

gesis

Leibniz-Institut
für Sozialwissenschaften

Schriftenreihe

Band 20

Matthias Sand

**Gewichtungsverfahren
in Dual-Frame-Telefon-
erhebungen bei Device-
Specific Nonresponse**

Gewichtungsverfahren in Dual-Frame-Telefonerhebungen

GESIS-Schriftenreihe

herausgegeben von GESIS – Leibniz-Institut für Sozialwissenschaften

Band 20

Matthias Sand

**Gewichtungsverfahren in Dual-Frame-Telefonerhebungen
bei Device-Specific Nonresponse**

Die vorliegende Arbeit wurde von Philosophischen Fakultät der Technischen Universität Dresden als Dissertation zur Erlangung des akademischen Grades eines Doktors der Sozialwissenschaften (Dr. phil.) angenommen.

Matthias Sand

Gewichtungsverfahren in Dual-Frame- Telefonerhebungen bei Device-Specific Nonresponse

Bibliographische Information Der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

ISBN	978-3-86819-031-1 (print)
ISBN	978-3-86819-032-8 (eBook)
ISSN	1869-2869

Herausgeber,

Druck u. Vertrieb: GESIS – Leibniz-Institut für Sozialwissenschaften
Unter Sachsenhausen 6-8, 50667 Köln, Tel.: 0221 / 476 94 - 0
info@gesis.org
Printed in Germany

Inhaltsverzeichnis

Abbildungsverzeichnis	VII
Tabellenverzeichnis	IX
Symbolverzeichnis	XV
Abkürzungsverzeichnis	XVII
1 Einleitung und Motivation	1
2 Telefonstichproben in Deutschland	7
2.1 Möglichkeiten und Grenzen telefonischer Befragungen	8
2.2 Historische Entwicklung der Strategie zur Ziehung von Telefonstichproben . . .	14
2.2.1 Die Entwicklung allgemein	15
2.2.2 Die Entwicklung in Deutschland	17
2.2.3 Mobiltelefonie und das veränderte Telekommunikationsverhalten	21
2.3 Die Dual-Frame-Telefonstichprobe	22
2.3.1 Erreichbarkeit nach Telekommunikationsmedium	23
2.3.2 Mobile-Only-Haushalte und die damit verbundenen Schwierigkeiten . .	25
2.3.3 Dual-Frame-Erhebungen zur Berücksichtigung der Only-Haushalte . . .	28
2.4 Dual-Frame-Erhebungen in Deutschland	31
2.4.1 Entwicklung einer Auswahlgrundlage für Mobilfunkerhebungen	31
2.4.2 Umsetzung von Dual-Frame-Erhebungen	33
2.4.3 Methodische Herausforderungen bei Dual-Frame-Erhebungen	37
2.4.4 Die <i>Cellphone and Landline</i> (CELLA)-Erhebungen	41
2.4.5 Die Erhebung Influenza 2014 des <i>Robert Koch-Institutes</i> RKI	42
2.5 HLR-Lookup-Verfahren zur Bereinigung von Mobilfunkstichproben	43
2.5.1 Aufbau von Mobilfunknetzen	45
2.5.2 Reliabilität und Validität von Lookup-Verfahren	47
3 Gewichtungsverfahren bei Dual-Frame-Erhebungen	55
3.1 Ziel der Gewichtung von Erhebungsdaten	55
3.2 Designgewichtung im Allgemeinen	57
3.3 Designgewichtung in Dual-Frame-Erhebungen	58
3.3.1 Single-Frame-Gewichte	60
3.3.2 Composite-Gewichte	64

3.3.3	Weitere Designgewichtungsverfahren für Dual-Frame-Erhebungen . . .	69
3.4	Kalibrierung und Anpassungsgewichtung	72
3.4.1	Poststratifizierung (nachträgliche Schichtung)	77
3.4.2	Raking	78
3.4.3	Der <i>GREG</i> -Schätzer und die Kombination von Design- und Anpassungs- gewichtung	79
3.4.4	Exkurs: Berücksichtigung der "Klumpung" über die Interviewer bei telefonischen Befragungen	81
3.5	Vergleich der Ansätze zur Designgewichtung und Kalibrierung	94
3.5.1	Vergleich unterschiedlicher Designgewichtungsverfahren	95
3.5.2	Vergleich der Kalibrierungsverfahren	99
4	Zwischenfazit	103
5	Einführung eines Composite-Modells zur Berücksichtigung von Nutzungs- gewohnheiten der Dual-User	113
5.1	Darstellung und Beurteilung des Nutzungsverhaltens von Dual- Usern nach Kommunikationsmedium	114
5.2	Einteilung der Dual-User nach Nutzungsgewohnheiten	118
5.3	Das Domain-Specific-Composite-Design-Weighting-Modell	121
6	Untersuchung des Composite-Modells anhand der Erhebungen CELLA 2 und Influenza 2014	127
6.1	Untersuchung anhand des synthetischen Datensatzes	127
6.2	Anwendung auf die Daten der CELLA 2-Erhebung	131
6.2.1	CELLA 2: Einteilung in Verwendungskategorien	132
6.2.2	CELLA 2: Berechnung der Composite-Faktoren und Schätzung	146
6.3	Anwendung auf die Daten der Erhebung Influenza	149
6.3.1	Influenza 2014: Einteilung nach Verwendungskategorie	149
6.3.2	Influenza 2014: Berechnung der Composite-Faktoren und Schätzung . .	168
6.4	Vergleich der Ergebnisse	171
7	Der Domain-Specific-Single-Frame-Ansatz	181
7.1	Alternative Gewichtungsverfahren zur Berücksichtigung der Nutzungsgewohnheiten	181
7.2	Anwendung des Domain-Specific-Single-Frame-Ansatz	184

8	Beurteilung und Vergleich der verschiedenen Gewichtungsmodelle	193
8.1	Varianzschätzung als Indikator der Genauigkeit	193
8.2	Vergleich des Mean Squared Errors für verschiedene Gewichtungsansätze	198
8.3	Beurteilung der verschiedenen Gewichtungsverfahren	203
9	Fazit/Ausblick	207
	Literatur	217
A	Anhang	229
A.1	Überprüfung der Variablen zur Einordnung in Subdomains auf Unabhängigkeit .	229
A.1.1	χ^2 -Tests auf Unabhängigkeit der Proxies zur Einteilung nach Subdomain im Rahmen von CELLA 2	229
A.1.2	χ^2 -Tests auf Unabhängigkeit der Variablen zur Einteilung nach Subdomain im Rahmen von Influenza 2014	230
A.2	Verteilung ausgewählter soziodemografischer Merkmale nach Verwendungskategorie	233
A.2.1	CELLA 2: Verteilung ausgewählter soziodemografischer Merkmale nach Verwendungskategorie	233
A.2.2	Influenza 2014: Verteilung ausgewählter soziodemografischer Merkmale nach Verwendungskategorie	235
A.2.3	Influenza 2014: Kontaktversuche nach Sub-Domain	236
A.3	Schätzergebnisse CELLA 2	238
A.4	Schätzergebnisse Influenza 2014	241
A.5	Schätzergebnisse DSF für CELLA 2	244
A.6	Schätzergebnisse DSF für Influenza 2014	250

Abbildungsverzeichnis

1	Kombination der bundesdeutschen Bevölkerung nach Erreichbarkeit über Festnetz und Mobilfunk	24
2	Vereinfachte Darstellung der UMTS-Netzwerkarchitektur	46
3	HLR-Lookup: Abfrage zum ersten Zeitpunkt nach Resultat und Netz	48
4	HLR-Lookup: Abfrage zum dritten Zeitpunkt nach Resultat und Netz	49
5	HLR-Lookup: Übereinstimmung des Status im Zeitverlauf	50
6	Potenzielle Fehlerquellen einer Stichprobe	56
7	Influenza 2014: Bearbeitete Fälle pro Interviewer	83
8	Influenza 2014: Durchschnittliche Anzahl der Kontaktversuche nach Interviewer und Anteil der Interviews	84
9	Influenza 2014: Mittlere Dauer der Befragung nach Interviewer und Anteil der Interviews	87
10	Influenza 2014: Designeffekt durch Klumpung	91
11	Neuzuordnung der Dual-Frame-Erhebung	120
12	Festnetz: Anzahl der Kontakte nach Modell 1	158
13	Mobilfunk: Anzahl der Kontakte nach Modell 1	160
14	Festnetz: Anzahl der Kontakte nach Modell 5	236
15	Mobilfunk: Anzahl der Kontakte nach Modell 5	237

Tabellenverzeichnis

2	CELLA 2: Aufteilung nach Art der Bruttostichprobe	41
3	Influenza 2014: Aufteilung nach Art der Stichprobe	43
4	HLR-Lookup: Ungleiche Resultate nach Lookup (für Substichprobe)	51
5	HLR-Lookup: Vergleich der zweiten Abfrage und der finalen Kontaktergebnisse	52
6	Parameter zur Bestimmung der Auswahlwahrscheinlichkeit bei Dual-Frame-Telefonstichproben	62
7	Median der Kontaktversuche nach Interviewer	85
8	Influenza 2014: Tage einer Rufnummer im Feld nach Interviewer-ID (in %) .	86
9	Merkmale der synthetischen Grundgesamtheit	95
10	Simulationsergebnisse nach Schätzverfahren	96
11	CELLA 2: Höchster allgemeiner Schulabschluss nach Schätzverfahren (in %) .	98
12	Influenza 2014: Alter nach Art der Kalibrierung (in %)	99
13	Influenza 2014: Haushaltsgröße nach Art der Kalibrierung (in %)	100
14	CELLA 2: Familienstand vor und nach Designgewichtung und Kalibrierung (in %)	114
15	CELLA 2: Kontaktergebnisse der Erhebung (Festnetzstichprobe)	116
16	CELLA 2: Kontaktergebnisse der Erhebung (Mobilfunk)	117
17	Zusammensetzung der Stichprobe nach Nutzungskategorie und Medium . . .	128
18	Vergleich der unterschiedlichen Schätzverfahren I (Anteilswerte)	129
19	Variable 2 nach Nutzungskategorie (Anteilswerte)	130
20	Vergleich der unterschiedlichen Schätzverfahren II (Anteilswerte)	131
21	CELLA 2: Unabhängigkeit der Items zur Einordnung von Nutzungsgewohnheiten	134
22	CELLA 2: Dual-User nach Verwendungskategorie	137
23	CELLA 2: Zusammensetzung innerhalb der Teilpopulation nach Modell und Stichprobe	138
24	CELLA 2: Geschlecht nach Nutzungsgewohnheit und Modell (in %)	139
25	CELLA 2: Alter nach Nutzungsgewohnheit und Modell (in %)	141
26	CELLA 2: Bildungsabschluss nach Nutzungsgewohnheit und Modell (in %) . .	143
27	CELLA 2: Composite-Faktoren nach Modell und Berechnungsgrundlage . . .	147
28	CELLA 2: Höchster allgemeiner Schulabschluss nach Modell und Methode 1 (in %)	147

29	CELLA 2: Höchster allgemeiner Schulabschluss nach Modell und Methode 2 (in %)	148
30	Influenza 2014: Unabhängigkeit der Items zur Einordnung von Nutzungsgewohnheiten	152
31	Influenza 2014: Dual-User nach Verwendungskategorie	155
32	Grundsätzliche Erreichbarkeit vs. Erreichbarkeit der letzten 2 Stunden (Mobilfunk)	156
33	Grundsätzliche Erreichbarkeit vs. Erreichbarkeit der letzten 2 Stunden (Festnetz)	156
34	Influenza 2014: Zusammensetzung innerhalb der Teilpopulation nach Modell und Stichprobe	161
35	Influenza 2014: Geschlecht nach Nutzungsgewohnheit und Modell (in %) . .	163
36	Influenza 2014: Alter nach Nutzungsgewohnheit und Modell (in %)	164
37	Influenza 2014: Bildungsabschluss nach Nutzungsgewohnheit und Modell (in %)	166
38	Influenza 2014: Composite-Faktoren nach Modell und Berechnungsgrundlage	169
39	Influenza 2014: Höchster allgemeiner Schulabschluss nach Modell und Methode 1 (in %)	170
40	Influenza 2014: Höchster allgemeiner Schulabschluss nach Modell und Methode 2 (in %)	170
41	Vergleich soziodemografischer Merkmale der Erhebungen CELLA 2 und Influenza 2014 (in %)	172
42	Influenza 2014: Vergleich der Einteilung nach Modell 1 und der Empfangsbereitschaft des Mobiltelefons (absolut)	177
43	Influenza 2014: Empfangsbereitschaft nach Domain in Modell 1 (in %) . . .	178
44	Vergleich der unterschiedlichen Schätzverfahren inklusive DSF-Ansatz für Variable 1 (in %)	185
45	Vergleich der unterschiedlichen Schätzverfahren inklusive DSF-Ansatz Variable 2 (in %)	186
46	CELLA 2: Höchster allgemeiner Schulabschluss mit DSF-Schätzer nach Modell 1 (in %)	188
47	Influenza 2014: Bildungsabschluss mit DSF-Schätzer nach Modell 1 (in %) .	190
48	MSE der Variablen im synthetischen Datensatz (in 100 Mio.)	200
49	Influenza 2014: MSE für Haushaltsgröße und Alter (in 10 Mrd.)	202
50	CELLA 2: Haushaltsgröße nach Nutzungsgewohnheit und Modell (in %) . . .	233

51	CELLA 2: Familienstand nach Nutzungsgewohnheit und Modell (in %)	234
52	Influenza 2014: Haushaltgröße nach Nutzungsgewohnheit und Modell (in %)	235
53	CELLA 2: Geschlecht nach Modell und Methode 1 (in %)	238
54	CELLA 2: Geschlecht nach Modell und Methode 2 (in %)	238
55	CELLA 2: Migrationshintergrund nach Modell und Methode 1 (in %)	238
56	CELLA 2: Migrationshintergrund nach Modell und Methode 2 (in %)	238
57	CELLA 2: Haushaltgröße nach Modell und Methode 1 (in %)	239
58	CELLA 2: Haushaltgröße nach Modell und Methode 2 (in %)	239
59	CELLA 2: Familienstand nach Modell und Methode 1 (in %)	239
60	CELLA 2: Familienstand nach Modell und Methode 2 (in %)	240
61	CELLA 2: Privates Nettoeinkommen nach Modell und Methode 1 (in %)	240
62	CELLA 2: Privates Nettoeinkommen nach Modell und Methode 2 (in %)	240
63	Influenza 2014: Geschlecht nach Modell und Methode 1 (in %)	241
64	Influenza 2014: Geschlecht nach Modell und Methode 2 (in %)	241
65	Influenza 2014: Migrationshintergrund nach Modell und Methode 1 (in %)	241
66	Influenza 2014: Migrationshintergrund nach Modell und Methode 2 (in %)	241
67	Influenza 2014: Haushaltgröße nach Modell und Methode 1 (in %)	242
68	Influenza 2014: Haushaltgröße nach Modell und Methode 2 (in %)	242
69	Influenza 2014: Alter nach Modell und Methode 1 (in %)	242
70	Influenza 2014: Alter nach Modell und Methode 2 (in %)	243
71	CELLA 2: Höchster allgemeiner Schulabschluss mit DSF-Schätzer nach Modell 2 (in %)	244
72	CELLA 2: Höchster allgemeiner Schulabschluss mit DSF-Schätzer nach Modell 3 (in %)	244
73	CELLA 2: Geschlecht mit DSF-Schätzer nach Modell 1 (in %)	244
74	CELLA 2: Geschlecht mit DSF-Schätzer nach Modell 2 (in %)	245
75	CELLA 2: Geschlecht mit DSF-Schätzer nach Modell 3 (in %)	245
76	CELLA 2: Migrationshintergrund mit DSF-Schätzer nach Modell 1 (in %)	245
77	CELLA 2: Migrationshintergrund mit DSF-Schätzer nach Modell 2 (in %)	245
78	CELLA 2: Migrationshintergrund mit DSF-Schätzer nach Modell 3 (in %)	246
79	CELLA 2: Haushaltgröße mit DSF-Schätzer nach Modell 1 (in %)	246
80	CELLA 2: Haushaltgröße mit DSF-Schätzer nach Modell 2 (in %)	246
81	CELLA 2: Haushaltgröße mit DSF-Schätzer nach Modell 3 (in %)	247
82	CELLA 2: Familienstand mit DSF-Schätzer nach Modell 1 (in %)	247
83	CELLA 2: Familienstand mit DSF-Schätzer nach Modell 2 (in %)	247
84	CELLA 2: Familienstand mit DSF-Schätzer nach Modell 3 (in %)	248
85	CELLA 2: Privates Nettoeinkommen mit DSF-Schätzer nach Modell 1 (in %)	248

86	CELLA 2: Privates Nettoeinkommen mit DSF-Schätzer nach Modell 2 (in %) .	248
87	CELLA 2: Privates Nettoeinkommen mit DSF-Schätzer nach Modell 3 (in %) .	249
88	Influenza 2014: Bildungsabschluss mit DSF-Schätzer nach Modell 2 (in %) .	250
89	Influenza 2014: Geschlecht mit DSF-Schätzer nach Modell 1 (in %)	250
90	Influenza 2014: Geschlecht mit DSF-Schätzer nach Modell 2 (in %)	250
91	Influenza 2014: Migrationshintergrund mit DSF-Schätzer nach Modell 1 (in %)	251
92	Influenza 2014: Migrationshintergrund mit DSF-Schätzer nach Modell 2 (in %)	251
93	Influenza 2014: Haushaltsgröße mit DSF-Schätzer nach Modell 1 (in %) . . .	251
94	Influenza 2014: Haushaltsgröße mit DSF-Schätzer nach Modell 2 (in %) . . .	252
95	Influenza 2014: Alter mit DSF-Schätzer nach Modell 1 (in %)	252
96	Influenza 2014: Alter mit DSF-Schätzer nach Modell 2 (in %)	252

Symbolverzeichnis

U^F	Menge der Elemente der Zielpopulation, die über das Festnetz erreichbar sind; Festnetzpopulation
U^C	Menge der Elemente der Zielpopulation, die über Mobilfunk erreichbar sind; Mobilfunkpopulation
$U^{F \cap C} = U^F \cap U^C$	Menge der Elemente der Zielpopulation, die sowohl über Festnetz- als auch Mobilfunk erreichbar sind; <i>Dual-User</i>
$U = U^F \cup U^C$	Zielpopulation einer Erhebung
$N = \sum_{i \in U} 1$	Anzahl der Elemente in einer Zielpopulation; Mächtigkeit von U
$N^F = \sum_{i \in U^F} 1$	Anzahl der Elemente in einer Zielpopulation, die über das Festnetz erreichbar sind; Mächtigkeit U^F
$N^C = \sum_{i \in U^C} 1$	Anzahl der Elemente in einer Zielpopulation, die über Mobilfunk erreichbar sind; Mächtigkeit U^C
S	Menge der Elemente einer Stichprobe der Zielpopulation U
S^F	Menge der Elemente einer Festnetzstichprobe der Zielpopulation U^F
S^C	Menge der Elemente einer Mobilfunkstichprobe der Zielpopulation U^C
R	Menge der Respondenten
r_i	Responseindikator der Erhebungseinheit i
η_i	Teilnahmewahrscheinlichkeit der Erhebungseinheit i
$n = \sum_{i \in S} 1$	Anzahl der Elemente einer Stichprobe S ; Stichprobenumfang von S
$n^F = \sum_{i \in S^F} 1$	Anzahl der Elemente einer Stichprobe S^F ; Stichprobenumfang von S^F
$n^C = \sum_{i \in S^C} 1$	Anzahl der Elemente einer Stichprobe S^C ; Stichprobenumfang von S^C
$def f$	Designeffekt (der Erhebung)
$n_{eff} = \frac{n}{def f}$	Effektiver Stichprobenumfang
$l \in L = \{1, 2, 3, 4, 5\}$	Index zur Zugehörigkeit einer <i>Nutzungskategorie</i> ; <i>Landline-Only</i> , <i>Landline-Mainly</i> , <i>True Dual-User</i> , <i>Mobile-Mainly</i> , <i>Mobile-Only</i>
π_i	Inklusionswahrscheinlichkeit erster Ordnung der Erhebungseinheit i
π_{ij}	Inklusionswahrscheinlichkeit zweiter Ordnung der Erhebungseinheiten i und j

M^F	Anzahl der Rufnummern innerhalb des Auswahlrahmens für Festnetzstichproben
M^C	Anzahl der Rufnummern innerhalb des Auswahlrahmens für Mobilfunkstichproben
m^F	Anzahl der Rufnummern innerhalb der Festnetzstichprobe
m^C	Anzahl der Rufnummern innerhalb der Mobilfunkstichprobe
k_i^F	Anzahl der Festnetzzufnummern, über die Erhebungseinheit i erreicht werden kann
k_i^C	Anzahl der Mobilfunkrufnummern, über die Erhebungseinheit i erreicht werden kann
z_i	Anzahl der erhebungsrelevanten Haushaltsmitglieder im Haushalt von Erhebungseinheit i
$d_i = \frac{1}{\pi_i}$	Designgewicht (des Horvitz-Thompson-Schätzers) der Erhebungseinheit i
$d_i^F = \frac{1}{\pi_i}$	Designgewicht (des Horvitz-Thompson-Schätzers) der Erhebungseinheit i aus der Festnetzstichprobe
$d_i^C = \frac{1}{\pi_i}$	Designgewicht (des Horvitz-Thompson-Schätzers) der Erhebungseinheit i aus der Mobilfunkstichprobe
g_i	Anpassungsgewicht der Erhebungseinheit i
$w_i = d_i * g_i$	Gesamtgewicht der Erhebungseinheit i
$\lambda, \phi, \mu \in [0,1]$	Compositenfaktoren (einzelner Subdomains)
ξ_i^F	Die von der Präferenz für die Verwendung von Festnetzanschlüssen ausgehende Teilnahmewahrscheinlichkeit der Erhebungseinheit i
c_i^F	Compositenfaktoren der Erhebungseinheit i aus der Festnetzstichprobe (über Domains)
c_i^C	Compositenfaktoren der Erhebungseinheit i aus der Mobilfunkstichprobe (über Domains)
$E(\bullet)$	Erwartungswert eines Schätzers eines Parameters/einer Variablen der Population
$Var(\bullet)$	Varianz eines Schätzers eines Parameters/einer Variablen der Population
$\hat{\bullet}$	Indikator eines Schätzwerters; i. d. R. basierend auf Erhebungsdaten für eine Zielpopulation
$Bias(\hat{Y}) = E(\hat{Y}) - Y$	Verzerrung (des geschätzten Totalwertes der Variablen y)
$MSE(\hat{Y}) = E((\hat{Y} - Y)^2)$	Mean Squared Error (des geschätzten Totalwertes der Variablen y)
SF	Gibt (als Index) an, dass der entsprechende Parameter mittels Single-Frame-Ansatz geschätzt wurde

<i>CW</i>	Gibt (als Index) an, dass der entsprechende Parameter mittels Composite-Ansatz geschätzt wurde
<i>DSC</i>	Gibt (als Index) an, dass der entsprechende Parameter mittels Domain-Specific-Composite-Design-Weighting-Ansatz geschätzt wurde
<i>DSF</i>	Gibt (als Index) an, dass der entsprechende Parameter mittels Domain-Specific-Single-Frame-Ansatz geschätzt wurde
$\forall$	Allquantor; "für alle bzw. jedes Element gilt..."

