr 2012 wurde ein anderes Vorgehen gewählt. Über einen anderen Stichprobenansatz hinaus bestanden weitere Gründe für ein von der Rekrutierung abweichendes Umfragedesign darin, neue Ansätze und Vorgehensweisen einer empirischen Überprüfung zu unterziehen. Hierfür wurde eine Stichprobe aus dem zentralen Einwohnermelderegister gezogen. Nach §21 des Melderechtsrahmengesetzes wird eine Gruppenauskunft erteilt, wenn dies im öffentlichen Interesse liegt. Zur Darlegung dieses Aspektes gegenüber der jeweils zuständigen Meldebehörde ist ein kurzes Exposé über das dahinterstehende Projekt notwendig.<sup>7</sup>

Tabelle 6.3 gibt eine Übersicht über die Final Disposition Codes nach der AAPOR (2015) sowie die sich daraus ergebende übergeordnete Antwortrate. Die AAPOR schlägt verschiedene Möglichkeiten zur Berechnung von Antwortraten vor. Hier wurde eine der konservativeren Berechnungsweisen gewählt, da die Partial Interviews als Ausfälle behandelt werden und somit im Nenner der Formel platziert sind. Diese setzt sich aus der Klassifikation der einzelnen Kontaktergebnisse zusammen:

$$ResponseRate1 = \frac{I}{(I + PI) + (R + NC + O) + (UH + UO)} \quad (6.1)$$

- I = Interview
- PI = Partial Interview
- R = Refusal and break off
- NC = Non Contact
- O = Other
- UH = Unknown Household
- UO = Unknown other

Der Blick auf Tabelle 6.3 zeigt, dass sich die übergeordnete Antwortrate mit 12 Prozent auf einem relativ niedrigen Niveau befindet. Dies liegt darin begründet, dass  $n = 1.825$  Personen in der Stichprobe keine Reaktion auf das Anschreiben gezeigt haben. Die Summe der geführten Interviews von  $n = 290$  erscheint vor dem Hintergrund der Größe der Ausgangsstichprobe zunächst einmal niedrig. Die Stichprobe setzt sich aus verschiedenen

---

<sup>7</sup> Im Rahmen einer Gruppenauskunft werden folgende Informationen der zuständigen Meldebehörde herausgegeben: Tag der Geburt, Geschlecht, Staatsangehörigkeiten, Anschriften, Tag des Ein- und Auszuges, Familienstand beschränkt auf die Angabe, ob verheiratet oder eine Lebenspartnerschaft führend oder nicht.

Tabelle 6.3: Final Disposition Codes und übergeordnete Antwortrate - PPSM Access Panel Nachrekrutierung 2012

<b>Neutrale Ausfälle</b>	
<i>CATI</i>	
Falsche Rufnummer	13
Fax-/Büroanschluss	3
Dauerhaft besetzt	2
Falsche Adresse, Brief zurückerhalten	33
$\Sigma$	51
<b>Relevante Ausfälle</b>	
<i>CATI</i>	
Verweigerung durch ZP/Abbruch durch ZP	96
Verweigerung KP	56
Kontaktausfälle	98
Annahme des Vorausbriefes verweigert	1
<i>CAWI</i>	
Abbruch mit unzureichender Beantwortung der Fragen	6
$\Sigma$	257
<b>Ungeklärte (relevante) Ausfälle</b>	
Keine Reaktion auf postalische Einladung	1825
<b>Interviews (CATI)</b>	117
<b>Interviews (CAWI)</b>	173
$\Sigma$	290
Gesamt	2423
<b>RR1</b>	<b>12%</b>

Quelle: Engel und Schmidt (2015, 268)

Anmerkung: ZP = Zielperson/ KP = Kontaktperson.

Untersuchungsgruppen zusammen, die in Tabelle 6.4 (Seite 102) dargestellt sind.<sup>8</sup>

In Kapitel 7.1.1 (Antwortraten und Moduspräferenzen) erfolgt eine lokale Betrachtung der erzielten Antwortraten in den Gruppen, die vereinzelt deutlich über 12 Prozent liegen.

Als Grundgesamtheit wurden in Bremen gemeldete Personen, die zum Zeitpunkt der Stichprobenziehung mindestens 18 Jahre sind, definiert. Der Bruttostichprobenansatz wurde auf  $n = 10.000$  Personen in Haushalten festgelegt. Dieser wurde so hoch angesetzt, damit zur Durchführung der Befragung ausreichend Rufnummern vorhanden sind. Um die jeweiligen Rufnummern der Personen zu eruieren, die über einen Telefonbucheintrag verfügen, hat ein Abgleich der digitalen Version des Telefonbuches mit der Stichprobendatenbank stattgefunden. Hierzu wurde ein entsprechender Matchingalgorithmus programmiert, der den Abgleich automatisiert vorgenommen hat. Das Matching ist in  $n = 844$  Fällen erfolgreich gewesen (Engel und Schmidt, 2013, 2). Wie beschrieben, ist das Telefonbuch als

<sup>8</sup> Die folgenden Ausführungen orientieren sich an Engel und Schmidt (2015).

Tabelle 6.4: Studiendesign Nachrekrutierung 2012

Registerstichprobe					
Rufnummer bekannt			Rufnummer unbekannt		
<b>A</b>		<b>B</b>		<b>C</b>	
Vorausbrief: Ankündigung des Anrufes/Möglichkeit Teilnahme online		Vorausbrief: Rückantwortschreiben/Möglichkeit Teilnahme online		Vorausbrief: Rückantwortschreiben/Möglichkeit Teilnahme online	
<b>D</b>	<b>E</b>	<b>F</b>	<b>G</b>	<b>H</b>	<b>I</b>
5 Euro	Kontrollgruppe	Verlosung	Kontrollgruppe	Verlosung	Kontrollgruppe

Auswahlrahmen dadurch, dass die Eintragungsdichte über die Zeit geringer geworden ist und sich Personen mit und ohne Eintrag systematisch voneinander unterscheiden, nicht mehr geeignet. Die zu erreichende Trefferquote reduziert sich notwendigerweise auf die vorherrschende Eintragungsdichte, die nach Schnell (2012) bereits bei unter 50 Prozent liegt.

Weiterhin konnte nicht zu jeder Person in der Stichprobe eine Rufnummer zugeordnet werden, auch wenn diese über einen Eintrag verfügt. Verursacht wird dies dadurch, dass sich die Schreibweisen von Straßennamen und/oder anderer Informationen in Abhängigkeit der jeweiligen Datenbank unterscheiden können. Eine valide Zuordnung konnte in solchen Fällen nicht immer stattfinden. Aus der Gruppe an Personen, für die keine Rufnummer eruiert werden konnte, wurden letztlich insgesamt  $n = 1.579$  Personen zufällig ausgewählt und angeschrieben. Dieses Vorgehen wurde aus Kosten- und Zeitgründen gewählt. Insgesamt wurden  $n = 2.423$  Personen kontaktiert.

In der Kontaktphase wurde an alle Personen in der Stichprobe ein postalisches Anschreiben versandt, das sich in Abhängigkeit der jeweiligen Untersuchungsgruppe unterschied.<sup>9</sup> Gemeinsam hatten diese eine persönliche Anrede der Zielperson, Informationen über die Hintergründe der Studie, den Befragungszeitraum sowie Hinweise zur Freiwilligkeit der Teilnahme und zum Datenschutz. Eine Variation bestand hinsichtlich der Möglichkeiten zur Teilnahme sowie der Hinweise auf vergebene Anreize. Die  $n = 844$  Personen, zu denen erfolgreich eine Rufnummer zugeordnet werden konnte, wurden zufällig den Untersuchungsgruppen *A* und *B* zugeordnet (siehe Tabelle 6.4). Personen in Gruppe *B* wurde, ebenso wie denjenigen in Gruppe *C*, die Möglichkeit angeboten, den Fragebogen entweder telefonisch oder online zu beantworten. Dadurch, dass für Gruppe *C* keine Rufnummern eruiert werden konnten, wurde dem Anschreiben eine Antwortkarte in Verbindung mit

<sup>9</sup> Die genauen Wortlaute der Anschreiben sind Anhang D zu entnehmen.



einem vorfrankierten Briefumschlag beigelegt. Die Antwortkarten sind zur späteren Identifizierung mit einer individuellen Personennummer versehen gewesen. Die potentiellen Teilnehmer hatten so die Möglichkeit, ihre Rufnummer über den Rückantwortbrief zur Teilnahme an der CATI-Befragung mitzuteilen. Darüber hinaus enthielt das Anschreiben einen Link zum Onlinefragebogen sowie ein individuelles Passwort für eine Teilnahme an der CAWI-Befragung.

Für Personen in Untersuchungsgruppe *B* konnte zwar eine Rufnummer eruiert werden, zum Zwecke der Vergleichbarkeit wurde hier dieselbe Art und Weise der Kontaktierung gewählt wie in Gruppe *C*. Daraufhin ist eine Vergleichbarkeit zwischen diesen beiden Gruppen entlang der Linie - Telefonbucheintrag vorhanden oder nicht vorhanden - möglich. Denn wie schon zuvor in Kapitel 2 (Fehlerquellen) beschrieben, sind zwischen diesen Gruppen systematische Unterschiede zu erwarten. Beachtet werden muss, dass innerhalb von Gruppe *C* Personen mit und ohne Telefonbucheintrag, aufgrund der bereits beschriebenen Problematik, konfundiert sind. Etwaige Unterschiede entlang dieser Linie werden tendenziell eher unterschätzt.

Personen in Untersuchungsgruppe *A* wurde mit dem postalischen Anschreiben ein wenige Tage nach Erhalt stattfindender Telefonanruf angekündigt. In diesem Anschreiben war ebenso ein Link sowie ein individuelles Passwort zur Teilnahme an der Onlinevariante enthalten, sodass die potentiellen Teilnehmer die Möglichkeit hatten, zwischen diesen beiden Modi (CATI/CAWI) wählen zu können.

### **Operationalisierung der Wirkung von Teilnahmeanreizen**

Letztlich wurden die Gruppen *A*, *B* und *C* zufällig in Experimental- und Kontrollgruppen zur Untersuchung der Wirkung von Teilnahmeanreizen aufgeteilt. Dem Anschreiben in Gruppe *D* lag ein Fünfeuroschein mit dem Hinweis bei, dass dies vorab ein kleines Dankeschön für die mit einer Teilnahme verbundene Mühe darstellt ( $n = 220$ , Form: NKPM - vgl. Kapitel 4.2, Tabelle 4.2). Der Wortlaut des Hinweises im Postskriptum der Anschreiben ist in folgender Textbox abgebildet.

P.S.: Bitte erlauben Sie uns, dass wir Ihnen ganz unabhängig davon, ob Sie an der Umfrage teilnehmen möchten oder nicht, ein kleines Dankeschön für Ihre Aufmerksamkeit beilegen.

Es wurde explizit darauf verwiesen, dass die fünf Euro auch dann behalten werden

dürfen, wenn man nicht an der Befragung teilnimmt. Die Anschreiben in den Gruppen *F* und *H* waren im Postskriptum jeweils mit einem Hinweis darauf versehen, dass mit der Teilnahme die Möglichkeit auf den Gewinn attraktiver Preise besteht ( $n = 193 / n = 807$ , Form: NKP - vgl. Kapitel 4.2, Tabelle 4.2). Unter allen Teilnehmern in diesen beiden Gruppen wurden zwei Smartphones sowie ein Tablet verlost. *E*, *G* und *I* bilden die jeweiligen Kontrollgruppen. In folgender Textbox findet sich die Formulierung des Hinweises auf die Verlosung.

P.S.: Ohne Antworten kann keine wissenschaftliche Umfrage funktionieren. Wir möchten uns deshalb bei allen, die uns für die Umfrage ein wenig ihrer kostbaren Zeit opfern, mit einer Verlosung attraktiver Preise bedanken. Teilnehmen kann daran jeder, der auch an der Umfrage teilgenommen hat.

Fraglich ist, ob die potentiellen Teilnehmer die Hinweise auf die Anreize im Postskriptum gelesen haben. Denn das Anschreiben hatte eine Vor- und Rückseite. Die Hinweise standen relativ am Ende der zweiten Seite, sodass diese schnell übersehen werden können. Dies spiegelt sich im Rahmen der Teilnahmequoten im Vergleich der einzelnen Gruppen im Ergebnis wieder, was im Ergebnisteil (Kapitel 7.1.1 - Antwortraten und Moduspräferenzen) wieder aufgegriffen wird.

### 6.1.3 Befragung Winter 2014

Die Datengrundlage wird durch die Befragungswelle des PPSM Access Panels aus dem Jahr 2014 komplettiert. Hierfür wurden von den befragungsbereiten Personen  $n = 846$  ausgewählt. Diese verteilten sich auf die beiden Modi CATI ( $n = 640$ ) und CAWI ( $n = 328$ ). Da die Erhebung als Lehrforschungsprojekt mit Studierenden der Universität Bremen realisiert worden ist, war eine Beschränkung der Telefonstichprobe in Abhängigkeit der zur Verfügung stehenden Interviewer notwendig. Die Onlinestichprobe umfasst alle Personen, die sich zu dieser Zeit im Onlinezweig befanden und zuvor weitergehende Teilnahmebereitschaft signalisiert haben. Das PPSM Access Panel wurde, wie schon in Kapitel 6.2 beschrieben, in verschiedene Streams aufgeteilt. Die Erhebungswelle aus dem Winter 2014 ist für einige der Personen bereits die fünfte oder sogar sechste Befragung.<sup>10</sup>

<sup>10</sup> Eine Übersicht der Panelwellen und Panelstreams ist in (Engel et al., 2012, 82 f.) zu finden.

Tabelle 6.5: Final Disposition Codes und Antwortrate - PPSM Access Panel Befragung Winter 2014

<b>Neutrale Ausfälle</b>	
<i>CATI</i>	
Falsche Rufnummer	68
Fax-/Büroanschluss	5
Kein Kontakt	155
ZP in Feldzeit nicht zu erreichen	22
ZP möchte online teilnehmen	4
Technische Probleme	2
$\Sigma$	256
<b>Relevante Ausfälle</b>	
<i>CATI</i>	
Verweigerung durch ZP/Abbruch durch ZP	57
Verweigerung KP	16
Kontaktausfälle	28
Verweigerung, unklar ob durch KP/ZP	24
<i>CAWI</i>	
Abbruch mit unzureichender Beantwortung der Fragen	22
Nicht auf Einladungs-E-Mail reagiert	142
$\Sigma$	289
<b>Interviews (CATI)</b>	259
<b>Interviews (CAWI)</b>	164
$\Sigma$	423
Gesamt	968
<b>RR1</b>	<b>58,7%</b>

Anmerkung: ZP = Zielperson/ KP = Kontaktperson.

Inhaltlich wurden Fragen zum Thema aktueller Krisenherde in der Welt, wie zum Bürgerkrieg in Syrien, sowie zu religionssoziologischen Aspekten gestellt. Der Fokus lag darauf, inwiefern eine Religion bestimmte Lebensweisen vorgeben und in den Alltag von Menschen eingreifen darf oder nicht. Tabelle 6.5 gibt einen Überblick über die Final Disposition Codes sowie die sich daraus ergebende Antwortrate. Es haben sich 58,7 Prozent der kontaktierten Personen an der Befragung beteiligt. Innerhalb des PPSM Access Panels reiht sich dieser Wert im unteren Bereich der von Welle zu Welle variierenden Antwortraten ein. Ein Grund dafür kann sein, dass zwischen der Befragung aus dem Jahr 2014 und der vorherigen Welle aus organisatorischen Gründen circa ein Jahr vergangen ist. Je mehr Zeit zwischen den einzelnen Befragungen in einem Panel vergeht, umso höher ist die Wahrscheinlichkeit von Ausfällen aufgrund falscher Kontaktinformationen wie durch Umzüge oder Ähnlichem (Schnell, 2012).

Mit dem feldexperimentellen Design wurde der Versuch unternommen, herauszufinden,

ob Befragte das Anbieten von Teilnahmeanreizen als Bezahlung oder kleine Aufmerksamkeit wahrnehmen. Von Interesse ist dabei, ob sich die subjektive Wahrnehmung seitens der Befragten diesbezüglich ab einem bestimmten Wert verändert. Um dies zu untersuchen, wurden zufällige Untersuchungsgruppen gebildet.

### **Operationalisierung der Wirkung von Teilnahmeanreizen**

In der Befragungswelle des PPSM Access Panels im Jahr 2014 wurde unter anderem ein feldexperimentelles Design implementiert, um die Wirkung von Teilnahmeanreizen nach dem theoretisch beschriebenen Vorgehen von Dillman et al. (2014) zu untersuchen, die vor allem den Charakter des sozialen Austausches betonen. Ergänzend zu den Feldexperimenten aus der Nachrekrutierung 2012 soll dieses Vorgehen die subjektive Wahrnehmung auf Seiten der potentiellen Teilnehmer in den Fokus rücken. Denn in Kapitel 4.2 (Teilnahmeanreize) ist deutlich geworden, dass Anreize in ihrer Wirkung auf die Antwortrate bereits häufig untersucht worden sind und die empirischen Evidenzen diesbezüglich als gesichert angesehen werden können. Bis heute ist relativ wenig untersucht, wie Befragte Anreize subjektiv wahrnehmen und einschätzen.

Von Interesse ist daran anknüpfend, welche Anreizhöhe subjektiv als angemessen eingeschätzt wird. Um dies empirisch zu untersuchen wurde im Fragebogen aus dem Jahr 2014 eine hypothetische Situation konstruiert, in der den Befragten der folgende Fragetext vorgelesen (CATI) oder angezeigt (CAWI) wurde:

Als nächstes würden wir gerne zu einem anderen Thema kommen. Institute aus Markt- und Sozialforschung zahlen häufig Geld, um Personen zur Teilnahme an Umfragen zu bewegen. Stellen Sie sich vor, Sie erhalten von einem solchen Institut Post. Dem Schreiben liegt ein Geldbetrag bei, den Sie behalten dürfen, egal ob Sie teilnehmen oder nicht. Würden Sie sich davon in Ihrer Entscheidung, teilzunehmen oder nicht teilzunehmen, beeinflussen lassen?

Ja/ Nein/ vielleicht, kommt darauf an.

Fluchtkategorien wie „keine Angabe“ und „weiß nicht“ sind ebenfalls im Fragebogen angegeben. Die Interviewer sind für die Erhebung mittels der CATI-Variante in allen Umfragen des PPSM Access Panels angewiesen, diese Kategorien nicht explizit vorzulesen. In der CAWI-Variante wurden diese Kategorien auf der jeweiligen Fragebogenseite angezeigt

und farblich in Graustufen gehalten. Hiermit sollten diese optisch von den inhaltlichen Kategorien abgesetzt werden, um dem Begriff der Fluchtkategorie gerechter zu werden. Denn der Anteil an Antworten, wie „keine Angabe“ oder „weiß nicht“, sollen möglichst gering gehalten werden, da Werte in diesen Kategorien als Item-Nonresponse anzusehen sind und nicht für die Analyse verwendet werden können.

Im Zentrum des Interesses steht die subjektive Wahrnehmung nicht konditionaler monetärer Anreize. So war es notwendig innerhalb des Fragebogens eine hypothetische Situation zu konstruieren, die dieses Prinzip aufgreift. Die erste Frage zielt darauf ab zu operationalisieren, ob sich Befragte in ihrer Teilnahmeentscheidung von einem Anreiz beeinflussen lassen. Hierbei handelt es sich um eine Filterfrage, da in Abhängigkeit der Antwort hierauf die Folgefragen gestellt/angezeigt worden sind („Ja“, „vielleicht, kommt darauf an“) oder nicht („Nein“). Lassen sich Befragte in ihrer Entscheidung von einem Anreiz beeinflussen, erfolgte eine experimentelle Zuteilung zu einer von vier Untersuchungsgruppen und die Überleitung zur nächsten Frage. Für alle anderen Personen wurde dieser Teil des Fragebogens übersprungen.

Angenommen, Sie erhalten *5 Euro/10 Euro/20 Euro/40 Euro (Zufallssplit zu jeweils 1/4)*, was wäre Ihre spontane Reaktion? Würden Sie dies als **kleine Aufmerksamkeit** oder als **Bezahlung** empfinden? Auf einer Skala von 1 bis 7 bedeutet die 1, Sie empfinden es als kleine Aufmerksamkeit und die 7, dass Sie es als Bezahlung empfinden würden. Wo auf dieser Skala würden Sie sich einordnen? Mit den Werten dazwischen können Sie Ihr Urteil abstufen.

So wurde zufällig ein Betrag von 5, 10, 20 oder 40 Euro vorgelesen oder angezeigt und eine Antwortskala von 1 „kleine Aufmerksamkeit“ bis 7 „Bezahlung“ aufgespannt. Fett gedruckte Begriffe wurden von den Interviewern im Rahmen der CATI-Variante besonders betont und im CAWI-Fragebogen so angezeigt. Die Teilnehmer wurden gebeten, ihr spontanes subjektives Empfinden auf dieser Skala einzuordnen. Das Ziel dieses Vorgehens ist es herauszufinden, ob es einen Grenzwert gibt, ab dem Befragte einen Teilnahmeanreiz als Bezahlung wahrnehmen. Dies sollte nach Dillman et al. (2014) vermieden werden, da ein Betrag, der als Bezahlung wahrgenommen wird, schnell als unzureichend empfunden werden kann und eine Verweigerung der Teilnahme nach sich zieht. Als Ergänzung wurde dieselbe Frage in Verbindung mit offenen Antwortkategorien gestellt.

Bitte sagen Sie mir noch, welchen Eurobetrag Sie als **kleine Aufmerksamkeit** einstufen würden?

Betrag in Euro wäre eine kleine Aufmerksamkeit.

Und welchen Betrag würden Sie als **Bezahlung** wahrnehmen?

Betrag in Euro würde ich als Bezahlung empfinden.

Wenn Befragte einen Anreiz als kleine Aufmerksamkeit oder Bezahlung wahrnehmen, der in der Skala des Feldexperimentes nicht vorkommt, haben diese die Möglichkeit in der Folgefrage einen expliziten Betrag anzugeben. Daran anknüpfend wird eine inhaltsanalytische Auswertung der Antworten auf der Suche nach einem möglichen Grenzwert stattfinden.

### **Bereitschaft zur Teilnahme an einer mobilen CAWI-Befragung**

Ein wesentliches Erkenntnisinteresse besteht darin herauszufinden, welcher Personenkreis im PPSM Access Panel wissenschaftliche Umfragen nicht nur mit stationären Computern, sondern ebenso mit mobilen Endgeräten, wie Smartphones, beantworten würde. Denn vorliegende Studien konnten wiederholt einen gewissen Anteil an Teilnehmern beobachten, die ein mobiles Endgerät für die Beantwortung von Fragen genutzt haben (siehe Kapitel 2.1.1). Daran anknüpfend wurde eine Filterfrage gestellt, um herauszufinden, welche Teilnehmer ein Smartphone besitzen und dieses auch regelmäßig nutzen. Denn besitzen diese keines oder nutzen es nur unregelmäßig ist es nahezu auszuschließen, dass diese auch einen Fragebogen hierüber beantworten würden. Die Folgefragen wären somit nicht zutreffend.

Zum Schluss würden wir gerne noch von Ihnen wissen, ob Sie ein Smartphone besitzen und dieses auch regelmäßig nutzen?

Ja/ Nein

Beantworten die Teilnehmer die Frage mit „Ja“, wird die folgende Frage eingeblendet. In diesem Fall werden unterschiedliche Formen der mobilen Onlineumfrage differenziert. Zunächst wird die Variante abgefragt, bei der die potentiellen Teilnehmer einen Link zum Fragebogen per E-Mail mit einer Umfrageeinladung zugeschickt bekommen.

Wären Sie bereit, an einer Umfrage teilzunehmen, in der wir Sie bitten, für die Beantwortung der Fragen Ihr Smartphone zu verwenden?

Ja, wenn wir Ihnen dafür an Ihre E-Mail-Adresse einen Link zum Fragebogen schicken würden?

Ja, über eine sichere, von uns entwickelte „App“, die Sie sich vorher kostenlos herunterladen könnten?

Nein

Darüber hinaus gibt es heutzutage die Möglichkeit, Befragungen über Applikationen (App) auf mobile Endgeräte zu bringen. Applikationen müssen zuvor manuell heruntergeladen werden. Dieses Vorgehen wird vor allem in der Marktforschung genutzt.<sup>11</sup> Aus der wissenschaftlichen Umfrageforschung ist bisher kein Beispiel bekannt, in der Applikationen für Umfragezwecke eingesetzt werden. Die Kombination der Fragen nach der Nutzung eines Smartphones und der Bereitschaft, hierüber einen Fragebogen zu beantworten, soll einerseits eine Einschätzung über die Quantität der Nutzer sowie deren Teilnahmebereitschaft möglich machen. Andererseits soll die Gruppe an Personen in Abhängigkeit relevanter Merkmale beschrieben werden, die teilnehmen beziehungsweise nicht teilnehmen würden. So soll ein Beitrag zum Stand der Forschung geleistet werden, um mögliche Trends aufzuzeigen und das Potential mobiler Onlinebefragungen besser als bisher nutzbar zu machen.

#### **6.1.4 Operationalisierung weiterer zentraler Konstrukte aus den Jahren 2012 und 2014**

In diesem Kapitel wird die Operationalisierung weiterer Konstrukte beschrieben, die in der statistischen Analyse zur Anwendung kommen. Da die Formulierungen sowie die Skalierung im PPSM Access Panel für zentrale Fragen in der Regel konstant geblieben sind, um Vergleichbarkeit zu gewährleisten und teilweise dieselben Variablen aus den unterschiedlichen Wellen herangezogen werden, erfolgt in diesem Kapitel eine integrierte Beschreibung soziodemografischer Variablen und verfügbarer Parainformationen.

<sup>11</sup> Die Firma Pollion nutzt etwa Applikationen für die Durchführung marktwirtschaftlicher Umfragen - <http://pollion.com/>.

## Soziodemografie und Parainformationen

Mit soziodemografischen Dimensionen, wie dem Bildungsniveau, wurde in der Vergangenheit der Begriff des Bildungsbias geprägt, der meint, dass besser gebildete Personen in wissenschaftlichen Umfragen überrepräsentiert sind (Vehre, 2011). Das Bildungsniveau dient dieser Arbeit in seiner Funktion zur Überprüfung der betreffenden Hypothesen, die besagen, dass sich mehrheitlich schlechter gebildete Personen von Teilnahmeanreizen angesprochen fühlen.

Tabelle 6.6: Soziodemografische Variablen und Parainformationen im PPSM Access Panel

Variable	Merkmalsausprägungen	Erhebungsjahr
<i>Soziodemografie</i>		
Alter	in Jahren	2012, 2014
Geschlecht	Weiblich Männlich	2012, 2014
Bildungsniveau	Noch Schüler Schule beendet ohne Abschluss Volks-/Hauptschulabschluss Polytechnische Oberschule (Abschluss 8./9. Klasse) Mittlere Reife Polytechnische Oberschule (Abschluss 10. Klasse) Fachhochschulreife Abitur Erweiterte Oberschule (Abschluss 12. Klasse) Anderer Abschluss	2012, 2014
<i>Parainformationen CATI</i>		
Kontaktversuche	1 bis n	2012, 2014
Tageszeit Kontaktversuch	Vormittags (10:00 - 12:00 Uhr) Nachmittags (13:00 - 17:00 Uhr) Abends (17:01 bis 19:00 Uhr)	2012, 2014 2012, 2014 2012, 2014
Interviewer	Individueller Code je Interviewer	2012, 2014
Überzeugungsargumente	1 bis n	2012, 2014
Verweigerungsargumente	1 bis n	
Anzahl vereinbarter Termine mit KP/ZP	1 bis n	2012, 2014
Anrufbeantworter besprochen	ja/nein	2012, 2014
<i>Parainformationen CAWI</i>		
Art des Browsers		2012, 2014
Start- und Endzeitpunkt	hierüber: Befragungsdauer	2012, 2014
Verweildauer bei einer Frage		2012, 2014
<i>Regionalinformation</i>		
Postleitzahl	hierüber: Dummy-Therme	2012



Die soziodemografischen Variablen werden ferner dazu genutzt, um zu überprüfen, wie dieses mit der Nutzung von Smartphones sowie der Bereitschaft eine Umfrage auch mobil zu beantworten, zusammenhängt. Zur Analyse des Ausfallprozesses und möglichen Verzerrungen kann das Bildungsniveau nicht heran gezogen werden, da diese Information ausschließlich für Teilnehmer und somit für die Nettostichprobe zur Verfügung steht. Wie schon mehrfach erwähnt, kann es sich bei der Analyse des Ausfallprozesses ausschließlich um Variablen handeln, die entweder bereits im Auswahlrahmen und somit der Bruttostichprobe vorhanden sind oder softwaregestützt als Parainformationen mit erhoben werden und den Prozess der Erhebung selbst beschreiben. Im Kontext des PPSM Access Panels und der zugrundeliegenden Erhebungszeitpunkte kann durch die zur Anwendung gekommene Erhebungssoftware VOXCO (CATI) sowie LimeSurvey (CAWI) ein Set an Parainformationen zur Schätzung von Teilnahmewahrscheinlichkeiten herangezogen werden. Ferner stehen durch die Art und Weise der Stichprobenziehung auf Basis des Melderegisters der Stadt Bremen für die Erhebung aus dem Jahr 2012 Variablen unabhängig davon zur Verfügung, ob sich Personen an der Umfrage beteiligt haben oder nicht. Tabelle 6.6 gibt einen Überblick über die verfügbaren Variablen.

Die Variablen Alter und Geschlecht sind für die 2012er Befragung im Auswahlrahmen enthalten und können somit für Nonresponseanalysen herangezogen werden. Das Bildungsniveau wurde zu den zwei Befragungszeitpunkten 2012 und 2014 über den Fragebogen erhoben und steht somit für die Personen zur Verfügung, die teilgenommen haben.<sup>12</sup>

Als Parainformationen können für Personen, die angerufen worden sind, die Anzahl an Kontaktversuchen, Tageszeit des Kontakts, die individuelle Zuordnung von Interviewern sowie Überzeugungs- und Verweigerungsargumente für Analysezwecke herangezogen werden. Letztere stehen nur zur Verfügung, wenn ein telefonischer Kontakt zustande gekommen ist und der Hörer nicht sofort aufgelegt worden ist, sodass die Interviewer keine Möglichkeit haben, zögerliche Personen zu überzeugen. Im Anschluss an die jeweilige Befragung wurde von den Interviewern ein Kontakt- beziehungsweise Verweigerungsprotokoll ausgefüllt. Diese Protokolle dienen in der Aufbauphase des Panels dazu, um besonders gut wirkende Überzeugungsargumente zu identifizieren oder Befragte auszumachen, die sich keine Mühe bei der Beantwortung der Fragen gegeben haben und durch den Fragebogen „gerast“ sind.

Der individuelle Code, der für jeden Interviewer von der CATI-Software vergeben wird, kann dazu genutzt werden, um mögliche verzerrende Effekte, die auf Unterschiede zwischen

---

<sup>12</sup> Die Frageformulierungen sind Anhang C zu entnehmen.

diesen zurückzuführen sind, zu kontrollieren.

Für die CAWI-Variante wird von LimeSurvey automatisch die jeweilige Art des Browsers in der Datenmatrix hinterlegt, mit dem eine Person auf den Fragebogen zugreift. Darüber wird es möglich zu differenzieren, welche Personen den Fragebogen mit stationären oder mobilen Endgeräten ausgefüllt haben.

Zuletzt steht aufgrund der Stichprobenziehung auf Basis des allgemeinen Melderegisters aus dem Jahr 2012 die Postleitzahl der Zielpersonen zur Verfügung. Hierüber kann eine Charakterisierung des jeweiligen Stadtteils stattfinden unabhängig davon, ob diese teilgenommen haben oder nicht. Gebietsdaten aus der Landesstatistik können so dem Datensatz hinzugefügt und zu Analysezwecken weiter verwendet werden.

## 6.2 Statistische Analyseverfahren - Chronologie und Methoden

Zur Beantwortung der Forschungsfragen sowie empirischen Überprüfung der Forschungshypothesen kommen verschiedene Auswertungsverfahren zum Einsatz. Eine Übersicht über die jeweiligen Zielgrößen und Analyseverfahren ist in Tabelle 6.7 abgebildet. Zu Anfang werden die Ergebnisse hinsichtlich des in der 2012er Befragung eingesetzten Mixed-Mode Designs berichtet. Hierzu erfolgt ein Blick auf die deskriptive Verteilung der Befragten zu den einzelnen Umfragemodi in Form von relativen Häufigkeiten. Alsdann wird mittels eines logistischen Regressionsmodells unter Kontrolle relevanter Hintergrundvariablen überprüft, ob mögliche Zusammenhänge nicht nur in der Stichprobe zu beobachten sind, sondern ebenso auf die Grundgesamtheit zu übertragen sind.<sup>13</sup>

Des Weiteren werden Personen charakterisiert, die nicht nur ein Smartphone besitzen, sondern dieses ebenso regelmäßig nutzen. Es ist zu erwarten, dass erst ab einer bestimmten Regelmäßigkeit der Nutzung davon auszugehen ist, dass Personen hierüber einen Fragebogen mobil beantworten. Das passende Auswertungsverfahren ist auch hier die logistische Regression, da die Teilnehmer bei der Frage nach der Smartphone-Nutzung die dichotomen Antwortkategorien Ja/ Nein zur Auswahl hatten. Ob die Befragten außerdem dazu bereit wären, an einer mobilen Onlineumfrage teilzunehmen, wurde mit drei inhaltlichen Kategori-

---

<sup>13</sup> Für eine Übersicht des zugrundeliegenden Schätzverfahrens und der damit einhergehenden mathematischen Grundlagen der logistischen Regression, siehe Backhaus et al. (2015); Long und Freese (2006); Windzio (2013).

Tabelle 6.7: Zielgrößen und Analyseverfahren

Designaspekt	Zielgröße	Analyseverfahren
B - Mixed-Mode	Wahl des Umfragemodi	logistische Regression
B - Mixed-Mode	Smartphonennutzung	logistische Regression
B - Mixed-Mode	Teilnahme an mobiler Onlinebefragung	logistische Regression
D - Anreize	Wirkung von Teilnahmeanreizen auf die Antwortrate und Zusammensetzung der Teilnehmer	Als Prädiktor in den multivariaten Modellen zur Schätzung von $p_e$ und $p_k$
D - Anreize	Subjektive Wahrnehmung der Anreizhöhe und Grenznutzen	Qualitative Inhaltsanalyse Mittelwertvergleiche (T-Test)
A - Registerstichprobe und Paradata C - postalische Vorankündigung	Erreichbarkeit ( $p_e$ ) + Kooperationswahrscheinlichkeit ( $p_k$ ) = Teilnahmewahrscheinlichkeit ( $p$ )	logistische Regression Weiterverwendung als Propensity Scores zur Überführung in Random Response Model

en abgefragt. Diese Kategorien sind für die multivariate Analyse zusammengefasst worden, sodass die abhängige Variable dichotom ausgeprägt ist und eine logistische Regression gerechnet werden kann. Dieses Vorgehen wird dadurch notwendig, da Mehrfachantworten auf die Frage nach der Bereitschaft, eine Umfrage auch mobil zu beantworten, möglich waren. Die beiden Kategorien (Teilnahme über einen Link/ Teilnahme über eine App), die eine Bereitschaft zur Teilnahme operationalisieren, wurden entsprechend zusammengefasst.

Ein weiterer zentraler Bestandteil dieser Arbeit besteht nicht ausschließlich in der Analyse der Wirkung vergebener Teilnahmeanreize auf die Antwortrate, denn dieser Aspekt ist bereits oft untersucht und die Forschungslage entsprechend breit aufgestellt. Im Fokus des Interesses steht zudem die Frage, ob sich Anreize in Abhängigkeit der jeweiligen Form (monetär/ nicht monetär - für die unterschiedlichen Formen siehe Tabelle 4.2 in Kapitel 4.2 (Teilnahmeanreize)) auf die Zusammensetzung der Stichprobe derart auswirken, dass sich der Bildungsbias reduzieren lässt. Weiterhin werden die relevanten Variablen, welche die einzelnen Anreizdimensionen abbilden, mittels eines logistischen Regressionsmodells in die Schätzgleichung zur Berechnung von Propensity Scores einbezogen. Zur weiteren Analyse möglicher verzerrender Effekte werden diese in das RRM (Random Response Model) von Bethlehem (2002) eingebunden.

Ein weiterer Aspekt, bezüglich dessen die Forschungslage und deren Evidenzen uneinheitlich ausfallen, ist die Frage nach einem möglichen Grenznutzen von monetären nicht konditionalen Anreizen. Wie bereits in der Diskussion des Forschungsstandes beschrieben, lässt sich bisher kein eindeutiger Zusammenhang zwischen der Höhe des Anreizes

und seiner Wirkung auf die Antwortrate feststellen, insbesondere in Bereichen ab zehn beziehungsweise zwanzig Euro. Aufgrund der siebenstufigen Skala und dem metrischen Skalenniveaus bezüglich der Frage, ab wann die Befragten einen Betrag als kleine Aufmerksamkeit beziehungsweise Bezahlung wahrnehmen, wird ein Mittelwertvergleich (T-Test) zwischen den einzelnen Zufallsvarianten des Experimentes durchgeführt. Von Interesse ist dabei, ob sich der Mittelwert der Antworten in Richtung des Kontinuums „Bezahlung“ bewegt, je höher die zufällig gezeigten Beträge ausfallen.