Abkürzungsverzeichnis

ADM	<i>Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V.</i>	1
AAPOR	<i>American Association of Public Opinion Research</i>	33
BNetzA	<i>Bundesnetzagentur</i>	21
CATI	<i>Computer Assisted Telephone Interview</i>	9
CELLA	<i>Cellphone and Landline</i>	III
DB Telekom	<i>Deutsche Bundespost Telekom</i>	17
DFG	<i>Deutsche Forschungsgemeinschaft</i>	2
DTAG	<i>Deutsche Telekom AG</i>	18
HLR	<i>Home Location Register</i>	45
IMSI	<i>International Mobile Subscriber Identity</i>	45
LTE	<i>Long Term Evolution</i>	46
MAR	<i>Missing at Random</i>	73
MCAR	<i>Missing Completely at Random</i>	73
MSC	<i>Mobile Switching Center</i>	45
MSE	<i>Mean Squared Error</i>	72
MSISDN	<i>Mobile Station ISDN Number</i>	45
NMAR	<i>Not Missing at Random</i>	73
ONK	<i>Ortsnetzkennzahlen</i>	19
RKI	<i>Robert Koch-Institut</i>	8
RLD	<i>Randomized Last Digit</i>	16
RNS	<i>Radio Network Subsystem</i>	46
RDD	<i>Random Digit Dialing</i>	15
RR	<i>Responserate</i>	74
SIM	<i>Subscriber Identification Module</i>	22
SS7	<i>Signaling System 7</i>	46
UMTS	<i>Universal Mobile Telecommunication System</i>	46
VLR	<i>Visitor Location Register</i>	45

Danksagung

*The presence of those seeking the truth is infinitely to be preferred to
the presence of those who think they've found it.*

- Terry Pratchett, Monstrous Regiment

Obwohl ich diese Arbeit eigenständig geschrieben habe, waren an ihrer Entstehung eine Vielzahl von Personen direkt oder indirekt, wissentlich oder unwissentlich beteiligt. Diesen möchte ich an dieser Stelle meinen tiefsten Dank aussprechen.

Beginnend gilt dieser meinem Doktorvater Prof. Dr. Michael Häder, der mich sowohl inhaltlich als auch strukturell stets bei dieser Arbeit unterstützt hat und durch dessen Anregungen die vorliegende Arbeit die bestehende Ausrichtung genommen hat. Einen großen Einfluss sowohl in dem erfolgreichen Gelingen dieser Arbeit als auch in meiner wissenschaftlichen Entwicklung hatte weiterhin mein Betreuer PD Dr. Siegfried Gabler. Für seinen Einsatz und seine Geduld möchte ich mich daher gerne bedanken.

Für die Bereitstellung der Daten der Erhebung Influenza 2014 sowie dem Einräumen der Möglichkeit, im Rahmen dieser Erhebung eigene Fragen einzubringen, gilt mein Dank weiterhin Patrick Schmich. Ferner richtet sich mein Dank an dieser Stelle an Dr. Robert Neumann für die Unterstützung bei der Entwicklung dieser Fragen. Meinen Kollegen bei GESIS möchte ich ebenfalls für den bilateralen Austausch sowie für die Weitung meines Forschungshorizonts danken. Insbesondere sind an dieser Stelle Dr. Sabine Häder, Dr. Stefan Zins, Dr. Jan-Philipp Kolb, Alexander Murray-Waters, Verena Ortmanns und Britta Gauly zu nennen.

Caterina Mathiesen und Felix Hofmann dürfen Sie als Leser genauso danken wie ich als Verfasser dieser Arbeit. Beiden gilt mein Dank für ihr umfangreiches Proofreading sowie dem Verkürzen meiner deutlich zu langen Ausführungen.

Für ihre andauernde Unterstützung auch während meiner weniger erträglichen Zeiten, dem Ermöglichen meiner wissenschaftlichen Ausbildung sowie dem Vermitteln grundsätzlicher Einstellungen und Werte, gilt meinen Eltern Martina Sand, Uwe Gierstorfer und Georg Sand, meinem Bruder Martin Sand und meiner restlichen Familie, meine größte Anerkennung. Großer Dank gilt darüber hinaus Stefan Weithoffer und Julia Kißling.

Auch wenn in meiner Wertschätzung an erster Stelle, möchte ich mich zuletzt bei meiner (baldigen) Frau Jana (bald) Sand für ihre tägliche Unterstützung, Motivation und notwendige Zerstreuung in dieser Phase bedanken.

1 Einleitung und Motivation

Der Jahresberichtes 2016 des *Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V.* (ADM) berichtet, dass die (sozialwissenschaftliche) Forschung häufig telefonische Befragungen zur Gewinnung von Erhebungsdaten anwendet. So ist diesem zu entnehmen, dass im Berichtsjahr ca. 36 % aller vom ADM durchgeführten Befragungen über diesen Modus erfolgten (vgl. Arbeitskreis Deutscher Markt und Sozialforschungsinstitute ADM (2017), S. 15). Es zeigt sich, dass trotz der verbreiteten Anwendung von Online-Befragungen und der stets hohen Relevanz von Face-to-Face-Befragungen, Erhebungen anhand von telefonischen Befragungen noch immer einen deutlichen Anteil an der gesamten Menge an (sozialwissenschaftlichen) Umfragen ausmacht. Begründet werden kann dies unter anderem damit, dass mit einer telefonischen Befragung verschiedene Vorzüge in Verbindung gebracht werden können. Die Verwendung dieses Modus erlaubt eine ungeklumpte Befragung, die, bspw. mit Face-to-Face-Befragungen verglichen, zu relativ geringen Kosten durchgeführt werden kann. Weiterhin bietet die Anwendung dieses Modus, wenn die Durchführung innerhalb eines Telefonlabors erfolgt, i. d. R. weitere Vorzüge der Interviewerüberwachung.

Andererseits ist auch eine solche Art der Befragung wie die meisten Erhebungen, die auf einer Zufallsauswahl basieren, anfällig für eine Vielzahl von Fehlerquellen, die das Potenzial haben, Schätzergebnisse zu verzerren. Neben der Antwortverweigerung stehen diese oftmals in Verbindung mit der Auswahlgrundlage, die zur Stichprobenziehung verwendet wird. Limitierend für die Durchführung einer solchen Erhebung kann dabei die Beschränkung auf Festnetzstichproben sein. So zeigt sich gerade im deutschen Kontext, dass der Anteil der Haushalte, die noch immer über einen Festnetzanschluss erreicht werden können, im europäischen Vergleich mit ca. 86 % relativ hoch ist (vgl. Bundesnetzagentur (2013), S. 56). Jedoch ist auch bei diesem Anteil zu befürchten, dass reine Festnetzstichproben zu fehlerhaften Aussagen führen, wenn die Zielpopulation die gesamte deutsche Bevölkerung ist. Grundsätzlich liegt dies an demjenigen Teil der Bevölkerung, der mittlerweile ausschließlich über das Mobilfunknetz erreichbar ist. Unterscheidet sich diese Population systematisch von derjenigen, die über das Festnetz erreicht werden kann, so führt deren Ausschluss unter Umständen zu fehlerhaften Schätzungen. Sowohl Untersuchungen der Bundesnetzagentur als auch die tief greifende sozialwissenschaftliche Forschung, die seitens des ADMs, GESIS sowie der Technischen Universität Dresden bspw. durch Gabler und S. Häder (2009a), M. Häder (2000) oder Heckel (2007) durchgeführt wurde, zeigt indes, dass sich diese Teilpopulation durchaus in ihren soziodemografischen Charakteristika von Festnetzhaushalten unterscheidet. Auch in der englischsprachigen Literatur lässt sich hierfür eine Vielzahl von Belegen finden, wie bspw. die Arbeiten von Blumberg et al. (2008), Brick et al. (2005) oder Kennedy (2007). Um diese Population nicht auszuschließen, wird daher vielfach

die Anwendung sog. Dual-Frame-Erhebungen angeraten, die sich aus einer oder mehreren unabhängigen Festnetz- und Mobilfunkstichproben zusammensetzen.

Im Zuge der Anwendung eines solchen Vorgehens ergeben sich drei disjunkte Teilpopulationen, die sowohl in der Durchführung der Erhebung als auch in der späteren Schätzung von Ergebnissen, im Rahmen einer adäquaten Gewichtungsstrategie, bedacht werden müssen. Die Population lässt sich nun, aufgrund der Verwendung zweier Auswahlgrundlagen, in diejenigen untergliedern, die ausschließlich über das Festnetz (sog. Landline-Onlys), ausschließlich über Mobilfunk (sog. Mobile-Onlys) sowie über beide Auswahlgrundlagen (sog. Dual-User) in die Erhebung gelangen können. Ausführliche Auseinandersetzungen mit Problemstellungen dieser einzelnen Teilpopulationen, deren Beachtung in der Stichprobenziehung sowie in der Gewichtung zum Erlangen geeigneter Schätzwerte, finden sich bspw. in den Untersuchungen von S. L. Lohr (2011), Wolter et al. (2015) oder Gabler und Ayhan (2007). Gerade im Rahmen der Designgewichtung solcher Dual-Frame-Erhebungen gibt es eine Vielzahl von Untersuchungen, wie erwartungstreue Schätzer für diese entwickelt werden können. Problematisch in diesem Kontext ist häufig die Frage, wie diejenigen der Erhebung und der Population zu behandeln sind, die aus dem Überlappungsbereich der beiden Auswahlrahmen stammen und somit über beide Telekommunikationsmedien erreichbar sind.

Gerade die Erkenntnisse, die im Zuge der beiden CELLA-Erhebungen hervorgebracht wurden, deren Untersuchung als gemeinsames Projekt von GESIS sowie der Technischen Universität Dresden durch die *Deutsche Forschungsgemeinschaft* (DFG) finanziert wurde, liefern einen tieferen Einblick in die Stichprobenziehung und Schätzung für Dual-Frame-Erhebungen in Deutschland. Weiterhin wurde anhand der Ergebnisse der beiden Erhebungen ein besseres Verständnis der soziodemografischen Unterschiede gerade unter denjenigen, die ausschließlich über ein bestimmtes Medium kontaktiert werden können, vermittelt.

Bis zum heutigen Zeitpunkt wenig untersucht ist jedoch, ob und wie sich diejenigen Zielpersonen einer Erhebung, die sich im Überlappungsbereich beider Auswahlgrundlagen befinden, voneinander unterscheiden. So werden diese bspw. in den meisten Gewichtungsansätzen so behandelt, als ob, in Abhängigkeit der Auswahlwahrscheinlichkeit einer jeweiligen Zielperson, das Medium, über das ein Dual-User kontaktiert wurde, keinen oder nur einen sehr geringen Einfluss auf dessen Teilnahmeverhalten hat. Gerade diese **Device-Specific Nonresponse** sowie der Umgang mit dieser, sind zentraler Untersuchungsgegenstand der vorliegenden Arbeit. So zeigen erste Untersuchungen von bspw. Sand (2014), Hoops und Michael (2015) oder Brick (2013), dass das Antwort- und Teilnahmeverhalten von Dual-Usern durchaus mit dem gewählten Medium, über das diese kontaktiert wurden, beeinflusst wird. In einem solchen Kontext beschreibt bspw. Sand (2015a), dass das Teilnahmeverhalten von Dual-Usern auch durch deren Präferenzen und Nutzungsgewohnheiten beeinflusst werden kann. Diese Präferenzen lassen sich in der Teilpopulation der Dual-User durchaus als Kontinuum verstehen, das wiedergibt, ob eine

Zielperson dieser Population eher zu der Verwendung ihres Festnetz- oder Mobilfunkanschlusses geneigt ist. Demnach kann eine solche Präferenz beeinflussen, ob eine Zielperson entweder eher zu einer Teilnahme an einer Erhebung bereit ist und/oder mit geringerem Aufwand in die Erhebung gelangt, wenn diese über das präferierte Medium kontaktiert wird.

Da sowohl die Messung eines solchen Kontinuums der Präferenz für ein bestimmtes Telekommunikationsmedium als auch die spätere Verwendung für weitere Untersuchungen durchaus fehleranfällig ist, wird in der vorliegenden Arbeit ein Modell erarbeitet, dass dieses Spektrum in lediglich drei Teilpopulationen untergliedert. So unterscheidet dieses Modell unter den Dual-Usern lediglich diejenigen, die die Verwendung ihres Festnetzanschlusses präferieren (*Landline-Mainly*), diejenigen, die die Verwendung ihres Mobilfunkanschlusses präferieren (*Mobile-Mainly*) sowie diejenigen, bei denen keine eindeutige Präferenz festzustellen ist (*wahre Dual-User*).

Ein besonderes Analysepotenzial ergibt sich dann, wenn sich diese drei neu integrierten Teilpopulationen in der Zusammensetzung ihrer soziodemografischen Merkmale voneinander unterscheiden. Dies würde dann einmal eine weitere Möglichkeit zur (nachträglichen) Schichtung sowie zur Analyse gerade dieser Teilpopulationen hervorbringen.

Von gesteigertem Interesse ist jedoch vor allem die Frage, ob und wie die Kenntnis der Zugehörigkeit zu einer dieser Teilpopulationen und der damit verbundenen Kenntnis über die Präferenz für ein bestimmtes Telekommunikationsmedium dazu beitragen kann, die Schätzung auf der Basis von Dual-Frame-Erhebungen zu verbessern. Gerade dies ist der zentrale Untersuchungsgegenstand dieser Arbeit. Daher werden im Zuge der Auseinandersetzung mit der dargelegten Problemstellung einmal Modelle zur Einteilung nach Nutzungsgewohnheiten erarbeitet und vorgestellt. Im nächsten Schritt erfolgt dann die Auseinandersetzung mit den einzelnen Teilpopulationen sowie deren Zusammensetzung soziodemografischer Merkmale, um zuletzt unterschiedliche Gewichtungsansätze vorzustellen und zu untersuchen, mit der Absicht, gerade diese Präferenzen in der Schätzung besser zu berücksichtigen.

Grundlage dieser Auseinandersetzung bildet ein umfassendes Verständnis sowohl der persistenten Problematik der Only-Haushalte, die die Notwendigkeit der Verwendung von Dual-Frame-Erhebungen bei telefonischen Befragungen begründet, sowie der Stichprobenziehung innerhalb solcher Erhebungen. Letzteres ist deshalb von Interesse, da die Anwendung einer unterschiedlichen Stichprobenstrategie unter Umständen direkte Auswirkungen auf die Auswahlwahrscheinlichkeit einer Zielperson hat. Die korrekte Bestimmung der Auswahlwahrscheinlichkeit bildet indes die Basis einer genauen Schätzung interessierender Parameter der Zielpopulation auf der Grundlage von Erhebungsdaten. Daher setzt sich das folgende Kapitel mit der Entwicklung der Erhebung über den Modus telefonische Befragung im Rahmen der Umfrageforschung anhand einschlägiger Literatur auseinander. Ziel des Kapitels ist es, die Anforderungen insbesondere von Dual-Frame-Erhebungen gerade im deutschen Kontext genauer zu vermitteln. Diese werden am Beispiel von zwei ausgewählten Dual-Frame-Erhebungen im Detail beschrieben. Darüber hinaus

bilden diese Erhebungen die Grundlage zur Begutachtung des in dieser Arbeit vorgestellten Modells zur Einteilung nach Präferenzen sowie deren Anwendung in der Schätzung.

Gerade um diese Präferenzen in der Schätzung sinnvoll berücksichtigen zu können, bedarf es eines Verständnisses darüber, welche Gewichtungsverfahren (gerade bei Dual-Frame-Erhebungen) bereits existieren und welche Eigenschaften diese besitzen. Daher wird in Kapitel 3 erstmals eine grundlegende Unterscheidung zwischen Designgewichtung und Kalibrierung vorgenommen. Davon ausgehend werden unterschiedliche Ansätze der Designgewichtung bei Dual-Frame-Erhebungen sowie die Möglichkeiten der Kalibrierung erläutert. Diese werden anschließend auf geeignete Daten angewendet und im Abschluss dieses Kapitels bezüglich ihrer Eigenschaften und Genauigkeit miteinander verglichen.

Die zentralen Erkenntnisse der beiden Kapitel, die sich hauptsächlich mit der relevanten Literatur auseinandersetzen, werden im Rahmen von Kapitel 4 nochmals hervorgehoben.

Ausgehend von den bereits bestehenden Erkenntnissen stellt Kapitel 5 das Modell zur Unterscheidung von Dual-User anhand ihrer Präferenzen und Nutzungsgewohnheiten vor, das im Rahmen dieser Untersuchung erarbeitet wurde und zentraler Gegenstand dieser Arbeit ist. Dazu wird vorab eine genauere Untersuchung von Erhebungsdaten dargestellt, die die Anwendung einer solchen Einteilung begründet. Daraufhin wird die Einteilung nach Nutzungsgewohnheiten der Dual-User im Zuge dieses Modells genauer erläutert. Zuletzt stellt dieses Kapitel eine Möglichkeit dar, gerade diese Erkenntnisse über die Präferenz und Nutzungsgewohnheiten von Dual-Usern im Zuge des Gewichtungsprozesses zugänglich zu machen, um sie innerhalb der Schätzung zu berücksichtigen.

Das Modell wird dann innerhalb von Kapitel 6 auf die beiden in Kapitel 2 vorgestellten Erhebungen angewendet. Im Zuge dessen wird sowohl eine Strategie zur Anwendung des Modells genauer erläutert als auch eine Auseinandersetzung mit der Verteilung der soziodemografischen Merkmale der jeweiligen Teilpopulationen angeführt. Weiterhin wird die Genauigkeit der Schätzung, die sich durch den zuvor beschriebenen Gewichtungsansatz ergibt, mit derjenigen der gängigen Schätzverfahren in Dual-Frame-Erhebungen verglichen und analysiert.

Kapitel 7 bietet einen alternativen Ansatz zur Berücksichtigung der Präferenzen und Nutzungsgewohnheiten von Dual-Usern im Rahmen der Gewichtung. Dieser wird zu Beginn des Kapitels vorgestellt und erläutert. Der weitere Teil des Kapitels beschäftigt sich dann mit der Anwendung dieses Gewichtungsansatzes sowie den daraus resultierenden Schätzergebnissen im Vergleich zu der vorherigen Vorgehensweise.

Zuletzt werden die hier erarbeiteten Gewichtungsverfahren, die die Präferenzen der Dual-User berücksichtigen, ebenso wie die gängigen Ansätze, die in Kapitel 3 dargestellt wurden, bezüglich ihrer Genauigkeit und ihrer Präzision miteinander verglichen. Da gerade die Bestimmung der Varianz innerhalb von komplexen Stichprobenerhebungen problematisch ist, erfolgt in diesem Kapitel die Erläuterung einer möglichen Varianzapproximation, die hierfür angewendet werden

kann. Im Anschluss erfolgt sowohl die Schätzung dieser Varianz anhand der vorgestellten Approximation sowie eine erneute Beurteilung der unterschiedlichen Verfahren anhand der *Mittleren Quadratischen Abweichung* (MSE), die auf der Basis der vorliegenden Daten geschätzt wurde.

Den Abschluss dieser Arbeit bildet eine Zusammenfassung und Deutung der beobachteten Ergebnisse sowie ein Ausblick auf weitere Forschungsfragen, die sich auf der Grundlage dieser Arbeit ergeben.

2 Telefonstichproben in Deutschland

Das folgende Kapitel gibt einen Überblick zu der Entwicklung der Befragung über das Telefon in der Umfrageforschung mit dem Ziel, die derzeitigen Anforderungen an eine solche Erhebung im deutschen Kontext zu beschreiben. Hierzu werden in einem ersten Schritt die Möglichkeiten und Grenzen der telefonischen Befragung genauer erläutert. Da diese oftmals durch technische Entwicklungen sowie durch den Umgang mit diesem Medium begründet sind, folgt anschließend ein Überblick über die historische Entwicklung der telefonischen Erhebung. Des Weiteren ist die Auswahlgrundlage, die für die Stichprobenziehung bei einer solchen Erhebung erforderlich ist, von entscheidender Bedeutung. Sie ist eng mit der technischen Entwicklung des Mediums und dem Umgang mit diesem verwoben. Daher werden bei der historischen Entwicklung ebenfalls die Konstruktion unterschiedlicher Auswahlrahmen sowie Strategien zur Stichprobenziehung im historischen Kontext betrachtet.

Weiterhin zeigen Untersuchungen der amerikanischen und deutschen Umfrageforschung, dass die reine Beschränkung auf Festnetzanschlüsse bei der Stichprobenauswahl für telefonische Erhebungen gegenwärtig nicht mehr ausreichend ist, um ein unverzerrtes Abbild der Allgemeinbevölkerung zu erhalten. Die verbreitete Durchsetzung von Mobilfunkanschlüssen sowie die Tatsache, dass ein Teil der Bevölkerung mittlerweile ausschließlich über Mobilfunkanschlüsse erreichbar ist (*Mobile-Onlys*), bedingt die Notwendigkeit, telefonische Befragungen auch über das Mobiltelefon durchzuführen. Sogenannte *Dual-Frame-Ansätze* verwenden daher eine (oder mehrere) unabhängige Festnetz- und Mobilfunkstichprobe für telefonische Befragungen. Kern dieser Arbeit sind Gewichtungsverfahren in Dual-Frame-Erhebungen. Ein gesteigertes Interesse liegt diesbezüglich in der Berücksichtigung von Nutzungsgewohnheiten bestimmter Telekommunikationsmedien bei der Gewichtung dieser Erhebungen. Um ein tieferes Verständnis bezüglich der Entwicklung von Dual-Frame-Erhebungen zu schaffen sowie die derzeitige Problematik der *Mobile-Onlys* genauer darzustellen, ist die Ziehung von Dual-Frame-Stichproben sowie die Anwendung im deutschen Kontext zentraler Bestandteil dieses einführenden Kapitels.

Dazu wird zuerst die Telefonerhebung basierend auf Stichproben aus mehr als einer Auswahlgrundlage allgemein diskutiert sowie der bestehende Forschungsbedarf aufgezeigt. Anschließend erfolgt die Darstellung im deutschen Zusammenhang. Hierzu wird die Bildung einer Auswahlgrundlage für Mobilfunkstichproben eingehend erläutert, um dann die Umsetzung von Dual-Frame-Erhebungen in Deutschland zu schildern. Durch die Verwendung eines weiteren Telekommunikationsmediums können zusätzliche Herausforderungen für die Durchführung von telefonischen Befragungen entstehen. Diese werden daher ebenso behandelt. Weiterhin wird die Erhebung CELLA 2, die von Forsa im Auftrag von GESIS, dem Leibniz-Institut für Sozialwissenschaften, zusammen mit der Technischen Universität Dresden durchgeführt

wurde, vorgestellt. Als weiteres Beispiel für Dual-Frame Erhebungen in Deutschland folgt die Darstellung der Erhebung Influenza 2014, die durch das *Robert Koch-Institut* (RKI) realisiert wurde. Eng mit den Herausforderungen der Stichprobenziehung, die durch die Verwendung mehrerer Telekommunikationsmedien entstehen, ist die Qualität der Auswahlgrundlage. Diese kann dabei direkten Einfluss auf die Präzision der Schätzung der Parameter der Zielpopulation anhand von Erhebungsdaten haben. Sowohl bei einer verwendeten Stichprobe als auch bei der Auswahlgrundlage im Allgemeinen ist es wünschenswert, eine möglichst geringe Anzahl von Rufnummern vorzufinden, die zu nicht-geschalteten Anschlüssen führen. Eine technische Lösung für die Auswahlgrundlage von Mobilfunkstichproben stellen in diesem Kontext sog. *Home Location Register-Lookup-Verfahren* (HLR) dar. Solche Verfahren ermöglichen die Verifizierung von Mobilfunknummern und können somit die spätere Schätzung anhand einer vorliegenden Erhebung bezüglich ihrer Genauigkeit und Präzision positiv beeinflussen. Die Reliabilität und Validität dieser Ergebnisse ist jedoch nur wenig untersucht. Infolgedessen wurde diesen Eigenschaften durch GESIS im Rahmen des Projektes *Verifikation mobiler Informationen (VermIn)* nachgegangen. Eine Darstellung der Verifizierung anhand dieser Verfahren sowie der Ergebnisse des Projektes bildet daher den Abschluss dieses Kapitels. Dies dient dem Zweck, einen besseren Einblick in die bestehenden Möglichkeiten zur Verbesserung der Qualität der Stichprobenziehung zu vermitteln, die wiederum Einfluss auf die generelle Qualität der Schätzung anhand von Erhebungsdaten hat.

2.1 Möglichkeiten und Grenzen telefonischer Befragungen

Nicht zuletzt wegen der vergleichsweise einfachen und kostengünstigen Handhabung, der Möglichkeiten zur Kontrolle des Interviewgeschehens und dem Potenzial, regional ungeklumpte Stichproben¹ zu ziehen, erfreuen sich Telefonstichproben auch in der heutigen Zeit großer Beliebtheit. Obgleich Online-Befragungen in den vergangenen Jahren zunehmend an Bedeutung gewonnen haben, berichtet bspw. der ADM in seinem Jahresbericht für 2016, dass der Anteil der Telefonbefragungen mit ca. 36% an allen ADM-Erhebungen über die vergangenen Jahre relativ konstant und hoch bleibt. Befragungen über dieses Medium sind somit noch immer eine der am häufigsten gewählten Methoden der Umfrageforschung (vgl. Arbeitskreis Deutscher Markt und Sozialforschungsinstitute ADM (2017), S. 15.).

Vorzüge telefonischer Befragungen liegen, so hält Waksberg bereits 1978 fest, besonders in der wesentlich schnelleren Durchführbarkeit (vgl. Waksberg (1978), S. 40). Dies ermöglichte S. Häder (1994) zufolge, dass Stichproben mit einer hohen Anzahl von Erhebungseinheiten in einer kurzen Zeit bearbeitet werden können. Als Referenzmodus gilt hier die Face-to-Face-Befragung. Die Autorin erörtert des Weiteren, dass ein gewichtiger Vorzug von Telefonbefragungen die

¹ Eine Telefonstichprobe kann jedoch über einzelne Interviewer klumpen.

Kosten einer solchen Erhebung sind. Diese fallen in den meisten Fällen geringer aus als bei persönlich-mündlichen Befragungen (vgl. S. Häder (1994), S. 2). Ein weiterer Vorteil bei der Verwendung von Telefonerhebungen ist die Möglichkeit, territorial ungeklumpte Stichproben zu realisieren. Aufgrund der Kosten und der begrenzten Möglichkeit, Interviewer deutschlandweit zu unterschiedlichen Erhebungseinheiten auszusenden, ist bei Face-to-Face-Erhebungen oftmals nur eine begrenzte Auswahl von Erhebungsgebieten möglich. Andernfalls erfolgt die Auswahl der Erhebungseinheiten in einem zweistufigen Verfahren. Im Idealfall werden dabei auf der ersten Auswahlstufe Klumpen (sog. *Primary Sampling Units (PSU)*) ausgewählt, die bspw. Gemeinden entsprechen können. Innerhalb dieser Klumpen werden dann erst in einem zweiten Schritt die Erhebungseinheiten (*Secondary Sampling Units (SSU)*) gezogen, um einen möglichst effizienten Einsatz der Interviewer für ein bestimmtes Gebiet zu erreichen.

Da Telefonstichproben in einem zentralen Telefonlabor durchgeführt werden können und Rufnummern regional streuen, kann bei Telefonerhebungen auf eine solche Klumpung verzichtet und als Erhebungsgebiet die gesamte Region der Zielpopulation abgedeckt werden. Zusätzlich besteht die Möglichkeit, Vorwahlen einer Festnetzstichprobe als Schichtungsmerkmal heranzuziehen (vgl. Meier et al. (2013), S. 118f.).

Durch die Durchführung telefonischer Befragungen in einem zentralen Labor ergeben sich weiterhin Vorzüge bezüglich der Supervision der Interviewer sowie der Überwachung des gesamten Feldgeschehens. So können die Art und Weise, wie eine Frage vom Interviewer gestellt wird, sowie die Arbeitsweise des Interviewers im Allgemeinen besser beobachtet und kontrolliert werden. Zusätzlich bestehen bei einer solchen Art der Erhebung weniger Schwierigkeiten aufgrund des Standortes des Interviewers (vgl. Waksberg (1978), S. 40).

Die direkte Supervision der Interviewer birgt nach Schnell et al. (1999) noch weitere Vorteile. So erlaubt diese bspw. die direkte Klärung möglicher Rückfragen des Interviewers oder des Befragten, das Einleiten von Korrekturmaßnahmen, bspw. bei Problemen mit dem Fragebogen, die Ermittlung des (Nach-) Schulungsbedarfes bestimmter Interviewer, Personaleinsparungen bei der Supervision und eine Überprüfung von Interviews direkt nach deren Fertigstellung (vgl. Schnell et al. (1999), S. 367f.).

Der systematische Einsatz computergestützter Verfahren (sog. *Computer Assisted Telephone Interview (CATI)*-Erhebungen) kann eine größere Genauigkeit und eine höhere Qualität der Daten erreichen während der Interviewereinfluss kontrolliert bzw. reduziert wird (vgl. S. Häder (1994), S. 2). Ferner wird es dadurch ermöglicht, Daten direkt zu vercoden, Zwischenergebnisse abzurufen sowie das Feldgeschehen bezüglich Kontaktversuche, bei Nicht-Erreichen eines Anschlusses, und die Verwaltung der Zuordnung von Interviewern zu Befragten unmittelbar zu steuern (vgl. Schnell et al. (1999), S. 367f.).

Die Verwendung entsprechender Soft- und Hardware (bspw. *Predictive Dialer*) ermöglicht, eingegebene Daten von einer Rufnummer, die zum Kontakt einer bestimmten Einheit geführt hat,

zu trennen. Hierdurch wird die Reidentifikation bestimmter Einheiten erschwert (vgl. S. Häder (1994), S. 2). Da die Interaktion zwischen Interviewer und Erhebungseinheit lediglich aural geschieht und keine physische Präsenz des Interviewers im Rahmen der Befragung besteht, wird oftmals betont, dass die Befragungssituation als anonymer wahrgenommen wird als bei einer persönlich-mündlichen Befragung. Wie sich diese scheinbar anonyme Befragungssituation jedoch auf die Antwortqualität auswirkt, ist fraglich. Hier wird oftmals argumentiert, dass aufgrund des fehlenden visuellen Kontaktes, Befragte bei sensiblen Fragen ehrlicher und weniger nach sozialer Erwünschtheit antworten. Entsprechende Ergebnisse berichtet bspw. Groves (1979). In seiner Untersuchung stellt der Autor fest, dass Befragte durch die physische Präsenz eines Interviewers eher dazu geneigt sind, eine gefälligere Antwort auf bestimmte Fragen mitzuteilen (vgl. Groves (1979), S. 192f). Jedoch gibt es ebenso Untersuchungen, die genau das Gegenteil berichten (vgl. Meier (1999), S. 119.).

Bei der Befragung über das Medium Telefon sind jedoch auch Grenzen und Probleme einzuräumen. Diese lassen sich grundsätzlich entweder auf den Prozess der Auswahl von Erhebungseinheiten über dieses Telekommunikationsmedium oder auf Messprobleme innerhalb der Befragung zurückführen. Bezüglich der Auswahl der Erhebungseinheiten in einer Befragung ist die Abdeckung der Zielpopulation durch die Auswahlgrundlage von entscheidender Bedeutung. Bei Face-to-Face-Befragungen wird im Idealfall (bei einer bundesweiten Befragung) auf die Einwohnermelderegister zurückgegriffen um eine Zielpersonenauswahl zu treffen. Bei Telefonerhebungen steht ein solch detailliertes Register i. d. R. nicht (mehr) zur Verfügung.

Im Rahmen des Auswahlprozesses bei telefonischen Befragungen ergeben sich wiederum zwei potenzielle Probleme. Einmal ist eine solche Erhebung lediglich dann sinnvoll durchführbar, wenn die Nutzungsdichte entsprechend hoch ist und die Einheiten der Zielpopulation auch über dieses Medium kontaktiert werden können. Unter der Nutzungsdichte ist in diesem Kontext der Anteil derjenigen Haushalte (oder Personen) an der Menge der Gesamthaushalte der Zielpopulation gemeint, die ein bestimmtes Medium verwendet. Zusätzlich muss bei einer Auswahlgrundlage für eine solche Erhebung für alle Einheiten der Zielpopulation eine Ziehungswahrscheinlichkeit für eine Stichprobe gelten, die größer als Null ist. Daher ist es wichtig, dass die Auswahlgrundlage bei einer hinreichenden Verbreitung von Telefonanschlüssen bekannt ist und auch alle Einheiten der Zielpopulation beinhaltet.

Das erste Problem lässt sich bspw. nach der Wiedervereinigung Deutschlands ebenso wie mit der kommerziellen Einführung der Mobiltelefonie (kurzfristig) beobachten. Der Umgang mit diesen beiden Ereignissen und deren Einfluss auf die Erreichbarkeit der bundesdeutschen Bevölkerung wird in Kapitel 2.2 und 2.3 näher erläutert. Festzuhalten ist jedoch, dass beide Ereignisse derzeit die Verwendung von Telefonstichproben für bundesdeutsche Erhebungen nicht verhindern und entsprechende Lösungsansätze, bspw. durch die zusätzliche Verwendung von Mobilfunkstichproben, zu finden sind (vgl. Gabler et al. (1998), S. 10. ; vgl. Meier (1999), S.

120., vgl. Sand (2015a), S. 134ff.).

Das zweite Problem lässt sich gerade an den Anfängen der telefonischen Befragungen in Deutschland verdeutlichen. Zu Beginn galt das Telefonregister aufgrund der Eintragungspflicht als geeignete Auswahlgrundlage, da dieses (potenziell) alle Rufnummern deutscher Privathaushalte beinhaltet. Mit der Aufhebung der Eintragungspflicht zum Ende des Jahres 1991 entstanden jedoch Probleme in der Verwendung eines solchen Telefonregisters als Auswahlgrundlage. S. Häder (1996) wies bereits darauf hin, dass durch die reine Beschränkung auf ein Telefonverzeichnis gerade dann verzerrte Schätzergebnisse entstehen können, wenn sich Einheiten, die nicht in einem solchen Verzeichnis geführt werden, systematisch von denjenigen unterscheiden, die darin gelistet sind (vgl. S. Häder (1996), S. 46ff.). Daher wurde es bereits in den 1990er-Jahren notwendig, eine Auswahlgrundlage zu gestalten, die sowohl eingetragene als auch nicht eingetragene Rufnummern beinhaltet. Eine nähere Darstellung unterschiedlicher Verfahren zur Generierung einer solchen Auswahlgrundlage wird in Abschnitt 2.2 gegeben. Dennoch zeigt u. a. Sand (2015b), dass trotz der derzeitigen Bemühungen und der zur Zeit verwendeten Auswahlgrundlage Probleme aufgrund neuer Anbieter, neu emittierter Rufnummernbereiche und dem Eintragsverhalten entstehen, die zur Folge haben können, dass diese Auswahlgrundlage nicht alle stichprobenrelevanten Einheiten beinhaltet (vgl. Sand (2015b), S. 23ff.).

Mit der Auswahl der Erhebungseinheit eng verbunden sind ebenfalls Probleme des Antwortverhaltens telefonisch Befragter. Noelle-Neumann und Petersen (2000) zufolge ist die durchschnittliche Verweigerungsrate bei telefonischen Befragungen im Vergleich zu Face-to-Face-Interviews höher. Hierbei argumentieren die beiden Autoren, dass es bei Telefonerhebungen einfacher sei, ein Interview durch das "Auflegen des Hörers" zu verweigern oder abzubrechen, als dies bei einem visuellen Kontakt wäre (vgl. Noelle-Neumann und Petersen (2000), S. 193f.). Ähnliche Schwierigkeiten beschreibt auch Groves (2011) während der Anfänge der telefonischen Erhebungen im Zeitraum zwischen 1960 und 1990. Ebenso sieht der Autor das vermehrte Aufkommen partiell abgeschlossener Interviews darin begründet (vgl. Groves (2011), S. 865.). Auch jüngere Untersuchungen verweisen auf gerade diese Unterschiede in der Responserate. So halten bspw. U. Engel und Schmidt (2014) fest, dass telefonische Befragungen mehrheitlich geringere Antwortraten hervorbringen.

Eine weitere Möglichkeit, durch die die Unterschiede in der Responserate begründet sein können, liegt in der Stichprobenauswahl. Wie zuvor bereits geschildert, greifen persönlich-mündliche Erhebungen im Idealfall auf Registerstichproben zurück, bei denen weitere Informationen (bspw. ob eine Einheit zur Zielpopulation gehört) verfügbar sind. Bei telefonischen Befragungen besteht die Stichprobenauswahl häufig lediglich aus einer Rufnummer, deren Informationsgehalt über den dazugehörigen Anschluss und die damit verbundene Einheit sehr gering ist. Daher ist der Anteil von stichprobenneutralen Ausfällen in Telefonbefragungen oftmals deutlich höher, da bei Stichproben, die nicht rein auf einer Listenauswahl basieren, auch Kontakte

zu Anschlüssen erfolgen, deren dazugehörige Einheiten nicht Teil der Zielpopulation sind (bspw. Geschäftsanschlüsse).

Darüber hinaus ist der Anteil der Rufnummern, deren Zugehörigkeit unklar ist, nicht zu vernachlässigen. Dies liegt darin begründet, dass ohne einen zustande kommenden Kontakt nicht determiniert werden kann, ob eine Rufnummer zu einer Einheit der Zielpopulation führt (vgl. American Association for Public Opinion Research (2011), S. 15ff.; vgl. S. Häder et al. (2009a), S. 72; vgl. Schneiderat und Schlinzig (2011), S. 124).

Bei der Messung bestimmter Merkmale bei telefonischen Befragungen gilt ein besonderes Augenmerk dem Antwortverhalten der befragten Einheit. Die Antwortneigungen in Abhängigkeit verschiedener Modi wurde von Groves bereits 1979 anhand des *SRC Spring Omnibus Surveys* aus dem Jahre 1976 näher untersucht. Bei der Erhebung wurde sowohl telefonisch als auch persönlich-mündlich befragt, wobei die Zufriedenheit über den verwendeten Erhebungsmodus in Erfahrung gebracht wurde. Auch hier lag die Responserate im Festnetzbereich unterhalb derer der Face-to-Face-Stichprobe. Als Ergebnis hält der Autor fest, dass ein deutlich höherer Anteil mit der Befragungssituation bei der persönlich-mündlichen Befragung zufrieden war. Die Gründe hierfür können vielfältig sein. Die Divergenz lässt sich demnach evtl. anhand einer ungleichen Hingabe oder auch Verpflichtung des Befragten in Abhängigkeit des Modus oder einer sich unterscheidenden Identifikation mit dem Medium begründen (vgl. Groves (1979), S. 192f).

Ebenfalls mit dem Teilnahmeverhalten verbunden ist die Selektivität der an einer solchen Befragung teilnehmenden Einheiten. Diesbezüglich sind gerade die soziodemografischen Merkmale der befragten Einheiten im Vergleich zu deren Verteilung innerhalb der Zielpopulation von gesteigertem Interesse. Dieser Verteilung gingen bereits Jung (1990) sowie S. Häder (1994) in den 1990er Jahren nach. Die von Häder angeführten Ergebnisse der Untersuchung des Politbarometers durch Jung, bei dem in aufeinanderfolgenden Jahrgängen sowohl eine Face-to-Face-Befragung als auch eine Telefonerhebung durchgeführt wurde, zeigen in einigen Merkmalen erhebliche Unterschiede. So wurden bei der Telefonerhebung bspw. ein höherer Anteil an Mehrpersonenhaushalten sowie Personen mit höherem formalen Bildungsabschluss und einem höheren beruflichen Status erreicht. Einen Unterschied bei dem Merkmal Geschlecht konnte nicht festgestellt werden. Diese Befunde bleiben auch nach einer entsprechenden Designgewichtung bestehen (vgl. S. Häder (1994), S. 4; vgl. Jung (1990), S. 393ff.). Gerade dieser sog. *Bildungsbias*, der das Phänomen darstellt, dass formal höher Gebildete eher dazu geneigt sind, an (sozialwissenschaftlichen) Umfragen teilzunehmen, manifestiert sich in telefonischen Erhebungen besonders deutlich. Auch in jüngeren Untersuchungen, wie bspw. bei Graeske und Kunz (2009) oder Schneiderat und Schlinzig (2011), zeigt sich, dass ein erheblich größerer Anteil formal höher gebildeter Teilnehmer in der Erhebung zu finden sind (vgl. Schneiderat und Schlinzig (2011), S. 129.; vgl. Graeske und Kunz (2009), S. 69.).

Im Rahmen der Messproblematik bezüglich dieses Erhebungsmodus führt Groves (2011) an,

dass die zuvor beschriebene vermehrte Entstehung partieller Interviews bei Telefonbefragungen zur Folge hat, dass kürzere Messinstrumente eingeführt werden (vgl. Groves (2011), S. 865.). Eben diese Verknappung der Befragung über das Telefon kritisieren Noelle-Neumann und Petersen (2000) und stellen fest, dass mit dem Umstieg auf diese Methode wesentliche Qualitätsmerkmale der Befragung verloren gehen. Die Autoren argumentieren, dass es so zu einem "Verarmen" des Fragebogens und zu einer Reduktion der methodischen Vielfalt kommt, wodurch wesentliche Qualitätsmerkmale einer Umfrage verloren gehen.

Der fehlende visuelle Kontakt führt demnach dazu, dass der Einsatz von Hilfsmitteln im Rahmen der Befragung ausbleibt oder zumindest erschwert wird. So fehlt bspw. die Möglichkeit während des Interviews sog. *Showcards* einzusetzen. Bestimmte Befragungsarten werden dadurch nicht ermöglicht, was Noelle-Neumann und Petersen zufolge die Gefahr einer monotonen Befragungssituation birgt und daher einen negativen Einfluss auf die Datenqualität haben kann (vgl. Noelle-Neumann und Petersen (2000), S. 184ff.).

Weiterhin besteht bei einem visuellen Kontakt der Vorzug gegenüber einer rein akustischen Kommunikation darin, dass der Interviewer bestimmte Parameter, die für die Erhebung von Interesse sein können, bereits vor der Befragung beurteilen kann. So erlaubt die persönlich-mündliche Befragung bspw. eine Einschätzung des Interviewers über die Wohnverhältnisse, Nachbarschaft oder Ausstattung des Haushaltes (mit bestimmten Gütern) der befragten Einheit. Bei einer telefonischen Erhebung kann dies lediglich erfragt werden. Die Richtigkeit der Angaben liegt bei diesen Befragungen außerhalb der Kontrolle des Interviewers (vgl. Noelle-Neumann und Petersen (2000), S. 188.).

Ein weiterer Kritikpunkt knüpft an das Antwortverhalten der Befragten an. Hierbei ist zu befürchten, dass bei Telefonbefragungen flüchtiger geantwortet werden könnte und der Befragte "den Weg des geringsten Widerstandes" geht. Hiermit verknüpft sind bspw. *Recency-Effekte*, deren Vorkommen in telefonischen Erhebungen höher eingeschätzt wird (vgl. Noelle-Neumann und Petersen (2000), S. 194.; S. 196.). Die Recency-Problematik beschreibt, dass Befragte bei der akustischen Verarbeitung der vorgegebenen Antwortmöglichkeiten einer Frage genau diejenigen wählen, die zuletzt angeboten wurden. Diese sind noch am präsentesten und erfordern eine geringere kognitive Anstrengung. Eine Untersuchung von M. Häder und Kühne (2009a) zeigt jedoch, dass sich eine solche Annahme auf der Grundlage der Daten der Erhebung CELLA (siehe Kapitel 2.4.4) nicht nachweisen lässt (vgl. M. Häder und Kühne (2009a), S. 211f.).