Um differenzierter analysieren zu können und den Befragten die Möglichkeit zu eröffnen, einen Betrag unabhängig von dem im vorhergehenden Experiment gezeigten anzugeben, den sie ihrer Meinung nach als kleine Aufmerksamkeit oder Bezahlung empfinden, wurde darauffolgend mit zwei offenen Antwortfeldern nach expliziten Beträgen gefragt. Zur Analyse der Antworten und der Feststellung eines möglichen Grenzwertes kommt die qualitative Inhaltsanalyse zur Anwendung. Dabei erfolgt im betreffenden Abschnitt des Ergebnisteils zunächst einmal eine Beschreibung des entwickelten Kategoriensystems sowie eine Auswertung der relativen Häufigkeiten.

Im letzten Abschnitt des Ergebnisteils werden die Modelle zur Schätzung von der Teilnahmewahrscheinlichkeit ( $p$ ) vorgestellt. Zur Beantwortung der dahinterstehenden Forschungsfragen werden logistische Regressionsverfahren verwendet. So ist es möglich, Modelle für die relevanten Zielgrößen mit dichotomer Ausprägung (0\1) zu schätzen. Die verschiedenen Prädiktoren (unabhängige Variablen) werden schrittweise in die jeweilige Schätzgleichung der statistischen Modelle aufgenommen. Für jede Person in der Stichprobe kann daraufhin ein Anteilswert geschätzt werden, der die konkrete Wahrscheinlichkeit beinhaltet, dass diese Person in Abhängigkeit der einbezogenen Prädiktoren in der Zielvariable den Wert eins annimmt (0 = Nonresponse/ 1 = Response). Diese Wahrscheinlichkeitsbeträge werden im nächsten Schritt weiter dazu verwendet, um mit dem RRM von Bethlehem (2002) zu untersuchen, ob diese mit der Beantwortung der gestellten Fragen zusammenhängen. Ist dies der Fall und die Hintergrundvariablen erfüllen die notwendigen Kriterien, kann davon ausgegangen werden, dass sich vorliegender selektiver Nonresponse derart auswirkt, dass die Antworten verzerrt sind und eine entsprechende Gewichtung vorgenommen werden sollte.

# Kapitel 7

## Ergebnisse

Der Ergebnisteil dieser Arbeit gliedert sich in vier Unterkapitel. Zunächst werden die Ergebnisse bezüglich des Mixed-Mode Designs aus der Nachrekrutierung des Jahres 2012 beschrieben. Hierbei geht es um die erzielten Antwortraten, die Verteilung der Moduspräferenzen sowie der Frage danach, ob Personen ein Smartphone regelmäßig nutzen und darüber hinaus ebenso dazu bereit wären, hierüber einen Fragebogen auszufüllen.

Abschnitt 7.2 beschäftigt sich mit der Wirkung von Teilnahmeanreizen. Hier ist nicht nur der schon häufig untersuchte Effekt auf die Antwortrate von Interesse, sondern ebenso ein möglicher Einfluss der vergebenen Anreize auf die Zusammensetzung der Teilnehmer. Besonders interessant ist hierbei, ob der Befund von Vehre et al. (2013) zu replizieren ist, der besagt, dass die in wissenschaftlichen Umfragen unterrepräsentierten Personen mit einem niedrigen Bildungsniveau durch einen Anreiz eher zu einer Teilnahme bereit sind. Ferner wird dieser Abschnitt eine Antwort auf die Frage liefern, ob es einen Grenzwert beim Einsatz monetärer Anreize gibt und ab welchem Betrag dieser von den Befragten nicht mehr als „kleine Aufmerksamkeit“ sondern als Bezahlung wahrgenommen wird. Um diese Frage zu beantworten, werden die Ergebnisse der qualitativen Inhaltsanalyse berichtet.

Abschließend werden die Ergebnisse der multivariaten Analysen beschrieben, mit der die jeweiligen Teilnahmewahrscheinlichkeiten in Abhängigkeit der unterschiedlichen Designbedingungen der Nachrekrutierung aus dem Jahr 2012 geschätzt werden. Zudem wird im Rahmen des RRM's überprüft, inwiefern diese unterschiedlichen Elemente einen Einfluss auf das Antwortverhalten haben und in ausgewählten Umfragevariablen ein möglicher Bias entstanden ist.

## 7.1 Mixed-Mode

In der Befragung zur Nachrekrutierung aus dem Jahr 2012 ist ein „Concurrent Mixed-Mode Design“ (De Leeuw et al., 2008, 305) zur Anwendung gekommen. Die Befragten hatten in den unterschiedlichen Untersuchungsgruppen jeweils die Möglichkeit sich entweder für die CAWI- oder die CATI-Variante des Fragebogens zu entscheiden. Der Mixed-Mode Charakter dieser Umfrage ist hierbei doppelt gelagert, da der erste Kontakt und die Einladung zur Umfrage per Post erfolgt sind, der Fragebogen selbst dann in zwei unterschiedlichen Modi beantwortet werden konnte. Da die Befragten nicht zufällig zu einem Umfragemodus zugewiesen worden sind, ist es bei diesem Vorgehen im Nachhinein nicht möglich, zwischen Verzerrungen aufgrund von Selektions- und Moduseffekten zu differenzieren, denn diese sind konfundiert. Mit einer Befragung ausschließlich in einem der beiden Modi wäre mit nicht zu unterschätzenden Verzerrungen aufgrund von Coverage-Effekten zu rechnen gewesen. Aus diesem Grund wurde dieses Vorgehen in seiner praktischen Ausgestaltung gewählt. Denn einerseits würde die Datenqualität dadurch beeinflusst werden, dass der Anteil an Personen, deren Rufnummer in Deutschland über das Telefonbuch mit Mühe identifizierbar ist, bei lediglich circa 55 Prozent liegt und sich darin eingetragene Personen von darin nicht eingetragenen systematisch unterscheiden (Schnell, 2012, 268). Andererseits leiden CAWI-Befragungen unter ähnlichen Coverage-Effekten in Bezug zu Bevölkerungsgruppen, die hierüber nach wie vor nicht oder nur sehr schlecht zu erreichen sind. Ferner wurde aufgrund der Praktikabilität seinerzeit die Entscheidung getroffen, nicht zweistufig zu rekrutieren, sodass die Befragten zufällig zu den Umfragemodi zugeordnet werden können. Denn sollen Selektions- und Moduseffekte sauber statistisch voneinander getrennt analysiert werden, muss zwangsläufig so wie in der Aufbauphase des PPSM Access Panels rekrutiert werden: Im Rahmen des Erstkontaktes müssen die unterschiedlichen Modi abgefragt werden, in denen der Befragte potentiell bereit wäre teilzunehmen, um darauffolgend zufällig allokkieren zu können. Es wurde somit nicht vorgesehen, diese Effekte voneinander zu trennen (Engel und Schmidt, 2015).

### 7.1.1 Antwortraten und Moduspräferenzen

In der Nachrekrutierung aus dem Jahr 2012 konnte eine übergeordnete Antwortrate von 12 Prozent erzielt werden. Bei lokaler Betrachtung der einzelnen in Tabelle 7.1 abgebildeten Untersuchungsgruppen zeigt sich, dass die Antwortraten in den einzelnen

Tabelle 7.1: Antwortraten und Moduspräferenzen der einzelnen Teilgruppen in Prozent  
- Zielgröße: Wahl des Befragungsmodus

Designbedingung	Antwortrate	Präferenz CATI	Präferenz CAWI
A - Vorausbrief - Ankündigung Anruf	36,4	61,0	39,0
B - Vorausbrief - Rückantwortschreiben	9,5	25,0	75,0
C - Vorausbrief - Rückantwortschreiben	6,1	13,5	86,5
D - 5 Euro	49,5		
E - Kontrollgruppe	22,2		
F - Verlosung	8,9		
G - Kontrollgruppe	10,1		
H - Verlosung	5,6		
I - Kontrollgruppe	6,6		

Quelle: Engel und Schmidt (2015, 269); PPSM Access Panel Nachrekrutierung 2012, eigene Berechnungen.

Gruppen teilweise deutlich höher ausgefallen sind. Wurde in Untersuchungsgruppe *A* ein Vorausbrief in Kombination mit einem Telefonanruf eingesetzt, konnte mit diesem Vorgehen eine Antwortrate von 36,4 Prozent erzielt werden. Nachvollziehbar erscheint dieses Ergebnis vor dem Hintergrund, dass der Anruf nicht kalt erfolgt und in seiner Wirkung einem Reminder gleicht. Denn legen die potentiellen Teilnehmer den Vorausbrief erst einmal zur Seite, werden sie durch den Anruf wieder an die Befragung und die erbetene Teilnahme erinnert.

Die Antwortraten in den Gruppen *B* und *C* fallen deutlich geringer aus. Mit 9,5 und 6,1 Prozent haben sich weniger Personen für eine der angebotenen Umfragevarianten rekrutieren lassen. In diesen beiden Untersuchungsgruppen musste man für eine Teilnahme selbst aktiv werden. Entweder war dem Link zur CAWI-Umfrage zu folgen oder der Rückantwortbrief mit der jeweiligen Rufnummer zurückzusenden. Diese beiden Aspekte und der fehlende Telefonanruf haben offensichtlich dazu geführt, dass sich die Antwortraten der Gruppen *A*, *B* und *C* deutlich voneinander unterscheiden. Ein klares Bild zeigt sich ebenso mit Blick auf den Effekt der vergebenen Teilnahmeanreize. Während der nicht konditionale Einsatz des monetären Anreizes in Gruppe *D* die Antwortrate auf ein Niveau von 49,5 Prozent gehoben hat, war der Hinweis auf die Verlosung sogar kontraproduktiv. Die Antwortraten liegen hierbei in den Experimentalgruppen (*F/H*) unter denen der jeweiligen Kontrollgruppen (*G/I*). Dies kann einerseits in der Tatsache begründet liegen, dass der Hinweis auf der zweiten Seite des Anschreibens im Postskriptum überlesen wurde. Andererseits kann eine Verlosung auf potentielle Teilnehmer unseriös wirken oder die subjektiv wahrgenommene Wahrscheinlichkeit auf einen Gewinn als zu niedrig erscheinen.

Die Ergebnisse können darüber hinaus die Annahme stützen, dass sich die Modusprä-

ferenzen zwischen den Gruppen unterscheiden. Wurde in Gruppe *A* ein Anruf per Einladungsschreiben angekündigt, haben sich in der Folge mehr Personen für die CATI-Variante entschieden. Nachvollziehbar erscheint dieses Ergebnis insbesondere vor dem Hintergrund, dass man auf den Telefonanruf sozusagen lediglich warten muss. Mit der Teilnahme an der CAWI-Variante ist ein höherer Aufwand derart verbunden gewesen, dass die Teilnehmer den Link zur Webseite und ihr Zugangspasswort eingeben mussten. Dies stellt eine zusätzliche Hürde zur Teilnahme dar und erklärt die Unterschiede in dieser Gruppe.

Ähnlich verhalten sich die Moduspräferenzen in den Gruppen *B* und *C*. In diesen haben sich wiederum deutlich mehr Personen für die CAWI-Variante entschieden. Hier wurde mit dem Vorausbrief kein Anruf angekündigt, auch wenn die Rufnummer in Untersuchungsgruppe *B* bekannt war. Dieses Vorgehen wurde gewählt, um Personen mit und ohne Telefonbucheintrag, beispielsweise hinsichtlich der Teilnahmebereitschaft, miteinander vergleichen zu können. Erklärt werden können die Unterschiede hinsichtlich der Präferenzen dadurch, dass in diesen beiden Gruppen die Teilnahme an der CATI-Variante mit einem deutlich höheren Aufwand verknüpft gewesen ist. Denn dem Einladungsschreiben lag eine Karte sowie ein vorfrankierter Briefumschlag bei, den die potentiellen Teilnehmer dazu nutzen konnten, ihre Telefonnummer mitzuteilen. Diesen musste man ausfüllen und in den Briefkasten werfen, während man für die CAWI-Variante lediglich URL sowie Zugangspasswort eingeben musste.

### **Zusammenfassung der Ergebnisse**

Durch die Überprüfung der Hypothesen bezüglich der Präferenz für einen der angebotenen Teilnahmemodi CATI oder CAWI können die Hypothesen  $H_{C2}$  sowie  $H_{C3}$  angenommen werden. Tabelle 7.2 gibt eine Übersicht über die hierzu formulierten Hypothesen, die vermutete Richtung des Zusammenhangs sowie die Entscheidung darüber, ob diese durch die Ergebnisse angenommen werden können oder abgelehnt werden müssen.

Wurde mit dem Anschreiben ein Anruf angekündigt, hat sich die Mehrzahl an Personen für die Teilnahme an der CATI-Befragung entschieden (Hypothese  $H_{C2}$ ). Wurde durch das Anschreiben kein Anruf angekündigt hat sich erwartungsgemäß gezeigt, dass sich ein größerer Teil der Befragten für die CAWI-Befragung entschieden. Offen bleiben muss an dieser Stelle, ob der Zusammenhang auch für die Grundgesamtheit gelten kann, da die Ergebnisse in diesem Kapitel deskriptiv sind. Für die Stichprobe können die Hypothesen  $H_{C2}$  und  $H_{C3}$  angenommen werden.  $H_{A7}$  sowie  $H_{A8}$  werden in Kapitel 7.3.1 (Analyse der



Tabelle 7.2: Entscheidung über die Hypothesen - Zielgröße: Wahl des Befragungsmodus

Hypothese	Vermutete Richtung des Zusammenhangs	Status
$H_{C2}$ : Anruf angekündigt $\Rightarrow$ Entscheidung CATI	positiv	angenommen
$H_{C3}$ : Kein Anruf angekündigt $\Rightarrow$ Entscheidung CAWI	positiv	angenommen
$H_{A7}$ : Alter $\Rightarrow$ Entscheidung CATI	positiv	wird in Kapitel 7.3.1 überprüft
$H_{A8}$ : Geschlecht (männlich) $\Rightarrow$ Entscheidung CAWI	positiv	wird in Kapitel 7.3.1 überprüft

Bruttostichprobe mit verfügbaren Hintergrundvariablen) einer empirischen Überprüfung unterzogen.

Bei dem Einsatz unterschiedlicher Befragungsmodi innerhalb eines Mixed-Mode Designs hat sich in den letzten Jahren abgezeichnet, dass immer mehr Teilnehmer einen Fragebogen auch mit mobilen Endgeräten wie Smartphones oder Tablets beantworten (Fuchs und Busse, 2009; De Bruijne und Wijnant, 2014; Lugtig und Toepoel, 2016; Rivella et al., 2014). Der nachfolgende Abschnitt dieser Arbeit beschäftigt sich mit den Fragen, inwiefern sich Personen, die ein Smartphone besitzen und dies auch regelmäßig nutzen, von der Gruppe unterscheiden, die dieses nicht nutzen. Denn hierin liegt aktuell, ähnlich wie beim Vorhandensein eines stationären Computers und eines Internetanschlusses, ein großes Fehlerpotential dieser Befragungsform. Darüber hinaus wird untersucht, ob die Nutzer von Smartphones dazu bereit wären, einen Fragebogen über das mobile Internet zu beantworten.

### 7.1.2 Zur Bereitschaft eine Umfrage via Smartphone zu beantworten

In der Befragungswelle des PPSM Access Panels aus dem Jahr 2014 wurde nicht nur danach gefragt, ob die Teilnehmer ein Smartphone besitzen und dieses auch regelmäßig nutzen. Ferner wurden diejenigen, die hierauf mit „Ja“ geantwortet haben, in der darauffolgenden Frage danach gefragt, ob sie ihr Smartphone ebenso zur Beantwortung einer Umfrage in unterschiedlicher Form (beispielsweise über einen Link oder eine App) nutzen würden. Streng genommen ist die mobile Onlinebefragung für allgemeine Bevölkerungsumfragen aufgrund der bereits diskutierten Problematiken derzeit nicht geeignet. Denn ähnlich wie bei der CAWI-Befragung werden hierbei große Teile der Bevölkerung im Vorhinein systematisch ausgeschlossen, da diese kein mobiles Endgerät besitzen oder dieses nicht

in der Regelmäßigkeit nutzen, dass eine Teilnahme wahrscheinlich ist (Fuchs und Busse, 2009).

Vor diesem Hintergrund ist es von zentraler Bedeutung, den Trend der Smartphonennutzung weiter zu verfolgen, um das Potential und deren Entwicklung für die wissenschaftliche Umfrageforschung besser einschätzen zu können, denn die Verbreitung dieser sowie der mobilen Internetnutzung wird auch in Zukunft sukzessive zunehmen (siehe Tabelle 2.2 sowie ARD/ZDF-Onlinestudie (2015)). Dieser allgemeine Trend wird sich erst über die nächsten Jahre und Jahrzehnte weiterentwickeln müssen, sodass das Potential für Coverage-Effekte durch diese Entwicklung kleiner werden kann.

So ist es heute schon zwingend notwendig, einen Onlinefragebogen für die Darstellung auf mobilen Endgeräten zu optimieren. Tut man dies als Forscher nicht, ist mit unkontrollierbaren Messeffekten durch das Erhebungsinstrument selbst zu rechnen, da die Fragebögen nicht der Displaygröße entsprechend dargestellt werden. Letztlich bietet der Trend zum Besitz von Smartphones und der mobilen Internetnutzung bei der Anwendung eines Mixed-Mode Designs ein nicht zu unterschätzendes Potential für die wissenschaftliche Umfrageforschung. Denn wie eine Reihe vorliegender aktueller Studien bereits zeigen konnte, beantwortet ein Teil der Teilnehmer einen Onlinefragebogen bevorzugt auch über ein mobiles Endgerät (Fuchs und Busse, 2009; De Bruijne und Wijnant, 2014; Lugtig und Toepoel, 2016; Rivella et al., 2014). Ist man als Forscher in der Lage diesen Modus zusätzlich anzubieten, kann hiervon bereits heute eine Erhöhung der Antwortrate ausgehen. Denn diese Personen würden den Fragebogen vielleicht gar nicht ausfüllen, hätten sie nicht die mobile Option.

### **Verbreitung und Nutzung von Smartphones im PPSM Access Panel**

Auf Basis der Befragungswelle des PPSM Access Panels aus dem Jahr 2014 wurde ein logistisches Regressionsmodell mit der dichotomen abhängigen Variable „Smartphonennutzung“ gerechnet. In Tabelle 7.3 sind die entsprechenden Ergebnisse abgebildet.

Dieses und folgende multivariate Analysemodelle sind „genestet/geschachtelt“ aufgebaut. Die unabhängigen Variablen wurden schrittweise über die Modelle hinweg aufgenommen, um mittels des Likelihood-Ratio-Tests überprüfen zu können, inwiefern diese einen signifikant von null verschiedenen Erklärungsbeitrag für die Varianz der abhängigen Variable leisten. Als Koeffizienten sind Odd Ratios angegeben, die eine Veränderung im Chancenverhältnis zwischen den beiden Zuständen der abhängigen Variable um den

Tabelle 7.3: Logistische Regression mit der Smartphonennutzung (0 = Nein; 1 = Ja ) als abhängige Variable

	M1		M2		M3		M4		M5	
	OR	s.e.	OR	s.e.	OR	s.e.	OR	s.e.	OR	s.e.
Geschlecht (Ref. Weiblich)	2,09 ***	0,44	2,01 ***	0,44	2,37 ***	0,59	2,47 ***	0,61	2,55 ***	0,64
<i>Bildung</i> (Ref. Volks-/Hauptschula.)										
Mittlere R./Polyt. O.			1,49	0,55	1,14	0,49	0,99	0,43	0,99	0,43
Fachhochschulreife			2,76 *	1,25	1,46	0,76	1,29	0,68	1,32	0,69
Abitur/erw. Oberschule			3,32 ***	1,14	2,51 *	1,01	2,19 *	0,89	2,12 +	0,87
Alter in Jahren					0,96 ***	0,01	1,08	0,74	1,10	0,08
Alter in Jahren <sup>2</sup>							0,99 *		0,99 *	0,00
<i>Erwerbstätigkeit</i> (Ref. arbeitslos/-suchend)										
Erwerbstätige									0,48	0,27
Schüler/Studierende									1,98	1,71
Rentner									1,00	0,32
<i>Fallzahl (N)</i>	374		374		374		374		374	
<i>Modellgüte</i>										
Prob > chi <sup>2</sup>	***		***		***		***		***	
Pseudo R <sup>2</sup>	0,02		0,06		0,21		0,22		0,23	
LL	-252,44		-243,24		-203,89		-201,24		-201,24	
LR-Test			***		***		*			

\*  $p \leq 0,05$ , \*\*  $p \leq 0,01$ , \*\*\*  $p \leq 0,001$ , +  $p \leq 0,10$

Referenzkategorien: 1 Weiblich, 2 Volks-/Hauptschulabschluss, 3 arbeitslos/arbeitssuchend

OR = Odd Ratio

PPSM Access Panelwelle 2014, eigene Berechnungen

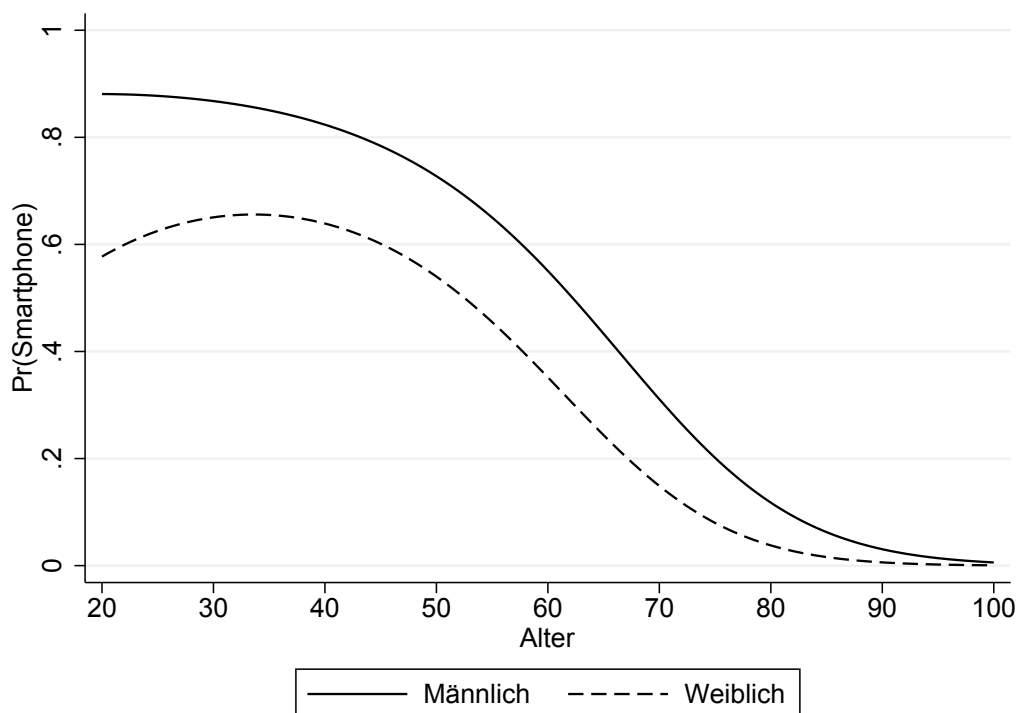
jeweiligen Faktor des Betrages angeben.

Der Blick auf Modell 1 zeigt, dass sich Männer und Frauen hinsichtlich der Smartphonennutzung signifikant voneinander unterscheiden. So haben Männer im Vergleich zu Frauen eine um 109 Prozent höhere geschätzte Wahrscheinlichkeit, ein Smartphone zu nutzen. Dieser Effekt ist zudem höchst signifikant und erklärt zwei Prozent der Varianz der abhängigen Variable. Dieser zeigt sich darüber hinaus als sehr robust. Der Betrag sowie das Signifikanzniveau verändern sich über die Modelle hinweg nur leicht. Auch unter Kontrolle weiterer Dimensionen bleibt der Effekt des Geschlechts bestehen.

Diese Niveauunterschiede zwischen Männern und Frauen spiegeln sich ebenso mit Blick auf Abbildung 7.1 wieder. Hier sind die geschätzten Wahrscheinlichkeiten der Smartphonennutzung differenziert nach Geschlecht und Alter für alle Personen in der Stichprobe abgebildet. Dabei zeigt sich, dass sich das Niveau der Verteilung hinsichtlich des Geschlechts derart unterscheidet, dass die geschätzten Wahrscheinlichkeiten für Männer im Vergleich zu Frauen im jeweils gleichen Alter durchweg höher ausfallen.

Modell 2 bezieht zusätzlich den Effekt der Bildung mit ein. Hier lässt sich im Allgemeinen erkennen, dass mit einem höheren Bildungsniveau einer Person auch höhere geschätzte

Abbildung 7.1: Geschätzte Wahrscheinlichkeiten der Zielgröße Smartphonennutzung differenziert nach Geschlecht und Alter



N = 374; PPSM Access Panelwelle 2014, eigene Berechnungen.

Wahrscheinlichkeiten für die Nutzung eines Smartphones im Vergleich zu Personen mit Volks- oder Hauptschulabschluss einhergehen. Der Effekt der mittleren Reife zeigt sich dabei als nicht signifikant.

Personen mit Fachhochschulreife besitzen im Vergleich zu denjenigen mit Volks- oder Hauptschulabschluss eine um 176 Prozent höhere Wahrscheinlichkeit, während solche mit Abitur oder einem Abschluss der erweiterten Oberschule gar eine um 232 Prozent höhere geschätzte Wahrscheinlichkeit haben, ein Smartphone zu besitzen und dies auch regelmäßig zu nutzen. Unter der Hinzunahme weiterer erklärender Prädiktoren zeigen sich die Effekte der Bildungs-Dummies allerdings als nicht robust. Die geschätzten Beträge der Koeffizienten werden einerseits unter Hinzunahme des Alters in Modell 3 nicht nur kleiner, sondern die Signifikanzen verändern sich andererseits teilweise deutlich. Hier scheint also insbesondere das Alter einen höchst signifikanten negativen Effekt zu haben, unter dessen Kontrolle lediglich noch die höchste Bildungskategorie einen positiven und signifikanten Effekt hat. Betrachtet man darüber hinaus die Modellgüte von Modell 3 wird deutlich, dass mit diesem gegenüber Modell 2 eine höchst signifikant gesteigerte Erklärungsleistung einhergeht. 21 Prozent der Varianz der abhängigen Variable können hiermit erklärt werden. Die Variable

Alter leistet einen Beitrag von 15 Prozent. Das Alter ist vor diesem Hintergrund einer der zentralen Faktoren, der die Verbreitung und Nutzung von Smartphones beeinflusst.

Aufgrund der Ausstattung von Haushalten mit Mobiltelefonen und der damit bedingten Nutzung von mobilem Internet ist zu vermuten, dass der Effekt des Alters auf das Nutzungsverhalten nicht linear verläuft. Um dies zu überprüfen enthält die Schätzgleichung von Modell 4 eine zusätzliche quadratische Effektkomponente. Der Blick auf Tabelle 7.3 offenbart, dass diese Vermutung durch die zugrundeliegenden Daten bestätigt werden kann. Lediglich der quadratische Effekt ist signifikant, was in Verbindung mit dem negativen Koeffizienten bedeutet, dass die Wahrscheinlichkeit der Smartphonennutzung mit steigendem Alter kurvenförmig abnimmt.<sup>1</sup> An Abbildung 7.1 lässt sich der Zusammenhang zwischen dem Alter und der Smartphonennutzung sehr gut erkennen. In der Gruppe der 20 bis 40-Jährigen bewegen sich die geschätzten Wahrscheinlichkeiten auf einem hohen Niveau von circa 80 bis 90 Prozent bei den Männern sowie circa 60 bis 80 Prozent bei den Frauen. Danach fallen diese steil und kurvenförmig ab, sodass die Wahrscheinlichkeiten bei den über 60-Jährigen bereits unter 50 Prozent liegen und mit zunehmendem Alter weiter abfallen. An dieser Stelle kann festgehalten werden, dass sich das Geschlecht in Verbindung mit dem Alter der Teilnehmer erwartungsgemäß als erklärungsrelevant und robust über die Modelle hinweg gezeigt haben.

Das letzte Modell 5 enthält zusätzlich die Dimension der Erwerbstätigkeit zur Kontrolle. Denn im aktuellen Bericht des Statistisches Bundesamt (2016) zur Nutzung des mobilen Internets konnten hier teilweise deutliche Unterschiede nachgewiesen werden. Dem liegt die logische Annahme zugrunde, dass sich Smartphonennutzung sowie die des mobilen Internets gegenseitig positiv bedingen. Nach dem Statistisches Bundesamt (2016) sind es insbesondere die Arbeitslosen, die das Internet seltener mobil nutzen. Der Blick in Tabelle 7.3 zeigt allerdings, dass mit den hier zugrundeliegenden Daten nicht davon ausgegangen werden kann, dass der Erwerbsstatus den Besitz und die Nutzung eines Smartphones bedingt. Modell 5 liefert keinen zusätzlichen Erklärungsbeitrag zur Varianz der abhängigen Variable und die Koeffizienten der einzelnen Erwerbs-Dummies sind allesamt nicht signifikant.

Als Zwischenfazit kann an dieser Stelle festgehalten werden, dass es vor allem die jüngeren Männer mit Abitur beziehungsweise erweitertem Oberschulabschluss sind, die Smartphones am häufigsten besitzen und auch regelmäßig nutzen. Das Alter hat einen

---

<sup>1</sup> In die Analyse wurde ebenso eine kubische Effektkomponente mit in die Schätzgleichung einbezogen, um zusätzlich überprüfen zu können, ob sich die Steigung über das Alter hinweg mehr als einmal verändert. Diese hat sich als nicht signifikant gezeigt und wurde somit nicht weiter berücksichtigt.

negativen Effekt auf die Wahrscheinlichkeit der Nutzung. Zudem hat sich das Geschlecht als erklärungsrelevant und robust über die einzelnen Modelle hinweg gezeigt.

### Zusammenfassung der Ergebnisse

Durch die Analyse zur Smartphonennutzung konnten die aufgestellten Hypothesen bestätigt werden. Tabelle 7.4 fasst die Ergebnisse zusammen und gibt eine Übersicht darüber, ob die Hypothesen angenommen werden können oder abgelehnt werden müssen.

Tabelle 7.4: Entscheidung über die Hypothesen - Zielgröße: Smartphonennutzung

Hypothese	Vermutete Richtung des Zusammenhangs	Status
$H_{S1}$ : Bildung $\Rightarrow$ Nutzung Smartphone	positiv	angelehnt
$H_{S2}$ : Alter $\Rightarrow$ Nutzung Smartphone	negativ	angenommen
$H_{S3}$ : Geschlecht (männlich) $\Rightarrow$ Nutzung Smartphone	positiv	angenommen

Der Prädiktor Bildung hat sich als erklärungsrelevant gezeigt. Anzumerken ist, dass lediglich der Effekt der höchsten Kategorie „Abitur/erweiterte Oberschule“ in Modell 4 signifikant ist. Die Ergebnisse stützen die erwartete Richtung des Zusammenhangs in  $H_{S1}$ . Diese können trotzdem nur bedingt gelten, da den anderen Bildungskategorien kein signifikanter Effekt auf die Nutzung von Smartphones nachzuweisen ist. In Modell 5 verliert auch diese ihre Übertragbarkeit auf die Grundgesamtheit. Deswegen muss  $H_{S1}$  zurückgewiesen werden.

Für die Variable Alter sowie Geschlecht konnten hingegen klare Zusammenhänge beobachtet werden. Die Nutzungsfrequenz steigt erwartungsgemäß mit zunehmendem Alter (Hypothese  $H_{S2}$ ). Letztlich bewegen sich die geschätzten Wahrscheinlichkeiten für Männer durchweg auf einem höheren Niveau, im Vergleich zu den Frauen (Hypothese  $H_{S3}$ ).

Daran schließt die Frage an, ob sich dieser Kreis an Personen auch an wissenschaftlichen Umfragen über ihr Smartphone beteiligen würden. Das folgende Kapitel greift dies auf.

### Bereitschaft, eine CAWI-Umfrage mobil zu beantworten

Die wissenschaftliche Umfrageforschung steht trotz zunehmender Verbreitung von Smartphones mit Internetzugang im Bereich der Onlinebefragungen, die auch oder ausschließlich mobil beantwortet werden können, noch relativ am Anfang. Dies liegt einerseits darin begründet, dass die Nutzung des mobilen Internets erst im Jahr 2015 überhaupt eine Quote

von über 50 Prozent (mit dem Smartphone) in Deutschland erreicht hat (ARD/ZDF-Onlinestudie, 2015). Hierbei sind es dann vor allem die Jüngeren, die das Internet mobil nutzen. Es ist allerdings, wie bereits diskutiert, zu erkennen, dass diese Form der Befragung in den letzten Jahren mehr in den Fokus wissenschaftlicher Forschung gerückt ist. Denn es ist davon auszugehen, dass sich der Trend hin zu einer weiteren Verbreitung mobiler Endgeräte in Zukunft weiter fortsetzen wird und diese hierdurch für wissenschaftliche Befragungen immer interessanter wird. Daran anknüpfend konnten vorliegende Studien bereits nachweisen, dass ein nicht zu ignorierender Teil von Personen einen Onlinefragebogen mobil beantwortet (Lugtig und Toepoel, 2016; Toepoel und Lugtig, 2015). Würde man diesen dann nicht für die Darstellung auf mobilen Endgeräten optimieren, entstehen unkontrollierbare Messeffekte durch das Erhebungsinstrument selbst. Vor diesem Hintergrund ist es von besonderem Interesse, die methodischen Besonderheiten weiter zu erforschen, um diese Form der Befragung für die Umfrageforschung nutzbar zu machen.

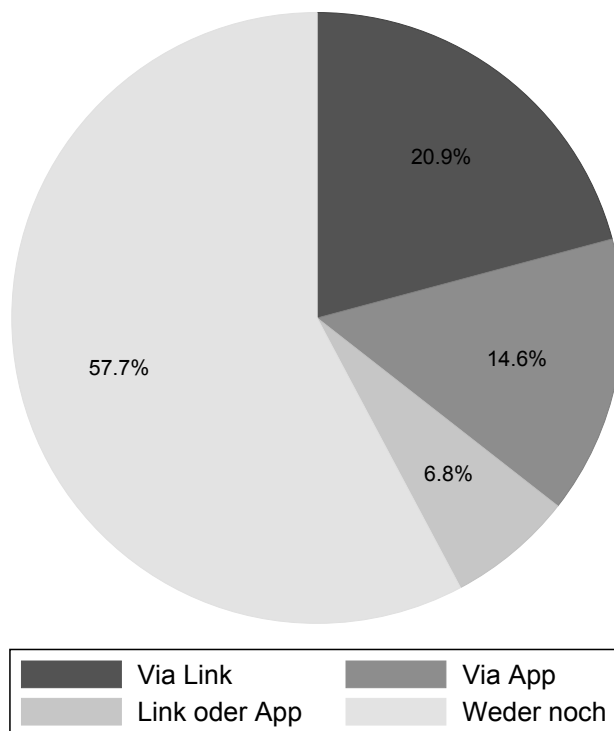
In der Welle des PPSM Access Panels aus dem Jahr 2014 wurde nicht nur danach gefragt, ob die Teilnehmer ein Smartphone besitzen und dies auch regelmäßig nutzen. Vielmehr wurde daraufhin die Bereitschaft erfragt, sich hierüber an einer Umfrage zu beteiligen. Um eine Befragung über ein mobiles Endgerät durchzuführen, hat man in der Praxis als Forscher mehrere Möglichkeiten. Das generelle Problem, dass ähnlich der allgemeinen Befragung über das Internet keine Auswahlgrundlage besteht, liegt hierbei nur bedingt vor. Wählt man die Variante, dass den potentiellen Teilnehmern ein Link zum Fragebogen per Short Message Service (SMS) zugesandt wird, existiert ein Stichprobenziehungsverfahren für die zufällige Generierung von Mobilfunknummern (Häder und Häder, 2009). Aktuell gibt es keine bekannte Studie, die eine Befragung in der Kombination dieser beiden methodischen Vorgehensweisen untersucht hat. Da es in dieser Arbeit lediglich um die Bereitschaft geht teilzunehmen, werden die methodischen Probleme, die mit diesem Vorgehen einhergehen, an dieser Stelle nicht weiter diskutiert.

Ferner können Befragungen über eine Applikation durchgeführt werden, die zunächst aus einer der gängigen Plattformen, wie dem Google-Playstore, heruntergeladen werden muss. Bekannt ist diese Vorgehensweise aus der Marktforschung. Eine bekannte Applikation zur Testung von Bioprodukten ist „biopinio“.<sup>2</sup> Für die wissenschaftliche Umfrageforschung kommt diese Variante eher nicht infrage, da hier keine Stichprobenziehung vorausgehen kann und die potentiellen Befragten sich somit selbst rekrutieren.

---

<sup>2</sup> Für weiterführende Informationen siehe <https://biopinio.de/>

Abbildung 7.2: Relative Häufigkeiten der Bereitschaft eine CAWI Umfrage mobil zu beantworten



N = 206; PPSM Access Panelwelle 2014, eigene Berechnungen.

Um die allgemeine Teilnahmebereitschaft zu operationalisieren, wurden im Fragebogen der Welle aus dem Jahr 2014 trotzdem beide Varianten abgefragt. Abbildung 7.2 stellt die Häufigkeitsverteilung dieser Frage als Anteilswerte dar.

Bei der Betrachtung fällt auf, dass der Anteil an regelmäßigen Smartphone-Nutzern, die weder zur Teilnahme in der einen noch der anderen Variante bereit wären, bei 57,7 Prozent liegt. Hier scheint also über die bereits angesprochenen Probleme hinaus eine noch geringe Akzeptanz vorzuherrschen. Worin dies begründet liegt, kann an dieser Stelle lediglich vermutet werden. Mobile Endgeräte bieten theoretisch die Möglichkeit, ein noch größeres Set an Parainformationen zu erheben. Hier wäre die Aufzeichnung von GPS-Koordinaten zu nennen. Denkbar ist, dass die Teilnehmer solche Aspekte bei der Beantwortung der Frage im Kopf hatten und dies ein Grund für den hohen Anteil derer sein kann, die weder über einen Link noch über eine Applikation partizipieren würden. Ferner kann diese Form der Befragung auch deswegen unter Akzeptanzproblemen leiden, weil sie heutzutage noch nicht weit verbreitet ist.