Der Interviewer kann ferner die Befragung weniger exakt steuern und bspw. die Anwesenheit dritter Personen nicht verhindern. Dies könnte sich unter Umständen ebenfalls negativ auf die Antwortqualität, gerade bei sensiblen Themen, sowie auf die generelle Teilnahmebereitschaft auswirken. Einige wenige Untersuchungen zeigen zudem, dass die Anwesenheit Dritter, gerade wenn diese dem Befragten vertraut sind, das Antwortverhalten beeinflussen kann (vgl. Zipp und Toth (2002), S. 195ff.; Beck et al. (2005)). Durch Schneiderat und Schlinzig (2009b) wurde

ein solcher Einfluss auf das Interviewgeschehen im Rahmen der CELLA-Studie untersucht. Hier wurden neben der direkten Frage nach der Anwesenheit Dritter auch Hintergrundgeräusche erfasst, die auf die Anwesenheit weiterer Personen hindeuten. Als Ergebnis halten die beiden Autoren fest, dass es keinen signifikanten Unterschied zwischen Befragungen mit und ohne Anwesenheit weiterer Personen bezüglich des Item-Response-Verhaltens gibt (vgl. Schneiderat und Schlinzig (2009b), S. 128f.).

Es bleibt jedoch unklar, ob die allgemeine Teilnahmebereitschaft durch die Anwesenheit weiterer Personen in diesem Rahmen beeinflusst wird oder ob das Antwortverhalten eher nach sozialer Erwünschtheit erfolgt. In diesem Bereich besteht ein Vorteil bei persönlich-mündlichen Befragungen: Durch den zusätzlichen visuellen Kontakt kann eine genauere Auskunft über die Anwesenheit und das Intervenieren weiterer Personen erfasst werden.

Trotz der geschilderten Schwierigkeiten kann allerdings der Aussage *„[e]ine Umfrageforschung, die sich überwiegend auf Telefoninterviews stützt, entwickelt sich auf ihre primitiven Anfangsformen zurück“*, von Noelle-Neumann und Petersen (2000) (S. 198) nicht entsprochen werden. So führt M. Häder (2009a) bspw. an, dass diese Probleme zwar bekannt seien, jedoch nicht die weite Verbreitung von telefonischen Umfragen verhindert hätten (vgl. M. Häder (2009a), S. 16.). Auch der Anteil der durch den ADM durchgeführten telefonischen Befragungen bestärkt dies (vgl. Arbeitskreis Deutscher Markt und Sozialforschungsinstitute ADM (2015) S. 20.). Demnach ist vielmehr der Aufbau einer geeigneten Auswahlgrundlage sowie die Wahl des richtigen Instrumentes für eine bestimmte Befragung ausschlaggebend.

2.2 Historische Entwicklung der Strategie zur Ziehung von Telefonstichproben

Der gezielte Einsatz des Telefons in der Umfrageforschung wurde in der US-amerikanischen Literatur bereits zu Beginn der 1960er Jahre beschrieben. Jedoch wurde diese Methode zunächst vornehmlich im privatwirtschaftlichen Bereich verwendet. Die Auswahl der Erhebungsteilnehmer beschränkte sich hauptsächlich auf Bestandslisten. Zusätzlich wurde diese Methode zuerst ausschließlich für spezielle Zielpopulationen angewendet (vgl. Groves (2011), S. 863f.). Mit wenigen Ausnahmen, wie bspw. die Untersuchung von Hochstim (1967), wurden innerhalb dieses Zeitraums lediglich eine geringe Anzahl von Untersuchungen der Vorzüge und Nachteile von Telefonumfragen im Vergleich zu anderen Erhebungsmodi durchgeführt. Aber auch solche Untersuchungen bieten nur eine eingeschränkte Vergleichbarkeit zu heutigen Vorgehensweisen, da diese sich oftmals auf Stichproben beschränkten, deren Adressen aus der damaligen Zensuspopulation stammten (vgl. Hochstim (1967), S. 978). Eine breitere Diskussion über die Vorzüge von Telefoninterviews wurde erst in den 1970er Jahren bspw. durch Rogers (1976) oder Groves

(1979) eingeleitet (vgl. Rogers (1976); Groves (1979); Hüfken (2000), S. 11.).

2.2.1 Die Entwicklung allgemein

Mit der rapiden Verbreitung des Festnetztelefons als Telekommunikationsmedium sowie durch das damals staatliche Monopol auf Telekommunikation wurde in den 1970er Jahren schließlich eine Grundlage geschaffen, die die Ziehung von Telefonstichproben für die USA ermöglichte (vgl. Groves (2011), S. 864).

Im selben Zeitraum erschien eine verbreitete Diskussion über die Bedeutung der Auswahlgrundlage, aus der eine Stichprobe für eine solche Erhebung zustande kommt. Hierbei gilt zu beachten, dass die Konstruktion einer Auswahlgrundlage oftmals eng mit dem Ziehungsverfahren aus dieser verwoben ist. Glasser und Metzger (1972) beschreiben bereits Probleme bei einer Auswahl von Zielpersonen basierend auf Einträgen in Telefonbüchern. Die Autoren sahen das hauptsächliche Problem in nicht-eingetragenen Telefonhaushalten, die durch dieses Vorgehen aus der Erhebung ausgeschlossen wurden. Daneben führen sie Schwierigkeiten bei Haushalten an, die über keinen Telefonanschluss verfügen. Überdies beschreiben sie Probleme mit doppelten Einträgen innerhalb und zwischen unterschiedlichen Verzeichnissen sowie Schwierigkeiten, ein geeignetes Verzeichnis für bestimmte Regionen zu finden. Den Autoren zufolge besaßen zum damaligen Zeitpunkt ca. 89,3 % aller Haushalte einen Telefonanschluss. Der Anteil der nicht-gelisteten Haushalte betrug jedoch schon 20 % bis 30 %. Der Vorschlag der Autoren war, ein *Random Digit Dialing* (RDD)-Verfahren einzuführen, das zumindest einige Stellen der Rufnummer als Zufallszahl generiert. Aus einer solchen Auswahlgrundlage können anschließend uneingeschränkt zufällig oder geschichtet Stichproben gezogen werden (vgl. Glasser und Metzger (1972), S. 60f.). Aufgrund der Gestaltung des amerikanischen Nummernraumes kann ein RDD-Verfahren laut Schnell (1999) vergleichsweise einfach implementiert werden. Im amerikanischen Fall geben die ersten drei Ziffern der zehnstelligen Rufnummer die Region und die folgenden drei die Vermittlungsstelle wieder. Um bei solchen Voraussetzungen einen entsprechenden Auswahlrahmen zu schaffen, muss daher lediglich das Wissen über alle aktiven Kombinationen von Vorwahlbereichen und Vermittlungsstellen vorhanden sein. Die letzten vier Ziffern werden dann zufällig generiert (vgl. Schnell et al. (1999), S. 280f.).

Dieses Verfahren kann den Autoren zufolge dem Problem der Nichtinklusion von ungelisteten Anschlüssen vorbeugen. Allerdings warnten sie zeitgleich vor einer fehlerhaften Implementierung dieses Verfahrens, da dies weitere Quellen der Verzerrung bei der Schätzung der Ergebnisse hinzufügen kann. Auch wenn ein solches Verfahren in den USA vergleichsweise einfach zu implementieren ist, zeigt sich, dass ein großer Teil der damals untersuchten Rufnummern (ca. 73 %) nicht geschaltet war. Neben der dadurch gestiegenen Notwendigkeit mehrere Kontaktversuche pro abgeschlossenem Interview zu unternehmen, beschreiben die Autoren ein weiteres Problem

darin, dass aufgrund der damaligen Netzwerkarchitektur oftmals eine Weiterleitung hin zu einer geschalteten Rufnummer erfolgte. Diese Rufnummer hat dann eine höhere Auswahlwahrscheinlichkeit als andere, was sich bei Nicht-Berücksichtigung verzerrend auf einen geschätzten Parameter auswirken kann. Das Fazit der Autoren bei diesem Verfahren war demnach, bei jedem Interview die gewählte Rufnummer zu verifizieren (vgl. Glasser und Metzger (1972), S. 59ff.).

Als mögliche Lösung, die hohe Anzahl von Kontaktversuchen zu umgehen, die durch nicht-vergebene Rufnummern in solchen RDD-Verfahren entstehen, schlugen Mitofsky (1970) und Waksberg (1978) vor, klusterbasierte Telefonstichproben zu ziehen. Die beiden Autoren erkannten in ihren Untersuchungen, dass eine Großzahl von vergebenen Rufnummern aus RDD-Verfahren aufeinanderfolgende Ziffernfolgen sind. Die Grundlage dieses Verfahrens ist noch immer die Auswahlgrundlage für RDD-Stichproben. Die Rufnummern werden dann in Blöcken aus bspw. 100 aufeinanderfolgenden Rufnummern zusammengefasst. Diese Blöcke bilden die *Primary Sampling Units (PSU)*. Bei dem vorgeschlagenen Ziehungsverfahren werden zuerst eine bestimmte Anzahl von Rufnummernblöcken uneingeschränkt zufällig (mit Zurücklegen) gezogen und aus den jeweiligen Blöcken eine der 100 Rufnummern ausgewählt und kontaktiert. Bei einer nicht-vergebenen Rufnummer wird dieser Block verworfen. Bei einem erfolgreichem Kontakt werden eine zuvor festgelegte Anzahl von weiteren Rufnummern aus diesem Block kontaktiert. Die Autoren zeigten, dass ein solches Verfahren noch immer gleiche Auswahlwahrscheinlichkeiten für jede Rufnummer zur Folge haben kann, während die Anzahl der Anwahlversuche reduziert wird (vgl. Waksberg (1978), S. 40ff.; Hansen (2008), S. 342.; Kalsbeek und Agans (2008), S. 34f.).

Weiterhin veränderte sich die Stichprobenstrategie im Laufe dieser Zeit hin zu der Verwendung von Auswahlgrundlagen, die sowohl Rufnummern aus Telefonverzeichnissen als auch nicht-gelistete Rufnummern beinhaltet und dennoch durch eine Listenauswahl gestützt wird. Ein Beispiel hierfür ist das *Randomized Last Digit (RLD)*-Verfahren. Bei diesem Verfahren wird zuerst aus einer bestehenden Liste, bspw. einem Telefonverzeichnis, eine Rufnummer uneingeschränkt zufällig ausgewählt. In einem zweiten Schritt werden die letzten ein bis zwei Ziffern durch zufällig generierte Ziffern ausgetauscht. Diese Nummer wird dann in der Stichprobe anstelle der initial ausgewählten Rufnummer verwendet.

In der praktischen Umsetzung der telefonischen Befragung wurde der verbreitete Einsatz von Telefonstichproben durch die Einbindung von Dateneingabegeräten begünstigt. Diese waren wiederum mit Computern verbunden und vereinfachten die Handhabung dieser Erhebungsmethode. Die ursprünglich aus der privatwirtschaftlichen Umfrageforschung stammenden CATI - Erhebungen ermöglichten es, dem Interviewer, bereits Ende der 1960er Jahre, den Fragetext an einem Bildschirm aufzuzeigen und die Eingabe des Interviewers weiterzuverarbeiten.

Zu Beginn der 1970er-Jahre wurden erstmals generelle Bedenken bezüglich der sinkenden Responseraten diskutiert. Auch wenn diese zum damaligen Zeitpunkt im Vergleich zur gegenwärtigen Situation noch sehr hoch waren, bot das Telefoninterview im Gegenzug zu Face-to-Face-

Befragungen durch das Auflegen eine relativ simple Möglichkeit, ein Interview zu verweigern oder abubrechen (vgl. Groves (2011), S. 864).

2.2.2 Die Entwicklung in Deutschland

In Deutschland verbreitete sich der gezielte Einsatz von Telefonstichproben in der Umfrageforschung deutlich langsamer. Hier ging der vermehrte Einsatz dieser Methode mit der flächendeckenden Versorgung an Festnetzanschlüssen des westdeutschen Bundesgebietes in den 1980er Jahren einher. Waren es bspw. 1983 lediglich 14 % aller Umfragen, die über das Telefon durchgeführt wurden, betrug dieser Wert 1989 bereits ca. 38 %. Als Grundlage zur Stichprobenziehung galt zum damaligen Zeitpunkt aufgrund der allgemeinen Eintragsverpflichtung das Telefonverzeichnis als adäquates Mittel zur Stichprobenziehung. Zeitgleich stieg der Anteil der Haushalte, die über einen Festnetzanschluss verfügten bis 1988 auf 92 % an (vgl. Gabler et al. (1998), S. 10; vgl. Anders (1990), S. 427; vgl. S. Häder et al. (2009a), S. 21.). Im Gegensatz dazu war die Versorgung mit Festnetzanschlüssen der im Folgejahr hinzukommenden neuen Bundesländer auf einem deutlich geringeren Niveau. So berichten Gabler et al. (1998), dass die Versorgung der ostdeutschen Haushalte mit Telefonanschlüssen 1990 lediglich 20 % betrug. Dieser Wert kann mit der Vergabep Praxis der ehemaligen DDR begründet werden. Die Verwendung von Telefonbefragungen in den frühen 1990er Jahren wurde durch diese Diskrepanz stark beeinträchtigt.

Verstärkt wurde das Problem durch die zu dem damaligen Zeitpunkt bestehenden soziodemografischen Unterschiede zwischen den Haushalten der neuen Bundesländer, die einen Festnetzzugang besaßen, und jenen, die nicht über einen Festnetzanschluss erreichbar waren. Eine flächendeckende Versorgung dieser Haushalte war somit trotz der Bemühungen der damals mit dem Netzausbau beauftragten *Deutsche Bundespost Telekom* (DB Telekom) nicht eher als Mitte der 1990er Jahre gewährleistet (1995: 53%; 1996: 67%). Der Angleich des Anteils der Haushalte, die über einen Festnetzanschluss verfügten, erfolgte jedoch erst zu Beginn der 2000er-Jahre (2000: 91 % (Ost), 96 % (West)) (vgl. Gabler et al. (1998), S. 11.; Heckel (2002), S. 11.).

Somit wurde auch hier der effektive Einsatz von Telefonbefragungen vornehmlich durch die bestehende Auswahlgrundlage determiniert. Die Aufhebung der Eintragspflicht, in Kraft getreten Ende des Jahres 1991, erschwerte die Durchführung der Methode. Der wachsenden Anzahl von Festnetzanschlüssen stand ein stetig geringer werdender Anteil an Haushalten gegenüber, die in Telefonverzeichnissen geführt wurden. Infolgedessen ergab sich für Telefonbefragungen das Problem, dass eine Stichprobenauswahl anhand von Telefonverzeichnissen aus mehreren Gründen nicht mehr möglich war. So beschreibt S. Häder (1996), dass aufgrund soziodemografischer Unterschiede der Haushalte mit und ohne Eintrag in das Telefonregister Einschränkungen zu befürchten sind, wenn sich der Zuwachs nicht-eingetragener Haushalte (1994: 3,7 %) verstetige.

Die Befürchtung, dass das Telefonverzeichnis als alleinige Auswahlgrundlage nicht mehr genüge, äußert Häder bereits 1996 in der Untersuchung der sog. *Non-Pubs*, also denjenigen Haushalten, die zwar über einen Festnetzanschluss verfügen, aber nicht in einem Telefonverzeichnis geführt werden. Der Autorin zufolge sei ein Telefonverzeichnis nur noch dann eingeschränkt zur Stichprobenziehung nutzbar, wenn sich der damals berichtete Trend des Eintragsverhaltens fortsetze und sich nicht-eingetragene Haushalte von denjenigen, die eingetragen sind, systematisch in ihren soziodemografischen Charakteristika unterscheiden (vgl. S. Häder (1996), S. 65f.).

Der zu diesem Zeitpunkt noch vergleichsweise geringe Anteil an Haushalten, die zwar über einen Festnetzanschluss verfügten, allerdings nicht in einem Telefonbuch verzeichnet waren, wird durch die *Kosten* des Austragens aus einem solchen Verzeichnis begründet. Demnach ist anzunehmen, dass das Beantragen auf Löschen einer bereits bestehenden Rufnummer aus einem solchen Verzeichnis häufig nicht dem Nutzen entsprach, da der dafür erforderliche Arbeitsaufwand als zu hoch angesehen wurde. Bei der Neuschaltung einer Rufnummer ist jedoch der Verzicht auf den Eintrag in das Telefonverzeichnis wahrscheinlicher (vgl. S. Häder (1996), S. 47ff.; Bundesnetzagentur (2011), S. 30.). Bestätigt wurde diese Annahme durch die Untersuchung von Heckel (2002). Sie vermerkt, dass in dem Untersuchungszeitraum (2002) ein nennenswerter Unterschied zwischen dem Eintragsverhalten von Ost- und Westdeutschen Haushalten bestand. Dieser wird durch die nach der Wiedervereinigung neu eingerichteten Festnetzanschlüsse, die nicht eingetragen wurden, begründet. Weiterhin sind deutlich mehr Haushalte in ländlichen Regionen eingetragen. Dabei besteht die Vermutung, dass das Eintragsverhalten eines Haushaltes mit dessen Mobilität zusammenhängt. Ein Haushalt, der häufig den Standort wechselt, könnte somit seltener in einem Telefonverzeichnis zu finden sein (vgl. Heckel (2002), S. 13f.).

Jüngere Untersuchungen zeigen, dass sich das Verhältnis von nicht-eingetragenen Haushalten in den Bundesgebieten angeglichen hat. Eine Untersuchung der von GESIS für Festnetzstichproben genutzten Auswahlgrundlage für das Jahr 2013 dokumentiert, dass dieser Wert für beide Bundesgebiete ähnlich gering ist. Dabei entsprach der Anteil der eingetragenen Haushalte in den neuen Bundesländern ca. 11,3 % und der in den alten Bundesländer ca. 11,6 %. Auch 2015 war der Anteil eingetragener Haushalte in den alten Bundesländern mit ca. 9,9 % mit dem Anteil der neuen Bundesländer von 9,8 %² vergleichbar. Es kann angenommen werden, dass sich das Eintragsverhalten der alten Bundesländer dem der neuen angeglichen hat.

Begünstigt wurde diese Entwicklung auch durch die Liberalisierung des Marktes für Festnetztelefonie, die durch die Aufhebung des Monopolstatus der *Deutsche Telekom AG* (DTAG) erwirkt wurde. Dem entspricht die Untersuchung von Heckel und Wiese (2011), die feststellt, dass ca. 97 % der noch in Telefonverzeichnissen geführten Haushalte aus dem Nummernbestand des ehemaligen Monopolisten DB Telekom stammen. Eine aktuellere Untersuchung der GESIS-Auswahlgrundlage zeigt jedoch, dass im Herbst 2013 ca. 13,4 % der eingetragenen Rufnummern

2 Beide Auswertungen wurden unter Ausschluss des Vorwahlbereiches "030" (Berlin) durchgeführt.

innerhalb des Telefonverzeichnisses bereits neuen Telekommunikationsanbietern zugeordnet werden konnten. Dennoch kann hieraus geschlossen werden, dass neben dem Wohnortswechsel auch der Wechsel des Anbieters Einfluss auf das Eintragsverhalten hat (vgl. Sand (2015b), S. 15f.; Heckel und Wiese (2011), 106.). Weiterhin berichtet die Bundesnetzagentur in ihrem Jahresbericht zu 2012/2013, dass für das Untersuchungsjahr damit zu rechnen ist, dass ca. 42 % aller Festnetzanschlüsse durch diese "neuen" Anbieter realisiert werden (vgl. Bundesnetzagentur (2013), S. 29). Dazu muss jedoch bemerkt werden, dass dieser Anteil jegliche Art von Festnetzzugang sowohl für private als auch für gewerbliche Anschlüsse beinhaltet.

In Anbetracht dieser Entwicklung kann demnach davon ausgegangen werden, dass das Telefonbuch als alleinige Auswahlgrundlage (in den meisten Fällen) nicht mehr hinreichend ist, um Telefonstichproben zu ziehen, die eine unverzerrte Schätzung (in den meisten Fällen) erlauben. Um dennoch Befragungen über das Telefon zu realisieren, die eine möglichst unverzerrte Schätzung von interessierenden Parametern (der deutschen Bevölkerung) erlauben, wurden mehrere Ansätze diskutiert.

Aufgrund der Gestaltung des deutschen Nummernraumes kann das in der amerikanischen Umfrageforschung oftmals bevorzugte RDD-Verfahren nicht ohne Einschränkungen verwendet werden. Schnell et al. (2013) merken diesbezüglich an, dass ein solches Vorgehen durch die Gestaltung des amerikanischen Rufnummernraumes begünstigt wird. In der Anwendung auf den deutschen Rufnummernraum kann ein solches Verfahren jedoch als ineffizient betrachtet werden, da die Länge einer Rufnummer inklusive ihrer Vorwahl zwischen acht und zwölf Stellen variiert. Für das von Mitofsky und Waksberg vorgeschlagene Verfahren gelten ähnliche Probleme. Beide Verfahren erleiden daher aufgrund der unterschiedlichen Längen deutscher (Festnetz-) Rufnummern erhebliche Effizienzverluste (vgl. Sand (2015b), S. 32f.). Meier (1999) stellt dazu fest, dass sich für den deutschen Rufnummernraum keine Systematik finden lässt, die es erlaubt, mit einem akzeptablen Aufwand Rufnummern vollständig aus Zufallsziffern zu generieren (vgl. Meier (1999), S. 98.).

S. Häder (2000) verdeutlicht dies sehr nachdrücklich anhand eines Beispiels bezogen auf die 5.200 deutschen Vorwahlbereiche (*Ortsnetzkennzahlen* (ONK)). Dazu erläutert die Autorin, dass bei der Durchführung eines RDD-Verfahrens in einem ersten Schritt die Spannweite dieser Vorwahlbereiche ermittelt werden müsste. Diese liegt zwischen drei und sechs Stellen.³ Danach müssten alle möglichen Ziffernkombinationen erzeugt werden, die in Kombination mit der Vorwahl mindestens acht und maximal zwölf Stellen haben. Nach Abschätzung der Autorin würde ein solches Vorgehen bereits im Jahre 2000 nur zu 0,3 % auf eingetragene Rufnummern treffen, während das Verhältnis der geschalteten Rufnummern zu der Anzahl der gewählten Rufnummern (*Hitrate*) zu diesem Zeitpunkt dann nur bei ca. einem Prozent läge (vgl. S. Häder (2000), S. 6.; vgl. Gabler und Häder. (2002), S. 340f.). Da die Anzahl der Festnetzanschlüsse seit

³ Inklusive der nationalen Vorwahl 0.

der damaligen Untersuchung deutlich angestiegen ist, kann vermutet werden, dass ein solches Verfahren zum heutigen Zeitpunkt deutlich schlechter abschneiden würde. Demnach ist ein reines RDD-Verfahren aufgrund des personellen und finanziellen Aufwandes in Deutschland nicht denkbar.