Mit 20,9 gegenüber 14,6 Prozent wäre außerdem ein größerer Anteil an Befragten zu einer Teilnahme über einen zugesandten Link im Vergleich zur Applikation bereit.



Tabelle 7.5: Logistische Regression mit der Bereitschaft, eine Umfrage mobil zu beantworten (0 = Nein; 1 = Via App/Via Link/Sowohl als auch) als abhängige Variable

	M1		M2		M3	
	OR	s.e.	OR	s.e.	OR	s.e.
Geschlecht (Ref. Weiblich)	2,13 **	0,68	2,16 **	0,69	2,27 **	0,74
<i>Bildung</i> (Ref. Volks-/Hauptschulabschluss)						
Mittlere Reife/Polytechnische Oberschule			1,09	0,68	1,19	0,77
Fachhochschulreife			0,91	0,62	1,06	0,74
Abitur/erweiterte Oberschule			0,79	0,45	0,91	0,53
Alter in Jahren					0,97 *	0,01
<i>Fallzahl (N)</i>	182		182		182	
<i>Modellgüte</i>						
Prob > chi <sup>2</sup>	**				*	
Pseudo R <sup>2</sup>	0,02		0,03		0,05	
LL	-121,03		-120,66		-117,84	
LR-Test					**	

\*  $p \leq 0,05$ , \*\*  $p \leq 0,01$ , \*\*\*  $p \leq 0,001$ , +  $p \leq 0,10$   
PPSM Access Panelwelle 2014, eigene Berechnungen

Nachvollziehbar erscheint dieses Ergebnis, da für die Teilnahme über eine Applikation zusätzlicher Aufwand dadurch entsteht, dass diese im Vorhinein heruntergeladen und installiert werden muss. Im Gegenteil dazu müsste ein Link lediglich geöffnet und der Fragebogen könnte direkt ausgefüllt werden. Interessant wäre es, an dieser Stelle zukünftig anzusetzen und eine Stichprobe mit Mobilfunknummern zu generieren und einen Link zu einem Fragebogen zu versenden, um das tatsächliche Teilnahmeverhalten zu untersuchen. Letztlich liegt der Anteil derer, die in beiden Varianten partizipieren würden, bei 6,8 Prozent und ist somit am geringsten ausgeprägt.

An die Bereitschaft eine Umfrage mobil in unterschiedlicher Weise zu beantworten knüpft die Frage an, wie die beiden Gruppen zu charakterisieren sind. Die hierzu aufgestellten Vermutungen werden im Folgenden mittels multivariater Analyse einer empirischen Überprüfung unterzogen. Tabelle 7.5 bildet die entsprechenden Ergebnisse logistischer Regressionsmodelle ab. Hierbei repräsentiert der Zustand „1“ den Sachverhalt, dass die Befragten unabhängig von der jeweiligen Form dazu bereit wären, eine Umfrage mobil zu beantworten.

Das Geschlecht zeigt sich hierbei, ähnlich wie bei der Nutzung von Smartphones, über alle Modelle hinweg als signifikanter Prädiktor. Wie vermutet sind es mehrheitlich die Männer, die Umfragen auch mobil beantworten würden. Das Bildungsniveau hat entgegen der Erwartung keinen Effekt und führt nicht zu einer verbesserten Erklärungsleistung des Modells. Dies hatte sich bereits bei der Analyse der Smartphonennutzung abgezeichnet,

da diese eine Prämisse für die hier abgefragte hypothetische Bereitschaft ist. Modell 3 überprüft den Effekt des Alters. Dieser ist erwartbar negativ ausgeprägt, was bedeutet, dass die Wahrscheinlichkeit der hypothetischen Teilnahmebereitschaft mit zunehmendem Alter abnimmt. Hierbei hat sich lediglich die lineare Effektkomponente als signifikant erwiesen. Modell 3 erklärt fünf Prozent der Varianz der abhängigen Variable, was als gering zu bezeichnen ist.

### Zusammenfassung der Ergebnisse

Was festgehalten werden kann ist, dass es vor dem Hintergrund der hier vorliegenden Ergebnisse im Speziellen die Männer (Hypothese  $H_{MO3}$ ) sowie die jüngeren Personen (Hypothese  $H_{MO2}$ ) sind, die eine Umfrage mit einem mobilen Endgerät beantworten würden. Dem Bildungsniveau konnte mit der Analyse kein signifikanter Einfluss nachgewiesen werden (Hypothese  $H_{MO1}$ ). Der Trend hin zu einer weiteren Verbreitung mobiler Endgeräte mit Internetzugang wird sich in Zukunft weiter fortsetzen, sodass diese für die wissenschaftliche Umfrageforschung interessanter werden.

Tabelle 7.6: Entscheidung über die Hypothesen - Zielgröße: Teilnahme an mobiler Onlineumfrage

Hypothese	Vermutete Richtung des Zusammenhangs	Status
$H_{MO1}$ : Bildung $\Rightarrow$ Teilnahme	positiv	abgelehnt
$H_{MO2}$ : Alter $\Rightarrow$ Teilnahme	negativ	angenommen
$H_{MO3}$ : Geschlecht (männlich) $\Rightarrow$ Teilnahme	positiv	angenommen

Unsere heutige Gesellschaft befindet sich in einem Trend der Digitalisierung, der sich in fast allen Bereichen bemerkbar macht. In diesem Kontext ist davon auszugehen, dass sich die Befragung über mobile Endgeräte in der Zukunft weiter etablieren wird und die Akzeptanz in der Bevölkerung steigt. Dieser Befragungsmodus mit seinem Potential Umfragen in kürzester Zeit zu beantworten unabhängig davon, wo sich Personen befinden, birgt für die wissenschaftliche Umfrageforschung ganz neue Möglichkeiten. Somit sollte die Forschung in diesem Bereich weiter voran getrieben werden, um dem Digitalisierungstrend sowie steigender Zahlen mobiler Nutzung des Internets gerecht werden zu können und das Potential für wissenschaftliche Befragungen nicht ungenutzt zu lassen.

## 7.2 Wirkung von Teilnahmeanreizen

Die Wirkung unterschiedlicher Anreize auf die Antwortrate ist bereits durch zahlreiche Studien untersucht und belegt worden. Auch in dieser Arbeit konnte im Rahmen von Kapitel 7.1.1 (Antwortraten und Moduspräferenzen) bereits gezeigt werden, dass sich der Einsatz eines nicht konditionalen monetären Anreizes positiv auf die Antwortrate auswirkt.

Deswegen wird der Fokus des folgenden Kapitels auf andere Zielgrößen gerichtet. Zunächst wird untersucht, ob sich die Vergabe von Anreizen auf die Zusammensetzung der Teilnehmer auswirkt und es hierdurch zu möglichen Einflüssen auf die Qualität der Daten kommt. Darüber hinaus besteht ein weiteres Erkenntnisinteresse darin, einen möglichen Grenznutzen monetärer Anreize zu operationalisieren. Dieses Thema wird in Verbindung mit der Nachrekrutierung von Teilnehmern im Jahr 2012 noch einmal im Rahmen von Kapitel 7.3.1 (Zur Wirkung der eingesetzten Anreize und der Art der Kontaktierung) aufgegriffen.

### 7.2.1 Auswirkungen auf die Zusammensetzung der Teilnehmer und potentielle Verzerrungen

Wie bereits in Kapitel 4.2 (Teilnahmeanreize) beschrieben wurde, gibt es bis heute nur wenige Studien, die Selektionseffekte in Verbindung mit dem Einsatz von Anreizen untersuchen. Mann kann aus diesem Grund noch nicht von gesicherten empirischen Evidenzen sprechen. Die Ergebnisse lassen allerdings zumindest die Vermutung zu, dass sich durch den Einsatz nicht konditionaler monetärer Anreize insbesondere die Bevölkerungsgruppen mit niedrigen Bildungsabschlüssen zu einer Teilnahme bewegen lassen (Martin und Winters, 2001; Nederhof, 1983; Petrolia und Bhattacharjee, 2009; Singer und Ye, 2013; Vehre et al., 2013). In diesem Sinne wäre dies auf der einen Seite ein positiver Effekt für die Repräsentativität einer Umfrage, denn Personen mit niedrigen Bildungsabschlüssen sind in der Regel unterrepräsentiert (Engel und Schmidt, 2014; Reuband, 2015). Auf der anderen Seite kann hierdurch eine zusätzliche Quelle der Verzerrung entstehen. Denn wie Engel und Köster (2014) zeigen konnten, investieren Personen, die von einer Teilnahme überzeugt werden müssen, weniger kognitive Mühe in die Beantwortung von Fragen. Daran anknüpfend kann man den Einsatz eines Anreizes als eine Strategie betrachten, Personen zu einer Teilnahme zu bewegen, die sich ohne diesen dagegen entschieden hätten. Eine geringe intrinsische Motivation kann somit in Verbindung mit einem Anreiz zu einer Teilnahme führen, in

Tabelle 7.7: Vergleich der Bildungsabschlüsse in den experimentellen Untersuchungsgruppen mit dem Mikrozensus 2013 in Prozent

	Mikrozensus 2012	Gesamt CAWI + CATI	Kein Anreiz CAWI	Anreiz CATI	5 Euro CAWI	5 Euro CATI	Verlosung CAWI	Verlosung CATI
Haupt- oder Volksschulabschluss	42,5	30,6	11,8	31,2	15,0	40,0	7,6	44,4
Mittlere Reife	22,1	25,2	20,2	29,2	19,0	23,3	15,1	22,2
Fachhochschul- oder Hochschulreife	27,3	32,5	67,7	33,3	66,0	35,0	75,5	33,3
$\Sigma$	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Bias		20,2	72,7	24,4	69,1	11,4	90,1	8,0

PPSM Access Panel Nachrekrutierung 2012, eigene Berechnungen

Anmerkung: Ohne Abschluss/Noch Schüler/Andere Abschlüsse sowie Item-Nonresponse werden in der Tabelle nicht gesondert dargestellt, sind aber Teil der Prozentuierungsbasis.

der Folge allerdings ein höheres Ausmaß an Item-Nonresponse nach sich ziehen und die Datenqualität hierdurch negativ beeinflussen. Die hierzu vorliegenden Studien konnten allerdings keinen Effekt nachweisen, der in diese Richtung geht (Göritz, 2006; Singer und Ye, 2013; Vehre et al., 2013).

Tabelle 7.7 zeigt die Verteilung der Bildungsabschlüsse in der Nachrekrutierung aus dem Jahr 2012. Unterschieden wird nach den experimentellen Anreizgruppen sowie dem Modus der Teilnahme. Um den Grad der Verzerrung einschätzen zu können sind als Referenz die entsprechenden Hochrechnungen des Mikrozensus aus dem Jahr 2012 mit angegeben. In der letzten Zeile von Tabelle 7.7 sind die Abweichungen zu der jeweiligen Kategorie des Mikrozensus angegeben. Damit sich positive und negative Abweichungen nicht gegenseitig aufheben, sind diese jeweils mit positiven Vorzeichen aufsummiert.

Betrachtet man die Gesamtverteilungen in den ersten beiden Spalten so fällt auf, dass Personen mit Haupt- oder Volksschulabschluss unterrepräsentiert sind. Vehre et al. (2013) sind in einer der Vorstudien zum PPSM Access Panel mit einer Befragung in der Stadt Bremen zu einem ähnlichen Ergebnis gekommen. Allerdings sind die Personen mit höheren Bildungsabschlüssen nicht in dem Maße überrepräsentiert, wie es zu erwarten gewesen wäre. Trotzdem stützt dieses Ergebnis zumindest die These, dass Personen mit niedrigen Abschlüssen, im Vergleich zur Bildungsverteilung des Mikrozensus, in wissenschaftlichen Umfragen weniger vertreten sind.

In den einzelnen experimentellen Untersuchungsgruppen, insbesondere in Abhängigkeit des Befragungsmodus, weichen die Verteilungen teilweise sehr deutlich vom Mikrozensus ab. Allgemein kann festgestellt werden, dass in der Gruppe der Personen, die sich an der CAWI-Variante beteiligt haben, die größten Abweichungen vorliegen. Unabhängig davon,

ob ein Anreiz vergeben wurde oder nicht, sind Personen mit Haupt- oder Volksschulabschluss deutlich unterrepräsentiert. Auffallend überrepräsentiert sind dafür Personen mit Fachhochschul- oder Hochschulreife. Womit sind diese Unterschiede zu erklären? Diesbezüglich können an dieser Stelle lediglich hypothetische Vermutungen angestellt werden. Einerseits kann dies ein über das Alter transportierter Effekt sein, denn ältere Teilnehmer verfügen in der Regel über die niedrigeren Bildungsabschlüsse (Stoop et al., 2010) und nutzen das Internet nicht in dem Ausmaße, dass eine Beteiligung an der CAWI-Variante hierdurch wahrscheinlich wird (ARD/ZDF-Onlinestudie, 2015). Andererseits kann ein weiterer Grund hierfür sein, dass die niedriger gebildeten Personen wissenschaftliche Umfragen eher nicht als eine Möglichkeit im Sinne einer Bürgerpflicht ansehen, in deren Rahmen sie ihre Meinung ausdrücken und somit ihren Teil zu Wissenschaft und Forschung beitragen können. Daran anknüpfend stellt sich zwangsläufig die Frage danach, warum die Verteilung in den CATI-Gruppen durchweg einen kleineren Grad der Verzerrung im Vergleich zum Mikrozensus aufweist. Der alleinige Einsatz der CAWI-Variante des Fragebogens in Form eines Single-Mode Designs unabhängig davon, ob ein Anreiz eingesetzt wird oder nicht, hätte eine massive Verzerrung bezüglich des Bildungsniveaus nach sich gezogen.

Die Unterschiede in der Verteilung bezüglich der beiden Umfragemodi können im Design der Rekrutierung begründet liegen. Denn in dem Falle, dass keine Rufnummer zugewiesen werden konnte, haben die Personen in der Stichprobe lediglich ein Einladungsschreiben erhalten. Wie bereits gezeigt werden konnte, hat sich in der Folge der Großteil der Teilnehmer für den CAWI-Fragebogen entschieden. Geht man von der Prämisse aus, dass die intrinsische Motivation zur Teilnahme bei niedriger gebildeten Personen eher geringer ausgeprägt ist, kann man daraus schlussfolgern, dass diese mit dem Einladungsschreiben auch eher nicht zu einer Teilnahme zu bewegen gewesen sind. Konnte eine Rufnummer eruiert werden, wirkt dem die Tatsache entgegen, dass die Personen angerufen werden. Der Anruf hat eine Funktion wie eine Art Reminder und die Interviewer sind darauf geschult, zögerliche Personen mit den passenden Argumenten von einer Teilnahme zu überzeugen. In Verbindung mit einem Anreiz hat dies offensichtlich dazu geführt, dass insbesondere der Anteil an niedriger gebildeten Personen erhöht werden konnte. Differenziert danach, ob die Rufnummer bekannt war oder nicht, liegen daran anknüpfend bezüglich des Bildungsniveaus systematische Unterschiede vor. Diese Tatsache unterstützt die Vermutung, dass die Art und Weise der Rekrutierung, in Verbindung mit dem Vorhandensein eines Telefonbucheintrages sowie dem Einsatz eines Anreizes, zu den Unterschieden zwischen

den beiden Befragungsmodi geführt hat. Der Frage danach, ob weitere systematische Unterschiede zwischen eingetragenen und nicht eingetragenen Personen bestehen, wird in Kapitel 7.4 (Unterschiede zwischen eingetragenen und nicht eingetragenen Personen) weiter nachgegangen.

Betrachtet man nun die einzelnen Gruppen der CATI-Befragung so fällt auf, dass die Verzerrung im Vergleich zu allen anderen Gruppen deutlich geringer ausfällt. Auch wenn kein Anreiz vergeben wurde, liegt dieser mit einer Abweichung von 20,2 Prozent auffallend unter dem Niveau der CAWI-Variante. Betrachtet man die Verteilungen und Abweichungen in den einzelnen experimentellen Anreizgruppen, bleiben diese Unterschiede zwischen den beiden Modi bestehen und fallen dabei noch deutlicher aus. Denn offensichtlich zieht der Einsatz von Anreizen in der Gruppe der CATI-Teilnehmer, wie erwartet, Selektionseffekte derart nach sich, dass die Abweichungen von der tatsächlichen Bildungsverteilung ersichtlich niedriger ausfallen.

Hervorzuheben sind die CATI-Gruppen, die einen Fünfeuroschein sowie die Aussicht auf einen Gewinn erhalten haben. Im Vergleich zum Mikrozensus liegen hier mit 11,4 (Fünfeuroschein) sowie 8 Prozentpunkten (Verlosung) die geringsten Abweichungen vor. Warum die Verlosung eine derart positive Auswirkung hat, kann an dieser Stelle lediglich vermutet werden. In den vorherigen Kapiteln ist deutlich geworden, dass der Einsatz der Verlosung sogar einen negativen Effekt auf die Antwortrate hatte. Daraus ergibt sich für die Teilnehmer in dieser Gruppe eine sehr geringe Fallzahl. Dies legt die Vermutung nahe, dass es sich hierbei um eine besondere Gruppe an Teilnehmern handelt, die sich deutlich vom Rest unterscheidet. An dieser Stelle müssen weitere Untersuchungen stattfinden, denn hierbei ist es nicht unwahrscheinlich, dass es sich um ein zufälliges Ergebnis oder ein statistisches Artefakt handelt.

Dieses Ergebnis wirft zwangsläufig die Frage danach auf, ob die Teilnehmer, die zunächst durch den Einsatz eines Anreizes zu einer Teilnahme bewegt werden müssen, aufgrund ihrer vermuteten niedrigen intrinsischen Motivation weniger Mühe in die Beantwortung der gestellten Fragen investieren. Dem entgegen steht die Annahme von Brehm (1994, 57): „[...] respondents become more talkative, more involved with the interview, when a reward has been offered.“ Brehm (1994) bezieht seine Aussagen allerdings auf Face-to-Face-Interviews in Verbindung mit offenen Fragen, womit durch den persönlichen Kontakt vor Ort und die Art und Weise der Frageformulierung bereits eine andere Befragungssituation gegeben ist und der Effekt des Anreizes hier nicht eins zu eins mit anderen Befragungsmodi zu

Tabelle 7.8: Auswirkung der vergebenen Anreize auf die Datenqualität: Lineare Regression mit der absoluten Häufigkeit von Item-Nonresponse als abhängige Variable

	b	s.e.
5 Euro	-0,01	1,65
Verlosung	2,27	1,22
<i>Bildung</i>		
(Ref. Volks./Hauptschulabschluss)		
Mittlere Reife	0,65	0,52
Fachhochschulreife	1,14	0,81
Abitur	0,80	0,66
<i>Interaktionseffekte</i>		
5 Euro x Bildung	-0,39	0,22
Verlosung x Bildung	-0,02	0,27
<i>Fallzahl (N)</i>	101	
<i>Modellgüte</i>		
Prob > F	0,66	
R <sup>2</sup>	0,05	

\*  $p \leq 0,05$ , \*\*  $p \leq 0,01$ , \*\*\*  $p \leq 0,001$

PPSM Access Panel Nachrekrutierung 2012, eigene Berechnungen

vergleichen ist. Um einen möglichen Einfluss auf die Datenqualität zu überprüfen, wurde für die Nachrekrutierung aus dem Jahr 2012 ein Index für die beobachtete Frequenz von Item-Nonresponse gebildet. Dieser summiert die absoluten Häufigkeiten für nicht beantwortete Fragen jedes Teilnehmers auf. Diese Variable dient nun als Zielgröße in einem linearen Regressionsmodell, deren Ergebnisse in Tabelle 7.8 abgebildet sind. Wie auch schon Vehre et al. (2013) herausgefunden haben, hat der Einsatz von Anreizen keinen negativen Einfluss auf die Datenqualität. Weder der Fünfeuroschein noch die Verlosung führen zu einer höheren Frequenz von Item-Nonresponse. Auch in der Interaktion mit dem Bildungsniveau verändert sich nichts an diesem Ergebnis. Das Modell ist nicht auf die Grundgesamtheit zu übertragen und erklärt lediglich drei Prozent der Varianz der abhängigen Variable. Somit ist aufgrund dieses Ergebnisses nicht von einem negativen Effekt auf die Datenqualität auszugehen, wenn Anreize eingesetzt werden.

### Zusammenfassung der Ergebnisse

Die aufgestellten Hypothesen  $H_{D5}$  und  $H_{D7}$  können angenommen werden. Die Erkenntnisse der hierzu vorliegenden Studien konnten dahingehend weiter erhärtet werden, dass auch mit der Nachrekrutierung des PPSM Access Panels aus dem Jahr 2012 der nicht

konditionale Einsatz von fünf Euro dazu geführt hat, dass der Anteil an niedriger gebildeten Personen unter den Teilnehmern der Umfrage erhöht werden konnte.

Tabelle 7.9: Entscheidung über die Hypothesen - Zielgrößen: Kooperationswahrscheinlichkeit und Datenqualität

Hypothese	Vermutete Richtung des Zusammenhangs	Status
$H_{D5}$ : Monetärer nicht konditionaler Anreiz $\Rightarrow$ Kooperationswahrscheinlichkeit niedrig gebildeter Personen	positiv	angenommen
$H_{D7}$ : Teilnahmeanreize $\Rightarrow$ Item-Nonresponse	gleichbleibend	angenommen

In diesem „methodischen Kniff“ liegt das Potential, den Bildungsbias in wissenschaftlichen Umfragen zu verringern. Ferner führt eine zu vermutende geringer ausgeprägte intrinsische Motivation, auf Seiten niedrig gebildeter Personen, nicht zu einer Verschlechterung der Datenqualität in Form von einem größeren Anteil an Item-Nonresponse. Letztlich darf dabei allerdings nicht außer Acht gelassen werden, dass durch den Einsatz monetärer Anreize hohe Kosten entstehen, die dem Ertrag für die Repräsentativität der Ergebnisse in einem angemessenen Verhältnis gegenüber stehen müssen.

Zunächst soll es im folgenden Kapitel um eine weitere zentrale Frage bezüglich des Einsatzes monetärer Anreize gehen, ob es einen möglichen Grenznutzen beim Einsatz dieser gibt. Gemeint ist hiermit speziell die Frage danach, ob monetäre Anreize einen linearen Effekt auf die Antwortrate haben oder ob sich dieser ab einem bestimmten Betrag verändert.

## 7.2.2 Grenznutzen monetärer Teilnahmeanreize

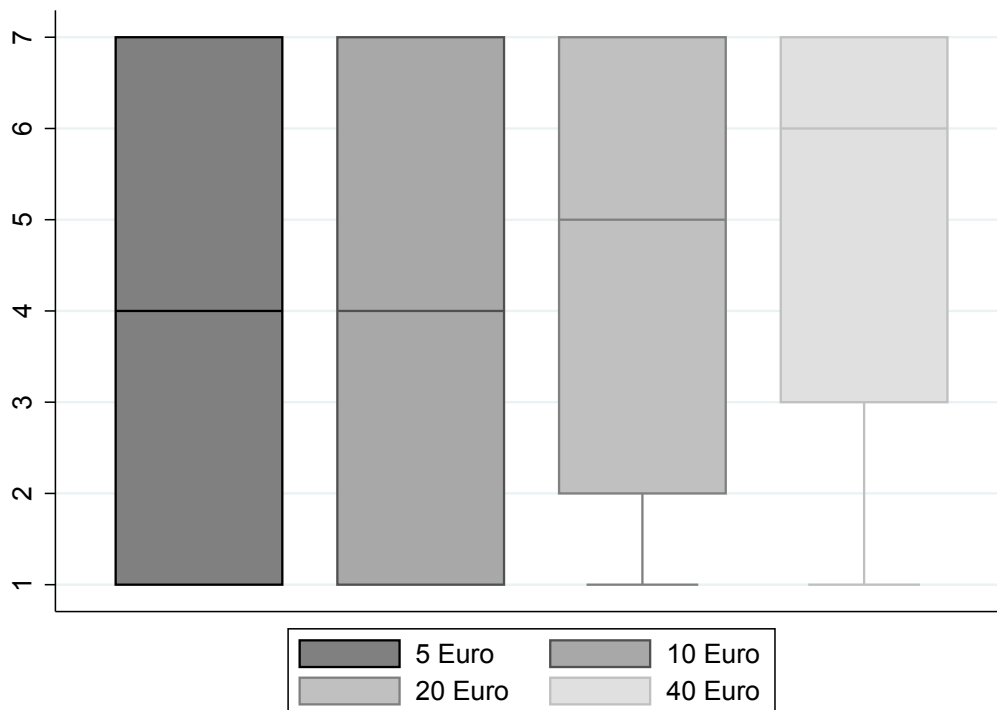
Der Einsatz unterschiedlicher Teilnahmeanreize ist schon häufig mittels experimenteller Designs untersucht worden.<sup>3</sup> Geht es um die Frage, inwiefern sich diese in unterschiedlicher Form auf die Antwortrate auswirken, ist die Forschungslage bereits breit aufgestellt und man kann von gesicherten Erkenntnissen sprechen.

Richtet man den Fokus auf einen möglichen Grenznutzen, insbesondere in Bezug zu monetären Anreizen, die unabhängig von einer Teilnahme vergeben werden (nicht konditional), sind die vorliegenden Ergebnisse uneindeutig. Vor allem ab größeren Beträgen von über zehn Euro gehen die Resultate teilweise in unterschiedliche Richtungen. Im Folgenden sollen die Ergebnisse des experimentellen Designs beschrieben werden, mittels

<sup>3</sup> Für eine Übersicht relevanter Studien siehe Kapitel 4.2.



Abbildung 7.3: Subjektive Wahrnehmung der Incentivehöhe - felddexperimentelles Design mit 5, 10, 20 und 40 Euro



Skala: 1 = „kleine Aufmerksamkeit“, 7 = „Bezahlung“; PPSM Access Panelwelle 2014, eigene Berechnungen.

dessen in der 2014er Welle des PPSM Access Panels der Versuch unternommen wurde, einen möglichen Grenznutzen zu operationalisieren. Die Teilnehmer der CATI- beziehungsweise CAWI-Befragung wurden zufällig zu vier Untersuchungsgruppen zugeordnet und haben daraufhin unterschiedliche Beträge von 5, 10, 20 und 40 Euro im Fragebogen angezeigt bekommen. Mit einer Likert-Skala wurde in Anlehnung an Dillman et al. (2014) eine Skala von „1 = kleine Aufmerksamkeit“ bis „7 = Bezahlung“ aufgespannt. Die Teilnehmer wurden daraufhin gebeten, sich in Abhängigkeit des Betrages auf der Skala einzuordnen. In Abbildung 7.3 ist die Verteilung der betreffenden Variablen in den experimentellen Gruppen in Form von Box-Plots abgebildet.

Die ersten beiden Boxen, welche die Verteilung der Gruppen repräsentieren, die 5, beziehungsweise 10 Euro als hypothetischen Anreiz erhalten haben, unterscheiden sich nicht voneinander. Sowohl der Interquartilsabstand als auch der Median liegen auf dem exakt gleichen Niveau. Dies kann als erster Hinweis darauf angesehen werden, dass sich die subjektive Wahrnehmung auf Seiten der Befragten weder in die Richtung „kleine Aufmerksamkeit“ noch in Richtung „Bezahlung“ geht, wenn Beträge von 5 oder 10 Euro vergeben werden. Der Median liegt in beiden Boxen genau in der Mitte der Skala bei einem

Wert von vier. Für die Praxis bedeutet dies, dass sich die subjektive Wahrnehmung bei kleineren Beträgen nicht verändert, sehr wohl aber die Auswirkung auf die Antwortrate, wie vorliegende Studien bereits zeigen konnten (Engel et al., 2004; Jobber et al., 2004; Millar und Dillman, 2011; Scherpenzeel und Toepoel, 2012; Singer et al., 1999, 2000; Toepoel, 2012). Dies ist ein Hinweis darauf, dass man als Forscher im Bereich von bis zu zehn Euro nicht Gefahr läuft, dass die Teilnehmer diesen als Bezahlung ansehen. Letztlich kann an dieser Stelle keine klare Aussage getroffen werden, da einige der Befragten auch bei 5 oder 10 Euro angegeben haben, dass sie dies als Bezahlung empfinden würden. Es kommt darauf an, wie bereits Dillman et al. (2014) deutlich gemacht hat, dass der Einsatz eines Anreizes bei den Befragten kein Denken in der Form auslösen darf, dass sie für die Teilnahme bezahlt werden. Denn dann kann ein Betrag schnell als unzureichend für die zu investierende Zeit empfunden werden und eine Teilnahme in der Folge unwahrscheinlicher machen.

Die Verteilung hinsichtlich der beiden Gruppen, die 20 beziehungsweise 40 Euro im Fragebogen angezeigt bekommen haben, unterscheiden sich dagegen untereinander sowie von den anderen Beträgen. So verschiebt sich die subjektive Wahrnehmung bei 20 Euro in Richtung Bezahlung. Der Median liegt hier bereits bei fünf. Das zweite Quartil verhält sich ähnlich dazu und steigt von eins auf zwei, womit die Streuung insgesamt etwas kleiner wird. Betrachtet man die letzte der vier Boxen, die einen hypothetischen Betrag von 40 Euro repräsentiert, steigt der Median auf einen Wert von sechs sowie das zweite Quartil auf drei. Werden die Beträge also sukzessive erhöht, ist an dieser Stelle für die Stichprobe zu schlussfolgern, dass sich die subjektive Wahrnehmung der Teilnehmer, im Vergleich zu 5 und 10 Euro, ab 20 beziehungsweise 40 Euro immer weiter Richtung Bezahlung verschiebt.

Daran anknüpfend sind in Tabelle 7.10 die Ergebnisse von Mittelwertvergleichen (T-Tests) zwischen den einzelnen Experimentalgruppen abgebildet.<sup>4</sup>

Die Mittelwerte unterscheiden sich im Vergleich von 5 zu 10 Euro sowie von 20 zu 40 Euro nicht signifikant voneinander. Dieses Ergebnis, insbesondere in Bezug zu den beiden 5 und 10 Euro Gruppen, ist mit Blick auf die Verteilung in Abbildung 7.3 zu erwarten gewesen. Denn diese unterscheiden sich sowohl in Schwerpunkt als auch Streuung nur in Nuancen voneinander. Richtet man den Blick in die vorletzte Spalte der Tabelle so ist zu erkennen, dass sich sowohl der Mittelwert in der 5 Euro Gruppe als auch der in der 10 Euro Gruppe signifikant von dem in der 20 Euro Gruppe unterscheidet. Hier kann also

---

<sup>4</sup> Für die mathematischen Grundlagen des T-Tests siehe Kühnel und Krebs (2001).

Tabelle 7.10: Mittelwertvergleiche der experimentellen Bedingungen mit unterschiedlichen Anreizbeträgen

	$\bar{x}$	n	5 Euro	10 Euro	20 Euro	40 Euro
5 Euro	4,1	98	-	n.s.	*	**
10 Euro	4,0	83	-	-	*	***
20 Euro	4,7	115	-	-	-	n.s.
40 Euro	4,9	109	-	-	-	-

\*  $p \leq 0,05$ , \*\*  $p \leq 0,01$ , \*\*\*  $p \leq 0,001$

PPSM Access Panelwelle 2014, eigene Berechnungen

bereits von einer signifikanten Verschiebung der subjektiven Wahrnehmung in Richtung Bezahlung gesprochen werden.

Eindeutiger werden diese Unterschiede, wenn der Betrag noch einmal verdoppelt wird. Während zwischen 20 und 40 Euro noch kein signifikanter Unterschied besteht, ist die Differenz zwischen 5 und 20 Euro bereits hoch sowie zwischen 10 und 40 Euro höchst signifikant. Es scheint also durchaus einen Unterschied auszumachen, wie Befragte einen Anreiz, in Abhängigkeit von dessen Höhe, wahrnehmen. Schließlich wurde den Teilnehmern die Möglichkeit gegeben, mittels zweier offener Antwortfelder explizite Beträge anzugeben, die sie ihrer Meinung nach als kleine Aufmerksamkeit beziehungsweise Bezahlung wahrnehmen würden. Tabelle 7.11 gibt einen Überblick über die Ergebnisse der qualitativen Inhaltsanalyse.<sup>5</sup>

Die von den Befragten genannten Beträge wurden dabei zwölf Kategorien zugeordnet. Fast alle haben hierbei, wie vorgesehen, absolute Beträge angegeben oder diese in der CAWI-Variante des Fragebogens selbst eingegeben. Die unterschiedlichen Fallzahlen in den beiden Variablen ist Item-Nonresponse geschuldet. Da einige Teilnehmer sehr hohe Beträge von über hundert Euro angegeben haben, wurde hierfür eine separate Kategorie mit aufgenommen. Nur wenige Personen haben hier längere Antworten gegeben. Am häufigsten haben diese geantwortet, dass sie jeden Betrag als Bezahlung wahrnehmen würden und im Umkehrschluss auch keinen Betrag als „kleine Aufmerksamkeit“ definieren. Mit  $n = 83$  (22,1 Prozent) beziehungsweise  $n = 62$  (15,9 Prozent) Nennungen hat diese Gruppe eine zu berücksichtigende Größe. Längere Antworten haben darüber hinaus Personen gegeben, die nicht auf monetäre Anreize reagieren, unabhängig von der Höhe dieser. Hier wurden Antworten gegeben wie etwa „Gar keinen, freundliche Worte ja gut aber kein Geld“, „Sachpreis z.B Buch oder Karte“ oder „keinen, würde Geld nicht annehmen“.

<sup>5</sup> Für die wissenschaftlichen Grundlagen der qualitativen Inhaltsanalyse siehe Mayring (2010) oder Früh (2015).

Tabelle 7.11: Ergebnisse der qualitativen Inhaltsanalyse  
- „kleine Aufmerksamkeit“ vs. „Bezahlung“

	Aufmerksamkeit		Bezahlung	
	n	in Prozent	n	in Prozent
1 bis 5 Euro	117	31,2	32	8,2
6 bis 10 Euro	70	18,7	43	11,0
11 bis 15 Euro	13	3,5	18	4,6
16 bis 20 Euro	44	11,7	48	12,3
21 bis 25 Euro	4	1,1	16	4,1
26 bis 30 Euro	4	1,1	8	2,1
31 bis 35 Euro	2	0,5	4	1,0
36 bis 40 Euro	6	1,6	5	1,3
> 40 Euro	27	7,2	103	26,3
≥ 100 Euro	3	0,8	37	9,5
alle Beträge	2	0,5	62	15,9
keinen Betrag	83	22,1	15	3,8
$\Sigma$	375	100,0	391	100,0

PPSM Access Panelwelle 2014, eigene Berechnungen

Richtet man den Blick auf die erste Spalte von Tabelle 7.11 so fällt auf, dass die meisten Personen einen Betrag von bis zu fünf Euro als Aufmerksamkeit einstufen. Insgesamt fast die Hälfte (49,9 Prozent) hat hier einen Betrag von bis zu zehn Euro angegeben. Dieses Ergebnis kann als Hinweis dahingehend gesehen werden, dass kleinere Beträge auch tatsächlich so wie vom Forscher intendiert wahrgenommen werden: Als eine kleine Aufmerksamkeit, die einen Reziprozitätseffekt derart auslöst, dass bei den Befragten das subjektive Gefühl ausgelöst wird „etwas zurückgeben zu wollen“. Konträr dazu verhält sich die Verteilung der genannten Beträge wenn danach gefragt wird, welcher als Bezahlung wahrgenommen wird. Hier sind es 35,8 Prozent, die einen Betrag von über vierzig beziehungsweise hundert Euro angegeben haben. Immerhin fast ein Fünftel (19,2 Prozent) der Teilnehmer hat dennoch Beträge von bis zu zehn Euro eingetragen. So geht das Ergebnis der qualitativen Inhaltsanalyse zwar in die vermutete Richtung, ist allerdings nicht eindeutig.

### Zusammenfassung der Ergebnisse

Die in Kapitel 5 (Hypothesen zum Rekrutierungsinterview - Tabelle 5.4) aufgestellten Forschungshypothesen zur subjektiven Wahrnehmung von Anreizen sowie einem möglichen Grenznutzen, konnten durch die empirische Analyse teilweise bestätigt werden. Die Hypothesen  $H_{D4}$  muss abgelehnt werden. Die Ergebnisse zeigen zwar in die vermutete Richtung,

dass Anreize von bis zu zehn Euro eher als kleine Aufmerksamkeit wahrgenommen werden. Die Befundlage ist allerdings nicht eindeutig genug. Die Mittelwerte der feldexperimentellen Bedingungen in Tabelle 7.10, für die fünf und zehn Euro Gruppen, unterscheiden sich nicht signifikant voneinander. Diese liegen mit 4,1 beziehungsweise 4,0 in der Mitte der siebenstufigen Skala und zeigen somit nicht eindeutig in eine Richtung der beiden Skalenendpunkte. Betrachtet man das Ergebnis der qualitativen Inhaltsanalyse in Tabelle 7.11 ist zu erkennen, dass fast die Hälfte der Befragten einen Betrag von bis zu zehn Euro angegeben haben, den sie als kleine Aufmerksamkeit wahrnehmen würden. Es gibt also durchaus Hinweise darauf, dass Hypothese  $H_{D4}$  empirisch haltbar ist.