Da darüber hinaus zwischen den jeweiligen Blöcken von Rufnummern große Lücken bestehen können, die keine geschaltete Rufnummer enthalten, wurde ein Verfahren notwendig, das lediglich auf Rufnummernbereiche zurückgreift, in denen sich geschaltete Anschlüsse befinden. Ein reines RDD-Verfahren, wie es zuvor beschrieben wurde, ist allerdings ebenfalls mit Problemen behaftet. So diskutieren Gabler und S. Häder (1997), dass Rufnummern, die nach einem solchen Verfahren erzeugt würden, mit unterschiedlichen Wahrscheinlichkeiten gezogen würden. Die Wahrscheinlichkeit, eine Rufnummer aus einer solchen Auswahlgrundlage zu ziehen, ist daher abhängig davon, wie viele Rufnummern desselben Blocks in einem Telefonverzeichnis zu finden sind. Zudem ist die Wahrscheinlichkeit mit der eine Rufnummer ausgewählt würde für eine gegebene Stichprobe nur schwer zu ermitteln (vgl. Gabler und S. Häder (1997), S. 8.). Ein weiteres Problem dieses Verfahrens, das von der Anzahl der eingetragenen Rufnummern abhängt, besteht darin, dass es zu viele eingetragene und zu wenige nicht-ingetragene Rufnummern für eine Stichprobe zieht (vgl. Meier (1999), S. 98.).

Aufgrund dieser Schwierigkeiten wurde von Gabler und Häder Mitte der 1990er Jahre ein Verfahren entwickelt, das es ermöglicht, eine Stichprobenauswahl in einem handhabbaren Aufwand zu treffen. Das Verfahren beinhaltet sowohl eingetragene als auch nicht-ingetragene Rufnummern. Das Gabler-Häder-Verfahren, das für Festnetzstichproben in Deutschland mittlerweile gängige Praxis ist, basiert in der Gestaltung der Auswahlgrundlage zwar auf Telefonbucheinträgen, ermöglicht jedoch die Ziehung von Telefonnummern mit zuvor bekannten, gleichen Wahrscheinlichkeiten. Da dieses Verfahren auf eingetragenen Rufnummern beruht, werden ausschließlich Nummern aus Bereichen gezogen, in denen mindestens ein geschalteter Festnetzanschluss existiert (vgl. M. Häder und Neumann (2015), S. 5.; S. Häder (2000), S. 6f.; S. Häder et al. (2009a), S. 30.).

Zur Erstellung einer solchen Auswahlgrundlage werden in einem ersten Schritt 100er-Blöcke anhand der eingetragenen Rufnummern eines Telefonverzeichnisses gebildet. Diese 100er-Blöcke sind als 100 aufeinanderfolgende Rufnummern zu verstehen. Zur Bildung dieser Blöcke werden alle Rufnummern eines Telefonverzeichnisses, die einem privaten Haushalt zuzuordnen sind, aus einem Telefonverzeichnis ermittelt. Darauf folgend werden diese Einträge trunziert. In diesem Schritt werden die letzten beiden Ziffern des Eintrags entfernt, um den eigentlichen Rufnummernblock zu bilden. Um zu gewährleisten, dass alle Blöcke die gleiche Auswahlwahrscheinlichkeit haben, werden anschließend alle Mehrfachnennungen einzelner Blöcke aus dem Datensatz entfernt. Nach mehreren weiteren Schritten der Bereinigung⁴ entsteht letztlich eine

⁴ Eine detaillierte Beschreibung zur Vorgehensweise ist bspw. bei Sand (2015b) nachzulesen.

Auswahlgrundlage, aus der dann uneingeschränkt zufällig oder geschichtet gezogen werden kann. Die Stichprobenauswahl aus einem solchen Auswahlrahmen erfolgt durch die Ziehung der jeweiligen Blöcke und deren anschließende Erweiterung um zwei Zufallszahlen zwischen "0" und "9" (vgl. S. Häder (2000), S. 6f.; vgl. S. Häder et al. (2009a), S. 30ff.).

Sand (2015b) beschreibt, dass aufgrund des derzeitigen Eintragsverhaltens in Kombination mit neuen Anbietern für Festnetztelekommunikation und der (damit verbundenen) gestiegenen Anzahl bereitgestellter Blöcke durch die *Bundesnetzagentur* (BNetzA) die Gefahr besteht, dass das Telefonverzeichnis als alleinige Grundlage zur Bildung von Rufnummernblöcken evtl. nicht mehr ausreichend ist. Der Grund hierfür liegt in dem bereits zuvor durch bspw. Heckel (2002) beschriebenen Verhalten von Anschlussaltern. So kann zum Beispiel der Wechsel des Anbieters dazu führen, dass eine neu-zugeteilte Rufnummer nicht mehr eingetragen wird. Bei einem *neuen* Anbieter, der von der BNetzA neue Rufnummernbereiche zugeteilt bekommen hat, kann dieses Verhalten im Aggregat dazu führen, dass neu emittierte Blöcke nicht mehr in Telefonverzeichnissen zu finden sind. Aufgrund dessen kann es erfolgversprechend sein, neben einem Telefonverzeichnis noch weitere Datenquellen, wie die Liste aller emittierten Rufnummernbereiche der Bundesnetzagentur, bei der Blockbildung zu bemühen (vgl. Heckel (2002), S. 11.; vgl. Sand (2015b), S. 23ff.). Darüber hinaus diskutiert unter anderem der ADM über die sinnvolle Kombination aus kleineren und größeren Blöcken (bspw. der Kombination aus 10er- und 100er-Blöcken) bei der Erstellung von Auswahlgrundlagen, um eine höhere Hitrate bei der Stichprobenziehung zu erzielen (vgl. Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute ADM (2012), S. 12.).

2.2.3 Mobiltelefonie und das veränderte Telekommunikationsverhalten

Die Idee der Telekommunikation, die nicht an einen stationären Festnetzanschluss gebunden ist, wurde erstmals mit der Einführung des Autotelefons 1983 realisiert. Dadurch kam das erste kommerziell genutzte Mobilfunknetz zustande. Eine weite Verbreitung der Mobiltelefonie wurde jedoch in Deutschland erst 1992 mit der Einführung des *D-Netzes* ermöglicht. Der asymmetrische Ausbau der Netzabdeckung für Festnetztelefonie zwischen den neuen und alten Bundesländern bewirkte, dass in einigen Regionen der neuen Bundesländer die Verbreitung von Mobiltelefonen den Ausbau von Festnetzanschlüssen überholte. Das Mobiltelefon als Massenkommunikationsmedium gibt es jedoch erst gegen Ende der 1990er Jahre (vgl. Schneiderat und Schlinzig (2009a), S.101; vgl. Bertschi (2007), S.25ff.).

Mit dessen Durchsetzung ergaben sich neue Herausforderungen für die Umfrageforschung. So berichtet Häder (2009), dass sich bereits in einer 1998 geführten Delphi-Befragung abgezeichnet hat, dass das vermehrte Aufkommen von Mobiltelefonen einen Rückgang an Haushalten bewirkt, die über einen Festnetzanschluss erreichbar sind (vgl. M. Häder (2009a), S. 13.). Die Bundesnetz-

agentur berichtet für Mobilfunkanschlüsse im Jahr 1999 eine Penetrationsrate von ca. 29 %, während die Europäische Verbrauchsstichprobe aus dem Jahr 1998 angibt, dass ca. 11 % aller deutschen Haushalte mindestens einen Mobilfunkanschluss besitzen (vgl. Bundesnetzagentur (2007), S. 24.; vgl. Statistisches Bundesamt (Destatis) (2016)).

Die Delphi-Untersuchung lud erstmals in Deutschland Experten aus Wissenschaft und Wirtschaft ein und befragte diese zu mehreren Zeitpunkten bezüglich ihrer Einschätzung künftiger *Telekommunikationsgewohnheiten* projiziert auf das Jahr 2005. Hierbei vermutete der Autor, basierend auf der Einschätzung der Experten, dass die zukünftige Entwicklung der Mobiltelefonie direkte Implikationen für eine Vielzahl von wissenschaftlichen Disziplinen, bspw. für die Sozial- und Kommunikationswissenschaft, haben wird. Ein wichtiger Befund dieser Erhebung ist, dass die Ausbreitung der Mobiltelefonie einen direkten Einfluss auf die Verbreitung von Festnetzanschlüssen haben kann. Auch wenn die Expertenmeinungen stark divergieren, wurde tendenziell davon ausgegangen, dass die Anzahl der Festnetzanschlüsse bis 2005 abnehmen würde, während der Anteil der Personen mit Mobilfunkanschluss im gleichen Zeitraum im Median auf 50 % geschätzt wurde (vgl. M. Häder (2000), S. 3ff.). Häder stellt für das Jahr 2005 fest, dass *„mit Mobiltelefonen [voraussichtlich] häufiger telefoniert wird, dass mehrere Handys im Haushalt vorhanden sein werden und dass die Nutzung des Mobilfunks zur Übermittlung von Kurznachrichten steigende Bedeutung erlangen wird“* (M. Häder (2000), S. 11.).

2.3 Die Dual-Frame-Telefonstichprobe

Tatsächlich wurde die Entwicklung der Mobiltelefonie innerhalb der Delphi-Untersuchung sehr gut vorhergesehen. Jedoch bleibt die Einschätzung der Experten noch hinter den erreichten Kennzahlen. Die BNetzA berichtete für das Jahr 2005 ca. 79 Mio. Mobilfunkanschlüsse bzw. eine Marktpenetration von ca. 96% (vgl. Bundesnetzagentur (2008), S. 81.). Der jährlich durch *TNS-Infratest*⁵ durchgeführte f2f-Bus ergab für das gleiche Jahr, dass ca. 75% aller Befragten mittlerweile über Mobilfunk erreichbar seien. Zur Differenz dieser Angaben hält Glemser (2007) fest, dass diese großteils dadurch zustande kommen, dass die BNetzA lediglich das Aggregat aller berichteten Anschlüsse der Mobilfunkanbieter als Berechnungsgrundlage verwende. Die *Penetrationsrate*, die von der BNetzA wiedergegeben wird, berechnet sich durch die Anzahl der vergebenen *Subscriber Identification Module* (SIM)-Karten geteilt durch die Anzahl der in Deutschland lebenden Personen. Diese Angaben berücksichtigen daher nicht, dass Personen über mehrere Mobilfunkanschlüsse verfügen oder Anschlüsse nicht mehr nutzen (vgl. Glemser (2007), S. 9ff.).

Als Kritik an der Verwendung einer solchen Rate führen Meier et al. (2013) an, dass sie die

5 Seit September 2016 treten diese unter Kantar TNS auf.

Individualebene nicht richtig berücksichtige und die Anzahl der Haushalte, die über einen Mobilfunkanschluss verfügen, dadurch überschätze. Dennoch kann sie als Indiz für die Verwendung von Mobiltelefonen herangezogen werden (vgl. Meier et al. (2013), S. 130., vgl. Heckel und Wiese (2011), S. 109.).

2.3.1 Erreichbarkeit nach Telekommunikationsmedium

Bezüglich der Anzahl von Mobilfunkanschlüssen lässt sich anhand der Angaben der BNetzA auch nach dieser Periode ein deutlicher Anstieg auf einen Wert von ca. 114,3 Mio. Mobilfunkanschlüssen bis 2011 feststellen. In der darauf folgenden Periode stagniert dieser Wert jedoch und beträgt im zweiten Quartal 2015 ca. 112,4 Mio. (vgl. Bundesnetzagentur (2015), S. 38). Für Erhebungen basierend auf den zuvor dargestellten Festnetzstichproben ist die Entwicklung des Mobilfunksektors nach Meier et al. (2014) lediglich dann von Bedeutung, wenn der Zuwachs von Haushalten mit Mobilfunkanschluss gleichzeitig zu einer Abnahme der Haushalte führt, die über einen Festnetzanschluss verfügen (vgl. Meier et al. (2013), S. 130.). In der Untersuchung von Glemser (2007) zeigt sich, dass zwar in dem Beobachtungszeitraum von 1999 bis 2006 der Anteil der Haushalte, die nur noch über einen Mobilfunkanschluss erreichbar waren (Mobile-Only-Haushalte), auf einen Wert von 7,6 % anstieg. 90 % aller bundesdeutschen Haushalte verfügten 2006 jedoch noch über einen Festnetzanschluss. Trotz verstärkter Mobilfunknutzung verzichtet demnach lediglich ein Bruchteil aller Haushalte auf einen Festnetzanschluss (vgl. Glemser (2007), S. 9ff.).

Derzeitig können nach Angaben der BNetzA für das Jahr 2013 noch ca. 86 % der bundesdeutschen Bevölkerung über das Medium Festnetz erreicht werden. Im europäischen Vergleich liegt dieser Wert deutlich höher als der Gesamtdurchschnitt von ca. 70 % (vgl. Bundesnetzagentur (2013), S. 56). Die jüngsten Veröffentlichungen des ADM weisen vergleichbare Ergebnisse auf. So schätzt der Forschungsbericht, den die Arbeitsgruppe "Dual-Frame-Ansätze" im Zuge ihrer Arbeit 2012 veröffentlicht hat, den Anteil der Mobile-Onlys auf 12,4 % (vgl. Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute ADM (2012), S. 10.). Auch jüngere Untersuchungen des ADMs zeigen hier wenig Veränderung. So geben Glemser et al. (2014) den Anteil der Mobile-Only-Haushalte mit einem Wert von ca. 11 %⁶ an (vgl. Glemser et al. (2014), S. 182.). Anhand der jüngsten Daten des ADMs kann die Erreichbarkeit nach Art des Anschlusses für das Jahr 2014 auf Haushaltsebene mittels Abbildung 1 für die jeweiligen Auswahlgrundlagen dargestellt werden.⁷

⁶ Die Daten sind dabei aus einer Erhebung von 2011 gewonnen. Die Schätzung wurde unter variierenden Mobilfunkanteilen vorgenommen.

⁷ Abbildung 1 liegt die vereinfachte Annahme zugrunde, dass der jeweilige Auswahlrahmen alle Einheiten der entsprechenden Subpopulation beinhaltet.

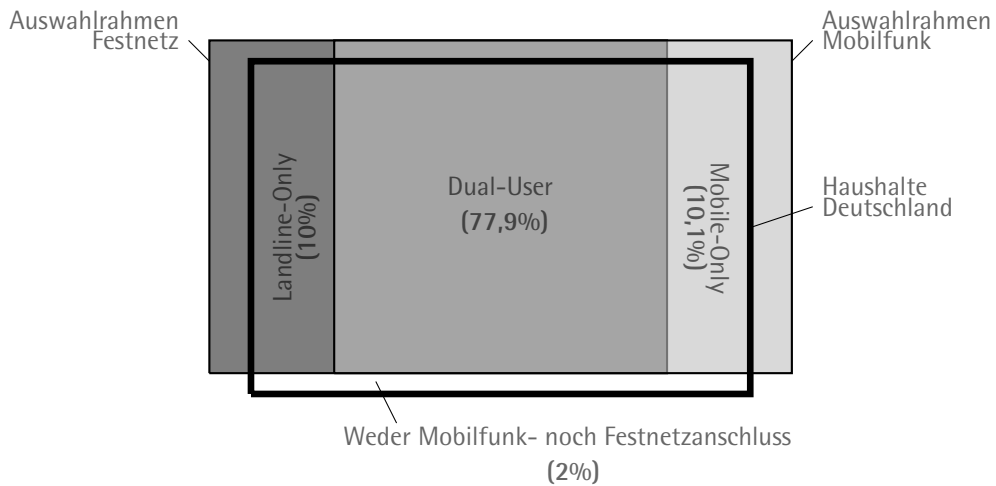


Abbildung 1 Kombination der bundesdeutschen Bevölkerung nach Erreichbarkeit über Festnetz und Mobilfunk

Quelle: Beckmann (2015)

Zur Bestimmung des Anteils der Only-Populationen an der Gesamtbevölkerung räumen Kuusela et al. (2008) erhebliche Probleme ein, wenn Schätzungen lediglich auf der Basis von Daten der Anbieter für Telekommunikationsdienste vorgenommen werden. Den Autoren zufolge kann demnach zwar der Anteil der Haushalte, die über Mobilfunk oder über Festnetz erreichbar sind, berechnet werden, jedoch sind diese Daten keine adäquate Grundlage, um den Anteil derjenigen Haushalte zu determinieren, die ausschließlich über Festnetz (Landline-Only) oder Mobilfunk erreichbar sind. Um dies zu erreichen, muss eine Erhebung, die genau diesem Zweck folgt, konzipiert werden (Kuusela et al. (2008), S. 88.).

Neben den umfangreichen Untersuchungen der beiden CELLA-Erhebungen versucht der f2f-Bus von *TNS-Infratest*, der jährlich persönlich-mündlich durchgeführt wird, (unter anderem) der Frage der jeweiligen Anteile der Only-Populationen auf den Grund zu gehen. Jüngere Ergebnisse dieser Erhebung zeigen, dass der Anteil an Mobile-Only-Haushalten derzeit vergleichsweise konstant bei ca. 10 % bis 11 % bleibt (vgl. Beckmann (2015)). Der Anteil der *Landline-Onlys*, also diejenigen Haushalte, die ausschließlich über einen Festnetzanschluss verfügen, hat darüber hinaus abgenommen. Die Bundesnetzagentur berichtet dazu, basierend auf den Daten des Eurobarometers 2014, dass ca. 91 % aller Haushalte mindestens einen Mobilfunkanschluss aufweisen (vgl. Bundesnetzagentur (2015), S. 48.).

Nach jüngeren Schätzungen des Eurobarometers 2015 liegt der Anteil der Mobile-Only-Haushalte jedoch bei ca. 15 %. Im innereuropäischen Vergleich ist dieser Anteil allerdings noch immer als gering einzuschätzen, vergewärtigt man, dass europaweit bereits 33 % aller

Haushalte ausschließlich über Mobilfunk erreichbar sind (vgl. European Commission (2016), S. 19.).

2.3.2 Mobile-Only-Haushalte und die damit verbundenen Schwierigkeiten

Anhand des dargestellten Anteils von Haushalten, die über einen Festnetzanschluss erreichbar sind, könnte evtl. noch immer davon ausgegangen werden, dass der Anteil der Mobile-Only-Haushalte in der Bevölkerung lediglich einen geringen Einfluss auf die Qualität der Schätzergebnisse hat. Hierzu wäre es wichtig, dass sich Mobile-Only-Haushalte in ihren soziodemografischen Charakteristika nicht von anderen Haushalten unterscheiden (vgl. Sand (2014), S. 10; vgl. Graeske und Kunz (2009), S. 58ff.).

Eine nähere Betrachtung der Angaben der Bundesnetzagentur zeigt hingegen, dass lediglich ca. 80 % aller Einpersonen-Haushalte bzw. nur 41 % aller Einpersonen-Haushalte unter 30 Jahren, einen Festnetzanschluss besitzen (vgl. Bundesnetzagentur (2015), S. 48.). Somit sind erhebliche Unterschiede bei Mobile-Only-Haushalten zu vermuten. Der Verzicht, gerade auch diese jüngeren Einpersonen-Haushalte bei einer bundesdeutschen Telefonerhebung zu befragen, birgt demnach das Potenzial, verzerrte Schätzwerte für diese (Teil-) Population zu erzeugen.

Dass im Hinblick auf soziodemografische Merkmale deutliche Unterschiede zwischen Mobile-Only-Haushalten und denjenigen, die über das Festnetz erreichbar sind, bestehen, kann ebenfalls anhand der Untersuchung von Schneiderat und Schlinzig (2011) dargestellt werden. Im Rahmen der CELLA 2-Erhebung wurde beobachtet, dass Mobile-Onlys vorwiegend alleinstehende männliche Personen jüngeren bis mittleren Alters sind. Zusätzlich verfügen diese über ein niedriges Einkommen und haben einen mittleren Bildungsstand (vgl. Schneiderat und Schlinzig (2011), S. 132.). Ähnliche Befunde weisen auch Graeske und Kunz (2009) bei der Untersuchung der Ergebnisse der Vorgängerstudie CELLA 1 aus (vgl. Graeske und Kunz (2009), S. 59f.). Diese Befunde decken sich zum größten Teil mit jenen der amerikanischen Umfrageforschung (vgl. Link et al. (2007), S. 827f.; Blumberg et al. (2008), S. 64f.). Der f2f-Bus des Jahres 2012 bestätigt diese Resultate und ergänzt, dass Mobile-Onlys vermehrt in den neuen Bundesgebieten und Berlin vorzufinden sind (Meier et al. (2013), S. 134f.).

Aufgrund der unterschiedlichen soziodemografischen Merkmale der Mobile-Only-Haushalte im Vergleich zu Haushalten, die über das Festnetz erreichbar sind, ergeben sich daher mehrere Implikationen für telefonische Erhebungen. So kann angenommen werden, dass diese eine andere Antwortneigung und ein anderes Teilnahmeverhalten an den Tag legen. Wichtiger jedoch ist, dass gerade diese Haushalte sich in ihrer Einstellung zu bestimmten Themengebieten sowie in ihrem Verhalten deutlich von denjenigen unterscheiden können, die über eine "gewöhnliche" Festnetzstichprobe erreicht werden. Der Ausschluss gerade dieser Subpopulation durch die Beschränkung auf eine reine Festnetzstichprobe kann daher verzerrend auf die Schätzung von

bestimmten Merkmalen in der Bevölkerung oder von Prävalenzen wirken (vgl. M. Häder (2009a), S. 16.; Heckel und Wiese (2011), S. 113.; Von Der Lippe et al. (2011), S. 105., Sand (2015a), S. 135.).

Bezüglich der Mobile-Only-Problematik beschreiben S. Häder et al. (2009a), dass sich bereits zum Untersuchungszeitpunkt die Tendenz abzeichne, dass die alleinige Verwendung von Festnetzstichproben als unzureichend betrachtet werden kann, um der Zielpopulation der deutschen Privathaushalte zu entsprechen (vgl. S. Häder et al. (2009a), S. 22.). In einer früheren Untersuchung von Fuchs im Jahr 2002 wird ebenso gezeigt, dass sich bereits Probleme für die Schätzung von Parametern basierend auf einer reinen Festnetzstichprobe ergeben können, obgleich der Anteil der Mobile-Only-Haushalte zum damaligen Zeitpunkt lediglich 5,8 % betrug (vgl. Fuchs (2002), S.122.). Wenn sich Mobile-Only-Haushalte anhand ihrer soziodemografischen Charakteristika von denjenigen Haushalten unterscheiden, die über einen Festnetzanschluss erreicht werden können, besteht die Gefahr verzerrter Schätzergebnisse aufgrund von *Undercoverage* (vgl. Sand (2015a), S. 134.).