Tabelle 7.12: Entscheidung über die Hypothesen - subjektive Wahrnehmung der Anreizhöhe

Hypothese	Vermutete Richtung des Zusammenhangs	Status
$H_{D4}$ : Anreizbetrag $\leq$ 10 Euro $\Rightarrow$ Wahrnehmung eher „kleine Aufmerksamkeit“	positiv	abgelehnt
$H_{D5}$ : Anreizbetrag $\geq$ 10 Euro $\Rightarrow$ Wahrnehmung eher Bezahlung	positiv	angenommen

Werden Beträge von über zehn Euro vergeben (Hypothese  $H_{D5}$ ) unterscheiden sich die Mittelwerte der Antworten von 20 und 40 Euro signifikant von der fünf und zehn Euro Bedingung. Mit Mittelwerten von 4,7 (20 Euro) und 4,9 (40 Euro) findet eine Verschiebung der Wahrnehmung in Richtung Bezahlung statt. Auch bei der offenen Frage hat die Mehrheit der Teilnehmer einen Betrag angegeben, der über zehn Euro liegt. Die mit  $H_{D5}$  aufgestellte Vermutung, dass sich die Wahrnehmung bei Beträgen von über zehn Euro eher in Richtung Bezahlung verschiebt, kann durch die empirische Analyse bestätigt werden.

In den folgenden Kapiteln wird der Ausfallprozess der Nachrekrutierung aus dem Jahr 2012 näher beleuchtet. Hier steht die Frage danach im Fokus, ob ein möglicher Bias durch Nonresponse mit Hilfe von zusätzlichen Hintergrundinformationen aus dem Auswahlrahmen operationalisiert werden kann und welche Ausmaße dieser gegebenenfalls annimmt. Noncoverage ist, wie bereits diskutiert, an dieser Stelle auszuschließen, da die Stichprobe auf Basis des Melderegisters gezogen und postalisch kontaktiert wurde. Dies impliziert, dass alle Elemente der Grundgesamtheit die gleiche Wahrscheinlichkeit hatten, gezogen zu werden und darüber hinaus zwei Modi zur Teilnahme angeboten worden sind.

### 7.3 Das Melderegister als Auswahlrahmen - Analyse des Ausfallprozesses

Ein zentrales Erkenntnisinteresse dieser Arbeit besteht darin, den Ausfallprozess einer Umfrage zu operationalisieren, um mögliche systematische Selektionseffekte aufzudecken, die zu einem Bias durch Nonresponse führen können. Um diesem Aspekt nachzugehen, ist zunächst noch einmal auf den Unterschied zwischen der Brutto- sowie Nettostichprobe einzugehen, denn daran knüpfen sich unterschiedliche Teilgruppen sowie verfügbare Informationen.

Wie bereits zuvor beschrieben, gibt es durch das methodische Vorgehen bei der Nachrekrutierung des PPSM Access Panels aus dem Jahr 2012 eine Reihe von Variablen, die für alle Personen in der Stichprobe zur Verfügung stehen.<sup>6</sup> Mit diesen Variablen kann eine Analyse des Ausfallprozesses für die Bruttostichprobe, also unabhängig von einer Teilnahme, stattfinden. Gerade hierbei kann sich möglicherweise die Stärke des angewandten methodischen Vorgehens bei der Nachrekrutierung zeigen. Allerdings ist hierfür notwendig, dass die Hintergrundvariablen die von Olson (2013, 144 f.) formulierten Kriterien zur Operationalisierung eines Bias durch systematischen Nonresponse erfüllen.

Mit der Analyse sowie der folgenden Beschreibung der Ergebnisse dieser wird folgendermaßen vorgegangen: Zunächst wird ein Grundlagenmodell für alle Personen in der Stichprobe mit allen verfügbaren Variablen aus dem Auswahlrahmen geschätzt. Daraufhin werden die jeweiligen Modelle für die einzelnen Teilgruppen der Nachrekrutierung beschrieben. Die zu treffende Unterscheidung ergibt sich jeweils aus den zur Verfügung stehenden Variablen für die gesamte Stichprobe sowie die einzelnen Teilgruppen in Abbildung 7.13. Denn die Parainformationen für die Personen, die angerufen worden sind, stehen auch nur für diese zur Verfügung. Letztlich werden die geschätzten Wahrscheinlichkeiten aus den genesteten logistischen Regressionsmodellen in das Random Response Model nach Bethlehem (2002) überführt. Im Zuge dessen sollen mögliche Zusammenhänge zwischen der Teilnahmewahrscheinlichkeit und den Umfragevariablen, in Abhängigkeit von unterschiedlichen Designbedingungen, untersucht werden. Werden die Variablen schrittweise in die Schätzgleichung aufgenommen, können hiermit zwei Dinge überprüft werden. Einerseits kann festgestellt werden, ob die zusätzlich aufgenommene (Hintergrund-)Variable einen signifikanten Erklärungsbeitrag für die Varianz der abhängigen Variable (Teilnah-

---

<sup>6</sup> Für eine Übersicht verfügbarer Parainformationen im PPSM Access Panel siehe Tabelle 6.6.

mewahrscheinlichkeit  $p$ ) leistet und somit ein Kriterium nach Olson (2013) erfüllt. Wird andererseits von Modell zu Modell jeweils eine neue Variable in die Schätzgleichung mit aufgenommen, kann daraufhin mit den geschätzten Wahrscheinlichkeiten im Rahmen des RRM Models überprüft werden, ob diese Variable den Zusammenhang mit den gegebenen Antworten ( $Y$ ) verändert. Kann ein signifikanter Zusammenhang nachgewiesen werden, ist dies ein Indikator für einen Bias durch selektiven Nonresponse, bezogen auf die jeweils zusätzlich mit aufgenommene Variable. So kann überprüft werden, ob Selektionseffekte vorliegen und ob diese mit dem methodischen Vorgehen in der 2012er Nachrekrutierung zusammenhängen oder nicht.

### 7.3.1 Teilnahme am Rekrutierungsinterview

Für eine Teilnahme an der Nachrekrutierung des PPSM Access Panels aus dem Jahr 2012 sind insbesondere zwei Konstrukte von Bedeutung. Zunächst muss erfolgreich Kontakt zur Zielperson hergestellt werden (Zielgröße: Erreichbarkeit  $p_e$ ). Daraufhin ist es von einer Reihe weiterer Merkmale abhängig, ob diese kooperiert oder nicht (Zielgröße: Kooperationswahrscheinlichkeit  $p_k$ ). Diese beiden Faktoren und die dahinterstehenden Variablen ergeben in ihrer Summe schlussendlich die übergeordnete Wahrscheinlichkeit  $p$ , ob sich eine Person am Rekrutierungsinterview beteiligt oder nicht.

In der Analyse selbst werden diese beiden Zielgrößen nicht getrennt voneinander untersucht. Es werden Variablen mit einbezogen, die sowohl die Erreichbarkeit als auch die Kooperationswahrscheinlichkeit beeinflussen. Letztlich geht es darum zu untersuchen, inwiefern diese die übergeordnete Teilnahmewahrscheinlichkeit  $p$  beeinflussen. Diese Unterscheidung wird auch nur dann relevant, wenn sich die Analysen auf die Teilgruppe an Personen beschränken, die in der Nachrekrutierung aus dem Jahr 2012 angerufen worden sind. Denn Prädiktoren, wie die Anzahl an Kontaktversuchen, wirken beispielsweise vermittelt über die Erreichbarkeit auf die Wahrscheinlichkeit einer Teilnahme.

Da es allerdings Unterschiede hinsichtlich der verfügbaren Informationen in Abhängigkeit davon gibt, ob sich die Zielperson an der Umfrage beteiligt hat oder nicht, muss dieser Aspekt bei der Analyse berücksichtigt und bei Betrachtung der Ergebnisse im Hinterkopf bleiben. Zunächst wird das Modell geschätzt, dass die Grundlagen für alle weiteren Analysen bildet. Die hierbei zur Anwendung kommenden Variablen stehen für die komplette Stichprobe unabhängig davon zur Verfügung, ob sich eine Person beteiligt

Tabelle 7.13: Studiendesign Nachrekrutierung 2012 - Wiederholung von Tabelle 6.4

<b>Registerstichprobe</b>						
<b>Rufnummer bekannt</b>			<b>Rufnummer unbekannt</b>			
<b>A</b>			<b>B</b>			<b>C</b>
Vorausbrief: Ankündigung des Anrufes/Möglichkeit Teilnahme online			Vorausbrief: Rückantwortschreiben/Möglichkeit Teilnahme online			Vorausbrief: Rückantwortschreiben/Möglichkeit Teilnahme online
<b>D</b>	<b>E</b>	<b>F</b>	<b>G</b>	<b>H</b>	<b>I</b>	
5 Euro	Kontrollgruppe	Verlosung	Kontrollgruppe	Verlosung	Kontrollgruppe	

hat oder nicht.

Zur besseren Orientierung für den Leser wird mit Tabelle 7.13 noch einmal das Rekrutierungsdesign aus der 2012er Nachrekrutierung innerhalb des PPSM Access Panels abgebildet. Im Zusammenspiel mit den hierzu stattfindenden Analysen in den folgenden Kapiteln ist so eine Zuordnung der Modelle zu den sich daraus ergebenden Teilgruppen möglich und es wird nachvollziehbar, worauf sich die einzelnen Modelle jeweils beziehen.

### **Analyse der Bruttostichprobe mit verfügbaren Hintergrundvariablen**

Zunächst wird ein Modell mit allen Variablen, die über den Auswahlrahmen zur Verfügung stehen, geschätzt. Die Ergebnisse der logistischen Regression sind in Tabelle 7.14 abgebildet. Neben Alter und Geschlecht steht zusätzlich die Information zur Verfügung, in welchem Stadtteil eine Person lebt. Diese Variablen werden als Dummy-Terme zur Kontrolle in die Schätzgleichung mit aufgenommen, damit dieser Effekt bei der Interpretation der weiteren Koeffizienten konstant gehalten werden kann.

Für die gesamte Stichprobe hat das Geschlecht, entgegen der formulierten Erwartung hierzu, einen signifikanten Einfluss auf die Umfrageteilnahme. Dieser bleibt auch unter der Hinzunahme weiterer Variablen über die Modelle hinweg höchst signifikant. Männer haben daran anknüpfend eine deutlich höhere geschätzte Wahrscheinlichkeit, sich am Rekrutierungsinterview zu beteiligen. In Abbildung 7.4 sind die Wahrscheinlichkeiten differenziert nach Geschlecht und Modus der Teilnahme grafisch abgebildet.

Der Blick hierauf zeigt ebenso relativ deutlich, inwiefern sich Männer und Frauen unterscheiden. Männer bewegen sich auf einem durchgehend höheren Niveau und besitzen im Alter von circa 60 Jahren die höchste Teilnahmewahrscheinlichkeit von über 20 Prozent. Hierbei zeigt sich, dass die Wahrscheinlichkeiten für eine Teilnahme insgesamt nicht hoch ausfallen. Dieses Ergebnis ist vor dem Hintergrund der übergeordneten Antwortrate von



Tabelle 7.14: Logistische Regression mit der Teilnahme am Rekrutierungsinterview (0 = Nein; 1 = Ja ) als abhängige Variable für die Bruttostichprobe

	M1		M2		M3		M4	
	OR	s.e.	OR	s.e.	OR	s.e.	OR	
Geschlecht (Ref. Weiblich)	2,95 ***	0,41	2,98 ***	0,42	2,93 ***	0,42	2,99 ***	0,43
Alter in Jahren			1,01 ***	0,00	1,11 ***	0,03	1,12 ***	0,03
Alter in Jahren <sup>2</sup>					0,99 ***	0,00	0,99 ***	0,00
<i>Stadtteile</i> (Ref. Walle)								
Blumenthal							0,83	0,31
Burglesum							2,27 *	0,92
Häfen							1,36	0,59
Neustadt							1,32	0,52
Oberneuland							2,22 *	0,74
Obervieland							1,37	0,43
Osterholz							1,36	0,37
Östliche Vorstadt							2,11 **	0,60
Schwachhausen							2,04 **	0,56
Vahr							1,27	0,33
Veogesack							0,72	0,23
Woltmershausen							1,36	0,35
<i>Fallzahl (n)</i>	2.423		2.423		2.423		2.423	
<i>Modellgüte</i>								
Prob > chi <sup>2</sup>	***		***		***		***	
Pseudo R <sup>2</sup>	0,04		0,05		0,06		0,07	
LL	-854,56		-844,76		-835,02		-822,29	
LR-Test			***		***		*	

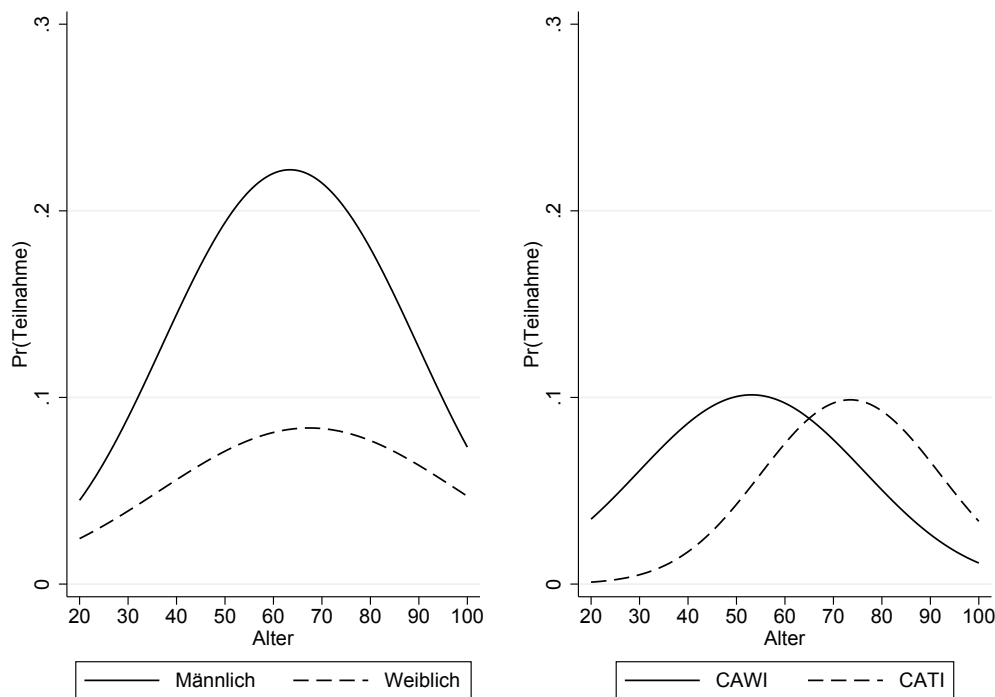
\* p ≤ 0,05, \*\* p ≤ 0,01, \*\*\* p ≤ 0,001, + p ≤ 0,10

PPSM Access Panel Nachrekrutierung 2012, eigene Berechnungen

zwölf Prozent, die in der 2012er Nachrekrutierung erzielt werden konnte, zu erwarten gewesen. Wird das Alter im Rahmen von Modell 2 und Modell 3 mit in die Schätzgleichung aufgenommen, verbessert sich die Erklärungsleistung. Der Zusammenhang offenbart sich dabei als nicht linear, denn der quadratische Alterseffekt in Modell 3 ist ebenso höchst signifikant. Die kubische Effektkomponente hat sich als nicht signifikant und somit nicht erklärungsrelevant gezeigt. Die Teilnahmewahrscheinlichkeit steigt mit zunehmendem Alter deutlich an, erreicht bei circa 60 Jahren ihren Höhepunkt und nimmt danach wieder ab, bis sie im hohen Alter schließlich ein Niveau von unter zehn Prozent erreicht.

Diese Verteilung kann in mehreren Faktoren begründet liegen. Effekte durch Noncoverage sind auszuschließen, da die zugrundeliegende Stichprobe auf Basis des Melderegisters gezogen worden ist und jede Person via Vorausbrief kontaktiert wurde. Darüber hinaus konnten sich die Personen entscheiden, ob sie lieber die CAWI-Version oder die CATI-Version des Fragebogens ausfüllen wollten. Hätte man als Forscher an dieser Stelle lediglich eine Möglichkeit angeboten, wäre mit einem anderen Effekt des Alters zu rechnen gewesen. Gestützt wird diese These von Abbildung 7.4. Hier sind im rechten Teil die geschätzten

Abbildung 7.4: Geschätzte Wahrscheinlichkeiten der Umfrageteilnahme differenziert nach Alter, Geschlecht und Modus



$n = 2.324 / n = 2.306$  (CATI),  $n = 2.250$  (CAWI); PPSM Access Panel Nachrekrutierung 2012, eigene Berechnungen.

Wahrscheinlichkeiten in Abhängigkeit der Teilnahmemodi abgebildet.<sup>7</sup> Diese beziehen sich damit zwar nicht mehr auf die Bruttostichprobe, zeigen aber deutlich die Altersunterschiede und das Potential für Verzerrungen, wenn man lediglich „single-mode“ befragen würde. Die Teilnehmer an der CATI-Variante sind deutlich älter. Im Alter von 20 bis 40 Jahren liegt die geschätzte Wahrscheinlichkeit für die Teilnahme an der CAWI-Variante bei unter fünf Prozent. Hier zeigt sich das immer noch vorhandene generelle Problem von CAWI-Befragungen: Ältere Personen sind aus den verschiedenen bereits diskutierten Gründen unterrepräsentiert, auch wenn zunächst per Vorausbrief Kontakt hergestellt wird. Die CATI-Version des Fragebogens wurde vor allem von älteren Personen über 50 Jahren beantwortet. Dieses Ergebnis kann wiederum im praktischen Vorgehen begründet liegen, denn angerufen wurden vor allem diejenigen, die über einen Eintrag im Telefonbuch verfügen.

Die Verteilungen des Alters in Abhängigkeit davon, in welchem Modi sich eine Person beteiligt hat, geht in die erwartete Richtung. Es sind allerdings noch größere Unterschiede

<sup>7</sup> Die zugrundeliegenden Modelle in Tabelle 7.14 wurden hierfür getrennt für die beiden Umfragemodi berechnet. Die Ergebnistabellen sind dem Anhang zu entnehmen. Die sich daraus ergebenden Wahrscheinlichkeiten sind in Abbildung 7.4 dargestellt.

zwischen beiden Varianten zu erwarten gewesen. Die Teilnehmer an der CAWI-Variante wurden in Bezug zu den Ergebnissen aktueller Studien als noch jünger erwartet. Die Kombination beider Modi scheint ein probates Mittel zu sein, um Personen unabhängig von ihrem Alter für eine Teilnahme zu gewinnen. Dies kann man als ersten Hinweis darauf ansehen, dass mit dem angewendeten Mixed-Mode Design das Potential für Verzerrungen, hinsichtlich des Alters der Zielpersonen, verringert werden kann.

Es scheint auf der einen Seite so zu sein, dass ältere Personen, wie vermutet, aufgrund ihres Rentenstatus häufiger zu Hause und somit über das Telefon besser zu erreichen sind und sich dies positiv auf die Teilnahmewahrscheinlichkeit auswirkt. Auf der anderen Seite besteht die Annahme, dass ältere Personen aufgrund von größerer Angst vor Verbrechen oder ähnlichem, unbekanntem Anrufern eher skeptisch gegenüber stehen und sich aus diesem Grund auch eher seltener an Umfragen beteiligen. Dies kann durch die vorliegenden Ergebnisse nicht bestätigt werden, denn es sind hier insbesondere die Älteren, die sich an der CATI-Variante beteiligt haben. Worin dies genau begründet liegt, kann an dieser Stelle lediglich vermutet werden. So kann eine höhere Skepsis gegenüber unbekanntem Anrufern durch bessere Erreichbarkeit nivelliert werden. Bei den Teilnehmern der CATI-Variante handelt es sich darüber hinaus um Personen, die über einen Telefonbucheintrag verfügen. Dies legt die Vermutung nahe, dass Personen, die selbst aktiv werden, um darin verzeichnet zu sein, im Umkehrschluss auch eher dazu bereit sind, eine Umfrageeinladung per Telefon von einer zunächst für sie unbekanntem Person anzunehmen beziehungsweise dieser offener gegenüberzutreten. Letztlich ist zu erwarten, dass sich das Ankündigungsschreiben in dieser Gruppe positiv ausgewirkt hat. Denn etwaige Bedenken können so bereits im Vorhinein ausgeräumt werden, wenn sich die später anrufende Institution zunächst einmal vorstellt und die potentiellen Teilnehmer von ihrer Seriosität überzeugt.

Das Grundlagenmodell wird durch die Information darüber, in welchem Stadtteil eine Person lebt, komplettiert. Es ist von Bedeutung, die Stadtteilinformationen zur Kontrolle mit in das Grundlagenmodell aufzunehmen. Denn wie mit Blick auf Modell 5 zu erkennen ist, gibt es teilweise signifikante Unterschiede zwischen den einzelnen Stadtteilen. Diese können in demografischen Unterschieden, in höherer Arbeitslosigkeit oder Unterschieden hinsichtlich des vorherrschenden Bildungsniveaus und dem damit zusammenhängenden monatlichen Haushaltseinkommen begründet liegen.

Eine signifikant höhere Teilnahmewahrscheinlichkeit konnte für Personen nachgewiesen werden, die in Burglesum, der östlichen Vorstadt oder im Stadtteil Schwachhausen wohnen

(im Vergleich zu Personen, die in Walle wohnen). Der Effekt für Oberneuland ist mit einer Irrtumswahrscheinlichkeit von zehn Prozent signifikant, was somit lediglich als Hinweis auf einen Zusammenhang gewertet werden kann. Gemeinsamkeiten weisen diese Stadtteile in Bezug zur Sozialstruktur der Bewohner auf. Hier leben eher Personen mit höheren Einkommen. In Oberneuland ist der Anteil an Personen mit hohen Jahreseinkommen besonders ausgeprägt (Statistisches Landesamt Bremen, 2016). Der Stadtteil Burglesum besteht aus fünf Ortsteilen. Bei einem (Werderland) muss man einschränkend anmerken, dass hier eher Personen mit niedrigen Einkommen leben (Statistisches Landesamt Bremen, 2016). Insgesamt lassen die Ergebnisse in Tabelle 7.14 den Schluss zu, dass sich Personen aus Stadtteilen, in denen mehrheitlich höhere Einkommen erzielt werden, im Vergleich zu Stadtteilen mit einer anderen Sozialstruktur, eher an Umfragen beteiligen.

Allerdings erklärt das finale Modell in Tabelle 7.14 lediglich sieben Prozent der Varianz der abhängigen Variable. Im Vergleich zu Modell 3 ist dies eine marginale aber signifikante Verbesserung. Die Stadtteilvariablen können in den folgenden Analysen nicht in alle Modelle mit aufgenommen werden. In den weiteren Kapiteln werden Bezüge auf einzelne Teilgruppen gesetzt, wodurch sich die Fallzahlen deutlich verringern. Somit können die Fallzahlen in den einzelnen Kategorien der Stadtteilvariablen sehr klein werden. Dies hat zur Folge, dass die Schätzung der einzelnen Koeffizienten mit einer relativ großen Unsicherheit belastet ist und die Standardfehler in der Folge zu groß sind. Eine Interpretation des jeweiligen Effektes kann beziehungsweise sollte somit nicht vorgenommen werden.

### **Zusammenfassung der Ergebnisse**

In diesem Teil des Ergebniskapitels wurde das Grundlagenmodell mit allen verfügbaren Variablen für die Bruttostichprobe geschätzt. Es wurde vermutet, dass sich das Alter negativ auf die Teilnahmewahrscheinlichkeit auswirkt (Hypothese  $H_{A5}$ ). Denn ältere Menschen stehen unbekannt anrufenden Personen aus verschiedenen Gründen (vor allem aus Angst vor Kriminalität) eher skeptisch gegenüber, sind häufiger aus gesundheitlichen Gründen nicht in der Lage teilzunehmen oder nutzen darüber hinaus das Internet nicht in dem Ausmaß, wie das jüngere Personen tun (ARD/ZDF-Onlinestudie, 2015; Couper und Groves, 1996; Goyder, 1987; Haunberger, 2011).

Die Ergebnisse haben mit Blick auf Abbildung 7.4 gezeigt, dass sich das Alter nicht linear auf die Wahrscheinlichkeit der Teilnahme auswirkt. Vielmehr muss hier von einem umgekehrt U-förmigen Zusammenhang gesprochen werden: Die Wahrscheinlichkeit kommt

Tabelle 7.15: Entscheidung über die Hypothesen - Zielgrößen: Teilnahmewahrscheinlichkeit und Wahl des Umfragemodi

Hypothese	Vermutete Richtung des Zusammenhangs	Status
$H_{A5}$ : Alter $\Rightarrow$ Teilnahmewahrscheinlichkeit	negativ	abgelehnt
$H_{A6}$ : Geschlecht $\Rightarrow$ Teilnahmewahrscheinlichkeit	gleichbleibend	abgelehnt
$H_{A7}$ : Alter $\Rightarrow$ Entscheidung CATI	positiv	angenommen
$H_{A8}$ : Geschlecht (Männlich) $\Rightarrow$ Entscheidung CAWI	positiv	angenommen

von einem niedrigen Niveau bei den 20-30 Jährigen, erreicht bei circa 60 Jahren ihren Höhepunkt und fällt danach wieder deutlich ab. Hypothese  $H_{A5}$  muss somit abgelehnt werden.

Hypothese  $H_{A6}$  bezieht das Geschlecht der Personen mit ein und vermutet, dass sich dieses nicht auf die Teilnahmewahrscheinlichkeit auswirkt und der Effekt für Männer sowie Frauen gleich ist. Auch diese Hypothese muss abgelehnt werden, da sich Männer im Vergleich zu Frauen deutlich häufiger am Rekrutierungsinterview beteiligt haben. Der Koeffizient für das Geschlecht kann Effekte relevanter Drittvariablen mit transportieren, die nicht im Modell enthalten sind. Da Männer das Internet häufiger nutzen (TNS Infratest, 2014) und in der Rekrutierungsbefragung die Möglichkeit zur Teilnahme an der CAWI-Variante bestand, kann der Effekt der Internetnutzung den Geschlechterkoeffizienten in Tabelle 7.14 beeinflussen. Eine Kontrolle der Internetnutzung ist an dieser Stelle nicht möglich, da sich die Analyse auf die Bruttostichprobe bezieht.

Die Hypothesen  $H_{A7}$  (Alter) und  $H_{A8}$  (Geschlecht) beziehen sich auf die Entscheidung der Teilnehmer für die CATI-Variante oder der CAWI-Variante des Fragebogens. Durch Abbildung 7.4 lassen sich die Altersunterschiede der Teilnehmer im Vergleich der beiden angebotenen Umfragevarianten gut erkennen. Die Teilnehmer der CATI-Variante sind eher älter. Hypothese  $H_{A7}$  kann somit als vorläufig bestätigt angesehen werden. Die statistischen Modelle zu Abbildung 7.4 und Hypothese  $H_{A8}$  sind dem Anhang A zu entnehmen. Diese zeigen, dass sich Männer häufiger an der CAWI-Variante des Fragebogens beteiligt haben. Begründet kann dieser Effekt darin liegen, dass Männer das Internet häufiger im Vergleich zu Frauen nutzen. Die aufgestellte Hypothese  $H_{A8}$ , kann als vorläufig bestätigt angesehen werden.

Im folgenden Abschnitt werden die unterschiedlichen Designbedingungen der Nachrekrutierung hinsichtlich ihres Effektes auf die Wahrscheinlichkeit der Teilnahme untersucht.

Diese Modelle und die sich daraus ergebenden Teilnahmewahrscheinlichkeiten bilden die Grundlage für die Operationalisierung eines möglichen Bias durch selektiven Nonresponse mit dem RRM nach Bethlehem (2002). So kann überprüft werden, ob die einzelnen Designaspekte einen systematischen Selektionsprozess ausgelöst und in der Folge einen Bias verursacht haben oder nicht.

### **Zur Wirkung der eingesetzten Anreize und der Art der Kontaktierung**

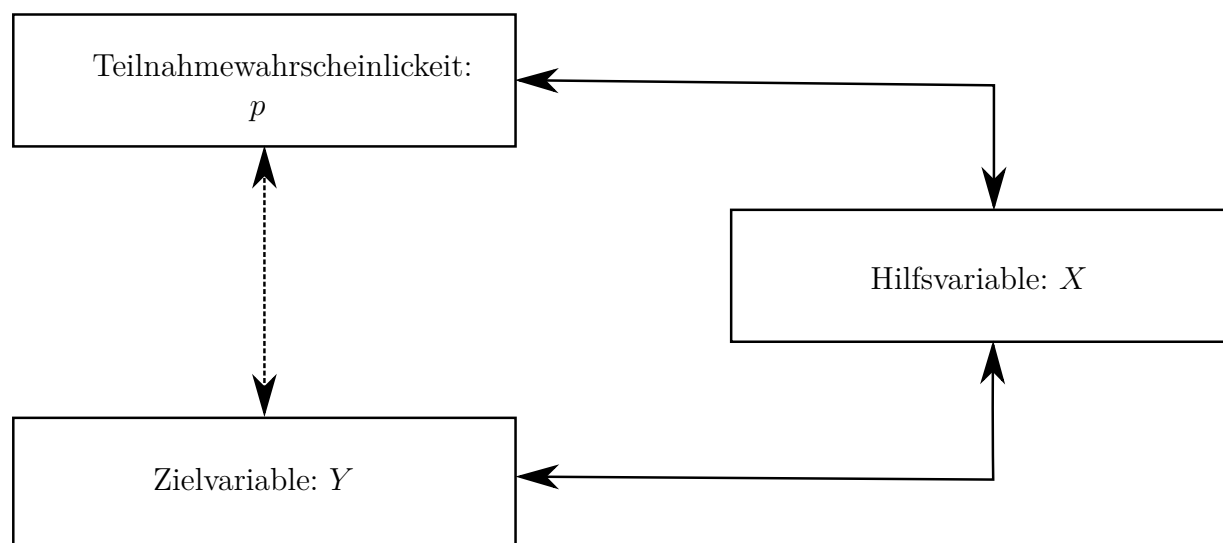
Wie bereits bei der Betrachtung der erzielten Antwortraten im Rahmen der Nachrekrutierung im Jahr 2012 deutlich geworden ist, hat sich die Vergabe von Anreizen in Form von fünf Euro (unabhängig von der Teilnahme) sowie die Vorankündigung des Telefonanrufes positiv auf die Höhe der Antwortrate ausgewirkt. Die Verlosung hatte einen negativen Effekt. Von Interesse ist darüber hinaus die Frage danach, ob sich diese Effekte einerseits im Rahmen multivariater Analysemodelle unter Kontrolle weiterer Variablen bestätigen lassen. Andererseits können daran anknüpfend, wie in den vorhergehenden Kapiteln bereits diskutiert, Selektionseffekte in Kraft treten, wodurch Fragen anders beantwortet werden können. Selektieren sich die beiden Gruppen der Respondenten und Nonrespondenten beispielsweise durch den Einsatz eines Anreizes systematisch und nicht zufällig voneinander, liegt kein MAR mehr vor und es kann ein Bias in den Antworten entstehen. Der Vergleich der Bildungsverteilung in Kapitel 7.2 (Zur Wirkung von Teilnahmeanreizen) hat bereits gezeigt, dass sich durch den Einsatz monetärer Anreize, die unabhängig von der Teilnahme vergeben werden, das Potential für einen Bias durch systematische Ausfälle reduzieren lässt. Zusätzlich konnte nachgewiesen werden, dass sich die Vergabe von Anreizen nicht derart auf die Qualität der Daten ausgewirkt hat, dass eine zusätzliche Fehlerquelle durch einen höheren Anteil an Item-Nonresponse entstanden wäre.

Die Analysen dieses Kapitels knüpfen unter anderem an dieser Stelle an und versuchen, den Effekt der einzelnen Designbedingungen der Nachrekrutierung aus dem Jahr 2012 im Rahmen einer multivariaten Analyse in Bezug zur Bruttostichprobe zu untersuchen. Hierdurch wird das Spektrum an zur Verfügung stehenden Variablen für die Untersuchung aber deutlich eingeschränkt (Engel und Schmidt, 2015).

Eine Teilfragestellung dieser Arbeit bezieht diesen Aspekt explizit mit ein fragt danach, ob ein möglicher Bias durch systematische Ausfälle zu operationalisieren ist. Die dahinterstehende Systematik und der Zusammenhang zwischen der Teilnahmewahrscheinlichkeit ( $p$ ), Zielvariable ( $Y$ ) sowie Hintergrundvariable ( $X$ ) ist zur Veranschaulichung in von

Abbildung 7.5 nochmals abgebildet. Zur Operationalisierung eines möglichen Bias durch systematische Ausfälle können, wie bereits ausführlich diskutiert, lediglich Variablen mit in die Schätzgleichung aufgenommen werden, die für alle Personen in der Stichprobe zur Verfügung stehen. Denn als abhängige Variable wird hierfür, wie auch schon im Ausgangsmodell, das letztliche Teilnahmeverhalten ( $p$ , dichotom ausgeprägt - 0 = Nein; 1 = Ja) herangezogen. Dass die sogenannten Hintergrundvariablen ( $X$ ) für alle Personen in der Stichprobe zur Verfügung stehen müssen, ist allerdings nicht die einzige Eigenschaft, die auf den zur Anwendung kommenden Kreis an Hintergrundvariablen ( $X$ ) zutreffen sollte. Damit das Potential für eine Operationalisierung sowie die Adjustierung eines möglichen Bias durch systematische Ausfälle ausgenutzt werden kann, sollten weitere Eigenschaften zutreffen. Wie bereits im Rahmen von Kapitel 4.5 (Statistische Konzepte zur Kompensation von Verzerrungen) beschrieben, sollte die Variable einen starken Zusammenhang mit der interessierenden Zielvariable ( $Y$ ) aufweisen sowie gleichzeitig ein starker Prädiktor für die Teilnahmewahrscheinlichkeit ( $p$ ) sein (Olson, 2013, 144 f.).

Abbildung 7.5: Zusammenhang zwischen Zielvariable, Teilnahmeentscheidung und Hintergrundvariable - Wiederholung von Abbildung 4.1



Quelle: Bethlehem et al. (2011, 123)

Um den Zusammenhang zwischen  $p$  und  $X$  zu untersuchen sowie eine individuelle Teilnahmewahrscheinlichkeit ( $p$ ) für jede Person zu schätzen, wurden logistische Regressionsmodelle mit den zur Verfügung stehenden Hintergrundvariablen gerechnet, deren Ergebnisse in Tabelle 7.16 sowie 7.17 abgebildet sind.

Die einzelnen Modelle beschränken sich dabei auf die im Rahmen von Abbildung 7.13

dargestellten Teilgruppen der Nachrekrutierung aus dem Jahr 2012. Die Modelle beziehen dabei die jeweiligen Designbedingungen der Incentivierung (fünf Euro nicht konditional/Verlosung konditional), die Art und Weise der Kontaktaufnahme sowie die Tatsache, ob eine Rufnummer festgestellt werden konnte oder nicht, in die Schätzgleichung mit ein. Durch den Bezug auf die unterschiedlichen Teilgruppen unterscheiden sich daran anknüpfend die Fallzahlen zwischen den einzelnen Modellen.

Tabelle 7.16: Logistische Regressionen mit der Teilnahme am Rekrutierungsinterview (0 = Nein; 1 = Ja) als abhängige Variable für jeweils unterschiedliche Designbedingungen

	Rufnummer bekannt (vs. unbekannt)		Kontaktierung 1 (A vs. B)		Kontaktierung 2 (B vs. C)	
	OR	s.e.	OR	s.e.	OR	s.e.
Geschlecht (Ref. Weiblich)	2,36 ***	0,35	1,90 **	0,39	2,64 ***	0,52
Alter in Jahren	1,08 ***	0,03	1,12 **	0,05	1,07 *	0,03
Alter in Jahren <sup>2</sup>	0,99 ***	0,00	0,99 **	0,00	0,99 *	0,00
<i>Stadtteile</i> (Ref. Walle)						
Blumenthal	0,79	0,61	0,89	0,44	1,34	0,61
Burglesum	2,41 *	1,00	2,37	1,49	2,50 +	1,28
Häfen	1,28	0,56	2,58 +	1,41	0,71	0,55
Neustadt	1,39	0,56	0,83	0,53	1,89	0,90
Oberneuland	1,91 +	0,66	2,08 +	0,91	1,42 +	0,76
Obervieland	1,38	0,44	1,26	0,55	1,34	0,61
Osterholz	1,27	0,36	1,43	0,52	0,80	0,36
Östliche Vorstadt	1,89 *	0,88	1,73	0,70	2,30 *	0,86
Schwachhausen	2,44 *	0,88	2,11 +	0,84	2,45 *	0,88
Vahr	1,24	0,33	1,19	0,42	1,09	0,41
Vege sack	0,68	0,22	0,81	0,33	0,39 +	0,22
Woltmershausen	1,39	0,37	0,97	0,36	1,70	0,58
Rufnummer bekannt	3,59 ***	0,52				
Anruf angekündigt (Ref. Anruf nicht angekündigt)			5,50 ***	1,10		
Anruf nicht angekündigt trotz bekannter Nummer (Ref. Anruf nicht angekündigt trotz unbekannter Nummer)					1,34	0,28
<i>Fallzahl (N)</i>	2423		844		2000	
<i>Modellgüte</i>						
Prob > chi <sup>2</sup>	***		***		***	
Pseudo R <sup>2</sup>	0,12		0,14		0,06	
LL	-464,67		-392,59		-464,95	

\*  $p \leq 0,05$ , \*\*  $p \leq 0,01$ , \*\*\*  $p \leq 0,001$ , +  $p \leq 0,10$

PPSM Access Panel Nachrekrutierung 2012, eigene Berechnungen

Anmerkung: Die Modelle beziehen sich jeweils auf unterschiedliche Teilgruppen der Stichprobe. Diese sind daher nicht miteinander zu vergleichen.