Für eine lineare Statistik kann nach Groves (1990) das Ausmaß der Verzerrung aufgrund dieser *Undercoverage* als das Produkt des Anteils der Bevölkerung, der aufgrund einer reinen Festnetzstichprobe nicht erfasst wird und der Differenz im interessierenden Merkmal zwischen dem Schätzwert der Festnetzstichprobe und des Schätzwertes für die Subpopulation, die nicht Teil des Auswahlrahmens ist, beschrieben werden (Groves (1990), S. 229).⁸

Die Verzerrung infolge des Auslassens der Population der Mobile-Only-Haushalte kann deshalb in Abhängigkeit des untersuchten Merkmals durchaus unterschiedlich ausfallen. Somit ist es nicht verwunderlich, dass selbst in jüngeren Untersuchungen unterschiedliche Ergebnisse und Empfehlungen bezüglich der Mobile-Only-Problematik bestehen. So sehen bspw. Hunsicker und Schroth (2014) nach ihrer Untersuchung der Ergebnisse des Politbarometers, dass es keine Notwendigkeit gibt, Haushalte, die ausschließlich über Mobilfunk erreichbar sind, zusätzlich zu berücksichtigen. Begründet wird dies mit der hohen Datenqualität, die Festnetzstichproben i. d. R. aufweisen sowie den bestehenden Möglichkeiten zur Schichtung (vgl. Hunsicker und Schroth (2014), S. 7 & S. 23.).

Im Gegensatz hierzu warnen bspw. Gabler und S. Häder (2009a) vor dem Ausschluss gerade dieser Haushalte aufgrund der zuvor geschilderten Unterschiede zwischen reinen Mobilfunknutzern und denjenigen, die auch über das Festnetz erreichbar sind. Das Auslassen dieser

8 Demnach kann die Verzerrung einer linearen Statistik, wie bspw. einem Mittelwert (der Zielpopulation) ($\bar{Y}$) einer interessierenden Variable y , wie folgt dargestellt werden:

$$\bar{Y}^F - \bar{Y} = \frac{N - N^F}{N} (\bar{Y}^F - \bar{Y}^{C \setminus F}).$$

N entspricht dabei der Anzahl der Einheiten der Zielpopulation, N^F der Anzahl der Einheiten der Population, die über das Festnetz erreicht werden können, $\bar{y}^F$ ist der Mittelwert der Festnetzstichprobe und $\bar{y}^{C \setminus F}$ der Mittelwert der (nicht befragten) Mobile-Onlys. $\bar{y}^C$ würde dem Mittelwert aller Zielpersonen der Mobilfunkstichprobe entsprechen.

Subpopulation aufgrund ihrer abweichenden Altersstruktur und ihres Mobilitätsverhaltens kann systematische Verzerrungen in Schätzwerten bspw. für die bundesdeutsche Bevölkerung hervorrufen (vgl. Gabler und S. Häder (2009a), S. 240.). Zu vergleichbaren Ergebnissen im deutschen Kontext kommen Heckel und Wiese (2011) oder Meier et al. (2013) (vgl. Heckel und Wiese (2011), S. 113.; Meier et al. (2013), S. 130f.). Schneiderat und Schlinzig (2011) zeigen anhand der Untersuchung der Persönlichkeitsmerkmale, die durch die *Big Five* (vgl. Rammstedt et al. (2017)) gemessen wurden, dass sich Mobile-Onlys in deren Ausprägungen deutlich von Festnetzhaushalten unterscheiden. So sind Mobile-Onlys bspw. (durchschnittlich) eher pragmatischer und extrovertierter als Personen, die ausschließlich über das Festnetz erhoben werden können (vgl. Schneiderat und Schlinzig (2011), S. 133f.)

Auch in der amerikanischen Literatur finden sich viele Beispiele, anhand derer sich die negativen Effekte des Auslassens der Mobile-Only-Haushalte auf die Schätzung von interessierenden Parametern darlegen lassen. So stellen mehrere Untersuchungen, die im Rahmen der Veröffentlichungen der *American Association for Public Opinion Research (AAPOR) Cell Phone Task Force* (2010) hervorgebracht wurden, heraus, dass Mobile-Only-Haushalte in den USA mit Deutschland vergleichbare soziodemografische Charakteristika aufweisen (AAPOR Cell Phone Task Force (2010), S. 19f.). Gemäß Wolter et al. (2015) ist der Anteil der Mobile-Only-Haushalte in den USA aber deutlich höher als im deutschen Vergleich. So berichten die Autoren, dass dieser bereits 2010 etwa 20% bis 25% aller amerikanischen Haushalte ausmacht. Zum Zeitpunkt ihrer Untersuchung schätzen die Autoren den Anteil auf mehr als 30% (vgl. Wolter et al. (2015), S. 400.).

Unterschiedlich jedoch ist der höhere Anteil von Personen mit Migrationshintergrund der Mobile-Only-Population in den USA. Der Untersuchung von Schneiderat und Schlinzig (2011) zufolge trifft dies für die deutsche Mobile-Only-Population nicht zu. Ein höherer Anteil von Personen mit Migrationshintergrund ist in Deutschland eher unter Landline-Only-Haushalten zu finden (vgl. Schneiderat und Schlinzig (2011), S. 132f.).

Eine weitläufige Untersuchung der Unterschiede zwischen der Mobile-Only-Population in den USA und denjenigen, die auch über das Festnetz erreichbar sind, wurde 2010 von Christian et al. durchgeführt. Die Autoren untersuchten insgesamt elf Erhebungen, innerhalb derer auch Mobile-Only-Haushalte in der Stichprobenziehung Berücksichtigung fanden. Sie verglichen insgesamt 72 Items, die u. a. Fragen zur politischen Einstellung, religiösen und sozialen Werte sowie technologiebasierte Fragestellungen beinhalteten. Das Ergebnis der Untersuchung zeigt, dass die Unterschiede zwischen den ausschließlichen Mobilfunknutzern und Festnetzhaushalten oftmals zwar gering sind, aber in die erwartete Richtung gehen. So wurden u. a. Unterschiede bezüglich politischer Einstellung, der Nutzung digitaler Medien sowie religiöser und sozialer Werte festgestellt. Darüber hinaus halten die Autoren fest, dass, auch wenn diese Unterschiede in den meisten Items lediglich gering sind, ein Undercoverage-Bias in allgemeinen Bevölker-

ungsumfragen ohne Berücksichtigung der reinen Mobilfunknutzer regelmäßig auftritt. Selbst eine geringe Verzerrung in solchen Erhebungen kann hierbei einen weitgreifenden Einfluss auf politische und soziale Implikationen haben (vgl. Christian et al. (2010), S. 2ff.). Ein solches Beispiel stellt die Untersuchung des *National Health Interview Surveys (NHIS)* von Blumberg und Luke (2012) dar. Die Autoren zeigen, dass Mobile-Only-Haushalte eine deutlich höhere Prävalenz für gesundheitsgefährdendes Verhalten (bspw. hoher Alkoholkonsum und Rauchen) haben als Haushalte, die über das Festnetz erreicht werden können. Des Weiteren finden sich bei diesen deutlich häufiger bestimmte Krankheitsbilder wie bspw. das Vorkommen von HIV und psychologische Probleme. Der Ausschluss dieser Population würde somit zu einer gravierenden Unterschätzung gerade dieser Prävalenzen führen (vgl. Blumberg und Luke (2012), S. 3).

Das wohl wichtigste Argument in der Diskussion über die Inklusion der Mobile-Only-Haushalte in Telefonbefragungen und dem damit verbundenen Effekt auf die Schätzung von Parametern (bei allgemeinen Bevölkerungsumfragen) äußern Peytchev et al. (2009). Ihnen zufolge ist es durchaus möglich, dass einige Punktschätzungen durch die Exklusion der Mobile-Only-Haushalte nicht oder nur in einem geringen Umfang betroffen sein können. Allerdings kann die Verzerrung sich für verschiedene Variablen unterschiedlich stark auswirken. Weiterhin muss der Ausschluss der Mobile-Only-Haushalte nicht zwangsläufig zu einer verzerrten Punktschätzung führen. Es kann aber dennoch inferenziell-statistische Aussagen negativ beeinflussen. So erläutern die Autoren, dass durch das Auslassen gerade dieser Population die Varianzschätzung verzerrt wird. Infolgedessen kann es zu einer mangelhaften Präzision von Schätzwerten und zu fehlerhaften Signifikanztests kommen (vgl. Peytchev et al. (2009), S. 288.).

2.3.3 Dual-Frame-Erhebungen zur Berücksichtigung der Only-Haushalte

Aufgrund des Potenzials, verzerrte Schätzergebnisse zu erzeugen, wird daher in der deutsch- und englischsprachigen Literatur angeraten, die Mobile-Only-Population bei Telefonerhebungen mit zu berücksichtigen. Dies sollte gewöhnlich durch die Verwendung von Dual-Frame-Erhebungen, bestehend aus einer unabhängigen Festnetz- und Mobilfunkstichprobe, erfolgen.

Bereits zu Beginn der 1960er Jahre wurde die Möglichkeit von Dual-Frame-Erhebungen in der amerikanischen Literatur diskutiert. Einer der ersten Vertreter in diesem Bereich war Hartley (1962, 1974). Allerdings wurde in diesem Kontext die Dual-Frame-Erhebung hauptsächlich unter dem Aspekt der Kostenersparnis und weniger unter der Berücksichtigung von Undercoverage betrachtet. Hartley zufolge könnte ein solcher Ansatz eine Genauigkeit aufweisen, die mit einer Erhebung, die auf einem einzelnen, vollständigen Auswahlrahmen basiert, vergleichbar ist. Darüber hinaus kann ein Dual-Frame-Ansatz zu einer erheblichen Kostensenkung beitragen. Der Autor legt dies anhand eines Beispiels dar, bei dem ein vollständiger und ein unvollständiger Rahmen besteht. Letzterer ist in diesem Beispiel jedoch mit geringeren Erhebungskosten verbun-

den. In einem solchen Szenario kann Hartley zufolge ein Dual-Frame-Ansatz der Kostenersparnis dienen, ohne die Genauigkeit der Schätzung zu beeinträchtigen (vgl. Hartley (1962); vgl. S. L. Lohr und J. N. K. Rao (2000), S. 271.). Gerade bezogen auf die Schätzung in solchen Ansätzen wurde dieses Vorgehen bspw. durch Burmeister (1972) weiterentwickelt.

Erste Untersuchungen im Bereich von Dual-Frame-Erhebungen bei Telefonerhebungen fokussierten sich auf die Verbesserung von Responseraten. Traugott et al. (1987) beschreiben bspw. eine erste Erhebung, bei der ein klassisches RDD-Verfahren mit einer durch kommerzielle Anbieter vertriebenen Listenauswahl kombiniert wurde. In dieser Untersuchung wird demnach der Auswahlrahmen für RDD-Stichproben, der, wie in Abschnitt 2.2 dargestellt, alle potenziellen Rufnummern enthalten sollte, mit einer unvollständigen Listenauswahl kombiniert. Ein damit verbundener Vorzug stellt die Kostenreduktion im Rahmen der Erhebung dar. Die Gesamtkosten einer Erhebung werden dadurch reduziert, dass einerseits der Anteil nicht-geschalteter Rufnummern in einer Listenauswahl geringer ist und andererseits ein höherer Anteil von Rufnummern zu Personenhaushalten führt. Zur Steigerung der Responserate wurde in dieser Untersuchung überdies noch ein personalisiertes Ankündigungsschreiben an die Haushalte der Listenauswahl gesendet. Für die Auswahl der RDD-Stichprobe waren lediglich *kalte Kontakte* ohne eine vorherige Ankündigung möglich. Die Autoren stellen fest, dass ein solches Vorgehen die Responserate durchaus steigern kann, während eine Kostenreduktion durch weniger Kontaktversuche zu nicht-geschalteten Rufnummern erfolgt (vgl. Traugott et al. (1987), S. 524f.; S. 530ff.).

Eine für die Mobile-Only-Problematik interessantere Untersuchung geht auf Haines und Pollock (1998) zurück. In ihrer Untersuchung wird die Verwendung zweier unvollständiger Auswahlrahmen beschrieben: einer Listenauswahl und einer Flächenstichprobe zur Schätzung von Tierbeständen. Im Gegensatz zu vorherigen Untersuchungen war dabei nicht die eine Auswahlgrundlage Teilmenge des anderen, vollständigen Auswahlrahmens. Vielmehr konnte die Zielpopulation ausschließlich durch die Kombination beider Auswahlgrundlagen abgedeckt werden, da beide für sich unvollständig gewesen sind. Analog zur derzeitigen Situation im Rahmen von Telefonstichproben ergeben sich dabei drei Teilpopulationen. Zwei Bereiche der Zielpopulation (*Domains*) können ausschließlich über eine bestimmte Stichprobe in die Erhebung gelangen (*Onlys*). Jedoch gibt es ebenso Einheiten, die über beide Stichproben in die Erhebung gelangen können. Zur Schätzung der Tierbestände wählten die Autoren einen sog. *Screening Ansatz*, bei dem einmal eine Stichprobe aus der bereits bestehenden Liste gezogen wurde, während die Flächenstichprobe zusätzlich nach Einheiten screente, die nicht Bestandteil der bereits existierenden Liste sind (vgl. Haines und Pollock (1998), S. 247f.).

Ein weiterer Vorzug von Dual-Frame-Ansätzen, der oftmals in der Literatur hervorgehoben wird, besteht in der besseren Erreichbarkeit jener Personen, die zwar über einen Festnetzanschluss verfügen, allerdings vornehmlich ihren Mobilfunkanschluss nutzen (vgl. Blumberg und Luke (2012), S. 3f.; vgl. Sand (2015a), S.144ff.). Boyle et al. (2009) erläutern an ihrer Untersuchung

des NHIS, dass diese Mobile-Mainly-Population unter Umständen Verzerrungen der geschätzten Parameter aufgrund von Nonresponse hervorrufen können, falls diese Parameter nicht entsprechend Berücksichtigung fänden. Als Mobile-Mainlys sind dabei diejenigen Haushalte zu verstehen, die zwar grundsätzlich über beide Telekommunikationsmedien verfügen⁹, jedoch aus unterschiedlichen Gründen (bspw. ihrer Mobilität) hauptsächlich ihr Mobilfunkgerät nutzen und daher einfacher über dieses zu erreichen sind. Das Ergebnis ihrer Untersuchung zeigt, dass, auch wenn der Anteil dieser Mainly-Population zum damaligen Zeitpunkt noch gering ausfiel, eine Nichtberücksichtigung verzerrend wirken kann (vgl. Boyle et al. (2009)). Dual-Frame-Erhebungen können demnach nicht nur der Berücksichtigung der Mobile-Only-, sondern auch der Mobile-Mainly-Haushalte dienen. Fuchs (2002) stellt ebenfalls fest, dass durch die Einbeziehung von Mobilfunkteilnehmern evtl. auch Haushalte besser erreicht werden können, die aufgrund ihrer Mobilität sonst nur unter großem Aufwand in eine Festnetzstichprobe gelangen. Gerade in urbanen Regionen ist es dem Autor zufolge sinnvoll, eine Mobilfunkstichprobe in die Erhebung mit einzubeziehen (vgl. Fuchs (2002), S.122f.).

Hoops und Michael (2015) zeigen in ihrer Untersuchung zum Wahlverhalten und zu bestimmten soziodemografischen Charakteristika, dass Mobile-Mainly-Haushalte in vielen Eigenschaften mit Mobile-Only-Haushalten vergleichbar sind. Dennoch bestehen deutliche Unterschiede bspw. bezüglich des Einkommens oder des intendierten Wahlverhaltens. Die Autoren halten fest, dass sowohl Mobile-Only- als auch Mobile-Mainly-Haushalte erheblich von der restlichen Bevölkerung abweichen, was eine Verwendung von Mobilfunkstichproben unumgänglich macht (vgl. Hoops und Michael (2015), S. 2ff.).

Bis zu diesem Zeitpunkt wenig untersucht ist jedoch der Einfluss gerade dieser Mobile-Mainly-Haushalte auf die Schätzung von Parametern. Zudem bleibt die Frage offen, ob sich im Spektrum der Dual-User nicht ebenfalls Haushalte finden lassen, die überwiegend ihren Festnetzanschluss verwenden und entsprechend besser darüber zu erreichen sind (Landline-Mainly). Mit einer solchen *dominanten* Nutzung eines bestimmten Telekommunikationsmediums sollte außerdem eine erhöhte Erreichbarkeit verbunden sein. Ob und wie sich diese gesteigerte Erreichbarkeit auf die Auswahlwahrscheinlichkeit der Dual-User auswirkt, ist ebenfalls wenig erforscht. Erste Untersuchungen zu diesen Fragestellungen wurden bspw. von Sand (2015a) anhand der CELLA-Daten unternommen. Eine solche erhöhte Auswahlwahrscheinlichkeit aufgrund der dominanten Verwendung eines bestimmten Telekommunikationsmediums sollte dann im Rahmen der Schätzung auch im Gewichtungsansatz der Daten Berücksichtigung finden. Gerade dies stellt den Gegenstand dieser Arbeit dar. Daher wird im Folgenden erst die Erhebung über Dual-Frame-Ansätze in Deutschland dargestellt, um im anschließenden Kapitel allgemein die Gewichtung bei solchen Erhebungen zu beschreiben. Die Kapitel 4 bis 8 beschäftigen sich dann mit dem Umgang der Mainly-Populationen in Dual-Frame-Erhebungen sowie mit den damit verbundenen

9 Somit sind Mobile-Mainly-Haushalte grundsätzlich eine Teilpopulation der Dual-User.

Anforderungen an die Gewichtungsprozeden und der Schätzung von relevanten Parametern.

2.4 Dual-Frame-Erhebungen in Deutschland

Spätestens seit den CELLA-Erhebungen, auf die in Abschnitt 2.4.4 näher eingegangen wird, finden, zumindest für bundesweite Telefonbefragungen, Dual-Frame-Erhebungen häufige Anwendung, um der Mobile-Only-Problematik zu begegnen. Zur Realisierung solcher Erhebungen bedarf es jedoch einer konkreten Auswahlgrundlage zur Ziehung von Mobilfunkstichproben. Aufgrund des Eintragsverhalten von Mobilfunknutzern und der Gestaltung von Mobilfunknummern lässt sich diese jedoch nicht uneingeschränkt durch das zuvor dargestellte Gabler-Häder-Verfahren umsetzen. Daher beschreibt Abschnitt 2.4.1 kurz die Entwicklung hin zu einer angemessenen Auswahlgrundlage und die derzeitige Realisierung von Mobilfunkstichproben. Anschließend werden die Überlegungen zur Umsetzung einer solchen Dual-Frame-Erhebung in Abschnitt 2.4.2 dargestellt.

Da sich das Telekommunikationsverhalten über ein Mobilfunkgerät durchaus von demjenigen über einen Festnetzanschluss unterscheiden kann und darüber hinaus durch die Hinzunahme einer Mobilfunkstichprobe neue Herausforderungen an die technische Umsetzung und die Befragung gestellt werden können, werden diese darauf folgend diskutiert. Des Weiteren kann sich die Befragungssituation über ein mobiles Endgerät grundsätzlich von einer Befragung über Festnetzanschlüsse unterscheiden, wodurch weitere Anforderungen an die Durchführung einer Befragung gestellt werden. Eine Diskussion dieser und weiterer methodischer Problemstellungen erfolgt daher in Abschnitt 2.4.3. Den Abschluss bildet die Vorstellung der beiden Erhebungen CELLA und Influenza 2014 als Beispiel für Dual-Frame-Erhebungen in Deutschland. Die beiden Erhebungen sind darüber hinaus Datengrundlage für die späteren Untersuchungen der vorliegenden Arbeit.

2.4.1 Entwicklung einer Auswahlgrundlage für Mobilfunkerhebungen

In Deutschland kamen erste Überlegungen zur Befragung über das Mobilfunknetz zu Beginn der 2000er Jahre auf. Neben der bereits erwähnten Untersuchung über die Auswirkung der Mobiltelefonie von M. Häder (2000) sowie erster Untersuchungen bspw. von Fuchs (2002) wurden Vorschläge seitens des ADMs und durch GESIS (ehemals *Zentrum für Umfragen, Methoden und Analysen (ZUMA)*) hervorgebracht. Grund schildert bereits im Jahr 2002 den Bedarf der Integration von Mobilfunkteilnehmern bei der Befragung über das ADM-Telefonstichprobensystem und der dafür notwendigen Schaffung einer Auswahlgrundlage (vgl. Grund (2002), S. 115ff.). Der ADM selbst berichtet, dass dieser seit 2005 eine solche Auswahlgrundlage aufbaut. Die methodische Verknüpfung zu dem System für Festnetzstichproben wurde jedoch erst 2011 mit den Unter-

suchungen des Forschungsprojektes "Dual-Frame-Ansätze" (vgl. Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute ADM (2012), S. 4.) abgeschlossen. Eine erste Auswahlgrundlage zur Generierung von Mobilfunkstichproben schaffte der ADM allerdings bereits 2001 anhand der eingetragenen Mobilfunknummern im Telefonverzeichnis. Da diese jedoch sehr selektiv scheint und nur ein geringer Anteil an Mobilfunknummern tatsächlich eingetragen ist, wurde ab 2005 ein gestütztes Verfahren verwendet. Dieses basiert auf einem RDD-Ansatz, der anhand externer Quellen erweitert/verändert wird, um bspw. Rufnummernbereiche auszuschließen, die nicht vergeben sind oder rein geschäftlich genutzt werden (vgl. Glemser (2007), S. 16.; vgl. Heckel (2007), S. 26f.).

Seitens ZUMA/GESIS wurde beginnend mit den Vorstudien des gemeinsamen Kooperationsprojektes CELLA (siehe Abschnitt 2.4.4), das mit der Technischen Universität Dresden verfolgt wurde, eine Grundlage zur Ziehung von Mobilfunkstichproben geschaffen. Zur Gestaltung eines solchen Auswahlrahmens beschreiben Schneiderat und S. Häder (2007) diverse Schwierigkeiten, die u. a. auf das Eintragsverhalten der Mobilfunknutzer in ein Telefonverzeichnis zurückzuführen sind. Zum Untersuchungszeitpunkt waren dabei lediglich 2,2 Mio. der ca. 74 Mio. Mobilfunkanschlüsse in einem Telefonverzeichnis geführt (vgl. Schneiderat und S. Häder (2007), S. 64.). Auch spätere Untersuchungen, wie bspw. durch Heckel und Wiese (2011) oder Sand (2014), zeigen, dass zwar die Anzahl der Mobilfunkanschlüsse in den darauf folgenden Jahren deutlich gestiegen ist,¹⁰ die Anzahl der eingetragenen Mobilfunknummern jedoch vergleichsweise konstant bei einem Wert von ca. 2 Mio. bleibt (vgl. Heckel und Wiese (2011), S. 114.; vgl. Sand (2014), S. 13.; vgl. Bundesnetzagentur (2015), S. 38.). Daher ist ein Verfahren, das eine Auswahlgrundlage schafft, die analog zu Festnetzstichproben in Deutschland nach der Methode von Gabler und Häder vorgeht, nicht umsetzbar. Eine Bildung von Rufnummernblöcken anhand der eingetragenen Mobilfunkanschlüsse könnte erhebliche Verzerrungen in der Schätzung interessierender Parameter aufgrund von Undercoverage mit sich ziehen. Überdies ist anzunehmen, dass sich, ebenso wie im Falle von Festnetzstichproben, Einheiten, die nicht in Telefonverzeichnissen geführt werden, in ihren soziodemografischen Merkmalen stark von anderen Einheiten unterscheiden.¹¹ Im Gegensatz zu Festnetzstichproben kann jedoch aufgrund der geringen Anzahl der gelisteten Anschlüsse keine sinnvolle Blockbildung gewährleistet werden, die sowohl eine akzeptable Hitrate als auch eine Abdeckung aller (oder der meisten) geschalteten Anschlüsse verspricht.

Allerdings ist die Gestaltung der Mobilfunkrufnummern deutlich übersichtlicher als diejenige der Festnetzzufnummern. Die 17 Mobilfunkvorwahlen der derzeit drei Netzbetreiber für Mobilfunk¹² sind vierstellig, während die tatsächliche Rufnummer in Abhängigkeit der Vorwahl zwischen sieben und acht Stellen aufweist. Hierdurch wäre ein reines RDD-Verfahren prinzipiell

10 Für das Jahr 2015 sind es ca. 112 Mio. Mobilfunkanschlüsse bzw. eine Marktpenetration von ca. 138 %.

11 Es kann bspw. angenommen werden, dass viele der in Telefonverzeichnissen geführten Mobilfunknummern geschäftliche Anschlüsse sind.

12 Seit der Übernahme des Betreibers *Eplus* durch *Telefonica O2*.

möglich.

Das von GESIS und dem ADM genutzte Verfahren, bei dem nicht-geschaltete Rufnummernbereiche entfernt werden, stellt hierzu eine Erweiterung dar. Dazu werden 10.000er-, und auch 1.000er-Blöcke gebildet, die ähnlich der Methode von Gabler und Häder, zur Stichprobenziehung herangezogen werden. Ein Auswahlrahmen, der alle potenziellen Rufnummern abdeckt, würde demnach ca. 448 Mio. Rufnummern beinhalten und eine geringe Hitrate aufweisen. Daher werden externe Quellen, wie bspw. die Gajek-Liste (vgl. Gajek (2015).) oder die Veröffentlichungen der Bundesnetzagentur (vgl. Bundesnetzagentur (2017b).) sowie die eingetragenen Rufnummern des Telefonverzeichnisses, herangezogen, um diesen Auswahlrahmen einzuschränken (vgl. S. Häder et al. (2009a), S. 35ff.; vgl. Heckel et al. (2014), S. 147ff.). GESIS bietet eine solche Stichprobenziehung seit 2008 an. Für das Jahr 2016 betrug die Anzahl potentieller Rufnummern, die aus dem Auswahlrahmen für Mobilfunkstichproben erzeugt werden können, ca. 329 Mio. Rufnummern.

2.4.2 Umsetzung von Dual-Frame-Erhebungen

Zur Durchführung von Dual-Frame-Erhebungen weiterhin relevant ist der Umgang mit den einzelnen Erhebungseinheiten und deren Erreichbarkeit über die jeweiligen Auswahlrahmen. Wie bereits in Abschnitt 2.3 dargestellt, unterscheidet man allgemein zwischen *Dual-Usern*, also Einheiten, die sowohl über das Festnetz, als auch über den Mobilfunk erreichbar sind, und *Landline-* bzw. *Mobile-Onlys*, also Einheiten, die ausschließlich über ein Medium erreicht werden können. Für die Only-Populationen ist der Umgang vergleichsweise einfach, da diese lediglich über eine bestimmte Auswahlgrundlage berücksichtigt werden können. In Verbindung mit Haushalten, die über beide Medien erreichbar sind, ergibt sich jedoch die Schwierigkeit, wie beide Stichproben kombiniert und behandelt werden können, um eine vollständige Abdeckung der Zielpopulation zu erhalten.

S. Häder et al. (2009a) beschreiben hierbei in Anlehnung an die von der *American Association of Public Opinion Research* (AAPOR) veröffentlichten *“Considerations and Guidelines”* zwei unterschiedliche Stichprobenansätze zur Kombination beider Stichproben. In Telefonbefragungen, die sich auf Dual-Frame-Erhebungen stützen, besteht einmal die Möglichkeit einen *Screening-Ansatz* oder einen *Overlap-Ansatz* (oder auch *Nonscreening-Ansatz*) zu wählen. Beide Ansätze haben dabei Einfluss auf die *Designgewichtung*, die in Abschnitt 3 näher beschrieben wird.

Bei einem Screening-Ansatz finden zwar beide Stichproben Verwendung, jedoch wird innerhalb einer der beiden Stichproben ausschließlich nach Only-Haushalten gescreent, während die restlichen Einheiten (Only-Haushalte der anderen Auswahlgrundlage und Dual-User) über die andere Stichprobe in die Erhebung gelangen. Im Rahmen einer Dual-Frame-Telefonstichprobe wäre es demnach denkbar, bspw. in der Mobilfunkstichprobe nach Mobile-Only-Haushalten zu

screenen und nur diese dann über das Medium Mobilfunk zu befragen, während Landline-Onlys und Dual-User ausschließlich über das Medium Festnetz in die Erhebung gelangen. Ein Overlap-Ansatz hingegen verwendet beide Stichproben und führt eine Befragung anhand des Mediums durch, über das eine Einheit in die Erhebung gelangt ist. Bei einer solchen Vorgehensweise muss jedoch die erhöhte Auswahlwahrscheinlichkeit der Dual-User berücksichtigt werden, die sich dadurch begründet, dass diese über beide Stichproben in die Erhebung gelangen können (siehe dazu: Abschnitt 3) (vgl. S. Häder et al. (2009a), S. 25.; vgl. AAPOR Cell Phone Task Force (2010), S. 30f.).

Kennedy (2007) erörtert, dass ein Screening-Ansatz die Vorteile der erhöhten Coverage durch die Verwendung beider Auswahlrahmen nicht negiert, da lediglich Dual-User bei einem solchen Ansatz ausgescreent würden, die ohnehin auch über eine Auswahlwahrscheinlichkeit innerhalb der Festnetzstichprobe verfügen. Wenn Dual-User aus der Festnetzstichprobe ähnlich bzw. vergleichbar mit denjenigen der Mobilfunkstichprobe sind, so besteht bei einem solchen Ansatz die Möglichkeit, die Kosten zu senken, ohne eine Reduktion der Datenqualität zu riskieren. Eine solche Erhebung kann dann als eine Art geschichtete Stichprobe betrachtet werden, bei der die jeweiligen *Nutzungskategorien* der Telekommunikationsmedien (Landline-Only, Dual-User und Mobile-Only) als einzelne Schichten betrachtet werden. Allerdings hebt die Autorin hervor, dass ein solcher Ansatz das Potenzial birgt, eine stärkere Verzerrung aufgrund von Nonresponse mit sich zu bringen, wenn sich das Antwortverhalten der Dual-User, die über Mobilfunk erreicht werden, von demjenigen der Dual-User, die über Festnetz an einer Erhebung teilnehmen, abweicht (siehe Abschnitt 5). Weiterhin besteht die Möglichkeit, dass sich der Messfehler nach Medium unterscheidet (vgl. Kennedy (2007), S. 750ff.).

Ein Unterschied bezüglich des Responseverhaltens wird in einer Untersuchung des *JPSM – Practicum Survey 2004* von Brick et al. (2006) festgehalten. Allerdings führt der Autor die Hauptkomponente dieses Fehlers auf die mangelnde Erreichbarkeit in Abhängigkeit der Auswahlgrundlage zurück. Hierbei hält der Autor fest, dass die Erreichbarkeit von Mobile-Only-Haushalten über das Mobilfunknetz durchaus als günstig eingeschätzt werden kann, während diese für Dual-User in einigen Fällen schwieriger einzuschätzen ist. Umgekehrt beschreibt der Autor eine gute Erreichbarkeit der Dual-User über das Festnetz. Aufgrund dessen sieht Brick et al. (2006) keine Beschränkung in der Durchführung des Screening-Ansatzes (vgl. Brick et al. (2006), S. 785ff.). Ähnliche Ergebnisse berichten auch Lambert et al. (2010). Diese halten fest, dass sich auch bei denjenigen Dual-User-Haushalten, die vorwiegend ihr Mobilfunkgerät verwenden, eine Erreichbarkeit durch die Steuerung der Anzahl der Anwahlversuche ermöglichen lässt (vgl. Lambert et al. (2010), S. 7.).

Grundlage zur Entscheidung für einen solchen Ansatz sind nach Kennedy (2007) die Varianz eines daraus resultierenden Schätzwertes, die durch einen Ansatz entstehende Verzerrung (bspw. aufgrund von Nonresponse) sowie die Kosten der Durchführung. Bezüglich der Verzerrung

der geschätzten Parameter merkt die Autorin an, dass ohne eine Anpassung an das Medium, das zur Durchführung der Befragung eines Dual-Users verwendet wurde, das Ausscreenen derjenigen Haushalte aus der Mobilfunkstichprobe, die über einen Festnetzanschluss verfügen, Verzerrungen der Schätzung mit sich führt (vgl. Kennedy (2007), S. 752 & S. 769.). Darüber hinaus kann ein solcher Ansatz Peytchev und Neely (2013) zufolge zu einer erheblichen Steigerung der Kosten führen bzw. bei gegebenem Budget die Stichprobengröße negativ beeinflussen und eine höhere Varianz mit sich bringen. Begründet wird dies damit, dass Einheiten zu einem Zeitpunkt ausgescreent werden, an dem anzunehmen ist, dass sie an einer Erhebung teilnehmen würden. Demnach ist ein Screening-Ansatz einem Overlap-Ansatz unterlegen, wenn die Interviewkosten im Mobilfunkrahmen sich nicht stark von denjenigen des Festnetzes unterscheiden oder das erfolgreiche Kontaktieren einer Einheit mehr Zeit als das tatsächliche Interview einnimmt (vgl. Peytchev und Neely (2013), S. 3.).

Weiterhin führen S. Häder et al. (2009a) bezogen auf die deutsche Situation an, dass aufgrund des geringen Anteils von Mobile-Only-Haushalten in der bundesdeutschen Bevölkerung mit einer geringen Hitrate zu rechnen ist, wodurch der Aufwand der gesamten Dual-Frame-Erhebung in einem Screening-Ansatz erheblich steigt. Darüber hinaus befürchten die Autoren eine eingeschränkte Erreichbarkeit derjenigen Dual-User, welche nur eines ihrer Telekommunikationsmedien tatsächlich nutzen (vgl. S. Häder et al. (2009a), S. 27.). Aufgrund dessen wird im deutschen Kontext i. d. R. die Durchführung eines Overlap-Ansatzes bevorzugt. Da Dual-User bei einem solchen Ansatz eine erhöhte Wahrscheinlichkeit aufweisen, in der Stichprobe enthalten zu sein und diese entweder über die Festnetz- oder die Mobilfunkstichprobe in die Erhebung gelangen können, ist es notwendig, unabhängig von der verwendeten Stichprobe, den Telefonstatus für beide Auswahlgrundlagen zu erfassen (vgl. AAPOR Cell Phone Task Force (2010), S. 31.). Des Weiteren gilt zu beachten, dass ein solcher Ansatz Gefahr läuft, die Erreichbarkeit derjenigen Dual-User zu überschätzen, die bspw. ihr Mobilfunkgerät nur äußerst selten eingeschaltet haben (vgl. Sand (2015a), S. 136.).

Wichtig für die Durchführung von Dual-Frame-Erhebungen ist darüber hinaus die Auswahl von Erhebungseinheiten in Abhängigkeit der verwendeten Stichprobe. Die Auswahl der Erhebungseinheit bzw. Zielperson erfolgt hierbei in Abhängigkeit des Telekommunikationsmediums unterschiedlich. Bei einer Festnetzstichprobe, bei der Aussagen auf Personenebene getroffen werden sollen, handelt es sich i. d. R. um ein zweistufiges Verfahren, bei dem durch das "Wählen der Rufnummer" erstmals Kontakt mit einem Haushalt¹³ hergestellt wird. Im Anschluss daran erfolgt erst die Ermittlung der tatsächlichen Erhebungsperson, die die Befragung durchführen soll. Begründet wird dies mit der Annahme, dass alle Personen des Haushaltes einen Festnetzanschluss nutzen (vgl. Gabler et al. (2011), S. 150.).¹⁴ Zur Ermittlung der Auswahlwahrscheinlichkeit

¹³ Repräsentiert wird dieser durch die Kontaktperson, die den Anruf entgegen nimmt.

¹⁴ Eine kritische Diskussion dieser Annahme findet sich in Sand und Gabler (2018).

jeder einzelnen Einheit des Haushaltes ist es daher wichtig, die Anzahl der *erhebungsrelevanten* Haushaltsmitglieder zu erfahren. Erhebungsrelevanz lässt sich in diesem Kontext durch die Zugehörigkeit zur Zielpopulation festlegen. Wichtig ist, dass jede der erhebungsrelevanten Personen des Haushaltes eine von Null unterschiedliche (im besten Fall gleiche) Wahrscheinlichkeit hat, in die Stichprobe zu gelangen.

Die Auswahl einer erhebungsrelevanten Einheit kann anhand mehrerer Verfahren vorgenommen werden. Gängig sind entweder ein *Kish-Grid* oder ein *Last (oder Next) Birthday-Verfahren*. Die praktikabelste der beiden Vorgehensweisen ist dabei eines der Birthday-Verfahren. Zur Ermittlung der Zielperson wird dabei diejenige erhebungsrelevante Einheit durch die Kontaktperson ausgewählt, die als letztes oder als nächstes Geburtstag hat (vgl. S. Häder (2000), S. 8f.).

Ein mögliches Problem dieses Verfahrens besteht nach S. Häder et al. (2009a) darin, dass einige Personen der Zielpopulation unter Umständen keine von Null unterschiedliche Auswahlwahrscheinlichkeit haben. Gerade bei sich wiederholenden Erhebungen, bei denen die Befragung immer im gleichen Zeitraum stattfindet, kann ein solches Vorgehen dazu führen, dass bestimmte Personen aufgrund des Datums ihres Geburtstages in Kombination mit der Haushaltszusammensetzung keine Möglichkeit haben, für die Stichprobe ausgewählt zu werden. Daher empfehlen die Autoren das Durchführen einer randomisierten Geburtstagsauswahl, bei der ein zufälliges Datum erzeugt wird und davon ausgehend die erhebungsrelevante Person, die als nächstes oder letztes Geburtstag hat, ausgewählt wird. Bei einem solchen Vorgehen ist den Autoren zufolge eine annähernd gleiche Ziehungswahrscheinlichkeit gegeben (vgl. S. Häder et al. (2009a), S. 34). Das Kish-Verfahren auf der anderen Seite erfüllt im strengen Sinne die Vorgabe, dass alle erhebungsrelevanten Personen des Haushaltes die gleiche Auswahlwahrscheinlichkeit haben. Jedoch sind zur Durchführung dieser Methode Informationen über alle Personen des Haushaltes erforderlich, die von der Kontaktperson eingeholt werden müssen, bevor eine tatsächliche Auswahl der Zielperson erfolgt. Da dieses Vorgehen aufwendig und invasiv ist, wird oftmals berichtet, dass es vermehrt zu Abbrüchen führt, weswegen es seltener angewendet wird (vgl. S. Häder (1994), S. 15f.).

Bei Mobilfunkstichproben hingegen wird i. d. R. von einem einstufigen Verfahren ausgegangen, bei dem die Kontaktperson der Zielperson entspricht. Ein Mobiltelefon wird dabei als personalisierter Gegenstand verstanden, der eindeutig einer Einheit zuzuordnen ist, wodurch sich dieses Vorgehen begründet. Dahin gehend divergieren einige frühere Ansätze, bei denen auch eine Haushaltskomponente bei der Befragung über das Mobilfunknetz inkludiert wurde (vgl. Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute ADM (2012), S. 23.). Jedoch zeigt u. a. die Untersuchung von Fuchs und Busse (2013) anhand des *Experimental Mobile Phone Panels*, dass zwar ein sog. Cell Phone-Sharing, also das Verwenden eines Mobilfunkgerätes durch mehrere Personen, durchaus zu beobachten ist, dass allerdings für einen Großteil aller Mobilfunknummern ein Hauptnutzer ausfindig gemacht werden kann (vgl. Fuchs und Busse

(2013), S. 6ff; vgl. Sand (2015a), S. 137f.).

2.4.3 Methodische Herausforderungen bei Dual-Frame-Erhebungen

Durch die Hinzunahme eines zusätzlichen Telekommunikationsmediums bei Telefonstichproben, dessen Anschlüsse Teil einer eigenständigen Auswahlgrundlage sind, können sich neben den zuvor bereits geschilderten methodischen Schwierigkeiten weitere Probleme im Rahmen der Erhebung ergeben.

Analog zu den Grenzen einer *gewöhnlichen* Festnetzbefragung sind diese zum Teil auf Probleme, die im Zusammenhang mit der Stichprobenziehung stehen, zurückzuführen. Zusätzlich können Schwierigkeiten in der Umsetzung aufgrund technischer Barrieren oder Besonderheiten und Verhaltensweisen, die die Messung über dieses Telekommunikationsmedium beeinträchtigen, eintreten.

Probleme in der technischen Umsetzung solcher Dual-Frame-Erhebungen können bspw. in der Verwendung der Hard- und Software für Telefonstichproben bei der Befragung über das Mobilfunknetz entstehen. Gewöhnlich verwenden größere Umfrageinstitute zur besseren Feldsteuerung sog. *Predictive Dialer*. Diese sind in der Lage, automatisiert Rufnummern anzuwählen, Kontaktergebnisse zu interpretieren und geschaltete Anschlüsse an Interviewer weiterzuleiten. In diesem Kontext beschreiben Heckel et al. (2014), dass im Mobilfunk teilweise andere, vom Festnetz unterschiedliche, Signale verwendet werden, die darüber hinaus auch abhängig vom Netzerkanbieter sein können. Dadurch kann es bei der Nutzung solcher Geräte zu Fehlinterpretationen der Rufnummer und einem Mehraufwand im Rahmen der Interviewdurchführung kommen (vgl. Heckel et al. (2014), S. 165.).

Da es sich bei Mobiltelefonen um einen personalisierten Gegenstand handelt und sich eine Mobilfunkstichprobe grundsätzlich als Personenstichprobe auffassen lässt, können weiterhin Probleme bezüglich des Alters der kontaktierten Person entstehen. Im Rahmen einer Festnetzstichprobe ist es, da ein Festnetzanschluss zuerst einen Haushalt kontaktiert, eher unwahrscheinlich, ausschließlich Minderjährige anzutreffen. Bei Mobilfunkstichproben kann dies jedoch durchaus häufiger zutreffen, da die Kontaktperson der Zielperson entspricht. Je nach Definition der Zielpopulation kann dies zu einer erhöhten Anzahl stichprobenneutraler Ausfälle und damit zu einer sinkenden Hitrate führen. Aber auch wenn Minderjährige Teil der Zielpopulation sind, ist es bei einer solchen Erhebung notwendig, eine entsprechende Einwilligung der Erziehungsberechtigten zu erhalten. Der ADM gibt diesbezüglich in seinen *Richtlinien für die Befragung Minderjähriger* eine Anleitung, ab welchem Alter die Einwilligung der Eltern erforderlich ist. Demnach ist eine solche Einwilligung grundsätzlich für Kinder unter 11 Jahren einzuholen. Bei Minderjährigen zwischen 11 und 13 Jahren obliegt es der Verantwortung des befragenden Institutes. Die Entscheidung über die Einsichtsfähigkeit eines Minderjährigen über

die Tragweite zur Teilnahme an einer Erhebung, liegt in diesem Fall bei der befragenden Instanz (vgl. Arbeitskreis Deutscher Markt und Sozialforschungsinstitute ADM (2016), S. 1).

Einen Einblick in die Verbreitung von Mobilfunkstichproben bietet bspw. die *KIM-Studie* des Medienpädagogischen Forschungsverbundes Südwest.¹⁵ Diese schätzt den Anteil der Sechs- bis 13-Jährigen, die über ein Mobilfunkgerät verfügen auf ca. 46 % (vgl. Feierabend et al. (2015), S. 9.). Aufgrund der technischen Aspekte und der Mobilität eines solchen Anschlusses kann dies zu weiteren Problemen führen (vgl. AAPOR Cell Phone Task Force (2010), S. 46 & S. 86.).