Das zieht die Konsequenz nach sich, dass die Ergebnisse zwischen den Modellen nicht miteinander vergleichbar sind. Aus den Modellen in Tabelle 7.16 ergibt sich für jede Person eine spezifische Teilnahmewahrscheinlichkeit ( $p$ ), die dann weiterverwendet wird, um mittels des RRM (Random Response Model) (Bethlehem, 2002) einen möglichen Zu-



sammenhang mit den gegebenen Antworten (Y) innerhalb der Befragung operationalisieren zu können.<sup>8</sup> Kann ein Zusammenhang nachgewiesen werden, der auch in der Grundgesamtheit signifikant von Null verschieden ist, ist dies ein Indikator für eine Über- oder Unterschätzung der interessierenden Zielgröße durch systematische Ausfälle in vorliegenden Modellen.

### Rufnummer bekannt vs. Rufnummer unbekannt

Wie erwartet hat die Tatsache, ob für eine Person eine Rufnummer festgestellt werden konnte oder nicht, einen positiven und höchst signifikanten Effekt darauf, ob sich eine Person an der Befragung beteiligt oder nicht (Hypothese  $H_{B1}$ ).

Tabelle 7.17: Fortsetzung von Tabelle 7.16

	Verlosung (F/H vs. G/I)		5 Euro (D vs. E)	
	OR	s.e.	OR	s.e.
Geschlecht (Ref. Weiblich)	2,75 ***	0,54	1,91 *	0,49
Alter in Jahren	1,07 *	0,03	1,16 **	0,07
Alter in Jahren <sup>2</sup>	0,99 *	0,00	0,99 **	0,00
<i>Stadtteile</i> (Ref. Walle)				
Blumenthal	1,34	0,61	0,31	0,23
Burglesum	2,48 +	1,27	4,00	3,57
Häfen	0,72	0,56	2,11	1,57
Neustadt	1,87	0,90	0,51	0,38
Oberneuland	1,45	0,78	2,17	1,18
Obervieland	1,33	0,60	1,43	0,75
Osterholz	0,81	0,36	1,66	0,75
Östliche Vorstadt	2,35 *	0,88	1,59	0,81
Schwachhausen	2,44 *	0,88	1,83	0,95
Vahr	1,09	0,41	1,28	0,55
Vegesack	0,39 +	0,22	0,87	0,42
Woltmershausen	1,69	0,58	0,82	0,37
Verlosung (Ref. keine Verlosung)	0,87	0,16		
Fünf Euro (Ref. keine fünf Euro)			3,78 ***	0,68
<i>Fallzahl (N)</i>	2000		423	
<i>Modellgüte</i>				
Prob > chi <sup>2</sup>	***		***	
Pseudo R <sup>2</sup>	0,06		0,12	
LL	-465,54		-243,30	

\* p ≤ 0,05, \*\* p ≤ 0,01, \*\*\* p ≤ 0,001, + p ≤ 0,10

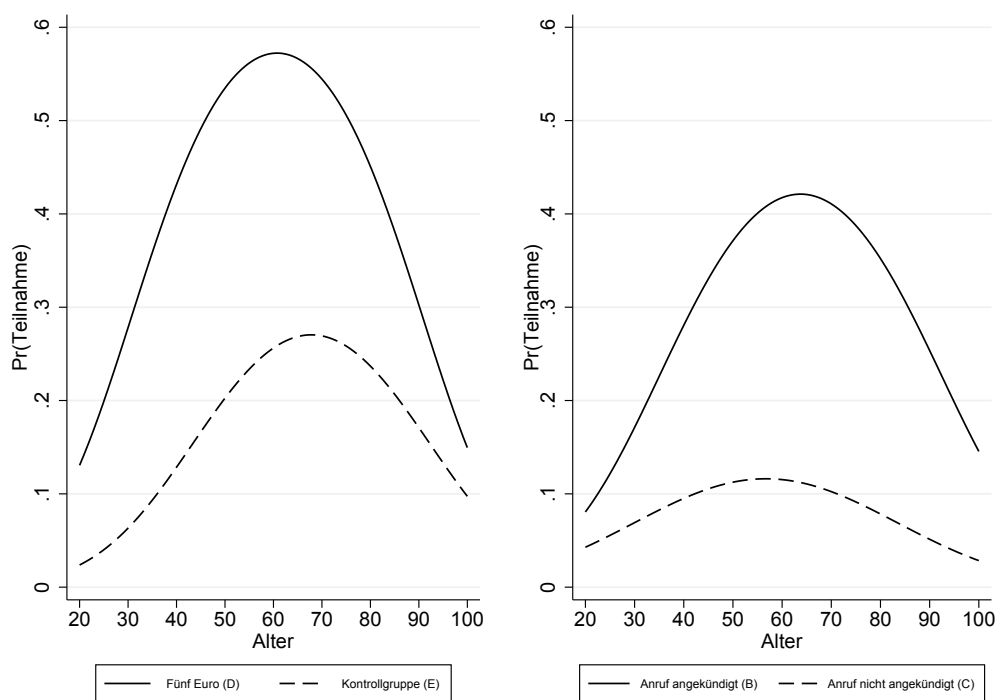
PPSM Access Panel Nachrekrutierung 2012, eigene Berechnungen

Anmerkung: Die Modelle beziehen sich jeweils auf unterschiedliche Teilgruppen der Stichprobe. Diese sind daher nicht miteinander zu vergleichen.

<sup>8</sup> Für die mathematischen Grundlagen des RRM siehe Kapitel 4.5 (Statistische Konzepte zur Kompensation von Verzerrungen).

Nicht zu allen Personen in der Stichprobe, die über einen Telefonbucheintrag verfügen, konnte die Rufnummer auch tatsächlich erfolgreich zugewiesen werden. So ist davon auszugehen, dass beide Untersuchungsgruppen konfundiert sind und in der Gruppe derer, deren Rufnummer nicht eruiert werden konnte, ebenso Personen enthalten sind, die über einen Eintrag verfügen. Etwaige Unterschiede werden tendenziell also eher unterschätzt (Engel und Schmidt, 2015). Hierbei darf allerdings nicht außer Acht gelassen werden, dass davon auszugehen ist, dass der Effekt des Rekrutierungsdesigns hierbei mit transportiert wird. Denn konnte eine Rufnummer zugewiesen werden, wurde in Untersuchungsgruppe A (siehe Tabelle 7.13) folglich auch angerufen.

Abbildung 7.6: Geschätzte Wahrscheinlichkeiten der Umfrageteilnahme für die felddexperimentellen Designbedingungen 5 Euro (vs. Kontrollgruppe)/Anruf angekündigt (vs. Anruf nicht angekündigt)



$N = 423$  (5 Euro vs. Kontrollgruppe) /  $N = 844$  (Anruf angekündigt vs. nicht angekündigt)  
PPSM Access Panel Nachrekrutierung 2012, eigene Berechnungen.

Dieser Anruf ähnelt in seiner Funktion einem klassischen Reminder und hat, wie bereits am Vergleich der erzielten Antwortraten zwischen den Untersuchungsgruppen deutlich geworden ist, zu einer Erhöhung der Antwortrate geführt. Denn die Personen in Untersuchungsgruppe A wurden im Gegensatz zu denjenigen in Untersuchungsgruppe B zusätzlich bis zu zehn mal angerufen. Darüber hinaus wurden in dieser Gruppe monetäre Anreize unabhängig von der Teilnahme eingesetzt, die weiterhin zu einer deutlichen Steigerung

der Antwortrate geführt haben. Somit kann man an dieser Stelle nicht ausschließen, dass eben diese in dem Modell nicht enthaltenen Drittvariablen (Telefonanruf und Anreiz) für Effektrichtung und -stärke mit verantwortlich sind.

### **Kontaktierung 1 - Untersuchungsgruppe A vs. B**

Wie der Blick auf das zweite Modell in Verbindung mit Abbildung 7.6 zeigt, trägt die Vorankündigung des Telefonanrufes in Untersuchungsgruppe A dazu bei, dass sich hier wesentlich mehr Personen an der Befragung beteiligt haben. Der Effekt ist zudem höchst signifikant. Somit lässt sich die Erhöhung der Antwortrate durch eben diese Designbedingung mit der multivariaten Analyse stützen. An dieser Stelle bestätigt sich, dass eine postalische Vorankündigung in Verbindung mit der Ankündigung eines Telefonanrufes ein probates Mittel zur Erhöhung der Antwortrate sowie Teilnahmewahrscheinlichkeit darstellt. Mit 14 Prozent kann dieses Modell zudem den größten Anteil der Varianz der abhängigen Variable aufklären. Vor dem Hintergrund der von Olson (2013) aufgestellten Kriterien kann allerdings weder der Aspekt der erfolgreichen Eruierung der Rufnummer noch die Ankündigung des Anrufes als starker Prädiktor der Teilnahmewahrscheinlichkeit bezeichnet werden. Somit wird das Potential zur Reduzierung eines Bias hierdurch notwendigerweise reduziert.

### **Kontaktierung 2 - Untersuchungsgruppe B vs. C**

Das darauffolgende Modell schließt den experimentellen Vergleich zwischen den Untersuchungsgruppen B und C mit ein. In Untersuchungsgruppe B wurde trotz bekannter Rufnummer der gleiche Zugang gewählt wie in Gruppe C. Hier wurde lediglich angeboten, die Umfrage online auszufüllen oder den vorfrankierten Rückantwortbrief für die Teilnahme an der telefonischen Variante zu nutzen. Daraus ergibt sich die Frage, ob sich Personen mit und ohne Eintrag im Telefonbuch in Bezug zu ihrem Teilnahmeverhalten unterscheiden. Im Rahmen der deskriptiven Betrachtung der Antwortraten wurde bereits festgestellt, dass sich diese voneinander unterscheiden und sich Personen in Untersuchungsgruppe B etwas häufiger beteiligt haben (B - 9,5 Prozent/ C - 6,1 Prozent) (Engel und Schmidt, 2013). Dieses Ergebnis wird durch die multivariate Analyse nicht bestätigt. Die betreffende Variable hat keinen signifikanten Effekt auf die Teilnahmewahrscheinlichkeit und das Modell kann sechs Prozent der Varianz der abhängigen Variable aufklären.

### **Anreiz: Verlosung - Untersuchungsgruppe F/H vs. G/I**

Wie aufgrund der Auswirkung der Verlosungsbedingung auf die Antwortrate in der Nachrekrutierung bereits zu erwarten gewesen ist, zeigt sich der Effekt im logistischen Modell als nicht signifikant. Auch in der multivariaten Analyse hat die Verlosung keinerlei Effekt auf die Wahrscheinlichkeit der Teilnahme und erklärt lediglich sechs Prozent der Varianz der abhängigen Variable. Einerseits kann dies darin begründet liegen, dass eine Verlosung als unseriös wahrgenommen wird. Andererseits ist der Hinweis auf die Verlosung im Postskriptum auf der zweiten Seite des Anschreibens verortet gewesen, was die Annahme nahelegt, dass der Hinweis von den potentiellen Teilnehmern häufig nicht zur Kenntnis genommen wurde.

### **Anreiz: 5 Euro - Untersuchungsgruppe D vs. E**

Die Vergabe von fünf Euro unabhängig davon, ob sich eine Person an der Befragung beteiligt oder nicht, hat mit Blick auf das entsprechende Modell in Tabelle 7.17 einen höchst signifikanten positiven Effekt auf die Teilnahmewahrscheinlichkeit. Diese Form der Incentivierung ist zwar relativ kostenintensiv, wirkt sich aber deutlich positiv auf die Teilnahmeentscheidung aus. Darüber hinaus konnte bereits gezeigt werden, dass sich hierdurch insbesondere geringer gebildete Personen angesprochen fühlen, die in der Regel bei der Durchführung wissenschaftlicher Umfragen unterrepräsentiert sind. Der Einsatz lohnt sich also gleich in doppelter Hinsicht, wie auch der Blick auf Abbildung 7.6 zeigt. Das Niveau der Teilnahmewahrscheinlichkeit liegt jeweils deutlich über dem der Kontrollgruppe, die keinen Anreiz erhalten hat. Inwiefern letztlich ein Zusammenhang zu den erhobenen Variablen innerhalb der Befragung besteht und ein Bias durch selektiven Nonresponse aufgrund dieser Designbedingung vorliegt, wird im folgenden Abschnitt diskutiert.

Die Designaspekte aus dem heuristischen Gesamtmodell (siehe Abbildung 5.1 auf Seite 80), wie die Art und Weise der Stichprobenziehung (A), das Mixed-Mode Design (B), die postalische Vorankündigung (C) und die Teilnahmeanreize (D), haben sich als probates Mittel zur Erhöhung der Teilnahmewahrscheinlichkeit sowie der erzielten Antwortrate erwiesen. Hiervon ausgenommen werden muss die Verlosung, der wie erwartet kein positiver Effekt nachgewiesen werden konnte. An dieses Ergebnis knüpft eine der zentralen Fragen der vorliegenden Arbeit an. Die einzelnen Designaspekte können trotz des positiven Effekts auf Antwortrate und Teilnahmewahrscheinlichkeit zu Verzerrungen in den beobachteten Antworten durch systematische Selektion führen. Denn es ist aufgrund vorliegender Studien

nicht davon auszugehen, dass es zufällig ist, welche Personen etwa auf Anreize oder eine Vorankündigung der telefonischen Kontaktaufnahme reagieren. Gleichzeitig muss untersucht werden, inwiefern etwa der Aspekt des Telefonbucheintrages zu systematischen Unterschieden führt und sich auf das Antwortverhalten auswirkt. Denn einerseits kann die Variation im Zugang sowie die Eigenschaften der Personen selbst einen Einfluss darauf haben, wie Fragen beantwortet werden.

Durch den Registerbezug der Stichprobenziehung und die Verfügbarkeit von Paradata wird es möglich, die vorliegenden Antworten genau auf diesen Aspekt hin zu untersuchen. An dieser Stelle setzt das folgende Kapitel an und versucht unter Anwendung des Random Response Model nach Bethlehem (2002) eine mögliche Unter- oder Überschätzung interessierender Umfragevariablen zu operationalisieren und auf die verschiedenen Designbedingungen der Nachrekrutierung aus dem Jahr 2012 zurückzuführen.

### **Nonresponse Bias: Zusammenhang zwischen Teilnahmewahrscheinlichkeit und vorliegenden Antworten**

Wie auch schon Kreuter und Olson (2011) festgestellt haben, ist es in der Praxis schwierig, einen Bias durch systematischen Nonresponse zu operationalisieren. Einerseits können lediglich Daten zur Anwendung kommen, die für alle Personen in der Stichprobe zur Verfügung stehen. Andererseits müssen diese Variablen sowohl in Verbindung zur Teilnahmewahrscheinlichkeit als auch zu den vorliegenden Antworten stehen, wenn ein Bias zusätzlich adjustiert werden soll. Das Set an Variablen beschränkt sich dadurch zumeist auf einen kleinen Kreis an Informationen, die zu diesem Zweck herangezogen werden können.

Vor diesem Hintergrund versucht dieses Kapitel zu operationalisieren, ob ein Bias durch systematische Ausfälle in Bezug zum eingesetzten Umfragedesign und den damit einhergehenden Variablen vorliegt oder nicht. Die Herangehensweise beschränkt sich dabei bewusst auf diesen Aspekt. Es wird hierbei nicht versucht, über Propensity Scores und ein Matchingverfahren zu gewichten oder Imputationsverfahren zur Anwendung zu bringen. Hierbei handelt es sich um andere Herangehensweisen an die Nonresponse-Problematik, die ihre Berechtigung haben, allerdings teilweise nicht eingesetzt werden können, weil die Voraussetzungen nicht erfüllt sind.<sup>9</sup>

Die spezifischen Teilnahmewahrscheinlichkeiten, die sich aus den multivariaten Modellen

---

<sup>9</sup> In Kapitel 4.5 (Statistische Konzepte zur Kompensation von Verzerrungen) wurden diese Verfahren und der diesbezügliche Stand der Forschung bereits ausführlich diskutiert.

des vorhergehenden Abschnitts ergeben haben, werden nun in das Random Response Model nach Bethlehem (2002) überführt. Die Ergebnisse sind in Tabelle 7.18 abgebildet. Hierbei ist zunächst einmal anzumerken, dass mit den gerechneten multivariaten Modellen, deren Ergebnisse in Tabelle 7.16 sowie 7.17 abgebildet sind, die Teilnahmewahrscheinlichkeit nicht in dem Maße aufgeklärt werden kann, wie von Olson (2013) beschrieben. Das Adjustierungspotential wird hierdurch letztlich eingeschränkt. Eine Einschätzung darüber, ob ein Bias in den vorliegenden Antworten vorliegt oder nicht, ist aber dennoch möglich. Denn an die unterschiedlichen Zugänge zu den Personen in der Nachrekrutierung schließt die Frage an, ob hiermit eine Variation im Antwortverhalten und somit ein Potential für eine Verzerrung einhergeht. Denn es können, wie bereits ausführlich diskutiert, bestimmte Designaspekte einer Umfrage unterschiedlich auf Personen wirken und somit systematische Selektionseffekte in Gang setzen, die es zu kontrollieren gilt.

Die mathematischen Grundlagen des RRM wurden in Kapitel 4.5 (Statistische Konzepte zur Kompensation von Verzerrungen) diskutiert. Diese Vorgehensweise wurde bereits in einer ersten Untersuchung von Engel und Schmidt (2015) erfolgreich genutzt. Das Verfahren wurde auf einen Teil der Fragen und die hierbei zustande gekommenen Antworten der Nachrekrutierung aus dem Jahr 2012 angewandt. Die folgenden Untersuchungen und Ergebnisse sind somit eine Fortsetzung des im Jahr 2015 erschienenen Sammelbandbeitrages in Schupp und Wolf (2015) (Nonresponse Bias: Qualitätssicherung sozialwissenschaftlicher Umfragen).

In der Untersuchung aus dem Jahr 2015 haben sich nur wenige Kovarianzen zwischen der Teilnahmewahrscheinlichkeit ( $p$ ) und den Zielvariablen ( $Y$ ) als signifikant erwiesen (Engel und Schmidt, 2015). Bei der Betrachtung der signifikanten Zusammenhänge hat sich allerdings ein Trend über die jeweiligen Modelle hinweg beobachten lassen: Die mittleren Werte in den untersuchten eurokritischen Aussagen ( $Y$ ) wurden in Abhängigkeit der Designbedingungen ( $x$ ) zum Teil unterschätzt und in den eurofreundlichen Aussagen ( $Y$ ) überschätzt (Engel und Schmidt, 2015, 274). Solch ein allgemeiner Trend lässt sich für die hier untersuchten Variablen nicht nachzeichnen.<sup>10</sup> Entgegen der Ergebnisse aus dem Jahr 2015, in der für die Bedingung der Rufnummereruierung (Rufnummer bekannt vs. nicht bekannt) fünf von neun Variablen eine Über- beziehungsweise Unterschätzung nachgewiesen werden konnte, wirkt sich diese auf die hier untersuchten Variablen nicht aus. Es ist allerdings nochmals anzumerken, dass es keine zu hundert Prozent eindeutige

---

<sup>10</sup> Die Fragetexte der in Tabelle 7.16 untersuchten Variablen sind dem Anhang zu entnehmen.

Tabelle 7.18: Nonresponse Bias in ausgewählten Variablen nach dem RRM Model von Bethlehem (2002)

		Auswahl- rahmen	Rufnummer bekannt	Kontaktierung 1 A vs. B	Kontaktierung 2 B vs. C	Verlosung F/H vs. G/I	Fünf Euro D vs. E
	$\bar{p}$	0,1197	0,1197	0,2299	0,0680	0,0680	0,3640
	$N$	2423	2423	844	2000	2000	423
EU10	$Cov_{pY}$	-0,0079	-0,0119	-0,0379	0,0039	0,0015	-0,1070 *
	$B(\bar{y}_R)$	-0,0659	-0,0996	-0,1649	0,0299	0,0225	-0,2940
	$n$	290	290	193	136	136	154
EU16	$Cov_{pY}$	-0,0085	-0,0040	-0,0498 *	0,0067	0,0068	-0,0098
	$B(\bar{y}_R)$	-0,0707	-0,0338	-0,2165	0,0509	0,0999	-0,0269
	$n$	271	271	175	133	133	138
EU17	$Cov_{pY}$	0,0161 +	0,0189	0,0165	0,0166 *	0,0173 *	0,0211
	$B(\bar{y}_R)$	0,1342	0,1577	0,0718	0,2436	0,2545	0,0579
	$n$	285	285	188	137	137	148
EU18	$Cov_{pY}$	0,0118	0,0039	0,0066	0,0046	0,0058	0,0227
	$B(\bar{y}_R)$	0,0982	0,0328	0,0287	0,0682	0,0859	0,0623
	$n$	285	285	188	137	137	148
EU19	$Cov_{pY}$	0,0098	0,0004	-0,0006	0,0081	0,0086	0,0154
	$B(\bar{y}_R)$	0,0817	0,0036	-0,0028	0,1193	0,1270	0,0422
	$n$	288	288	191	137	137	151
EU20	$Cov_{pY}$	-0,0014	0,0097	0,0140	-0,0214	0,0080	-0,0214
	$B(\bar{y}_R)$	-0,0117	-0,0808	0,0610	0,1025	0,1179	-0,0589
	$n$	288	288	191	137	137	151
S13	$Cov_{pY}$	0,0127	0,0127	-0,0128	0,0200	0,0215 +	0,0639
	$B(\bar{y}_R)$	0,1058	0,1058	-0,0556	0,2873	0,3168	0,1755
	$n$	288	288	191	136	136	152
S14	$Cov_{pY}$	0,0065	0,0298	-0,0181	0,0064	0,0079	0,0385
	$B(\bar{y}_R)$	0,0546	0,2492	-0,0785	0,0948	0,1167	0,1058
	$n$	290	290	193	137	137	153
S16	$Cov_{pY}$	-0,0184	-0,0083	-0,0570 *	0,0023	0,0044	0,0069
	$B(\bar{y}_R)$	-0,1533	-0,0696	-0,2480	0,0339	0,0642	0,0188
	$n$	288	288	191	137	137	151
S17	$Cov_{pY}$	-0,0070	-0,0033	-0,0453 +	0,0467	0,0081	-0,0064
	$B(\bar{y}_R)$	-0,0584	-0,0276	-0,1972	0,0833	0,1193	-0,0176
	$n$	287	287	190	137	137	150
S18	$Cov_{pY}$	0,0017	0,0037	-0,0484 +	0,0117	0,0141	0,0155
	$B(\bar{y}_R)$	0,0982	0,0328	-0,0287	0,0682	0,2078	0,0623
	$n$	290	290	193	137	137	153

\*  $p \leq 0,05$ , \*\*  $p \leq 0,01$ , \*\*\*  $p \leq 0,001$ , +  $p \leq 0,10$   
PPSM Access Panel Nachrekrutierung 2012, eigene Berechnungen

Zuordnung zu diesen beiden Gruppen geben konnte. Deswegen ist davon auszugehen, dass diese beiden Gruppen konfundiert sind und etwaige Unterschiede tendenziell unterschätzt werden (Engel und Schmidt, 2015). Auf diese möglichen Unterschiede geht das folgende Kapitel abschließend differenzierter ein.

Für Variable EU10, welche die subjektive Einschätzung der Befragten zur Betroffenheit von der Eurokrise repräsentiert, konnte eine signifikante Unterschätzung für die Bedingung nachgewiesen werden, ob fünf Euro vergeben worden sind oder nicht. Dabei handelt es sich allerdings um die einzige Verzerrung bezüglich dieser Designbedingung. In der Untersuchung aus dem Jahr 2015 wurden daran anknüpfend jeweils zwei mittlere Werte der eurofreundlichen Aussagen unterschätzt. Die Vergabe von fünf Euro unabhängig davon,

ob sich eine Person an der Umfrage beteiligt hat oder nicht, hat sich also nicht nur positiv auf die Antwortrate ausgewirkt, sondern gleichzeitig in drei von 20 untersuchten Variablen eine Verzerrungen verursacht.

Für die Variablen EU16 bis EU20 wurde im Fragebogen eine hypothetische Situation konstruiert und danach gefragt, inwiefern die Befragten bei Einführung eines Finanzausgleichs auf europäischer Ebene bereit wären, persönliche finanzielle Nachteile in unterschiedlichen Dimensionen in Kauf zu nehmen. Hierfür wurden ein späterer Renteneintritt oder erhöhte Steuern als Items zur Antwort auf einer siebenstufigen Skala von eins „auf keinen Fall auf mich nehmen“ bis sieben „auf jeden Fall auf mich nehmen“ angeboten. Für diese Gruppe an Variablen lassen sich über die Modelle hinweg lediglich vereinzelt signifikante Zusammenhänge zwischen der Teilnahmewahrscheinlichkeit ( $p$ ) und den gegebenen Antworten ( $Y$ ) sowie Unter- und Überschätzungen nachweisen. Die Ankündigung des Telefonanrufes (Kontaktierung 1 - Untersuchungsgruppen A vs. B) führt zu einer signifikanten Unterschätzung in der Variable EU16 (Bereitschaft, später in Rente zu gehen). Zusätzlich kann hier eine weitere Unterschätzung des mittleren Wertes für die Variable S16 (subjektive Einschätzung Glück auf elfstufiger Skala) nachgewiesen werden. Die Kombination aus der postalischen Ankündigung und des später erfolgten Telefonanrufes hat sich einerseits deutlich ausschöpfungssteigernd auf die Antwortrate ausgewirkt, während darüber hinaus in der Untersuchung aus dem Jahr 2015 eine signifikante Verzerrung nachgewiesen werden konnte (Engel und Schmidt, 2015). Somit konnte lediglich in drei von 20 untersuchten Variablen eine Verzerrung in Bezug auf diese Designbedingung beobachtet werden, was in Verbindung mit der Erhöhung der Antwortrate ein positives Ergebnis in doppelter Hinsicht ist.

Für das Modell das die Untersuchungsgruppen B und C miteinander vergleicht (Anruf nicht angekündigt, trotz bekannter Rufnummer - Untersuchungsgruppen B vs. C), liegt lediglich eine signifikante Kovarianz und daraus resultierende Überschätzung vor. In Kombination mit der vorhergehenden Untersuchung konnten für sechs von 20 Variablen Verzerrungen aufgrund dieser Designbedingung nachgewiesen werden. Letztlich sollte hiermit überprüft werden, ob sich Personen mit und ohne Eintrag im Telefonbuch hinsichtlich ihrer Teilnahmewahrscheinlichkeit unterscheiden. Die Antwortraten der beiden Gruppen (B - Rufnummer bekannt, 9,5 Prozent/ C - Rufnummer unbekannt, 6,1 Prozent) unterscheiden sich dabei signifikant voneinander (Engel und Schmidt, 2015). Mit dieser Bedingung gehen nicht zu ignorierende Verzerrungen einher, die auf systematische Unterschiede zwischen die-



sen beiden Gruppen zurückzuführen sind. Auch hier werden diese Unterschiede tendenziell unterschätzt. Wird ein solches Rekrutierungsdesign in der Praxis eingesetzt, muss dieser Aspekt bedacht und kontrolliert werden, denn andernfalls würden sich Verzerrungen in den Antworten und den letztlichen Ergebnissen niederschlagen. Der Einsatz der Verlosung hat sich als nicht vorteilhaft für die Qualität der Daten erwiesen. Einerseits konnte die Antwortrate hiermit nicht erhöht werden. Andererseits ist in sechs von 20 untersuchten Variablen eine Über- oder Unterschätzung der mittleren Werte in der interessierenden Population zu beobachten. Das Ergebnis für den Einsatz der Verlosung steht dem der fünf Euro somit relativ konträr gegenüber. Während sich die Vergabe der fünf Euro in doppelter Hinsicht als lohnenswert erwiesen hat, können der Verlosung keine positiven Effekte auf Antwortrate und Datenqualität nachgewiesen werden.

Vergleichbar mit der Untersuchung aus dem Jahr 2015 haben sich nicht viele Kovarianzen zwischen der Teilnahmewahrscheinlichkeit  $p$  und der Zielvariable  $Y$  als signifikant gezeigt. Es konnte für die relevanten Designaspekte überprüft werden, ob sich diese auf das spätere Antwortverhalten ausgewirkt haben. So kann eine bessere Bewertung dieser im Einzelnen stattfinden und über die Auswirkung auf die Antwortrate insgesamt ein differenzierteres Bild nachgezeichnet werden. Denn wie schon zu Anfang der vorliegenden Arbeit festgestellt wurde, sagt die Höhe der Antwortrate nicht notwendigerweise etwas darüber aus, wie sich gleichzeitig das Fehlerpotential dazu entwickelt. Von einem linearen Zusammenhang zwischen Antwortrate und Datenqualität kann man aufgrund vorliegender Studien und der hier diskutierten Ergebnisse nicht sprechen.

## Zusammenfassung der Ergebnisse

Mit den statistischen Analysen dieses Kapitels wurde speziell zwei der zentralen Fragen nachgegangen:

- Kann die Wirkung der eingesetzten feldexperimentellen Designelemente auf die Antwortraten in den jeweiligen Untersuchungsgruppen mittels multivariater Analyse bestätigt werden?
- Wie wirken sich die Designelemente auf die Qualität der Daten aus?

Die formulierten Forschungshypothesen konnten bestätigt werden. Die nicht konditionale Vergabe des Fünfeuroscheins hat sich auch in der multivariaten Analyse als ausschöpfungssteigernd erwiesen. Hypothese  $H_{D1}$  kann somit als vorläufig bestätigt betrachtet

werden. Durch den Einsatz von nicht konditionalen Anreizen lässt sich die Kooperationswahrscheinlichkeit und somit die Wahrscheinlichkeit einer Teilnahme erhöhen. Gleichzeitig konnte unter Anwendung des Random Response Models gezeigt werden, dass mit diesem Vorgehen keine zusätzlichen Verzerrungen einhergehen. Es hat sich in Abhängigkeit dieser Bedingung lediglich ein Zusammenhang zwischen Variable EU10 (Y) und der Teilnahmewahrscheinlichkeit (p) als signifikant gezeigt.

Tabelle 7.19: Entscheidung über die Hypothesen - Zielgrößen: Erreichbarkeit, Kooperationswahrscheinlichkeit, Teilnahmewahrscheinlichkeit

Hypothese	Vermutete Richtung des Zusammenhangs	Status
$H_{D1}$ : Fünfeuroschein (nicht konditional) $\Rightarrow$ Kooperationswahrscheinlichkeit	positiv	angenommen
$H_{D2}$ : Verlosung konditional $\Rightarrow$ Kooperationswahrscheinlichkeit	gleichbleibend	angenommen
$H_{B1}$ : Telefonbucheintrag $\Rightarrow$ Teilnahmewahrscheinlichkeit	positiv	angenommen
$H_{B3}$ : Telefonanruf $\Rightarrow$ Teilnahmewahrscheinlichkeit	positiv	angenommen
$H_{C1}$ : Postalische Vorankündigung $\Rightarrow$ Teilnahmewahrscheinlichkeit	positiv	angenommen

Ferner hat die Verlosung (Hypothese  $H_{D2}$ ) nicht zur Erhöhung der Teilnahmewahrscheinlichkeit beigetragen. Zusätzliche Verzerrungen konnten diesbezüglich in einer der untersuchten Variablen nachgewiesen werden. Des Weiteren wurde vermutet, dass Personen, die über einen Telefonbucheintrag verfügen auch eher dazu bereit sind, sich an Umfragen zu beteiligen (Hypothese  $H_{B1}$ ). Diese Annahme kann durch das erste Modell in Tabelle 7.16 bestätigt werden. Personen mit bekannter Rufnummer haben sich deutlich häufiger beteiligt. Letztlich stehen mit dieser Designbedingung keine zusätzlichen Verzerrungen in Verbindung, da die Anwendung des Random Response Models keine signifikanten Kovarianzen in Abhängigkeit von diesem Aspekt gezeigt hat. Die postalische Vorankündigung (Hypothese  $H_{C1}$ ) in Verbindung mit einem Telefonanruf (Hypothese  $H_{B3}$ ) hat sich als probates Mittel erwiesen, um Antwortrate und Teilnahmewahrscheinlichkeit zu erhöhen. Personen in dieser Untersuchungsgruppe haben sich deutlich häufiger beteiligt. Letztlich sind mit dieser Bedingung zwei signifikante Zusammenhänge in den Umfragevariablen verbunden. Die mittleren Werte in den Aussagen zu EU16 sowie S16 werden unterschätzt. Eine Erhöhung der Antwortrate durch die Elemente eines Umfragedesings kann sich folglich nicht nur positiv auf die Antwortrate auswirken, sondern auch zusätzliche systematische Selektionseffekte in Gang setzen, die zu Verzerrungen in den Antworten führen.

## Analyse der CATI-Befragung mit Parainformationen

Es stehen über die zuvor verwendeten Informationen aus dem Auswahlrahmen zur Schätzung der Teilnahmewahrscheinlichkeit hinaus weitere Parainformationen zur Verfügung. Diese sollen nicht unberücksichtigt bleiben, denn sie können dazu verwendet werden, das eigene Umfragedesign zu bewerten und anzupassen.

Die Personen in der Stichprobe hatten in jeder Untersuchungsgruppe die Möglichkeit, sich für einen der Befragungsmodi (CATI/CAWI) zu entscheiden. Betrachtet man nun die Gruppe derer, die zusätzlich zum Vorausbrief angerufen worden sind, bezieht dies vor allem die Personen in Untersuchungsgruppe A mit ein. Dies trifft aber nicht ausschließlich zu, denn jeder hatte hierbei ebenso die Möglichkeit den Fragebogen per Telefon zu beantworten, auch wenn die entsprechende Rufnummer nicht festgestellt werden konnte. Der Anteil an Personen, welche die Möglichkeit genutzt und den vorfrankierten Briefumschlag inklusive ausgefüllter Karte und Rufnummer zurückgesandt hat, ist mit  $n = 30$  allerdings relativ klein. Hierbei werden zur Schätzung der Teilnahmewahrscheinlichkeiten alle Informationen mit einbezogen, die unabhängig davon zur Verfügung stehen, ob sich eine Person letztlich an der CATI-Umfrage beteiligt hat oder nicht. Da sich diese Analysen auf eine Teilgruppe der Gesamtstichprobe beziehen, wird die Fallzahl in den einzelnen Modellen notwendigerweise kleiner. Die Dummy-Terme, welche die Informationen über die Stadtteile enthalten, können hierbei nicht berücksichtigt werden. Die Fallzahlen in den einzelnen Kategorien dieser Variablen sind teilweise sehr klein, sodass aufgrund hoher Standardfehler in diesen Modellen keine gesicherten Erkenntnisse gewonnen werden können.

Über die Befragungssoftware VOXCO werden Parainformationen wie die Häufigkeit von Kontaktversuchen oder der Vereinbarung von Terminen automatisiert erhoben, sodass diese zu Analysezwecken herangezogen werden können. Tabelle 7.20 zeigt die Ergebnisse schrittweiser logistischer Regressionen mit der Teilnahmewahrscheinlichkeit in Bezug zu allen Personen, die angerufen worden sind.

Das Geschlecht hat keinen Einfluss auf die Wahrscheinlichkeit der Teilnahme. In Bezug zur gesamten Bruttostichprobe war der Effekt noch höchst signifikant, was unter anderem in der deutlich größeren Fallzahl begründet liegen kann. Wie bereits zuvor vermutet, hat das Geschlecht für die Teilgruppe der angerufenen Personen in der Stichprobe keinen Effekt auf die Teilnahme. Darüber hinaus besteht in Modell 1 und Modell 2 ein Zusammenhang mit dem Alter, der seine Signifikanz allerdings bis hin zum finalen Modell 5 verliert. Unter der Kontrolle weiterer Variablen ist das Alter einer Person nicht mehr übertragbar auf die

Tabelle 7.20: Logistische Regression mit der Teilnahme am Rekrutierungsinterview (0 = Nein; 1 = Ja ) als abhängige Variable und Parainformationen aus der CATI-Befragung

	M1		M2		M3		M4		M5	
	OR	s.e.	OR	s.e.	OR	s.e.	OR	s.e.	OR	s.e.
Geschlecht (Ref. Weiblich)	1,52	0,41	1,56	0,42	1,62 +	0,45	1,64 +	0,46	1,51	0,43
Alter in Jahren	1,16 *	0,08	1,13 *	0,07	1,13 +	0,07	1,12 +	0,08	1,11	0,08
Alter in Jahren <sup>2</sup>	0,99 *	0,00	0,99 +	0,00	0,99	0,00	0,99	0,00	0,99	0,00
Rufnummer bekannt	0,03 ***	0,03	0,04 ***	0,04	0,04 ***	0,03	0,03 ***	0,03	0,02 ***	0,02
Anzahl Kontaktversuche	0,91 +	0,05	1,67 *	0,41	2,54 *	1,13	2,82 *	1,16	1,73	0,87
Anzahl Kontaktversuche <sup>2</sup>			0,93 *	0,03	0,82 *	0,08	0,81 *	0,08	0,87	0,10
Anzahl Kontaktversuche <sup>3</sup>					1,01 +	0,05	1,01 +	0,00	1,00	0,01
Anrufbeantworter besprochen (Ref. nicht besprochen)					1,63	0,53	1,61	0,52	2,04 *	0,72
Anreiz: Fünf Euro (Ref. kein Anreiz)							2,47 ***	0,63	2,42 ***	0,62
Summe Termin mit KP/ZP									1,62 **	0,30
<i>Fallzahl (N)</i>	378		378		378		378		378	
<i>Modellgüte</i>										
Prob > chi <sup>2</sup>	***		***		***		***		***	
Pseudo R <sup>2</sup>	0,08		0,10		0,11		0,14		0,16	
LL	-208,65		-204,66		-202,51		-195,95		-191,45	
LR-Test			**				***		**	

\* p ≤ 0,05, \*\* p ≤ 0,01, \*\*\* p ≤ 0,001, + p ≤ 0,10

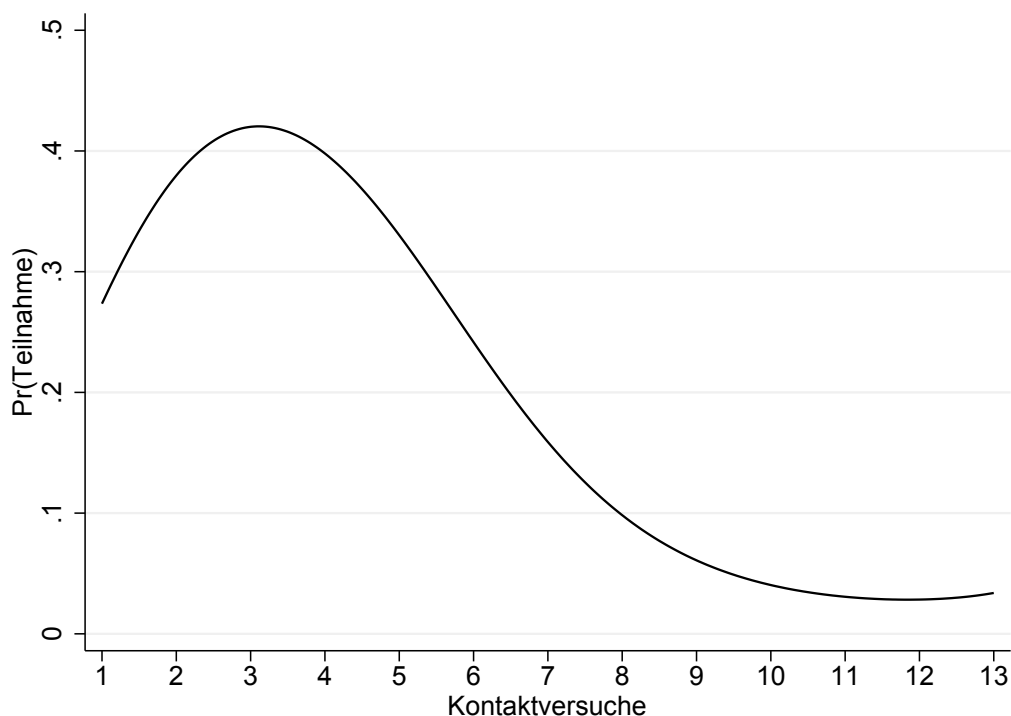
PPSM Access Panel Nachrekrutierung 2012, eigene Berechnungen

Anmerkung: KP = Kontaktperson; ZP = Zielperson

Grundgesamtheit. Auch unter Hinzunahme einer kubischen Effektkomponente hat sich daran nichts geändert.

Ob die Rufnummer zuvor durch die Recherche und das Matching der jeweiligen Datenbanken erfolgreich eruiert werden konnte, hat nach wie vor einen höchst signifikanten Einfluss. Hierbei muss man sich zunächst vergegenwärtigen, welche beiden Gruppen hierbei miteinander verglichen werden. Anders als zuvor bei der Analyse der gesamten Bruttostichprobe bezieht diese Variable nun zusätzlich die Information mit ein, welche Personen die Gelegenheit genutzt und den frankierten Briefumschlag inklusive Rufnummer zurückgesendet haben. Die Teilgruppe, deren Rufnummer bereits im Vorhinein festgestellt werden konnte, besitzt nun eine deutlich geringere geschätzte Teilnahmewahrscheinlichkeit. Der Effekt der entsprechenden Variable in Tabelle 7.20 ist zudem höchst signifikant. Sind die Personen in der Stichprobe also selbst aktiv geworden und haben ihre Rufnummer zurückgesendet, haben diese sich dann auch mehrheitlich an der Befragung beteiligt. Hier wäre weiterhin zu überprüfen, inwiefern sich diese beiden Gruppen in Bezug zu relevanten Merkmalen unterscheiden. Denn Personen, die für die Teilnahme an einer Befragung selbst aktiv werden, unterscheiden sich höchst wahrscheinlich systematisch vom Rest der

Abbildung 7.7: Geschätzte Wahrscheinlichkeiten der Umfrageteilnahme differenziert nach Kontaktversuchen



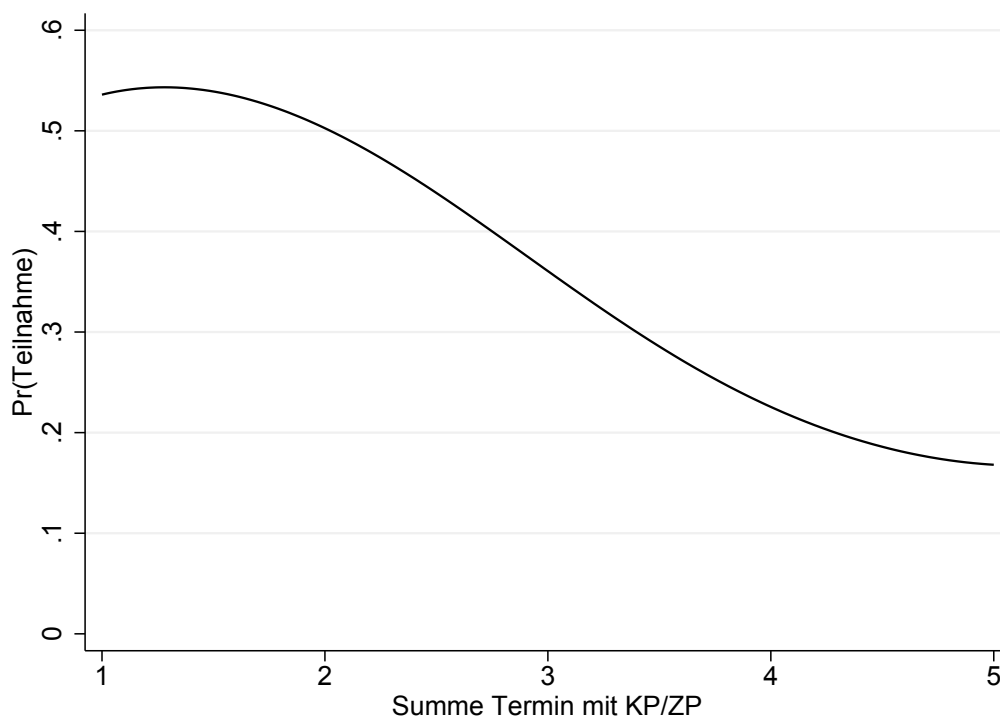
N = 378; PPSM Access Panel Nachrekrutierung 2012, eigene Berechnungen.

Stichprobe. Dieser Frage wird in Kapitel 7.4 weiter nachgegangen.

Modell 1 bis Modell 3 beziehen zusätzlich den Effekt der unternommenen telefonischen Kontaktversuche mit ein. Die lineare Komponente zeigt sich hierbei in Modell 1 als nicht signifikant. Wird in Modell 2 zusätzlich ein quadratischer Effekt mit einbezogen, sind beide auf die Grundgesamtheit übertragbar. In Abbildung 7.7 ist der Einfluss der Kontaktversuche auf die Wahrscheinlichkeit der Teilnahme grafisch abgebildet. Daran erkennt man, dass es insbesondere bei den ersten Kontaktversuchen am wahrscheinlichsten ist, dass die Personen erreicht werden und der erbetenen Teilnahme nachkommen. Liegt das Niveau der geschätzten Wahrscheinlichkeiten beim ersten Versuch noch bei unter 30 Prozent, wird es mit zunehmenden Kontaktversuchen immer wahrscheinlicher, dass sich eine Person beteiligt. Am wahrscheinlichsten ist diese dann bei drei stattgefundenen Versuchen, während diese daraufhin immer weiter abfällt und ab elf ihr niedrigstes Niveau erreicht. Dies lässt den Schluss zu, dass zehn Kontaktversuche ausreichend sind und für CATI-Befragungen hier eine Grenze gesetzt werden kann.

Engel (2013) hat den Einfluss der Kontaktversuche auf die Teilnahmewahrscheinlichkeit für die Rekrutierungsphase des PPSM Access Panels ebenso untersucht und ist zu ähnli-

Abbildung 7.8: Geschätzte Wahrscheinlichkeiten der Umfrageteilnahme differenziert nach der Summe an Terminen mit der Kontakt- oder Zielperson



N = 378; PPSM Access Panel Nachrekrutierung 2012, eigene Berechnungen.

chen Ergebnissen gekommen. Entgegen der formulierten Annahme, dass das Besprechen des Anrufbeantworters durch den Interviewer einen positiven Effekt hat, zeigt sich die entsprechende Variable in Modell 3 als nicht relevant. Die Erklärungskraft des Modells verbessert sich nicht und der Effekt ist zudem nicht auf die Grundgesamtheit übertragbar. Wobei anzumerken ist, dass die Anzahl an Fällen im Datensatz, deren Anrufbeantworter vom Interviewer besprochen wurde, relativ klein ist. Nichtsdestotrotz kann dem Besprechen des Anrufbeantworters im Kontaktverlauf kein Effekt auf die Teilnahme nachgewiesen werden.

In die erwartete Richtung geht der Einfluss des vergebenen Anreizes von fünf Euro. Dieser ist höchst signifikant und in die erwartete Richtung ausgeprägt: Personen, die fünf Euro als Anreiz unabhängig von der Teilnahme erhalten haben, besitzen eine um 142 Prozent höhere geschätzte Wahrscheinlichkeit, sich an der Befragung zu beteiligen.

Das finale Modell 5 der CATI-Teilgruppe wird durch die Variable ergänzt, welche die kumulierten Terminvereinbarungen mit der Kontakt- (KP) oder Zielperson (ZP) operationalisiert. Aus der Forschung zur initialen Teilnahme am PPSM Access Panel ist bereits bekannt, dass eine Terminvereinbarung im Kontaktverlauf die Wahrscheinlichkeit

der Teilnahme erhöht. Zur Veranschaulichung des Zusammenhangs ist dieser in Abbildung 7.8 grafisch dargestellt.

Es zeigt sich mit Blick auf Tabelle 7.20, dass die Vereinbarung eines Termins einen durchaus höchst signifikant positiven Effekt auf die Teilnahme sowie die Erklärungsleistung des Modells hat. Betrachtet man Abbildung 7.8, offenbart sich ein anderes Bild. Werden ein oder auch zwei Termine vereinbart, liegt die geschätzte Wahrscheinlichkeit der Teilnahme noch bei über 50 Prozent. Dies ist einerseits der bisher größte geschätzte Wert in dieser Kategorie und dieser Teilgruppe. Andererseits ist aber auch zu erkennen, dass die Wahrscheinlichkeit ab dem dritten Termin deutlich abnimmt und bei der fünften Vereinbarung bereits unter 20 Prozent liegt. Überraschend ist dieses Ergebnis allerdings nicht, denn in der Praxis liegt es auf der Hand, dass Personen, die mehr als zwei Termine für einen Rückruf vereinbaren, offensichtlich schlecht zu erreichen sind. Werden diese erreicht, scheinen sie offensichtlich nicht genug zeitlichen Freiraum für eine Teilnahme zu haben. Dessen ungeachtet ist es allerdings durchaus sinnvoll, Rückruftermine zu vereinbaren. Ein oder zwei Termine erhöhen die Wahrscheinlichkeit einer Teilnahme deutlich.

### Zusammenfassung der Ergebnisse

Entgegen der Erwartung, hat sich das Alter als nicht erklärungsrelevant erwiesen. Hypothese  $H_{A1}$  muss in Bezug zur CATI-Teilgruppe abgelehnt werden. Zu beachten ist, dass es sich hierbei um eine besondere Gruppe handelt. Denn es wurden fast ausschließlich Personen telefonisch kontaktiert, die über einen Eintrag in das Telefonbuch verfügen. Da es sich dabei vor allem um ältere Personen handelt, kann das Fehlen einer Variation in der Altersvariable dazu führen, dass kein erklärungsrelevanter Effekt nachgewiesen werden konnte.

Tabelle 7.21: Entscheidung über die Hypothesen - Zielgröße: Erreichbarkeit

Hypothese	Vermutete Richtung des Zusammenhangs	Status
$H_{A1}$ : Alter $\Rightarrow$ Erreichbarkeit	positiv	abgelehnt
$H_{A2}$ : Kontaktversuche $\Rightarrow$ Erreichbarkeit	negativ	angenommen
$H_{A3}$ : Anrufbeantworter $\Rightarrow$ Erreichbarkeit	negativ	abgelehnt
$H_{A4}$ : Termin mit KP/ZP $\Rightarrow$ Erreichbarkeit	positiv	angenommen

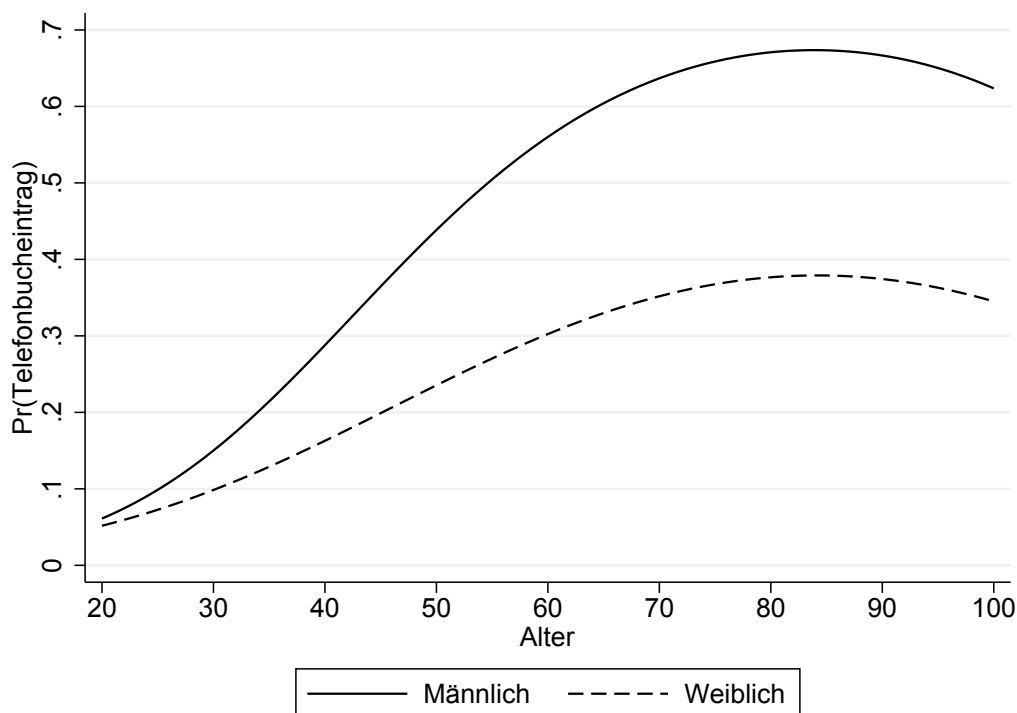
Die Kontaktversuche hängen nicht linear mit der Erreichbarkeit und der sich letztlich daraus ergebenden Wahrscheinlichkeit einer Teilnahme zusammen  $H_{A2}$ . Bis zum fünften

Kontaktversuch erhöht sich diese, um daraufhin kurvenförmig abzunehmen. Ob ein Anrufbeantworter in den Kontaktverlauf involviert war, hat keinen Einfluss darauf, ob die potentiellen Teilnehmer besser zu erreichen sind und sich somit auch häufiger beteiligen ( $H_{A3}$ ). Sehr wohl einen Unterschied macht es aus, ob mit der Zielperson oder Kontaktperson ein Termin für einen Rückruf vereinbart wurde oder nicht (Hypothese  $H_{A4}$ ). Liegt die Teilnahmewahrscheinlichkeit nach der Vereinbarung eines Termins noch bei über 50 Prozent, fällt diese mit zunehmender Anzahl an Vereinbarungen ab und erreicht schließlich bei fünf Terminen ein Niveau von unter 20 Prozent.

## 7.4 Unterschiede zwischen eingetragenen und nicht eingetragenen Personen

Abschließend wird der Frage nachgegangen, inwiefern sich vor dem Hintergrund der verfügbaren Informationen durch die registerbezogene Stichprobe Unterschiede zwischen Personen zeigen, die über einen Telefonbucheintrag verfügen oder nicht.

Abbildung 7.9: Geschätzte Wahrscheinlichkeiten des Telefonbucheintrages differenziert Geschlecht und Alter



N = 2423; PPSM Access Panel Nachrekrutierung 2012, eigene Berechnungen.



Zur Beantwortung dieser Frage wurden schrittweise logistische Regressionsmodelle mit dem Telefonbucheintrag als abhängiger Variable geschätzt. Die Ergebnisse in Tabelle 7.22 können allerdings aufgrund der nicht eindeutigen Zuordnung zu den beiden Gruppen der abhängigen Variable nur mit Einschränkungen interpretiert werden. Der Blick auf die Ergebnisse sowie Abbildung 7.9 zeigt, was zuvor bereits vermutet wurde. Es können sich deutliche Unterschiede zwischen Personen mit und ohne Telefonbucheintrag hinsichtlich Geschlecht und Alter nachgewiesen werden. Es sind eher die Männer, die über einen Eintrag verfügen. Der Blick auf Abbildung 7.9 zeigt die Niveauunterschiede zwischen den Geschlechtern deutlich. Auch die Abhängigkeit des Merkmals Telefonbucheintrag vom Alter der Personen ist zu erkennen.

Tabelle 7.22: Logistische Regression mit dem Eintrag im Telefonbuch (0 = Nein; 1 = Ja ) als abhängige Variable für die Bruttostichprobe

	M1		M2		M3		M4	
	OR	s.e.	OR	s.e.	OR	s.e.	OR	s.e.
Geschlecht (Ref. Weiblich)	2,46 ***	0,22	2,76 ***	0,26	2,71 ***	0,26	2,36 ***	0,06
Alter in Jahren			1,04 ***	0,00	1,14 ***	0,01	1,08 ***	0,03
Alter in Jahren <sup>2</sup>					0,99 ***	0,00	0,99 ***	0,00
<i>Stadtteile</i> (Ref. Walle)								
Blumenthal							0,99	0,23
Burglesum							0,89	0,30
Häfen							1,28	0,39
Neustadt							0,81	0,23
Oberneuland							1,76 *	0,45
Obervieland							0,99	0,22
Osterholz							1,19	0,23
Östliche Vorstadt							1,75 **	0,37
Schwachhausen							1,11	0,23
Vahr							1,17	0,21
Vegeack							1,07	0,21
Woltmershausen							0,95	0,17
<i>Fallzahl (N)</i>	2423		2423		2423		2423	
<i>Modellgüte</i>								
Prob > chi <sup>2</sup>	***		***		***		***	
Pseudo R <sup>2</sup>	0,03		0,12		0,13		0,13	
LL	-1512,98		-1374,65		-1358,03		-1349,77	
LR-Test			***		***			

\*  $p \leq 0,05$ , \*\*  $p \leq 0,01$ , \*\*\*  $p \leq 0,001$ , +  $p \leq 0,10$

PPSM Access Panel Nachrekrutierung 2012, eigene Berechnungen

Wie erwartet zeigt sich, dass die Wahrscheinlichkeit mit zunehmendem Alter deutlich ansteigt, über einen Eintrag im Telefonbuch zu verfügen. Mit den verfügbaren Hintergrundvariablen konnten systematische Unterschiede zwischen Personen mit und ohne Telefonbucheintrag nachgewiesen werden. Die Stadtteilvariablen haben sich als nicht erklärungsrelevant gezeigt.

## Zusammenfassung der Ergebnisse

Die aufgestellte Forschungshypothese  $H_{B2}$  konnte bestätigt werden. Systematische Unterschiede konnten hinsichtlich Geschlecht und Alter nachgewiesen werden: Personen, die über einen Eintrag im Telefonbuch verfügen, sind eher männlich und älter.

Tabelle 7.23: Entscheidung über die Hypothesen

Hypothese	Status
$H_{B2}$ : Personen mit Telefonbucheintrag $\neq$ Personen ohne Telefonbucheintrag	angenommen

Jüngere Personen sind hierüber nur sehr bedingt zu erreichen. Letztlich stützen die Ergebnisse die methodische Vorgehensweise bei der Nachrekrutierung aus dem Jahr 2012. Würde das Telefonbuch als einzige Auswahlgrundlage zur Stichprobenziehung dienen, ist mit systematischen Verzerrungen aufgrund von Noncoverage durch Unterschiede zwischen diesen beiden Gruppen zu rechnen.

## Kapitel 8

# Fazit und Handlungsempfehlungen

Die wissenschaftliche Umfrageforschung sieht sich aufgrund gesellschaftlicher Veränderungen, wie beispielsweise der zunehmenden Digitalisierung, immer wieder neuen Herausforderungen gegenüber. In den letzten Jahren hat sich ein Trend hinzu abnehmenden Antwortraten abgezeichnet, der dazu geführt hat, dass diese ein teilweise bedenklich niedriges Niveau erreicht haben (Engel und Schmidt, 2015). Niedrige Antwortraten sind für sich genommen nicht zwangsläufig ein Problem für die Repräsentativität von Umfrageergebnissen. In der Praxis selektieren sich Respondenten von Nonrespondenten häufig systematisch entlang bestimmter Eigenschaften voneinander, sodass nicht mehr von einem zufälligen Ausfallprozess gesprochen werden kann. Durch die systematischen Unterschiede zwischen diesen beiden Gruppen werden die gestellten Fragen in der Folge anders beantwortet, wodurch ein Bias entsteht, den es zu kontrollieren gilt. Dies wird dadurch verhindert, dass über die Personen, die sich nicht an einer wissenschaftlichen Umfrage beteiligt haben, nur sehr wenige bis keine Informationen vorliegen. Nonresponseanalysen sind dadurch entsprechend eingeschränkt.

Darüber hinaus sind wissenschaftliche Umfragen, die online durchgeführt werden, von Noncoverage betroffen. Denn es ist noch immer so, dass ältere Menschen das Internet eher nicht nutzen und es zusätzlich keine Auswahlgrundlage gibt, aus der Stichproben gezogen werden können. Hinzukommen neuere Entwicklungen bezüglich der Nutzung mobiler Endgeräte für die Beantwortung von Umfragen.

Mit dieser Arbeit wurde versucht, ein Umfragedesign zu entwickeln, mit dem es nicht nur gelingen soll Antwortraten zu erhöhen. Vielmehr wurde ein Stichprobenansatz gewählt, der es durch Zusatzinformationen über Nonrespondenten sowie Paradata erlaubt, den

Ausfallprozess und somit mögliche Verzerrungen aufgrund dessen zu untersuchen.

Es wurde ein Gesamtmodell aus insgesamt vier übergeordneten Designelementen entwickelt und einer empirischen Überprüfung unterzogen: (A) Registerstichprobe und Paradata, (B) Mixed-Mode Design, (C) postalische Vorankündigung und (D) Teilnahmeanreize. Die folgenden Ausführungen orientieren sich in ihrer Gliederung an den Ergebnissen dieser Arbeit.

### **Mixed-Mode**

Es ist heute aus verschiedenen Gründen schwierig, eine Befragung der allgemeinen Bevölkerung in einem Modus durchzuführen und dabei eine befriedigende Datenqualität zu erzielen. Im Vorhinein einer Befragung muss abgewogen werden, wie groß das Fehlerpotential einer Single-Mode Befragung gegenüber einer vergleichbaren Mixed-Mode Befragung ist. Erst wenn die zu erwartende Fehlerbilanz durch den Einsatz von Mixed-Mode geringer ist, sollte dieses Verfahren zur Anwendung kommen (Vannieuwenhuyze, 2014). Es gibt viele verschiedene Kombinationsmöglichkeiten der einzelnen Modi bei der Durchführung einer Umfrage.

Bezogen auf das Umfragedesign der Nachrekrutierung aus dem Jahr 2012 hat die Kombination aus CAWI-Fragebogen und CATI-Fragebogen in mehrfacher Hinsicht Sinn gemacht. Denn wäre die Umfrage in nur einem der beiden Modi durchgeführt worden, wäre mit Verzerrungen aufgrund von Noncoverage und Nonresponse zu rechnen gewesen. Die Fehlerquelle Noncoverage ist durch den registerbezug der Stichprobe auf dieser Stufe nicht relevant. Hätte man allerdings lediglich die Personen angerufen, die im Telefonbuch eingetragen sind, wären diejenigen, die nicht darin eingetragen sind, bereits vor der Befragung davon ausgeschlossen. Das damit einhergehende Fehlerpotential wird deutlich, wenn man die Unterschiede zwischen diesen beiden Gruppen untersucht. Über eine Registerstichprobe in Verbindung mit dem Telefonbuch lassen sich vor allem ältere Personen und Männer für eine Umfrage gewinnen, da diese wesentlich häufiger über einen Eintrag verfügen. Hätte man ausschließlich die Option zur Teilnahme an der CAWI-Variante zur Wahl gestellt, hätte man Personen ohne Affinität zum Internet oder ohne Anschluss nicht berücksichtigt. Zentral ist nun, dieses Vorgehen, wie in der Nachrekrutierung aus dem Jahr 2012, um weitere Teilnahmemöglichkeiten zu einem Mixed-Mode Design zu erweitern. Bietet man als zusätzliche Option eine CAWI-Variante an, lassen sich hierdurch mehrheitlich jüngere Personen gewinnen. Daran anknüpfend wird das Potential der Kombination unterschiedlicher

Teilnahmemodi deutlich. Die praktische Konsequenz aus diesem Vorgehen ist allerdings, dass Moduseffekte nicht mehr von Selektionseffekten sauber statistisch getrennt werden können, denn die Zuweisung zu den einzelnen Modi erfolgt nicht zufällig. Hier muss eine individuelle Abschätzung des Fehlerpotentials des jeweiligen Vorgehens stattfinden, um letztlich eine Entscheidung für eine der Alternativen treffen zu können. Denn in der Praxis wird es nicht möglich sein, alle Fehlerquellen zu kontrollieren. Sind Effekte aufgrund des Modus zu erwarten, wenn beispielsweise bei der Beantwortung der Fragen mit ausgeprägten Effekten aufgrund sozialer Erwünschtheit zu rechnen ist, sollte die Kontrolle dessen eher im Vordergrund stehen. Hier sollte eine zufällige Allokation zu einer der angebotenen Teilnahmealternativen erfolgen, da man diesen Effekt andernfalls im Nachhinein nicht mehr kontrollieren kann.

Der Mixed-Mode Charakter der Nachrekrutierung aus dem Jahr 2012 ist in Bezug zu den drei Phasen einer Umfrage (Kontaktphase, Feldphase, Follow-up Phase) doppelt gelagert. Denn die Kontaktierung hat in Form eines schriftlichen Vorausbriefes stattgefunden und die letztliche Erhebung in zwei unterschiedlichen Modi. Dabei hat sich die Kombination aus Vorausbrief und der Ankündigung eines Anrufes (bei bekannter Rufnummer) im Zusammenspiel mit der Möglichkeit, den Fragebogen online auszufüllen, als erfolgreichste Rekrutierungsstrategie erwiesen. Zu verbessern wäre für die zukünftige Forschung der Matchingalgorithmus, der die Stichprobendatenbank mit der elektronischen Version des Telefonbuches zusammengebracht hat. Eine Verfeinerung der dahinterstehenden Logik sollte zu einer höheren Trefferquote und somit zu einer zusätzlichen Erhöhung der Antwortrate führen, da mehr Personen angerufen werden können. Darüber hinaus würden bei einer eindeutigeren Trennung zwischen Personen, mit und ohne Eintrag im Telefonbuch, etwaige Unterschiede nicht mehr unterschätzt werden.

Ein vielversprechendes Verfahren, um das Potential noch besser auszunutzen, ist der Einsatz eines sequentiellen Mixed-Mode Designs (Dillman und Messer, 2010; Millar und Dillman, 2011). Die initiale Umfrageeinladung erfolgt hier in einem kostengünstigen Modus (zum Beispiel online). Andere, kostenintensivere Modi (zum Beispiel postalisch) werden erst nach einer bestimmten Zeit all denjenigen angeboten, die sich bis dahin noch nicht an der Umfrage beteiligt haben. Hiermit lässt sich nicht nur das Fehlerpotential durch die Erhöhung der Antwortrate verringern, sondern gleichzeitig die Kosten für die Durchführung einer Umfrage senken.

## Mobile Onlineumfragen

In den letzten Jahren konnte beobachtet werden, dass Onlineumfragen auch immer häufiger mit mobilen Endgeräten beantwortet werden (De Bruijne und Wijnant, 2014; Lugtig und Toepoel, 2016; Rivella et al., 2014). Hierbei handelt es sich nur indirekt um einen Moduswechsel. Denn die Befragung wird nach wie vor online durchgeführt und zur Beantwortung steht mehr als ein Endgerät zur Verfügung. So ist es zwingend notwendig, einen Onlinefragebogen für die Darstellung auf mobilen Endgeräten (Smartphone, Tablet, Laptop) zu optimieren. Aufgrund der unterschiedlichen Displaygrößen und Eingabelogiken wäre andernfalls mit Messeffekten zu rechnen, die im Nachhinein nur schwer zu kontrollieren sind. Diesbezüglich konnte gezeigt werden, dass es systematische Unterschiede hinsichtlich der Nutzung von Smartphones sowie der Bereitschaft, hierüber einen Fragebogen zu beantworten, gibt. Auffällig ist diesbezüglich, dass über die Hälfte der befragten Personen nicht zur Teilnahme an einer Umfrage über ihr mobiles Endgerät bereit wären. Es ist aber trotzdem davon auszugehen, dass sich die Befragung über mobile Endgeräte in der Zukunft weiter etablieren wird und die Akzeptanz in der Bevölkerung zunimmt. Denn mobile Endgeräte werden sich in unserer Gesellschaft immer weiter verbreiten, sodass die systematischen Unterschiede zwischen Personen, die diese nutzen und Personen, die diese nicht nutzen, sukzessive kleiner werden sollte. Die Entwicklung in diesem Bereich kann man mit der des stationären Internets vergleichen. Hier hat die Verbreitung in den letzten Jahren zwar rasant zugenommen (ARD/ZDF-Onlinestudie, 2015), vor allem ältere Menschen sind aber hier immer noch unterrepräsentiert.

Dieser Befragungsmodus mit seinem Potential Umfragen in kürzester Zeit zu beantworten unabhängig davon, wo sich Personen befinden, birgt für die wissenschaftliche Umfrageforschung ganz neue Möglichkeiten. So könnte über die Aufzeichnung von GPS-Koordinaten festgestellt werden, wo sich die Teilnehmer bei der Beantwortung des Fragebogens befinden. Denkbar wäre ebenso, dass es in Zukunft eher Kurzumfragen gibt, die einige wenige Fragen beinhalten und schnell über ein mobiles Endgerät beantwortet werden.

Dieses Set an Möglichkeiten zeigt, dass die Forschung in diesem Bereich weiter voran getrieben werden muss, um dem Digitalisierungstrend sowie steigender Zahlen mobiler Nutzung des Internets gerecht werden zu können und das Potential für wissenschaftliche Umfragen nicht ungenutzt zu lassen. Letztlich ist es für die Befragung der allgemeinen Bevölkerung über das Internet, die ausschließlich über mobile Endgeräte ausgefüllt werden können, durch die beschriebenen methodischen Probleme noch zu früh. Diese können

und sollten in der Praxis als zusätzliche Teilnahmeoption angeboten werden. Auch wenn diese Option nicht aktiv an die potentiellen Teilnehmer kommuniziert wird, sollte jeder Onlinefragebogen schon heute für die Darstellung auf mobilen Endgeräten optimiert sein.

### **Teilnahmeanreize und postalische Vorankündigung**

Der Einsatz von Anreizen, speziell in monetärer Form und unabhängig von einer Teilnahme, hat sich in der Vergangenheit wiederholt also probates Mittel erwiesen, Antwortraten zu erhöhen. Auch die vorliegende Arbeit konnte zeigen, dass sich Anreize, die in monetärer Form und unabhängig von der Teilnahme vergeben werden, positiv auf die Teilnahmewahrscheinlichkeit auswirken. Dem Einsatz einer Verlosung konnte im Gegensatz erwartungsgemäß keine positive Wirkung nachgewiesen werden. Selten untersucht wurden bis heute zwei Teilaspekte: Ein möglicher Grenznutzen und die Wirkung auf unterschiedliche gesellschaftliche Gruppen. Ein möglicher Grenznutzen wurde mittels eines Experiments sowie einer offenen Frage operationalisiert. Erfragt wurde die subjektive Wahrnehmung auf Seiten der Befragten, welchen Betrag sie als kleine Aufmerksamkeit und welchen als Bezahlung wahrnehmen würden.

Letztlich lässt die Betrachtung der Ergebnisse den Schluss zu, dass die Wahrnehmung in Bezug zu höheren Beträgen, insbesondere bei 20 beziehungsweise 40 Euro, in Richtung Bezahlung geht. Diese Beträge könnten von den potentiellen Teilnehmern schnell als unzureichend für die zu investierende Zeit eingestuft und eine Verweigerung somit wahrscheinlicher werden lassen. Die Frage nach einem möglichen Grenznutzen kann damit beantwortet werden, dass es diesen in Bezug zur subjektiven Wahrnehmung auf Seiten der Teilnehmer zu geben scheint. Denn wird der Betrag im ökonomischen Sinne als (unzureichende) Bezahlung wahrgenommen und zieht eine Verweigerung nach sich, kann sich dieser Sachverhalt in seiner Folge direkt negativ auf die zu erzielende Antwortrate einer Umfrage auswirken.

Ob es tatsächlich einen Grenznutzen in Form eines explizit zu benennenden Betrages gibt, kann an dieser Stelle nicht abschließend geklärt werden. Die Ergebnisse zeigen aber, dass je höher der Betrag ausfällt, sich die subjektive Wahrnehmung immer weiter in Richtung Bezahlung verschiebt. Nicht zuletzt ist an die entstehenden Kosten bei der Durchführung einer Umfrage zu denken, denn in der wissenschaftlichen Umfrageforschung stehen in der Regel begrenzte Budgets zur Verfügung, die optimal ausgenutzt werden sollten. Vor diesem Hintergrund ist für den praktischen Einsatz nicht konditionaler monetärer

Anreize zu raten, keine Beträge über zehn Euro zu vergeben. Die hier vorliegenden Ergebnisse lassen sogar den Schluss zu, nicht mehr als fünf Euro zu vergeben. Denn wie Abbildung 7.3 sowie Tabelle 7.10 gezeigt haben, unterscheidet sich die subjektive Wahrnehmung seitens der Teilnehmer zwischen fünf und zehn Euro nicht voneinander. Mit dem Feldexperiment konnte für fünf und zehn Euro allerdings nicht klar gezeigt werden, ob sich die subjektive Wahrnehmung eher in Richtung Bezahlung oder kleine Aufmerksamkeit verschiebt. Bei der darauffolgenden offenen Frage nach expliziten Beträgen haben allerdings die Mehrheit der Befragten einen Betrag von bis zu zehn Euro benannt, den sie als Aufmerksamkeit einstufen würden. Es ist weitere Forschung in diesem Bereich notwendig, um möglicherweise zu eindeutigeren Ergebnissen zu kommen.

Für die Praxis kann man daraus ableiten, dass nicht nur aufgrund immenser Kosten davon abzuraten ist, Anreize in der Höhe von 20 oder 40 Euro zu vergeben. Sollen die Befragten einen Anreiz eher als eine kleine Aufmerksamkeit wahrnehmen und hierdurch der von Dillman et al. (2014) beschriebene Reziprozitätseffekt einsetzen, ist es vor dem Hintergrund der vorliegenden Ergebnisse ratsam, einen Betrag von maximal 10 Euro zu vergeben. Ferner ist es zusätzlich von der Länge der potentiellen Befragung selbst abhängig, ob die Teilnehmer einen Betrag von 20 oder 40 Euro als angemessen für die zu investierende Zeit empfinden, wenn sie diesen eher als Bezahlung wahrnehmen. Was dieses Ergebnis durchaus zeigt ist die Tatsache, dass höhere Anreizbeträge die subjektive Wahrnehmung in Richtung Bezahlung verschieben.

Konditionale Anreize wie eine Verlosung, die abhängig von einer Teilnahme sind, führen nicht zu einer Erhöhung der Antwortrate. Diese haben sich sogar als kontraproduktiv gezeigt. Vom Einsatz von Verlosungen in dieser Form ist für die Praxis abzuraten. Weiterführend wurde der Frage nachgegangen, ob monetäre Anreize auf alle gesellschaftlichen Gruppen die gleiche Wirkung haben. Denn wie bereits Vehre et al. (2013) herausgefunden haben, lässt sich mit monetären Anreizen der Anteil niedrig gebildeter Personen unter den Teilnehmern einer Umfrage erhöhen. Da diese in der Regel unterrepräsentiert sind, kann sich dieses Vorgehen in doppelter Hinsicht positiv auswirken. Einerseits wird die Antwortrate erhöht, andererseits werden die Unterschiede zwischen Respondenten und Nonrespondenten kleiner und somit auch das Potential für einen möglichen Bias in den Antworten.

Die Ergebnisse von Vehre et al. (2013) konnte dahingehend bestätigt werden, dass die Verteilung der Bildungsvariable in der Gruppe, die fünf Euro unabhängig von der Teilnahme erhalten hat, der Verteilung des Zensus aus dem Erhebungsjahr der Nachrekrutierung mit



am nächsten kommt. Dies ist ein weiterer Hinweis dafür, dass sich dem vorherrschenden Bildungsbias in wissenschaftlichen Umfragen mit dem Einsatz von Anreizen entgegenwirken lässt. Aus dieser Erkenntnis ergibt sich die Frage, ob dies Folgen für die Qualität der Daten hat. Denn aus der Forschung ist bekannt, dass Personen, die zu einer Teilnahme motiviert werden müssen, in der Folge eher nicht dazu bereit sind, kognitive Mühen in die Beantwortung der Fragen zu investieren (Engel und Köster, 2014). Um dieser Frage nachzugehen, wurde Item-Nonresponse kumuliert und als abhängige Variable genutzt. Dabei hat sich gezeigt, dass sich die Anreizbedingung sowie das Bildungsniveau nicht derart auf die Qualität der Daten auswirken, dass hierdurch ein höherer Anteil an Item-Nonresponse zu beobachten ist.

Der Einsatz monetärer Anreize, die unabhängig von einer Teilnahme vergeben werden, erhöhen nicht nur die Antwortrate sondern ebenso den Anteil an niedrig gebildeter Personen unter den Befragten, ohne dabei die Datenqualität negativ zu beeinflussen. Als Schlussfolgerung kann man deswegen für die Praxis festhalten, dass Anreize unabhängig von der Teilnahme vergeben werden sollten. In der Höhe sind fünf Euro angemessen, um den erwünschten Effekt der Reziprozität zu erzielen. Negative Auswirkungen auf die Qualität der Daten ist, aufgrund der hier vorliegenden Ergebnisse, nicht zu erwarten.

### **Melderegister und Paradata**

Werden Stichproben aus dem Melderegister gezogen und computergestützte Erhebungsverfahren eingesetzt, können die zusätzlichen Hintergrundinformationen über die Bruttostichprobe dazu genutzt werden, um Nonresponse und dessen Auswirkung auf die vorliegenden Antworten zu untersuchen. Somit ergeben sich weitere Informationen wie Alter, Geschlecht oder der Wohnort über Personen, die nicht teilgenommen haben. Auch wenn die Vielfalt an Informationen beschränkt ist, eröffnet sich mit diesen Variablen die Möglichkeit zu einer Analyse des Ausfallprozesses und der Operationalisierung möglicher Verzerrungen aufgrund systematischer Unterschiede zwischen Respondenten und Nonrespondenten.

Der Einsatz dieses Verfahrens muss trotzdem kritisch gesehen werden. Die auf Basis des Melderegisters durchgeführte Nachrekrutierung von Teilnehmern für das PPSM Access Panel war regional beschränkt auf die Stadt Bremen. Die Ziehung der Stichprobe wurde von der zuständigen Meldebehörde übernommen. Deutschlandweite Befragungen auf Basis des Melderegisters sind weniger praktikabel, da man mit einer Vielzahl von Meldeämtern

in Kontakt treten müsste. Dies würde nicht nur viel Arbeit in Anspruch nehmen, sondern eine zusätzliche Fehlerquelle produzieren. Denn es ist zu erwarten, dass die Meldeämter unterschiedlich viel Zeit bis zur Herausgabe der Daten benötigen. Kontaktinformationen können deswegen veraltet sein, wenn die eine Behörde schnell liefert, die andere aber mehr Zeit benötigt. Auch wenn eine Registerstichprobe aus methodischer Sicht den besten Zugang zu den potentiellen Teilnehmern durch die Möglichkeit einer postalischen Kontaktierung bietet, beschränken sich diese in ihrer Praktikabilität eher auf einzelne Regionen oder Bundesländer.

### **Analyse des Ausfallprozesses**

Mit der Analyse des Ausfallprozesses konnten teilweise deutliche Unterschiede hinsichtlich der Teilnahmewahrscheinlichkeiten nachgewiesen werden. Die Effekte des Grundmodells, das alle im Auswahlrahmen enthaltenen Variablen beinhaltet, gehen teilweise in die erwartete Richtung. Etwas überraschend erscheint insbesondere der stark ausgeprägte Effekt des Geschlechts, denn erwartet wurde kein Effekt. Es ist allerdings nicht auszuschließen, dass die Effektstärken und Zusammenhänge von relevanten Drittvariablen beeinflusst werden, die nicht in den Modellen enthalten sind, weil sie entweder nicht oder ausschließlich für die Nettostichprobe zur Verfügung stehen. Dies zeigt gleichzeitig auf, dass Nonresponseanalysen aufgrund der eingeschränkten Informationen über alle Personen in der Stichprobe, limitiert sind. Trotzdem wird es in Kombination mit Parainformationen möglich, Teilnahmewahrscheinlichkeiten zu schätzen und hierüber messbar zu machen, ob eine Verzerrung in den vorliegenden Antworten zu beobachten ist oder nicht. Die eingesetzten Designelemente einer Umfrage können daraufhin überprüft werden, ob diese Selektionseffekte derart in Gang setzen, dass hierdurch weitere systematische Unterschiede zwischen Respondenten und Nonrespondenten entstehen oder nicht und sich diese auf das Antwortverhalten auswirken.

Die Analyse des Ausfallprozesses ist auf diesen Aspekt beschränkt. Denn soll mit den geschätzten Teilnahmewahrscheinlichkeiten eine Gewichtung vorgenommen werden, ist der Erfolg dieses Vorgehens von den Eigenschaften der verwendeten Hintergrundvariablen abhängig. Die von Olson (2013) definierten Kriterien, dass die Hintergrundvariablen wie das Geschlecht oder das Alter der Personen in der Stichprobe ein starker Prädiktor sowohl für die Zielvariable als auch die Teilnahmewahrscheinlichkeit sein sollten, kann für die hier vorliegenden Ergebnisse nur bedingt gelten. Denn die genannte Grenze von einem  $R^2$  von

0.48, konnte mit keinem der geschätzten Modell erreicht werden. Letztlich wird hierdurch das Potential für die Gewichtung eingeschränkt. Die Operationalisierung von möglichen Verzerrungen ist aber dennoch möglich und konnte mit dem Random Response Model nach (Bethlehem, 2002) in dieser Arbeit erfolgreich angewendet werden.

Für die Praxis ist anzuraten, die jeweils eingesetzten Designelemente einer Umfrage auf diese Aspekte hin zu untersuchen, um etwaige Verzerrungen aufzudecken. Festzuhalten bleibt aber auch, dass es hier noch weitergehender Forschung bedarf. Denn jede Umfrage besteht aus anderen inhaltlichen Fragen und methodischen Designelementen, die jeweils für sich genommen einer Untersuchung potentieller Verzerrung durch systematische Selektionseffekte notwendig macht.

### **Analyse des Ausfallprozesses mittels Parainformationen**

Die durch die computergestützte Befragungssoftware automatisiert erhobenen Parainformationen, wie die Anzahl an Kontaktversuchen oder Terminvereinbarungen, wurden im vorletzten Abschnitt des Ergebniskapitels in Bezug zur Teilnahmewahrscheinlichkeit untersucht. Als ein Fazit kann man diesbezüglich festhalten, dass es offensichtlich sinnvoll ist, maximal bis zu zehn telefonische Kontaktversuche zu unternehmen. Denn einerseits sinkt die Wahrscheinlichkeit, dass eine Person ein Interview gibt danach auf unter fünf Prozent. Andererseits sollte auf zu häufige Kontaktversuche verzichtet werden, da diese als störend empfunden werden und eine ablehnende Haltung gegenüber wissenschaftlichen Befragungen erzeugen und verstärken können. Befragt man innerhalb eines Access Panels - haben die Personen also bereits ihre Zustimmung signalisiert, sich regelmäßig befragen zu lassen - sollte man bei der Anzahl an Kontaktversuchen zusätzlich Vorsicht walten lassen. Denn versucht man bei der darauffolgenden Welle ebenso all diejenigen zu erreichen, die bei der Welle zuvor nicht erreicht werden konnten, können sich zu viele Kontaktversuche negativ auf die Teilnahmebereitschaft auswirken.

Mit dieser Arbeit konnte ein Beitrag rund um die Diskussion um niedrige Antwortraten und Noncoverage geleistet werden. Das *Ziel* bestand darin, Effekte durch Noncoverage und selektiven Nonresponse bei der Rekrutierung von Teilnehmern für sozialwissenschaftliche Panelbefragungen zu reduzieren und zu operationalisieren. Dieses Ziel konnte, mit dem hierfür entwickelten heuristischen Gesamtmodell und dem Einsatz des Random Response Models nach Bethlehem (2002), erreicht werden. Gleichzeitig lässt sich mit dieser Erkenntnis

die übergeordnete Forschungsfrage damit beantworten, dass sich selektiver Nonresponse (mit den bereits diskutierten Einschränkungen) operationalisieren und auf die durch den Auswahlrahmen zur Verfügung stehenden Hintergrundvariablen sowie die einzelnen Designelemente einer Umfrage zurückführen lässt. Das entwickelte Gesamtmodell zur Rekrutierung von Teilnehmern hat sich in der Praxis bewährt und zu einer teilweise deutlichen Erhöhung der Antwortrate geführt. Als beste Variante hat sich dabei die Registerstichprobe in Verbindung mit einer postalischen Umfrageeinladung, einem kurz danach erfolgten Telefonanruf, Teilnahmeanreizen und dem Einsatz eines Mixed-Mode Verfahrens, gezeigt. Mögliche Verzerrungen aufgrund von Noncoverage werden durch den Registerbezug der Stichprobenziehung nivelliert, denn alle Personen in der Grundgesamtheit haben die gleiche Wahrscheinlichkeit gezogen zu werden unabhängig davon, ob sie über einen Internetzugang oder ein Festnetztelefon verfügen. Wenn Personen trotz fehlendem Zugang zum Internet zur Umfrage eingeladen worden sind, hatten diese die Möglichkeit telefonisch zu partizipieren und vice versa. Das Mixed-Mode Verfahren sollte also zusätzlich zur Verringerung möglicher verzerrender Effekte beigetragen haben.

Besonders hervorzuheben bezüglich des Einsatzes von Teilnahmeanreizen (monetär und unabhängig von der Teilnahme) ist die Tatsache, dass sich hiermit der Bildungsbias in wissenschaftlichen Umfragen verringern lässt. Trotz der anfallenden Kosten handelt es sich um ein probates Mittel, um die Repräsentativität der erhobenen Daten zu erhöhen. Dieses Ergebnis muss durch weitere Untersuchungen erhärtet werden.

Die zukünftige Forschung sollte sich darüber hinaus verstärkt mit mobilen Endgeräten und ihren methodischen Besonderheiten auseinandersetzen, wenn diese zur Beantwortung wissenschaftlicher Umfragen genutzt werden. Zu denken wäre hier beispielsweise nicht an längere Fragebögen, deren Beantwortung einige Zeit in Anspruch nehmen. Vielmehr könnte man eine Art von Impulsbefragung durchführen mit einigen wenigen Fragen, die schnell zu beantworten sind. Vorstellbar wären Fragen zur Befindlichkeit von Personen wie Glück und Lebenszufriedenheit. Solche Dispositionen verändern sich häufig über die Zeit und sind über kleinere und höher frequentiertere Befragungen, die über ein mobiles Endgerät beantwortet werden können, besser abzubilden. Solche Ideen zeigen, welches Potential in dieser relativ neuen Form der Befragung liegt. Mobile Web und Mixe-Mode Befragungen sollten auch in Zukunft im Fokus wissenschaftlicher Forschung liegen, um das dahinterstehende Potential noch besser nutzbar zu machen und mögliche Fehlerquellen aufzudecken.

# Literaturverzeichnis

- AAPOR (2015): Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 8th Edition. *The American Association of Public Opinion Research*.
- Aarts, Kees/ Weßels, Bernhard (2005): Wahlbeteiligung in Deutschland und bei europäischen Nachbarn. In: Falter, Jürgen W./ Gabriel, Oscar W./ Wessels, Bernhard (Hrsg.), *Wahlen und Wähler*, Wiesbaden: Springer, 595–617.
- Adloff, Frank/ Mau, Steffen (2005): *Vom Geben und Nehmen*. Frankfurt/New York: Campus.
- Adua, Lazarus/ Shap, Jeff S. (2010): Examining Survey Participation and Response Quality: The Significance of Topic Salience and Incentives. *Survey Methodology*, 36(1), 95–109.
- Alanya, Ahu/ Wolf, Christof/ Sotto, Cristina (2015): Comparing Multiple Imputation and Propensity-Score Weighting in Unit-Nonresponse Adjustments: A Simulation Study. *Public Opinion Quarterly*, 79(3), 1–27.
- Albers, Ines (1997): Einwohnermelderegisterstichproben in der Praxis. In: Gabler, Siegfried/ Hoffmeyer-Zlotnik, Jürgen H. P. (Hrsg.), *Stichproben in der Umfragepraxis*, Stichproben in der Umfragepraxis, Opladen: Westdeutscher Verlag, 117–127.
- Allison, Paul D. (2000): Multiple Imputation for Missing Data: A Cautionary Tale. *Sociological Methods & Research*, 28(3), 301–309.
- Allison, Paul D. (2001): *Missing Data, Quantitative Applications in the Social Sciences*, 136. Thousand Oaks: SAGE Publications.
- Allison, Paul D. (2009): *Fixed Effects Regression Models, Quantitative Applications in the Social Sciences*, 160. Los Angeles: SAGE Publications.

- Andersen, Ronald/ Kasper, Judith/ Frankel, Martin R. (1979): *Total Survey Error*. San Francisco: Jossey-Bass.
- Arbeitskreis deutscher Markt- und Sozialforschungsinstitute e.V. (2016): Quantitative Interviews der Mitgliedsinstitute des ADM nach Befragungsart. Empfangen am: 20.06.2016. URL <https://www.adm-ev.de/zahlen/#c245>
- ARD/ZDF-Onlinestudie (2015): Entwicklung der Onlinenutzung in Deutschland. Empfangen am: 27.08.2015. URL <http://www.ard-zdf-onlinestudie.de/index.php?id=530>
- Arzheimer, Kai/ Klein, Markus (1998): Die Wirkung materieller Incentives auf den Rücklauf einer schriftlichen Panelbefragung. *ZA-Information/ Zentralarchiv für Empirische Sozialforschung*, (43), 6–31.
- Atkeson, Lonna Rae/ Tafoya, Lorraine M. (2008): Surveying Political Activists: The Effectiveness of a Mixed Mode Survey Design. *Journal of Elections, Public Opinion and Parties*, 18(4), 367–386.
- Aust, Folkert/ Schröder, Helmut (2009): Sinkende Stichprobenausschöpfung in der Umfrageforschung - ein Bericht aus der Praxis. In: Weichbold, Martin/ Bacher, Johann/ Wolf, Christof (Hrsg.), *Umfrageforschung - Herausforderung und Grenzen*, Wiesbaden: Springer Verlag, 195–212.
- Backhaus, Klaus/ Erichson, Bernd/ Plinke, Wulff/ Weiber, Rolf (2015): *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Wiesbaden: Springer Verlag.
- Bandilla, Wolfgang/ Kaczmirek, Lars/ Blohm, Michael/ Neubarth, Wolfgang (2009): Coverage- und Nonresponse-Effekte bei Online-Bevölkerungsumfragen. In: Jakob, Nikolaus/ Schoen, Harald/ Zerback, Thomas (Hrsg.), *Sozialforschung im Internet*, Wiesbaden: Springer Verlag, 129–143.
- Bandilla, Wolfgang/ Bosnjak, Michael/ Couper, Mick P./ Galesic, Mirta/ Haas, Iris/ Kaczmirek, Lars (2013): Sample Composition Discrepancies in Different Stages of a Probability-based Online Panel. *Field Methods*, 25(4), 339–360.
- Bartsch, Simone (2012): *... würden Sie mir dazu Ihre E-Mail-Adresse verraten?: Internetnutzung und Nonresponse beim Aufbau eines Online Access Panels*. Baden Baden: Nomos.

- Baruch, Yehuda (1999): Response Rate in Academic Studies - A Comparative Analysis. *Human Relations*, 52(4), 421–438.
- Bauman, S.L./ Merkle, D.M./ Lavrakas, P.J. (1992): Interviewer Estimates of Refusers Gender, Age, and Race in Telephone Surveys. *Midwest Association for Public Opinion Research Conference, Chicago*.
- Baur, Nina/ Florian, Michael J. (2009): Stichprobenprobleme bei Online-Umfragen. In: Jakob, Nikolaus/ Schoen, Harald/ Zerback, Thomas (Hrsg.), *Sozialforschung im Internet, Sozialforschung im Internet*, Wiesbaden: Springer Verlag, 109–128.
- Bautista, René (2012): An Overlooked Approach in Survey Research: Total Survey Error. In: Gideon, Lion (Hrsg.), *Handbook of Survey Methodology for the Social Sciences*, Wiesbaden: Springer Verlag, 37–49.
- Bech, Mickael/ Kristensen, Morten Bo (2009): Differential response rates in postal and Web-based surveys among older respondents. *Survey Research Methods*, 3(1), 1–6.
- Becker, Rolf/ Imhof, Regula/ Mehlkop, Guido (2007): Die Wirkung monetärer Anreize auf den Rücklauf bei einer postalischen Befragung und die Antworten auf Fragen zur Delinquenz. *Methoden - Daten - Analysen*, 1(2), 131–159.
- Berger, Fred (2006): Zur Wirkung unterschiedlicher materieller Incentives in postalischen Befragungen - Ein Literaturbericht. *ZUMA-Nachrichten*, 30(58), 81–100.
- Bethlehem, Jelke (2002): Weighting Nonresponse Adjustments Based on Auxiliary Information. In: Groves, Robert M./ Dillman, Don/ Eltinge, John L./ Little, Roderick J.A. (Hrsg.), *Survey Nonresponse*, New York: John Wiley & Sons, Inc., 275–302.
- Bethlehem, Jelke (2009): *Applied Survey Methods. A Statistical Perspective*. Hoboken, New Jersey: John Wiley & Sons, Inc., iD: 467.
- Bethlehem, Jelke (2015): Web Surveys in Official Statistics. In: Engel, Uwe/ Jann, Ben/ Lynn, Peter/ Scherpenzeel, Annette/ Sturgis, Patrick (Hrsg.), *Improving Survey Methods. Lessons from Recent Research*, New York: Routledge, Taylor & Francis Group, New York, 156–170.
- Bethlehem, Jelke/ Cobben, Fannie/ Schouten, Barry (2011): *Handbook of Nonresponse in Household Surveys*, 568. Hoboken, New Jersey: John Wiley & Sons, Inc.

- Biemer, Paul P./ Lyberg, Lars E. (2003): *Introduction to Survey Quality*. New York: John Wiley & Sons, Inc.
- Binson, Diane/ Catania, Joseph A. (2000): Random Selection in a National Telephone Survey: A Comparison of the Kish, Next-Birthday, and Last-Birthday Methods. *Journal of Official Statistics*, 16(1), 53.
- BMJV (2015): Bundesdatenschutzgesetz (BDSG).
- Brehm, John (1994): Stubbing Our Toes for a Foot in the Door? Prior Contact, Incentives and Survey Response. *International Journal of Public Opinion Research*, 6(1), 45–63.
- Brick, J. Michael (2011): The Future of Survey Sampling. *Public Opinion Quarterly*, 75(5), 872–888.
- Brick, J. Michael (2013): Unit Nonresponse and Weighting Adjustments: A Critical Review. *Journal of Official Statistics*, 29(3), 329–353.
- Brick, J. Michael/ Dipko, Sarah/ Presser, Stanley/ Tucker, Clyde/ Yuan, Yangyang (2006): Nonresponse Bias in a Dual Frame Sample of Cell and Landline Numbers. *Public Opinion Quarterly*, 70(5), 780–793.
- Bruijne, Marika De/ Wijnant, Arnaud (2013): Comparing Survey Results Obtained via Mobile Devices and Computers: An Experiment With a Mobile Web Survey on a Heterogeneous Group of Mobile devices versus a Computer-Assisted Web Survey. *Social Science Computer Review*, 31(4), 482–504.
- Buelens, Bart/ van den Brakel, Jan (2014): Measurement Error Calibration in Mixed-Mode Sample Surveys. *Sociological Methods & Research*, 44(3), 391–426.
- Busse, Britta/ Fuchs, Marek (2015): Telephone Surveys Using Mobile Phones. In: Engel, Uwe/ Jann, Ben/ Lynn, Peter/ Scherpenzeel, Annette/ Sturgis, Patrick (Hrsg.), *Improving Survey Methods. Lessons from Recent Research*, New York: Routledge, 51–63.
- Callegaro, Mario (2013): Paradata in Web Surveys. In: Kreuter, Frauke (Hrsg.), *Improving Surveys with Paradata - Analytic Uses of Process Information*, New York: John Wiley & Sons, Inc., 259–279.



- Christian, Leah Melani/ Dillman, Don A./ Smyth, Jolene D. (2005): The Effects of Mode and Format on Answers to Scalar Questions in Telephone and Web Surveys. Empfangen am: 26.05.2017.  
 URL [http://s3.amazonaws.com/academia.edu.documents/42149838/The\\_effects\\_of\\_mode\\_and\\_format\\_on\\_answer20160205-31195-1ee2x48.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1495806539&Signature=AJv2Lhn3GPBr2%2BrraHM%2FngoW9WQ%3D&response-content-disposition=inline%3B%20filename%3DThe\\_Effects\\_of\\_Mode\\_and\\_Format\\_on\\_Answer.pdf](http://s3.amazonaws.com/academia.edu.documents/42149838/The_effects_of_mode_and_format_on_answer20160205-31195-1ee2x48.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1495806539&Signature=AJv2Lhn3GPBr2%2BrraHM%2FngoW9WQ%3D&response-content-disposition=inline%3B%20filename%3DThe_Effects_of_Mode_and_Format_on_Answer.pdf)
- Church, Allan H. (1993): Estimating The Effect of Incentives on Mail Survey response Rates: A Meta-Analysis. *Public Opinion Quarterly*, 57(1), 62–79.
- Cobben, Fannie (2007): Mode Effects in a Basic Question Approach for the Dutch LFS. Discussion Paper 07003, empfangen am: 26.05.2017.  
 URL <https://www.cbs.nl/-/media/imported/documents/2007/16/2007-03-x10-pub.pdf>
- Cobben, Fannie/ Bethlehem, Jelke (2005): Adjusting Undercoverage and Nonresponse Bias in Telephone Surveys. Discussion Paper, empfangen am: 28.05.2017.  
 URL [https://www.researchgate.net/profile/Jelke\\_Bethlehem/publication/242207187\\_Adjusting\\_Undercoverage\\_and\\_Non-Response\\_Bias\\_in\\_Telephone\\_Surveys/links/00b7d53b4e942d86ed000000.pdf](https://www.researchgate.net/profile/Jelke_Bethlehem/publication/242207187_Adjusting_Undercoverage_and_Non-Response_Bias_in_Telephone_Surveys/links/00b7d53b4e942d86ed000000.pdf)
- Couper, Mick P. (2000): Review: Web Surveys: A Review of Issues and Approaches. *Public Opinion Quarterly*, 64(4), 464–494.
- Couper, Mick P. (2011): The Future of Modes of Data Collection. *Public Opinion Quarterly*, 75(5), 889–908.
- Couper, Mick P./ Bosnjak, Michael (2010): Internet Surveys. In: Marsden, Peter V./ Wright, James D. (Hrsg.), *Handbook of Survey Research*, Bingley: Emerald Group Publishing, 527–551.
- Couper, Mick P/ Groves, Robert M (1996): Household-level Determinants of Survey Nonresponse. *New Directions for Evaluation*, 1996(70), 63–79.
- Couper, Mick P./ Kreuter, Frauke (2013): Using Paradata to Explore Item Level Response Times in Surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(1), 271–286.

- Currivan, Douglas B. (2005): The Impact of providing Incentives to Initial Telephone Survey Refusers on Sample Composition and Data Quality.  
URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.525.4621&rep=rep1&type=pdf>
- Curtin, Richard/ Presser, Stanley/ Singer, Eleanor (2005): Changes in Telephone Survey Nonresponse Over the Past Quarter Century. *Public Opinion Quarterly*, 69(1), 87–98.
- De Bruijne, Marika/ Wijnant, Arnaud (2014): Mobile Response in Web Panels. *Social Science Computer Review*, 32(6), 728–742.
- De Heer, Wim (1999): International Response Trends: Results of an International Survey. *Journal of Official Statistics Stockholm*, 15, 129–142.
- De Leeuw, Edith/ Callegaro, Mario/ Hox, Joop/ Korendijk, Elly/ Lensvelt-Mulders, Gerty (2007): The Influence of Advance Letters on Response in Telephone Surveys: A Meta-Analysis. *Public Opinion Quarterly*, 71(3), 413–443.
- De Leeuw, Edith D. (2005): To Mix or Not to Mix Data Collection Modes in Surveys. *Journal of Official Statistics Stockholm*, 21(2), 233.
- De Leeuw, Edith D./ De Heer, Wim (2002): Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In: Groves, Robert M./ Dillman, Don/ Eltinge, John L./ Little, Roderick J.A. (Hrsg.), *Survey Nonresponse*, New York: John Wiley & Sons, Inc., 41–54.
- De Leeuw, Edith D./ Hox, Joop/ Huisman, Mark (2003): Prevention and Treatment of Item Nonresponse. *Journal of Official Statistics Stockholm*, 19(2), 153–176.
- De Leeuw, Edith D./ Hox, Joop J./ Dillman, Don A. (2008): Mixed-Mode Surveys: When and Why. In: De Leeuw, Edith D./ Hox, Joop J./ Dillman, Don A. (Hrsg.), *International Handbook of Survey Methodology*, New York: Erlbaum, 299–316.
- Deming, Edwards (1944): On Errors in Surveys. *American Sociological Review*, 9(4), 359–69.
- Deutschmann, Marc/ Häder, Sabine (2002): Nicht-Eingetragene in CATI-Surveys. *Telefonstichproben. Methodische Innovationen und Anwendungen in Deutschland*, New York, Waxmann Verlag.

- Diekmann, Andreas (2007): Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. 17. Aufl., Orig. *Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verlag*.
- Dillman, Don A./ Messer, Benjamin L. (2010): Mixed-Mode Surveys. In: Marsden, Peter V./ Wright, James D. (Hrsg.), *Handbook of Survey Research*, second edition , Bingley: Emerald Group Publishing, 551–575.
- Dillman, Don A./ Eltinge, John L./ Groves, Robert M./ Little, Roderick JA. (2002): Survey Nonresponse in Design, Data Collection, and Analysis. In: Dillman, John L., Don A. and Eltinge/ Groves, Robert M./ Little, Roderick JA. (Hrsg.), *Survey Nonresponse*, New York: John Wiley & Sons, Inc., 3–26.
- Dillman, Don A./ Phelps, Glenn/ Tortora, Robert D./ Swift, Karen/ Kohrell, Julie/ Berck, Jodie/ Messer, Benjamin L. (2009): Response Rate and Measurement Differences in Mixed-Mode Surveys Using Mail, Telephone, Interactive Voice Response (IVR) and the Internet. *Social science research*, 38, 1–18.
- Dillman, Don A/ Smyth, Jolene D/ Christian, Leah Melani (2014): *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. New York: John Wiley & Sons, Inc.
- Durrant, Gabriele B/ Steele, Fiona (2009): Multilevel Modelling of Refusal and Non-Contact in Household Surveys: Evidence From Six UK Government Surveys. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(2), 361–381.
- Engel, Uwe (2006): Anreizeffekte in Studien der Markt- und Sozialforschung. In: Faulbaum, Frank/ Wolf, Christof (Hrsg.), *Stichprobenqualität in Bevölkerungsumfragen*, Bonn: Informationszentrum Sozialwissenschaften, 185–195.
- Engel, Uwe (2013): Access Panel and Mixed-Mode Internet Survey. PPSM Panel Report. 2nd Edition.  
URL [https://www.methodenzentrum-bremen.de/uploads/PPSM\\_Panel\\_Report\\_Online.pdf](https://www.methodenzentrum-bremen.de/uploads/PPSM_Panel_Report_Online.pdf)
- Engel, Uwe (2014): *Gerechtigkeit ist gut, wenn sie mir nützt: Was den Deutschen wichtig ist - Eine Umfrage*. Frankfurt am Main: Campus Verlag.

- Engel, Uwe (2015): Response Behavior in an Adaptive Survey Design for the Setting-Up Stage of a Probability-Based Access Panel in Germany. In: Engel, Uwe/ Jann, Ben/ Lynn, Peter/ Scherpenzeel, Annette Clarine (Hrsg.), *Improving Survey Methods: Lessons from Recent Research*, New York: Routledge, Taylor & Francis Group, 207–222.
- Engel, Uwe/ Köster, Britta (2014): Response Effects and Cognitive Involvement in Answering Survey Questions. In: Engel, Uwe/ Jann, Ben/ Lynn, Peter/ Scherpenzeel, Annette/ Sturgis, Patrick (Hrsg.), *Improving Survey Methods. Lessons from Recent Research*, New York: Routledge, Taylor & Francis Group, 35–50.
- Engel, Uwe/ Köster, Britta (2015): Framing Effects. In: Engel, Uwe (Hrsg.), *Survey Measurements: Techniques, Data Quality and Sources of Error*, Frankfurt am Main: Campus Verlag, 58–75.
- Engel, Uwe/ Schmidt, Björn Oliver (2013): Ein Feldexperiment zur Wirkung von Befragungsanreizen innerhalb eines Mixed-Mode Designs. Empfangen am: 26.10.2015.  
URL [https://www.methodenzentrum-bremen.de/uploads/Engel\\_Schmidt\\_\[Feldexperiment\\_zur\\_Wirkung\\_von\\_Befragungsanreizen\]\\_20140805.pdf](https://www.methodenzentrum-bremen.de/uploads/Engel_Schmidt_[Feldexperiment_zur_Wirkung_von_Befragungsanreizen]_20140805.pdf)
- Engel, Uwe/ Schmidt, Björn Oliver (2014): Unit- und Item-Nonresponse. In: und Jörg Blasius, Nina Baur (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung*, Wiesbaden: Springer Verlag, 331–348.
- Engel, Uwe/ Schmidt, Björn Oliver (2015): Mixed-Mode Design, Incentivierung und Nonresponse Bias im Rahmen einer Einwohnermeldeamtsstichprobe. In: Schupp, Jürgen/ Wolf, Christof (Hrsg.), *Nonresponse Bias*, Wiesbaden: Springer Verlag, 255–286.
- Engel, Uwe/ Pötschke, Manuela/ Schnabel, Christiane/ Simonson, Julia (2004): *Nonresponse und Stichprobenqualität*. Frankfurt am Main: Verlagsgruppe Deutscher Fachverlag.
- Engel, Uwe/ Bartsch, Simone/ Schnabel, Christiane/ Vehre, Helen (2012): *Wissenschaftliche Umfragen: Methoden und Fehlerquellen*. Frankfurt am Main: Campus Verlag.
- Esser, Hartmut (1973): Kooperation und Verweigerung beim Interview. In: Erbslöh, Eberhard/ Esser, Hartmut/ Reschka, Willibald/ Schöne, Dino (Hrsg.), *Studien zum Interview*, Meisenheim am Glan: Verlag Anton Hain, 71–141.
- Esser, Hartmut (1986): Über die Teilnahme an Befragungen. *ZUMA-Nachrichten*, 10(18), 38–47.

- Esser, Hartmut (1990): Habits, Frames und Rational Choice. *Zeitschrift für Soziologie*, 19(4), 231–247.
- Fan, Weimiao/ Yan, Zheng (2010): Factors Affecting Response Rates of the Web Survey: A Systematic Review. *Computers in Human Behavior*, 26(2), 132–139.
- Faulbaum, Frank (2014): Total Survey Error. In: Baur, Nina/ Blasius, Jörg (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung*, Wiesbaden: Springer Verlag, 439–453.
- Felderer, Barbara/ Birg, Alexandra/ Kreuter, Frauke (2014): Paradata. In: Baur, Nina/ Blasius, Jörg (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung*, Wiesbaden: Springer Verlag, 357–365.
- Früh, Werner (2015): *Inhaltsanalyse: Theorie und Praxis*. Konstanz: UVK Verlagsgesellschaft mbH.
- Fuchs, Marek (2010): Herausforderungen der Umfrageforschung. In: Faulbaum, Frank/ Wolf, Christof (Hrsg.), *Gesellschaftliche Entwicklungen im Spiegel der empirischen Sozialforschung*, Wiesbaden: Springer Verlag, 227–252.
- Fuchs, Marek/ Busse, Britta (2009): The Coverage Bias of Mobile Web Surveys Across European Countries. *International Journal of Internet Science*, 4(1), 21–33.
- Gabler, Siegfried/ Ayhan, Öztas (2007): Gewichtung bei Erhebungen im Festnetz und über Mobilfunk: ein Dual Frame Ansatz. In: Gabler, Siegfried/ Häder, Sabine (Hrsg.), *Mobilfunktelefonie - Eine Herausforderung für die Umfrageforschung*, ZUMA Nachrichten Spezial, 13, Mannheim: ZUMA, 39–45.
- Gabler, Siegfried/ Ganninger, Matthias (2010): Gewichtung. In: Wolf, Christof/ Best, Henning (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse*, Wiesbaden: Springer Verlag, 143–164.
- Gabler, Siegfried/ Häder, Sabine (1997): Überlegungen zu einem Stichprobendesign für Telefonumfragen in Deutschland. *ZUMA Nachrichten*, 21(41), 7–18.
- Gabler, Siegfried/ Häder, Sabine (2002): *Telefonstichproben*. Münster, New York, München, Berlin: Waxmann.

- Geuzinge, Linda/ van Rooijen, Johan/ Bakker, Bart (2000): The Use of Administrative Registers to Reduce Non-Response Bias in Household Surveys. *Netherlands Official Statistics* 2, 15, 32–38.
- Giesselmann, Marco/ Windzio, Michael (2013): *Regressionsmodelle zur Analyse von Paneldaten*. Wiesbaden: Springer Verlag.
- Glantz, Alexander/ Michael, Tobias (2014): Interviewereffekte. In: und Jörg Blasius, Nina Baur (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung*, Springer Verlag, 313–322.
- Göritz, Anja (2004): The Impact of Material Incentives on Response Quantity, Response Quality, Sample Composition, Survey Outcome and Cost in Online Access Panels. *International Journal of Market Research*, 46(3), 327–346.
- Göritz, Anja (2006): Incentives in Web Studies: Methodological Issues and a Review. *International Journal of Internet Science*, 1(1), 58–70.
- Göritz, Anja (2010): Using Lotteries, Loyalty Points, and Other Incentives to Increase Participant Response and Completion. In: Gosling, Samuel/ Johnson, John (Hrsg.), *Advanced Methods for Conducting Online Behavioral Research*, Washington, DC: American Psychological Association, 219–233.
- Göritz, Anja (2015): Incentives Effect. In: Engel, Uwe/ Jann, Ben/ Lynn, Peter/ Scherpenzeel, Annette/ Sturgis, Patrick (Hrsg.), *Improving Survey Methods: Lessons from Recent Research*, Routledge, 339–350.
- Göritz, Anja/ Wolff, Hans-Georg (2007): Lotteries as Incentives in Longitudinal Web Studies. *Social Science Computer Review*, 25(1), 99–110.
- Göritz, Anja/ Wolff, Hans-Georg/ Goldstein, Daniel G. (2008): Individual Payments as a Longer-Term Incentive in Online Panels. *Behavior Research Methods*, 40(4), 1144–1149.
- Goyder, John (1987): *The Silent Minority: Nonrespondents on Sample Surveys*. Cambridge: Westview Press.
- Groves, Robert M. (2006): Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, 70(5), 646–675.

- Groves, Robert M./ Heeringa, Steven G. (2006): Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(3), 439–457.
- Groves, Robert M./ Lyberg, Lars (2010): Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly*, 74(5), 849–879.
- Groves, Robert M./ Peytcheva, Emilia (2008): The Impact of Nonresponse Rates on Nonresponse Bias - A Meta-Analysis. *Public Opinion Quarterly*, 72(2), 167–189.
- Groves, Robert M./ Singer, Eleanor/ Corning, Amy D. (2000): Leverage-Saliency Theory of Survey Participation. *Public opinion quarterly*, 64(3), 299–308.
- Groves, Robert M./ Presser, Stanley/ Dipko, Sarah (2004): The Role of Topic Interest in Survey Participation Decisions. *Public Opinion Quarterly*, 68(1), 2–21.
- Groves, Robert M./ Fowler Jr., Floyd J./ Couper, Mick P./ Lepkowski, James M./ Singer, Eleanor/ Tourangeau, Roger (2011): *Survey Methodology*. New York: John Wiley & Sons.
- Häder, Michael/ Häder, Sabine (2009): *Telefonbefragungen über das Mobilfunknetz - Konzept, Design und Umsetzung einer Strategie zur Datenerhebung*. Wiesbaden: Springer Verlag.
- Häder, Michael/ Häder, Sabine (2014): Stichprobenziehung in der quantitativen Sozialforschung. In: Baur, Nina/ Blasius, Jörg (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung*, Springer, 283–297.
- Häder, Sabine/ Gabler, Siegfried/ Heckel, Christiane (2009): Stichprobenziehung, Gewichtung und Realisierung. In: Häder, Michael/ Häder, Sabine (Hrsg.), *Telefonbefragungen über das Mobilfunknetz - Konzept, Design und Umsetzung einer Strategie zur Datenerhebung*, Wiesbaden: Springer Verlag, 21–45.
- Hansen, Kasper M./ Pedersen, Rasmus T. (2012): Efficiency of Different Recruitment Strategies for Web Panels. *International Journal of Public Opinion Research*, 24(2), 238–249.
- Haunberger, Sigrid (2011): *Teilnahmeverweigerung in Panelstudien*. Wiesbaden: Springer Verlag.

- Heckel, Christiane (2002): Erstellung der ADM-Telefonauswahlgrundlage. In: Gabler, Siegfried/ Häder, Sabine (Hrsg.), *Telefonstichproben*, Münster, New York, München, Berlin: Waxmann, 11–32.
- Von der Heyde, Christian (2014): Einwohnermeldeamtsstichproben (EWA-Stichproben). In: Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. (Hrsg.), *Stichproben-Verfahren in der Umfrageforschung*, Wiesbaden: Springer Verlag, 191–195.
- Hox, Joop/ De Leeuw, Edith D. (2002): The Influence of Interviewers Attitude and Behavior on Household Survey Nonresponse: An International Comparison. In: Groves, Robert M./ Dillman, Don/ Eltinge, John L./ Little, Roderick J.A. (Hrsg.), *Survey Nonresponse*, New York: John Wiley & Sons, Inc., 103–120.
- Hüfken, Volker (2014): Telefonische Befragung. In: Baur, Nina/ Blasius, Jörg (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung*, Wiesbaden: Springer Verlag, 631–641.
- Huinink, Johannes/ Brüderl, Josef/ Nauck, Bernhard/ Walper, Sabine/ Castiglioni, Laura/ Feldhaus, Michael (2011): Panel Analysis of Intimate Relationships and Family Dynamics (Pairfam): Conceptual Framework and Design. *Journal of Family Research*, 23(1), 77–101.
- James, Jeannine M./ Bolstein, Richard (1990): The Effect of Monetary Incentives and Follow-Up Mailings on the Response Rate and Response Quality in Mail Surveys. *Public Opinion Quarterly*, 54(3), 346–361.
- Jobber, David/ Saunders, John/ Mitchell, Vince-Wayne (2004): Prepaid Monetary Incentive Effects on Mail Survey Response. *Journal of Business Research*, 57(1), 21–25.
- Kaczmirek, Lars (2015): Conducting Web Surveys. In: Engel, Uwe/ Jann, Ben/ Lynn, Peter/ Scherpenzeel, Annette/ Sturgis, Patrick (Hrsg.), *Improving Survey Methods. Lessons from Recent Research*, New York: Routledge, 153–155.
- Kane, Emily W./ Macaulay, Laura J. (1993): Interviewer Gender and Gender Attitudes. *Public Opinion Quarterly*, 57(1), 1–28.
- Kaplowitz, Michael D./ Hadlock, Timothy D./ Levine, Ralph (2004): A Comparison of Web and Mail Survey Response Rates. *Public Opinion Quarterly*, 68(1), 94–101.



- Kersten, Hubert M. P./ Bethlehem, Jelke G. (1984): Exploring and Reducing the Non-response Bias by Asking the Basic Question. *Statistical Journal of the United Nations Economic Commission for Europe*, 2(4), 369–380.
- Kish, Leslie (1962): Studies of Interviewer Variance for Attitudinal Variables. *Journal of the American Statistical Association*, 57(297), 92–115.
- Klausch, Thomas/ Hox, Joop/ Schouten, Barry (2015): Selection Error in Single and Mixed Mode Surveys of the Dutch General Population. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178, 945–961.
- Klingler, Walter/ Müller, Dieter K (2001): MA 2001 Radio: Kontinuität bei Methode und Ergebnissen. Hörfunknutzung in Deutschland. *Media Perspektiven*, 9, 434–449.
- Koch, Achim (1997): ADM-Design und Einwohnermelderegister-Stichproben. In: Gabler, Siegfried/ Hoffmeyer-Zlotnik, Jürgen H. P. (Hrsg.), *Stichproben in der Umfragepraxis*, Opladen: Westdeutscher Verlag, Opladen, 99–117.
- Kreuter, Frauke (2013a): Facing the Nonresponse Challenge. *The Annals of the American Academy of Political and Social Science*, 645(1), 23–35.
- Kreuter, Frauke (2013b): *Improving Surveys With Paradata: Analytic Uses of Process Information*. New York: John Wiley & Sons inc.
- Kreuter, Frauke/ Carolina, Casas-Cordero (2010): Paradata. *Working Paper Series des Rates für Sozial- und Wirtschaftsdaten* 136, 4.
- Kreuter, Frauke/ Olson, Kristen (2011): Multiple Auxiliary Variables in Nonresponse Adjustment. *Sociological Methods & Research*, 40, 1–22.
- Kreuter, Frauke/ Olson, Kristen (2013): Paradata for Nonresponse Error Investigation. In: Kreuter, Frauke (Hrsg.), *Improving Surveys with Paradata - Analytic Uses of Process Information*, New York: John Wiley & Sons inc., 13–34.
- Kreuter, Frauke/ Couper, Mick/ Lyberg, Lars (2010): The Use of Paradata to Monitor and Manage Survey Data Collection. *Proceedings of the Joint Statistical Meetings, American Statistical Association*, 282–296.
- Krosnick, Jon A. (1991): Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*, 5(3), 213–236.

- Krosnick, Jon A. (1999): Survey Research. *Annual Review of Psychology*, 50(1), 537–567.
- Kühnel, Steffen-M./ Krebs, Dagmar (2001): *Statistik für die Sozialwissenschaften*. Reinbek bei Hamburg: Rowohlt-Verlag.
- Ladik, Dainel M./ Carrilat, Fracois A./ Solomon, Paul J. (2007): The Effectiveness of University Sponsorship in Increasing Survey Response Rate. *Journal of Marketing Theory and Practice*, 15(3), 263–271.
- Lavrakas, Paul J. (2010): Telephone Surveys. In: Marsden, Peter V./ Wright, James D. (Hrsg.), *Handbook of Survey Research*, Bingley: Emerald Group Publishing, 471–499.
- Leenheer, Jorna/ Scherpenzeel, Annette (2011): Does It Pay Off to Include Non-internet Household in an Internet Panel? *International Journal of Internet Science*, 8, 17–29.
- Lepkowski, James M./ Couper, Mick P. (2002): Nonresponse in the Second Wave Of Longitudinal Household Surveys. In: Groves, Robert M./ Dillman, Don/ Eltinge, John L./ Little, Roderick J.A. (Hrsg.), *Survey Nonresponse*, New York: John Wiley and Sons, Inc., 259–272.
- Lepkowski, James M./ Tucker, Clyde/ Brick, J. Michael/ De Leeuw, Edith D./ Japac, Lilli/ Lavrakas, Paul J./ Link, Michael W./ Sangster, Roberta L. (2007): *Advances in Telephone Survey Methodology*. New York: John Wiley & Sons Inc.
- Link, Michael W./ Battaglia, Michael P./ Frankel, Martin R./ Osborn, Larry/ Mokdad, Ali H. (2008): A Comparison of Address-Based Sampling (ABS) Versus Random-Digit Dialing (RDD) for General Population Surveys. *Public Opinion Quarterly*, 72(1), 6–27.
- Little, Roderick J./ Vartivarian, Sonya (2005): Does Weighting for Nonresponse Increase the Variance of Survey Means? *Survey Methodology*, 31, 161–168.
- Little, Roderick J. A./ Rubin, Donald B. (2002): *Statistical Analysis With Missing Data*. New York: John Wiley & Sons Inc.
- Long, J. Scott/ Freese, Jeremy (2006): *Regression Models for Categorical Dependent Variables Using Stata*. College Station, Tex: Stata press.
- Lugtig, Peter/ Toepoel, Vera (2016): The Use of PCs, Smartphones, and Tablets in a Probability-Based Panel Survey: Effects on Survey Measurement Error. *Social Science Computer Review*, 34, 78–94.

- Lynn, Peter (2003): PEDAKSI: Methodology for Collecting Data About Survey Non-Respondents.
- Lynn, Peter (2009): *Methodology of Longitudinal Surveys*. New York: John Wiley & Sons Inc.
- Lynn, Peter/ Kaminska, Olena (2012): The Impact of Mobile Phones on Survey Measurement Error. *Public Opinion Quarterly*, 77(2), 586–605.
- Maitland, Aaron/ Casas-Cordero, Carolina/ Kreuter, Frauke (2009): An Evaluation of Nonresponse Bias Using Paradata From a Health Survey. *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, 370–378.
- Marcus, Bernd/ Bosnjak, Michael/ Lindner, Steffen/ Pilischenko, Stanislav/ Schütz, Astrid (2007): Compensating for Low Topic Interest and Long Surveys: A Field Experiment on Nonresponse in Web Surveys. *Social Science Computer Review*, 25(3), 372–383.
- Martin, Elizabeth/ Winters, Franklin (2001): Money and Motive: Effects of Incentives on Panel Attrition in the Survey of Income and Program Participation. *Journal of Official Statistics*, 17(2), 267–284.
- Mavletova, Aigul (2013): Data Quality in PC and Mobile Web Surveys. *Social Science Computer Review*, 31(6), 725–743.
- Mayring, Philipp (2010): Qualitative Inhaltsanalyse. In: Mey, Günter (Hrsg.), *Handbuch qualitative Forschung in der Psychologie*, Wiesbaden: Springer Verlag, 601–613.
- Meier, Gerd/ Glemser, Axel/ Heckel, Christiane (2014): Random-Telefonstichproben. In: und Sozialforschungsinstitute e.V., ADM Arbeitskreis Deutscher Markt (Hrsg.), *Stichproben-Verfahren in der Umfrageforschung*, Wiesbaden: Springer Verlag, 117–135.
- Messer, Benjamin L./ Dillman, Don A. (2010): Using Address-Based Sampling to Survey the General Public by Mail vs. Web Plus Mail. *Social and Economic Sciences Research Center Technical Report*, Technical Report 10-13, empfangen am: 29.05.2017.
- URL <https://sesrc.wsu.edu/dillman/papers/2010/Messer%20Dillman%20WCSTechReport.pdf>

- Messer, Benjamin L./ Dillman, Don A. (2011): Surveying the General Public Over the Internet Using Address-Based Sampling and Mail Contact Procedures. *Public Opinion Quarterly*, 75(3), 429–457.
- Meulemann, Heiner/ Hagenah, Jörg (2006): *Sozialer Wandel und Mediennutzung in der Bundesrepublik Deutschland*. Berlin: LIT Verlag.
- Millar, Morgan M./ Dillman, Don A. (2011): Improving Response to Web and Mixed-Mode Surveys. *Public Opinion Quarterly*, 75(2), 249–269.
- Mohorko, A./ Leeuw, E. De/ Hox, J. (2013): Coverage Bias in European Telephone Surveys: Developments of Landline and Mobile Phone Coverage across Countries and over Time. *Survey Methods: Insights from the Field (SMIF)*, empfangen am: 29.05.2017.  
URL <http://surveyinsights.org/?p=828>
- Nederhof, Anton J (1983): The Effects of Material Incentives in Mail Surveys: Two Studies. *Public Opinion Quarterly*, 47(1), 103–112.
- Olson, Kristen (2013): Paradata for Nonresponse Adjustment. *The Annals of the American Academy of Political and Social Science*, 645(1), 142–170.
- Petrolia, Daniel R./ Bhattacharjee, Sanjoy (2009): Revisiting Incentive Effects Evidence from a Random-Sample Mail Survey on Consumer Preferences for Fuel Ethanol. *Public Opinion Quarterly*, 73(3), 537–550.
- Peytchev, Andy (2013): Consequences of Survey Nonresponse. *The ANNALS of the American Academy of Political and Social Science*, 645(1), 88–111.
- Peytchev, Andy/ Hill, Craig A. (2010): Experiments in Mobile Web Survey Design Similarities to Other Modes and Unique Considerations. *Social Science Computer Review*, 28(3), 319–335.
- Poggio, Teresio/ Bosnjak, Michael/ Weyandt, Kai (2013): Online Survey Participation via Mobile Devices: Implications for Nonresponse. *Conference Paper. ESRA Conference, Ljubljana*.
- Rao, Kumar/ Kaminska, Olena/ McCutcheon, Allan L. (2010): Recruiting Probability Samples for a Multi-Mode Research Panel with Internet and Mail Components. *Public Opinion Quarterly*, 74(1), 68–85.

- Reuband, Karl-Heinz (2001): Möglichkeiten und Probleme des Einsatzes postalischer Befragungen. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 53(2), 307–333.
- Reuband, Karl-Heinz (2015): Ausschöpfung und Nonresponse Bias in postalischen Befragungen. In: Schupp, Jürgen/ Wolf, Christof (Hrsg.), *Nonresponse Bias*, Wiesbaden: Springer Verlag, 209–251.
- Rivella, Melanie/ Toninelli, Daniele/ Ochoa, Carlos/ Loewe, German (2014): Do Online Access Panels Really Need to Allow and Adapt Surveys to Mobile Devices? *RECSM Working Paper Number 41*, empfangen am: 21.06.2017.  
URL [https://www.upf.edu/documents/3966940/3986764/RECSM\\_wp041.pdf](https://www.upf.edu/documents/3966940/3986764/RECSM_wp041.pdf)
- Roose, Henk/ Lievens, John/ Waege, Hans (2007): The Joint Effect of Topic Interest and Fallow-Up Procedures on the Response in a Mail Questionnaire: an Empirical Test of the Leverage-Saliency Theory in Audience Research. *Sociological Methods & Research*, 35(3), 410–428.
- Rosenbaum, Paul R./ Rubin, Donald B. (1983): The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1), 41–55.
- Rubin, Donald B. (1976): Inference and Missing Data. *Biometrika*, 63(3), 581–592.
- Saßenroth, Denise (2012): The Impact of Personality Traits on the Willingness to Cooperate in Surveys. *Methoden - Daten - Analysen*, 6(1), 12–44.
- Saßenroth, Denise (2013): *The Impact of Personality on Participation Decisions in Surveys: A Contribution to the Discussion on Unit Nonresponse*. Wiesbaden: Springer Verlag.
- Scherpenzeel, Annette/ Toepoel, Vera (2012): Recruiting a Probability Sample for an Online Panel Effects of Contact Mode, Incentives, and Information. *Public Opinion Quarterly*, 76(3), 470–490.
- Schnell, Rainer (1991): Wer ist das Volk? Zur faktischen Grundgesamtheit bei allgemeinen Bevölkerungsumfragen: Undercoverage, Schwererreichbare und Nichtbefragbare. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 43(1), 106–137.
- Schnell, Rainer (2012): *Survey-Interviews: Methoden standardisierter Befragungen*. Wiesbaden: Springer Verlag.

- Schonlau, Matthias/ Asch, Beth J./ Du, Can (2003): Web Surveys as Part of a Mixed-Mode Strategy for Publications That Cannot Be Contacted by E-Mail. *Social Science Computer Review*, 21(2), 218–222.
- Schonlau, Matthias/ van Soest, Arthur/ Kapteyn, Arie (2007): Are "Webographic" Attitudinal Questions Useful for Adjusting Estimates from Web Surveys Using Propensity Scoring? *Survey Research Methods*, 1(3), 155–163.
- Schouten, Barry (2004): Adjustment for bias in the integrated survey on household living condition (pols) 1998. *Discussion paper 04001*, empfangen am: 30.05.2017.  
 URL <https://www.cbs.nl/-/media/imported/documents/2004/10/discussion-paper-04001.pdf?la=nl-nl>
- Schupp, Jürgen (2014): Paneldaten für die Sozialforschung. In: Baur, Nina/ Blasius, Jörg (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung*, Wiesbaden: Springer, 925–939.
- Schupp, Jürgen/ Wolf, Christof (2015): *Nonresponse Bias: Qualitätssicherung sozialwissenschaftlicher Umfragen*. Wiesbaden: Springer Verlag.
- Shih, Tse-Hua/ Fan, Xitao (2008): Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis. *Field Methods*, 20(3), 249–271.
- Singer, Eleanor (2006): Introduction Nonresponse Bias in Household Surveys. *The Public Opinion Quarterly*, 70(5), 637–645.
- Singer, Eleanor/ Ye, Cong (2013): The Use and Effects of Incentives in Surveys. *The ANNALS of the American Academy of Political and Social Science*, 645(1), 112–141.
- Singer, Eleanor/ Van Hoewyk, John/ Maher, Mary P. (1998): Does the Payment of Incentives Create Expectation Effects? *Public Opinion Quarterly*, 152–164.
- Singer, Eleanor/ Groves, Robert M./ Corning, Amy D. (1999): Differential Incentives: Beliefs about Practices, Perceptions of Equity, and Effects on Survey Participation. *Public Opinion Quarterly*, 251–260.
- Singer, Eleanor/ Van Hoewyk, John/ Maher, Mary P. (2000): Experiments with Incentives in Telephone Surveys. *Public Opinion Quarterly*, 64(2), 171–188.

Smith, Tom W. (2011): The Report of the International Workshop on Using Multi-Level Data from Sample Frames, Auxiliary Databases, Paradata and Related Sources to Detect and Adjust for Nonresponse Bias in Surveys. *International Journal of Public Opinion Research*, 23(3), 389–402.

Spieß, Martin (2010): Der Umgang mit fehlenden Werten. In: Wolf, Christof/ Best, Henning (Hrsg.), *Handbuch der sozialwissenschaftlichen Datenanalyse*, Springer, 117–142.

Statistisches Bundesamt (2015a): Ausstattung privater Haushalte mit Informations- und Kommunikationstechnik - Deutschland. Empfangen am: 20.06.2015.

URL [https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/EinkommenKonsumLebensbedingungen/AusstattungGebrauchsguetern/Tabellen/Infotechnik\\_D.html](https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/EinkommenKonsumLebensbedingungen/AusstattungGebrauchsguetern/Tabellen/Infotechnik_D.html)

Statistisches Bundesamt (2015b): Pressemitteilung vom 12. Mai 2015. Empfangen am: 20.06.2015.

URL [https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2015/05/PD15\\_172\\_631.pdf;jsessionid=34E223203B1949854D7248CC0D6794FD.cae3?\\_\\_blob=publicationFile](https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2015/05/PD15_172_631.pdf;jsessionid=34E223203B1949854D7248CC0D6794FD.cae3?__blob=publicationFile)

Statistisches Bundesamt (2015c): Wirtschaftsrechnungen - Private Haushalte in der Informationsgesellschaft - Nutzung von Inforamtions und Kommunikationstechnologien. Empfangen am: 27.08.2015.

URL [https://www.destatis.de/DE/Publikationen/Thematisch/EinkommenKonsumLebensbedingungen/PrivateHaushalte/PrivateHaushalteIKT2150400157004.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Publikationen/Thematisch/EinkommenKonsumLebensbedingungen/PrivateHaushalte/PrivateHaushalteIKT2150400157004.pdf?__blob=publicationFile)

Statistisches Bundesamt (2016): Private Haushalte in der Informationsgesellschaft - Nutzung von Informations- und Kommunikationstechnologien. *Fachserie 15, Reihe 4*, empfangen am: 30.05.2017.

URL [https://www.destatis.de/DE/Publikationen/Thematisch/EinkommenKonsumLebensbedingungen/PrivateHaushalte/PrivateHaushalteIKT2150400157004.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Publikationen/Thematisch/EinkommenKonsumLebensbedingungen/PrivateHaushalte/PrivateHaushalteIKT2150400157004.pdf?__blob=publicationFile)

Statistisches Landesamt Bremen (2016): Stadtteil- und Ortsteiltabellen. Empfangen am: 08.06.2017.

URL [http://www.statistik.bremen.de/datenangebote/stadtteil\\_\\_und\\_\\_ortsteiltabellen-4529](http://www.statistik.bremen.de/datenangebote/stadtteil__und__ortsteiltabellen-4529)

- Stoop, Ineke (2005): *The Hunt For the Last Respondent: Nonresponse in Sample Surveys*. The Hague: Social and Cultural Planning Office of the Netherlands.
- Stoop, Ineke (2006): Access Panels and Online Surveys: Mystifications and Misunderstandings. In: Stoop, Ineke/ Wittenberg, Marion (Hrsg.), *Access Panels and Online Research, Panacea or Pitfall?*, Amsterdam: Aksant Academic Publishers, 5–18.
- Stoop, Ineke/ Billiet, Jaak/ Koch, Achim/ Fitzgerald, Rory (2010): *Improving Survey Response: Lessons Learned from the European Social Survey*. New York: John Wiley & Sons Inc.
- TNS Infratest (2014): D21-Digita-Index 2014 - Die Entwicklung der digitalen Gesellschaft in Deutschland. Empfangen am: 11.12.2015.  
URL <http://initiated21.de/app/uploads/2017/02/d21-digital-index-2014.pdf>
- Toepoel, Vera (2012): Effects of Incentives in Surveys. In: Gideon, Lior (Hrsg.), *Handbook of Survey Methodology for the Social Sciences*, Wiesbaden: Springer Verlag, 209–223.
- Toepoel, Vera/ Lugtig, Peter (2014): What Happens if You Offer a Mobile Option to Your Web Panel? Evidence From a Probability-Based Panel of Internet Users. *Social Science Computer Review*, 32(4), 544–560.
- Toepoel, Vera/ Lugtig, Peter (2015): Online Surveys are Mixed-Device Surveys. Issues Associated With the Use of Different (Mobile) Devices in Web Surveys. *Methods, Data, Analysis*, 9(2), 155–162.
- Tourangeau, Roger (2004): Survey Research and Societal Change. *Annual Review of Psychology*, 55, 775–801.
- Trussell, Norm/ Lavrakas, Paul J. (2004): The Influence of Incremental Increases in Token Cash Incentives on Mail Survey Response: Is There an Optimal Amount? *Public Opinion Quarterly*, 68(3), 349–367.
- Valliant, Richard/ Dever, Jill A./ Kreuter, Frauke (2013): *Practical Tools For Designing and Weighting Survey Samples*. Wiesbaden: Springer Verlag.



- Van Eimeren, Birgit/ Frees, Beate (2014): Ergebnisse der ARD/ZDF-Onlinestudie 2014: 79% der Deutschen online - Zuwachs bei mobiler Internetnutzung und Bewegtbild. *Media Perspektiven*, 45(7-8), 378–396.
- Vannieuwenhuyze, Jorre T. A. (2014): On the Relative Advantage of Mixed-Mode versus Single-Mode Surveys. *Survey Research Methods*, 8(1), 31–42.
- Vehovar, Vasja/ Slavec, Ana/ Berzelak, Nejc (2012): Costs and Errors in Fixed and Mobile Phone Surveys. In: Gideon, Lior (Hrsg.), *Handbook of Survey Methodology for the Social Sciences*, Wiesbaden: Springer Verlag, 277–295.
- Vehre, Helen (2011): *Sie wollen mir doch was verkaufen: Analyse der Umfrageteilnahme in einem offline rekrutierten Access Panel*. Wiesbaden: Springer Verlag.
- Vehre, Helen/ Bartsch, Simone/ Engel, Uwe (2013): The Education Bias in a Telephone Pretest with Incentives. Empfangen am: 30.05.2017.  
URL [https://www.methodenzentrum-bremen.de/uploads/Vehre\\_et\\_al\\_\[Education\\_Bias\].pdf](https://www.methodenzentrum-bremen.de/uploads/Vehre_et_al_[Education_Bias].pdf)
- Verbraucherzentrale Nordrhein-Westfalen (2012): Auswertung der Aktion Telefonwerbung und untergeschobene Verträge. Empfangen am: 17.06.2015.  
URL <http://www.vz-nrw.de/mediabig/218817A.pdf>
- Voogt, Robert J. J./ Saris, Willem E. (2005): Mixed Mode Designs: Finding the Balance Between Nonresponse Bias and Mode Effects. *Journal of Official Statistics*, 21(3), 367–387.
- Wagner, James (2010): The Fraction of Missing Information as a Tool for Monitoring the Quality of Survey Data. *Public Opinion Quarterly*, 74(2), 223–243.
- Wagner, James (2012): A Comparison of Alternative Indicators for the Risk of Nonresponse Bias. *Public Opinion Quarterly*, 76(3), 555–575.
- Wagner, Pia/ Hering, Linda (2014): Online-Befragung. In: Baur, Nina/ Blasius, Jörg (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung*, Wiesbaden: Springer Verlag, 661–673.
- Waksberg, Joseph (1978): Sampling Methods for Random Digit Dialing. *Journal of the American Statistical Association*, 73(361), 40–46.

- Watson, Nicole/ Wooden, Mark (2009): Identifying Factors Affecting Longitudinal Survey Response. In: Lynn, Peter (Hrsg.), *Methodology of Longitudinal Surveys*, New York: John Wiley & Sons Inc., 157–182.
- Weisberg, Herbert F. (2005): *The Total Survey Error Approach: A Guide to the New Science of Survey Sesearch*. Chicago: University of Chicago Press.
- Windzio, Michael (2013): *Regressionsmodelle für Zustände und Ereignisse: Eine Einführung*. Wiesbaden: Springer Verlag.

## Ehrenwörtliche Erklärung

Hiermit erkläre ich, dass

- die Arbeit ohne unerlaubte fremde Hilfe angefertigt ist;
- keine anderen als die angegebenen Quellen und Hilfsmittel benutzt wurden und
- die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht sind;
- eine Überprüfung der Dissertation mit qualifizierter Software im Rahmen der Untersuchung von Plagiatsvorwürfen gestattet ist.

---

Datum, Ort

---

Unterschrift

# Anhang

# Anhang A

## Weitere Ergebnistabellen

Tabelle A.1: Logistische Regression mit der Teilnahme an der CAWI-Variante der Befragung (0 = Nein; 1 = Ja ) als abhängige Variable für die Bruttostichprobe

	M1		M2		M3		M4	
	OR	s.e.	OR	s.e.	OR	s.e.	OR	
Geschlecht (Ref. Weiblich)	3,31 ***	0,61	3,31 ***	0,42	3,28 ***	0,61	3,41 ***	0,43
Alter in Jahren			1,00	0,00	1,12 ***	0,03	1,12 ***	0,03
Alter in Jahren <sup>2</sup>					0,99 ***	0,00	0,99 ***	0,00
<i>Stadtteile</i> (Ref. Walle)								
Blumenthal							0,67	0,34
Burglesum							2,22	1,11
Häfen							1,22	0,71
Neustadt							1,51	0,70
Oberneuland							2,63 *	1,07
Obervieland							1,15	0,48
Osterholz							0,85	0,33
Östliche Vorstadt							2,25 *	0,76
Schwachhausen							2,58 **	0,83
Vahr							1,38	0,44
Veogesack							0,52	0,23
Woltmershausen							1,51	0,48
<i>Fallzahl (n)</i>	2.306		2.306		2.306		2.306	
<i>Modellgüte</i>								
Prob > chi <sup>2</sup>	***		***		***		***	
Pseudo R <sup>2</sup>	0,04		0,04		0,06		0,08	
LL	-589,78		-589,78		-580,60		-564,55	
LR-Test					***		**	

\*  $p \leq 0,05$ , \*\*  $p \leq 0,01$ , \*\*\*  $p \leq 0,001$ , +  $p \leq 0,10$

PPSM Access Panel Nachrekrutierung 2012, eigene Berechnungen

Tabelle A.2: Logistische Regression mit der Teilnahme an der CATI-Variante der Befragung (0 = Nein; 1 = Ja ) als abhängige Variable für die Bruttostichprobe

	M1		M2		M3		M4	
	OR	s.e.	OR	s.e.	OR	s.e.	OR	s.e.
Geschlecht (Ref. Weiblich)	2,51 ***	0,52	2,61 ***	0,55	2,50 ***	0,53	2,50 ***	0,53
Alter in Jahren			1,04 ***	0,01	1,26 ***	0,08	1,26 ***	0,67
Alter in Jahren <sup>2</sup>					0,99 ***	0,00	0,99 ***	0,00
<i>Stadtteile</i> (Ref. Walle)								
Blumenthal							1,13	0,59
Burglesum							2,31	1,14
Häfen							1,64	1,00
Neustadt							1,04	0,69
Oberneuland							1,80	0,95
Obervieland							1,70	0,78
Osterholz							2,14 *	0,82
Östliche Vorstadt							2,02	0,97
Schwachhausen							1,24	0,61
Vahr							1,14	0,46
Ve gesack							1,02	0,45
Woltmershausen							1,14	0,75
<i>Fallzahl (n)</i>	2.250		2.250		2.250		2.250	
<i>Modellgüte</i>								
Prob > chi <sup>2</sup>	***		***		***		***	
Pseudo R <sup>2</sup>	0,02		0,08		0,10		0,11	
LL	-449,08		-442,74		-413,25		-408,60	
LR-Test			***		***			

\*  $p \leq 0,05$ , \*\*  $p \leq 0,01$ , \*\*\*  $p \leq 0,001$ , +  $p \leq 0,10$

PPSM Access Panel Nachrekrutierung 2012, eigene Berechnungen

## Anhang B

# Umfragevariablen Nachrekrutierung 2012

### [EU10]

Wie sehr fühlen Sie sich persönlich von der Eurokrise betroffen? Auf einer Skala von 1 = „gar nicht betroffen“ bis 11 = „sehr betroffen“, wo auf dieser Skala würden Sie sich einordnen?

Zusätzliche Antwortoptionen: weiß nicht/ Keine Angabe

### [EU16 - EU20]

Wenn ein dauerhafter Finanzausgleich zwischen den Euro-Ländern nur zum Preis persönlicher Nachteile eingerichtet werden könnte: Wären Sie bereit, den ein oder anderen im Folgenden genannten Nachteil auf sich zu nehmen? Auf einer Skala von 1 bis 7 soll die 1 bedeuten, „würde ich auf keinen Fall auf mich nehmen“ und die 7 „würde ich auf jeden Fall auf mich nehmen“. Mit den Zahlen dazwischen können Sie ihre Antwort abstufen.

Zusätzliche Antwortoptionen: weiß nicht/ Keine Angabe

1. Später in Rente oder Pension zu gehen

2. allgemeine finanzielle Nachteile, z. B. durch Inflation
3. Höhere Steuern auf Ihre monatlichen Einkünfte, z. B. 2,5 Prozent vom Bruttolohn
4. Höhere Mehrwertsteuern
5. Einführung eines Solidaritätszuschlag auf EU-Ebene, z. B. 1 Prozent vom Bruttolohn

[S13]

Auf einer Skala von 1 „überhaupt nicht zuversichtlich“ bis 11 „sehr zuversichtlich“:  
Wo auf dieser Skala würden Sie sich einordnen?

Zusätzliche Antwortoptionen: weiß nicht/ Keine Angabe

[S14]

Ist Ihr Lebensstandard sicher? Auf einer Skala von 1 „überhaupt nicht sicher“ bis 11 „sehr sicher“: Wo auf dieser Skala würden Sie sich einordnen?

Zusätzliche Antwortoptionen: weiß nicht/ Keine Angabe

[S16]

Auf einer Skala von 1 „sehr unglücklich“ bis 11 „sehr glücklich“: Wo auf dieser Skala würden Sie sich einordnen?

Zusätzliche Antwortoptionen: weiß nicht/ Keine Angabe

[S17]

Wie zufrieden sind Sie - alles in allem - mit ihrem gegenwärtigen Leben? Auf einer Skala von 1 bis 11 soll die 1 bedeuten, dass sie „sehr unzufrieden“ sind und die 11, dass sie „sehr zufrieden“ sind. Mit den Zahlen dazwischen können Sie Ihre Einschätzung



abstufen.

Zusätzliche Antwortoptionen: weiß nicht/ Keine Angabe

**[S18]**

Und auf derselben Skala gefragt: Wie zufrieden sind Sie mit ihrem gegenwärtigen Lebensstandard?

Zusätzliche Antwortoptionen: weiß nicht/ Keine Angabe

# Anhang C

## Soziodemografie

Welchen höchsten allgemeinbildenden Schulabschluss haben Sie?

Noch Schüler/ Schule beendet ohne Abschluss/ Volks- bzw. Hauptschulabschluss/ Polytechnische Oberschule mit Abschluss 8. oder 9. Klasse/ Mittlere Reife, REalschulabschluss bzw. Polytechnischer Oberschule mit Abschluss 10. Klasse/ Fachhochschulreife (Abschluss einer Fachoberschule etc.)/ Abitur bzw. Erweiterte Oberschule mit Abschluss 12. Klasse (Allg. oder fachgebundene Hochschulreife)/ Anderer Schulabschluss und zwar ...

# Anhang D

## Einladungsschreiben 2012

[Untersuchungsgruppe: B, C, F, G, H, I]

Sehr geehrte/r Frau/Herr Nachname,

die Eurokrise ist sicherlich das beherrschende Thema dieser Tage. Zur Lösung dieser Krise wird viel von Deutschland erwartet. Wie weitreichend aber kann, bzw. wie weitreichend sollte die Unterstützung für andere Länder des Euroraums ausfallen?

Deutschland in Europa: Die Politik steht vor wichtigen Entscheidungen. Wir haben einige Möglichkeiten zusammengetragen, wie es weitergehen könnte, und möchten Sie bitten, uns dazu Ihre Meinung zu sagen.

Unsere Umfrage zum Thema „Eurokrise“ wird im Zeitraum vom 26. November bis 21. Dezember 2012 stattfinden. Wir laden Sie herzlich ein, an dieser Umfrage teilzunehmen. Wenn Sie mögen, können wir Sie hierzu telefonisch kontaktieren. Dafür müssten Sie uns lediglich mittels des beigelegten Rückantwortbriefes Ihre Telefonnummer zurückzusenden. Dieser ist bereits ausreichend frankiert. Sie können uns auf unsere Fragen aber auch gerne „online“ im Internet antworten. Falls Ihnen diese Möglichkeit lieber sein sollte, bitten wir Sie, uns über den geschützten Link

[www.befragungsgemeinschaft.de/umfrage](http://www.befragungsgemeinschaft.de/umfrage)

zur Befragungsseite zu folgen. Die Umfrage ist dort bereits eingerichtet: Wenn Sie mögen, können Sie gleich heute an der Umfrage teilnehmen.

Hierfür müssen Sie lediglich dem geschützten Link folgen und nachfolgendes Zugangspasswort eingeben: .....

Die Teilnahme an der Umfrage ist selbstverständlich freiwillig. Aus einer Nichtteilnahme entstehen Ihnen keinerlei Nachteile. Alle Ihre Angaben werden streng vertraulich behandelt, ausschließlich für wissenschaftliche Zwecke verwendet und nicht an Dritte weitergegeben. Gerne informieren wir Sie, wenn die statistischen Ergebnisse vorliegen.

Über den Datenschutz informieren wir auf der Internetseite [www.befragungsgemeinschaft.de](http://www.befragungsgemeinschaft.de) und senden Ihnen die Datenschutzerklärung auf Wunsch gerne auch per Post zu.

Wir würden uns über Ihre Teilnahme sehr freuen!

Mit freundlichen Grüßen

gescannte - Unterschrift -

Prof. Dr. Uwe Engel

Für Rückfragen stehen wir Ihnen jederzeit zur Verfügung.

**[Optional in den Untersuchungsgruppen F und H:]**

P.S.: Ohne Antworten kann keine wissenschaftliche Umfrage funktionieren. Wir möchten uns deshalb bei allen, die uns für die Umfrage ein wenig ihrer kostbaren Zeit opfern, mit einer Verlosung attraktiver Preise bedanken. Teilnehmen kann daran jeder, der auch an der Umfrage teilgenommen hat.

[Untersuchungsgruppe: A, D, E]

Sehr geehrte/r Frau/Herr Nachname,

die Eurokrise ist sicherlich das beherrschende Thema dieser Tage. Zur Lösung dieser Krise wird viel von Deutschland erwartet. Wie weitreichend aber kann, bzw. wie weitreichend sollte die Unterstützung für andere Länder des Euroraums ausfallen?

Deutschland in Europa: Die Politik steht vor wichtigen Entscheidungen. Wir haben einige Möglichkeiten zusammengetragen, wie es weitergehen könnte, und möchten Sie bitten, uns dazu Ihre Meinung zu sagen.

Unsere Umfrage zum Thema „Eurokrise“ wird im Zeitraum vom 26. November bis 21. Dezember 2012 stattfinden. Wir laden Sie herzlich ein, an dieser Umfrage teilzunehmen, und möchten Sie dafür im besagten Zeitraum anrufen. Wenn es Ihnen recht sein sollte, können Sie das Interview dann gleich am Telefon geben, Sie können uns auf unsere Fragen aber auch gerne „online“ im Internet antworten. Falls Ihnen diese Möglichkeit lieber sein sollte, bitten wir Sie, uns über den geschützten Link

**[www.befragungsgemeinschaft.de/umfrage](http://www.befragungsgemeinschaft.de/umfrage)**

zur Befragungsseite zu folgen. Die Umfrage ist dort bereits eingerichtet: Wenn Sie mögen, können Sie gleich heute an der Umfrage teilnehmen.

Hierfür müssen Sie lediglich dem geschützten Link folgen und nachfolgendes Zugangspasswort eingeben: .....

Die Teilnahme an der Umfrage ist selbstverständlich freiwillig. Aus einer Nichtteilnahme entstehen Ihnen keinerlei Nachteile. Alle Ihre Angaben werden streng vertraulich behandelt, ausschließlich für wissenschaftliche Zwecke verwendet und nicht an Dritte weitergegeben. Gerne informieren wir Sie, wenn die statistischen Ergebnisse vorliegen.

Über den Datenschutz informieren wir auf der Internetseite [www.befragungsgemeinschaft.de](http://www.befragungsgemeinschaft.de) und senden Ihnen die Datenschutzerklärung auf Wunsch gerne auch per Post zu.

Wir würden uns über Ihre Teilnahme sehr freuen!

Mit freundlichen Grüßen

- gescannte Unterschrift -

Prof. Dr. Uwe Engel

Für Rückfragen stehen wir Ihnen jederzeit zur Verfügung.

**[Optional in Untersuchungsgruppe D]**

P.S.: Bitte erlauben Sie uns, dass wir Ihnen ganz unabhängig davon, ob Sie an der Umfrage teilnehmen möchten oder nicht, ein kleines Dankeschön für Ihre Aufmerksamkeit beilegen.