Erhebliche Probleme können auch bei der Verwendung von Mobilfunkstichproben in Erhebungen mit räumlichem Bezug entstehen. Derzeitig existieren 17 Mobilfunkvorwahlen, denen eine Rufnummer von sieben bis acht Stellen folgt. Im Gegensatz zu Festnetzrufnummern besteht jedoch keine Systematik, die es erlaubt, anhand der Kombination aus Mobilfunkvorwahl und Rufnummer einen Anschluss geografisch zuzuordnen. Daher sind Dual-Frame-Erhebungen bei Fragestellungen, die sich nur auf ein bestimmtes Gebiet innerhalb Deutschlands beziehen oder nur Einheiten einer bestimmten Region in ihrer Zielpopulation enthalten, nur schwer durchführbar (vgl. Heckel und Wiese (2011), S. 115.; Fuchs (2007), S. 107.). Versuche, wie bspw. von T. Engel (2013), anhand der Kombination von Vorwahl und Rufnummer Mobilfunkanschlüsse geografisch zu referenzieren, wurden zwar unternommen, führten jedoch nicht zu verwertbaren Ergebnissen (vgl. T. Engel (2013), S. 13ff.; vgl. Sand (2014), S. 32.). Demnach ist eine der einzigen Möglichkeiten derzeitig Mobilfunknummern in geografisch begrenzten Erhebungen zu verwenden, eine Mobilfunkstichprobe für das gesamte Bundesgebiet zu ziehen und nach der Region der Zielpopulation zu screenen.¹⁶ Ein solches Verfahren kann jedoch in Abhängigkeit der Größe des untersuchten Gebietes sehr ineffizient und kostenintensiv sein, da mitunter ein Großteil der erfolgreichen Kontakte ausgescreent wird.

Da ein Mobilfunkgerät im Gegensatz zu einem Festnetzanschluss nicht stationär an einen bestimmten Ort gebunden ist, ergeben sich aufgrund der Erreichbarkeit in unterschiedlichen Situationen und Orten weitere Aspekte, die die Befragungssituation (negativ) beeinflussen können. Die AAPOR (2010) führt diesbezüglich eine Liste an Ursachen auf, die sich aus der Mobilität eines solchen Anschlusses ableiten und einen Einfluss auf die Qualität der Befragung haben können. So kann die erweiterte Erreichbarkeit dazu führen, dass Befragungen in Situationen abgehalten werden, in denen der Befragte nur ungern auf bestimmte, bspw. sensible Fragestellungen, antworten möchte. So könnte zum Beispiel eine Befragung zu sensiblen Themen, im Gegensatz

¹⁵ Zu beachten gilt dabei jedoch, dass es sich hierbei um eine Quotenstichprobe und keine reine Zufallsauswahl handelt (vgl. Feierabend et al. (2015), S. 3f.). Demnach sind die Ergebnisse dieser Studie nur bedingt belastbar.

¹⁶ Hierbei wird davon ausgegangen, dass eine Mobilfunkstichprobe als Grundlage dient. Eine weitere Möglichkeit wäre, die Stichprobenziehung über die Einwohnermeldeämter durchzuführen und nach einem Initialkontakt (persönlich oder postalisch) die Mobilfunknummer zu erfragen. Da der Fokus dieser Arbeit jedoch auf Erhebungen liegt, deren Grundlage Stichproben aus den Auswahlrahmen für Telefonstichproben stammen, wird diese Möglichkeit nicht weiter diskutiert, da die Auswahlwahrscheinlichkeit einer solchen Einheit sich i. d. R. grundsätzlich von Telefonstichproben unterscheidet.

zu Festnetzbefragungen, auch an öffentlichen Orten wie einem Bus stattfinden (vgl. AAPOR Cell Phone Task Force (2010), S. 55.). Untersuchungen von M. Häder (2011) sowie M. Häder und Kühne (2009b) zeigen jedoch, dass auch bei Interviews über das Mobilfunkgerät ein Großteil der Interviews zu Hause durchgeführt wird. Ferner wird festgehalten, dass in beiden Untersuchungen kein negativer Einfluss der Umgebung auf die kognitive Leistungsfähigkeit des Befragten, die Höhe der Verweigerung für bestimmte Fragestellungen oder das Antwortverhalten nach sozialer Erwünschtheit zu bestehen scheint (vgl. M. Häder (2011), S. 253; vgl. M. Häder und Kühne (2009b), S. 195ff.).

Einen Vorzug der technischen Aspekte von Mobilfunkgeräten sehen Lynn und Kaminska (2011) in der Möglichkeit, die Anwesenheit dritter Personen des Haushaltes zu vermeiden, da sich der Respondent einfacher an einen Ort begeben kann, der außerhalb der Hörweite anderer Personen ist (vgl. Lynn und Kaminska (2011), S. 219).

Eine weitere Option, die einen Einfluss auf die durch Häder [2009] dargestellten Ergebnisse haben kann, sind technische Neuerungen und ein anderer Umgang mit Mobilfunkgeräten. So verfügt ein Mobiltelefon i. d. R. über eine Rufnummernerkennung, die evtl. dazu führen kann, dass Anrufe von einer dem Nutzer unbekannten Nummer generell oder in bestimmten Situationen nicht angenommen werden. Ein derartiges Verhalten kann unter Umständen als implizite Verweigerung aufgefasst werden. Da jedoch keine Verweigerung explizit ausgesprochen wird und auch andere Umstände, wie bspw. das Roaming eines Mobilfunkgerätes,¹⁷ sich nicht von diesen trennen lassen, ist eine Abschätzung dieses Effektes nur schwer möglich. Zudem kann nicht genau bestimmt werden, ob es sich bei einem Mailbox-Kontakt um einen aktiven Anschluss oder einen, der nicht mehr genutzt wird, handelt (vgl. S. Häder et al. (2009b), S. 73.).

Lynn und Kaminska (2011) führen weiterhin an, dass Erreichbarkeit in allen Situationen noch weitere Schwierigkeiten in der Erhebung über das Mobilfunknetz mit sich bringen kann. So besteht den Autoren zufolge eine höhere Wahrscheinlichkeit dafür, dass eine kontaktierte Einheit während der Befragung noch andere Tätigkeiten ausübt. Neben der Anwesenheit weiterer Personen kann hierbei die Ablenkung durch diese Tätigkeit(en) die Interviewqualität negativ beeinflussen (vgl. Lynn und Kaminska (2011), S. 219). Außerdem ist im Rahmen dieses Multitaskings besonders zu beachten, dass aus der Ablenkung, die das Interview darstellt, auch eine Gefahr für die interviewte Person resultieren kann. Denkbar wäre in diesem Kontext bspw. das Steuern eines Autos. Eine solche Situation kann daher entweder zur Verweigerung oder der Gefährdung

¹⁷ Roaming bezeichnet allgemein, wenn sich ein Mobilfunkgerät in ein Netz eingewählt hat, das nicht dem eigentlichen Netzwerkanbieter entspricht. Im speziellen wird dabei oftmals ein Zustand beschrieben, bei dem ein Mobilfunkgerät im Ausland in ein fremdes Netz eingewählt ist. Gespräche, bei denen sich ein Mobilfunkgerät im Roaming-Modus befindet, können daher höhere Kosten für den Nutzer verursachen (vgl. Heckel et al. (2014), S. 165.). Seit Juni 2017 sind diese Kosten jedoch zumindest innerhalb der Europäischen Union weitgehend abgeschafft (vgl. Bundesnetzagentur (2017a)). Eine Kostenabwägung zumindest für Personen die sich in einem Mitgliedsland der Europäischen Union befinden, ist daher in den meisten Fällen auszuschließen. Dennoch können weitere Faktoren, wie bspw. eine verminderte Erreichbarkeit bei Urlaubsreisen in diesem Kontext relevant sein.

einer befragten Einheit führen. Daher empfiehlt die AAPOR (2010), den Respondenten beurteilen zu lassen, ob er sich gerade in einer Situation befindet, in der er an der Befragung teilnehmen kann, ohne sich selbst dabei in Gefahr zu bringen. Eine solche Einschätzung sollte hierbei explizit erfragt werden (vgl. AAPOR Cell Phone Task Force (2010), S. 55; S. 80).

Unterschiedliche technische Voraussetzungen können ebenfalls zu methodischen Problemen in Dual-Frame-Erhebungen führen und die Datenqualität gerade bei Mobilfunkstichproben negativ beeinflussen. So ist bei einer Festnetzstichprobe i. d. R. nicht mit technischen Schwierigkeiten der Verbindung zu rechnen. Im Gegensatz hierzu ist dies bei Mobilfunkstichproben durchaus denkbar. Mangelnde Stromversorgung eines Endgerätes, fehlende Netzabdeckung und verminderte Sprachqualität können dabei die Datenerfassung erschweren und ungewollte Abbrüche hervorrufen (vgl. M. Häder und Kühne (2009c), S. 169).

Auch aufgrund der technischen Besonderheiten von Mobilfunkanschlüssen stellen mehrere Untersuchungen, wie bspw. von S. Häder et al. (2009b), Schneiderat und Schlinzig (2011) oder Sand (2015a), fest, dass sich die Kontaktierung von Mobilfunknutzern schwerer darstellt, als im Festnetz. Sand (2015a) bemerkt diesbezüglich bei der Untersuchung der Ergebnisse der CELLA 2 Erhebung, dass durchschnittlich weniger Kontaktversuche benötigt wurden, um einen erfolgreichen Kontakt mit einem Festnetzhaushalt herzustellen. Allerdings liegt bei einem erfolgreichen Kontakt auch der Anteil der Verweigerungen zur Teilnahme höher als bei Mobilfunkstichproben.¹⁸ Diese Beobachtung stimmt dabei mit früheren Ergebnissen von S. Häder et al. (2009b) überein. Demnach kann vermutet werden, dass Mobilfunkeinsteiger zwar schwieriger zu erreichen sind, dass ein positiver Kontakt jedoch eher zur Teilnahme an einer Erhebung führt (vgl. S. Häder et al. (2009b), S. 73; vgl. Schneiderat und Schlinzig (2011), S. 124.; vgl. Sand (2015a), S. 142.). Gerade wegen dieser (technischen) Gegebenheiten lässt sich eine durchschnittlich geringere Responserate¹⁹ in Mobilfunktstichproben im Vergleich zu reinen Festnetzstichproben berichten. Fuchs (2002) erweitert die Begründung der Unterschiede in den Kontaktversuchen um ein weiteres Argument. Dem Autor zufolge ist dies der Art der Stichprobe zuzusprechen. Bei einer Festnetzstichprobe wird in einem ersten Schritt ein Haushalt kontaktiert, der aus mehreren Personen bestehen kann. Dies kann dazu führen, dass ein erfolgreicher Kontakt zustande kommt, auch wenn einzelne Einheiten des Haushaltes zu einem Zeitpunkt nicht erreichbar sind. Im Mobilfunk ist dies, da es sich hierbei um eine direkte Personenstichprobe handelt, nicht zu vermuten (vgl. Fuchs (2002), S. 126ff.).

Bezüglich der Effekte auf die Datenqualität in Abhängigkeit des gewählten Modus halten M. Häder und Kühne (2009d) fest, dass Unterschiede zwischen Festnetz und Mobilfunk lediglich bei der Anzahl der Abbrüche des Interviews bestehen. Diese liegt demnach bei Mobilfunkbefra-

18 Ein vergleichbares Ergebnis zeigt sich ebenfalls bei der Betrachtung der AAPOR Contact Rates (1-3) sowie der Refusal Rates (1-3).

19 Bspw. die Responserate 3 der AAPOR-Standarddefinition (siehe: American Association for Public Opinion Research (2011)).

gungen höher als im Festnetz. Darüber hinaus merken die Autoren an, dass Mobilfunkbefragte weniger häufig nach sozialer Erwünschtheit antworten und eine unterschiedliche Antworttendenz bei Fragen mit Bezug auf das Thema Mobilfunk haben. Dabei wirkt der Modus Mobilfunk nach Untersuchung der Autoren positiv auf gerade solche Fragestellungen (vgl. M. Häder und Kühne (2009d), S. 242.).

2.4.4 Die CELLA-Erhebungen

Um die Vorzüge von Dual-Frame-Ansätzen in der deutschen Umfrageforschung näher zu untersuchen, wurden im Rahmen zweier Projekte, die durch die DFG finanziert wurden, die beiden Erhebungen CELLA 1 (2007/ 2008) und CELLA 2 (2010/ 2011) durchgeführt. Antragsteller beider Projekte waren Häder, M. (TU Dresden), Häder, S. (GESIS) und Gabler, S. (GESIS). Beide Erhebungen gingen Fragestellungen der Durchführung solcher Erhebungen, methodischer Unterschiede, der Gewichtung in Dual-Frame-Ansätzen (siehe hierzu: Abschnitt 3.3.1), der Zusammensetzung soziodemografischer Merkmale nach Medium sowie Mode-Effekten nach. Die Erhebung CELLA 2, die im weiteren Verlauf dieser Arbeit Verwendung finden wird, realisierte hierbei 1.500 Interviews im Festnetz und 1.507 im Mobilfunk, wobei die Bruttostichprobengrößen 30.360 Festnetz- und 46.966 Mobilfunkrufnummern beinhalteten. Eine kurze Darstellung der Ausschöpfung in CELLA 2 kann anhand von Tabelle 2²⁰ abgelesen werden.

Tabelle 2 CELLA 2: Aufteilung nach Art der Bruttostichprobe

	Insgesamt	Festnetz	Mobilfunk
Brutto-SP	77.326	30.360	46.966
Nicht geschaltete Rufnummern	34.057	13.004	21.053
Andere SP-neutrale Ausfälle	7.751	3.355	4.396
Kein Kontakt	21.451	6.239	15.212
Verweigerung	9.956	5.221	4.735
Abgeschlossene Interviews	3.007	1.500	1.507
AAPOR RR 3	13	14,8	11,7

Im Mittel wurden dabei 2,4 Anwahlversuche benötigt, um eine Rufnummer zu bearbeiten, wobei im Festnetz durchschnittlich 2,7 Anrufversuche und im Mobilfunk durchschnittlich 2,2 Versuche anfielen. Der Median liegt dabei bei beiden Stichproben bei dem Wert 1.²¹ Es wurden maximal 15

²⁰ Es handelt sich hierbei um eine verkürzte Darstellung, bei der die Kontaktergebnisse aufgrund des Auslassens bestimmter Kategorien nicht der Bruttostichprobe entsprechen.

²¹ Für realisierte Interviews liegt der Median bei einem Wert von 2.

Anwahlversuche unternommen (vgl. Schneiderat und Schlinzig (2011), S. 124f.; vgl. Sand (2015a), S. 142.).

2.4.5 Die Erhebung Influenza 2014 des Robert Koch-Institutes RKI

Zur Durchführung der Erhebung Influenza 2014 wurden dem Robert Koch-Institut im Zeitraum vom 03. März 2014 bis zum 26. Mai 2014 mehrere Festnetz- und Mobilfunkstichproben durch GESIS zur Verfügung gestellt. Die Erhebung beschäftigte sich mit dem Impfverhalten der bundesdeutschen Bevölkerung sowie deren Einstellung zu Gripeschutzimpfungen und weiteren gesundheitlichen Themen. Zusätzlich wurden im Rahmen dieser Erhebung auch Fragestellungen zu den Nutzungsgewohnheiten von Telekommunikationsmedien integriert, die von GESIS in Kooperation mit der Technischen Universität Dresden erstellt wurden. Die abgefragten Nutzungsgewohnheiten dienen einer besseren Einschätzung über die Präferenz bestimmter Telekommunikationsmedien bei der Kontaktaufnahme und somit einer verbesserten Anwendung des im Rahmen dieser Arbeit vorgestellten Gewichtungsansatzes.

Für die Dual-Frame-Erhebung wurden hierbei ca. 32.000 Festnetz- und ca. 11.000 Mobilfunkrufnummern brutto in der Zeit vom 21. März 2014 bis zum 27. Juni 2014 eingesetzt. Dieser Zeitraum untergliedert sich in eine Pretest- und die eigentliche Feldphase. Insgesamt realisierte die Erhebung dabei 1.519 Interviews. Da im Rahmen der Erhebung ein gesteigertes Interesse an dem Verhalten und den Einstellungen von Personen, die mindestens 60 Jahre alt sind, bestand, wurde hierzu ein Oversample anhand einer unabhängig gezogenen Festnetzstichprobe eingesetzt. Im Rahmen dieses Oversamples wurde ausschließlich nach Personen, die mindestens 60 Jahre alt sind, gescreent. Dadurch ergibt sich die in Tabelle 3²² dargestellte Aufteilung nach Stichprobentyp.

²² Es handelt sich hierbei um eine verkürzte Darstellung, bei der die Kontaktergebnisse aufgrund des Auslassens bestimmter Kategorien nicht der Bruttostichprobe entsprechen.

Tabelle 3 Influenza 2014: Aufteilung nach Art der Stichprobe

	Insgesamt	Oversample	Festnetz	Mobilfunk
Brutto-SP	42.952	20.978	11.000	10.974
Nicht geschaltete Rufnummern	18.934	10.596	5.628	2.710
Nicht Teil der Zielpopulation	4.772	3.912	494	366
Kein Kontakt	12.116	3.639	2.121	6.356
Verweigerung	2.526	1.136	1.122	268
Abgeschlossene Interviews	1.519	505	647	367
AAPOR RR 3	18,7	19,8	22,7	11,1

Im Mittel wurden hierbei fünf Kontaktversuche benötigt, bis es zu einem endgültigen Abschluss einer Rufnummer kam, wobei die durchschnittliche Anzahl der Kontakte bei beiden Festnetzstichproben ca. vier beträgt und im Mobilfunk sieben. Im Median liegt die Anzahl der Kontaktversuche bis zum Abschluss jedoch bei allen drei Stichproben bei ca. zwei und weist eine Spannweite von 1 bis 20 auf (vgl. Gramlich (2015), S. 4f.; vgl. Schmich et al. (2015), S. 2ff.).

Als Teilpopulation der bundesdeutschen Bevölkerung ab 18 Jahren lässt sich für die Einheiten, die über das Oversample in die Erhebung gelangt sind, vermuten, dass abweichende Nutzungsgewohnheiten von Telekommunikationsmedien bestehen. Daher wurde bei diesem Oversample auf die Abfrage der Items zur Nutzung von Telekommunikationsmedien verzichtet. Bei der Betrachtung der Erhebung Influenza 2014 werden daher im weiteren Verlauf dieser Arbeit lediglich die *reguläre Festnetzstichprobe* sowie die Mobilfunkstichprobe Berücksichtigung finden. Die Anzahl der insgesamt abgeschlossenen Interviews reduziert sich dadurch auf 1.014. Die gemeinsame Responserate (AAPOR RR3) dieser beiden Stichproben beträgt dann ca. 17 %.

2.5 HLR-Lookup-Verfahren zur Bereinigung von Mobilfunkstichproben

Wie bereits in Abschnitt 2.4.3 dargestellt, weisen Mobilfunkstichproben i. d. R. eine geringere Hitrate als Festnetzstichproben auf. Aufgrund der Gestaltung der Auswahlgrundlage für Mobilfunkstichproben, aber auch gesteuert durch die Nutzungsgewohnheiten und technischen Aspekte von Mobilfunkgeräten, wird somit die Kontaktaufnahme von Mobilfunknutzern erschwert. Weiterhin problematisch ist, dass dadurch eine hohe Bruttostichprobengröße gewählt werden muss. Daraus resultiert, neben den Erschwernissen im Feldgeschehen, eine geringere Inklusionswahrscheinlichkeit einer Einheit, die über eine Mobilfunkstichprobe kontaktiert wird (und

somit ein höheres Designgewicht).²³ Aufgrund ungeklärter Kontaktergebnisse kann weiterhin die Varianz des Schätzers negativ beeinflusst werden. Die Steigerung der Hitrate hat somit ebenfalls einen (direkten) Einfluss auf die Präzision der Schätzung. Zusätzlich kann eine Auswahlgrundlage mit niedriger Hitrate zu Verzerrungen in der Schätzung führen. Dies kann damit begründet werden, dass es durch den Overcoverage, der durch nicht geschaltete Rufnummern entsteht und dessen Ausmaß nur schwer abzuschätzen ist, zu einer niedrigeren Inklusionswahrscheinlichkeit der erhobenen Einheiten kommt. Dies wiederum führt zu höheren Designgewichten dieser Einheit, was dann zu einer Überschätzung bestimmter Parameter führen kann. Die Bereinigung der Auswahlgrundlage und der Stichprobe um nicht geschaltete Rufnummern, gerade vor dem Hintergrund ungeklärter Kontaktergebnisse, kann somit zu einer Reduktion der Verzerrung und einer Erhöhung der Präzision führen. Eine technische Lösung zur Steigerung der Hitrate kann daher einen direkten Einfluss auf die Qualität der Schätzung haben. In den vergangenen Jahren haben sich dafür technische Lösungswege etabliert, die gerade diesem Zweck dienen können. Solche Verfahren bieten die Möglichkeit, netzwerkbasierend Rufnummern zu verifizieren, ohne eine Rufnummer tatsächlich anzuwählen (vgl. Sand (2014), S. 14.).

Bereits Struminskaya et al. (2011) zeigen durch die nachträgliche Verifizierung der Mobilfunknummern, die im Rahmen der CELLA 2-Erhebung eingesetzt wurden, dass ein solches Verfahren eine genauere Bestimmung von Responseraten ermöglicht und darüber hinaus dazu beitragen kann, Interviewkosten zu sparen. Eine Untersuchung von Kunz und Fuchs (2011) berichtet vergleichbare Ergebnisse. Die Autoren beider Untersuchungen bemerken, dass durch ein solches Verfahren die Arbeitslast der Interviewer in einer Dual-Frame-Erhebung gemindert werden kann. Darüber hinaus wird bemerkt, dass ein solches Verfahren dazu beitragen kann, die Kontakt- und Interviewraten zu erhöhen, da weniger Anwahlversuche pro komplettiertem Interview benötigt würden (vgl. Struminskaya et al. (2011), S. 3f.; vgl. Kunz und Fuchs (2011), S. 5596ff.). Begründet werden kann dies durch das Wegfallen nicht-geschalteter Anschlüsse in der Stichprobenauswahl der Mobilfunkstichprobe.

Jedoch wird ebenfalls berichtet, dass sich bei der Anwendung einer strengen Screening-Regel eine falsch negative Selektion einstellen kann. Das bedeutet, dass Rufnummern aus der Stichprobenauswahl entfernt werden, obwohl sich geschaltete Anschlüsse dahinter verbergen. Eine ähnliche Beobachtung konnte bei der Bearbeitung der Kontaktergebnisse der RKI-Erhebung Influenza 2014 festgestellt werden. Im Rahmen dieser Erhebung wurde zuvor für den Mobilfunkteil ein solches HLR-Lookup-Verfahren durchgeführt. Um die Zuverlässigkeit dieser Verfahren zu überprüfen, wurden alle Mobilfunknummern in der Feldzeit verwendet und nochmals anhand des eingesetzten *Predictive Dialers* überprüft. Dabei lag zwischen Lookup und Feldzeit ein Zeitraum von ca. acht bis zwölf Wochen. Die Auswertung der Dialer-Ergebnisse, weist ca. 73 % der Rufnummern, die durch ein Lookup ausgescreent werden, als geschalteten Anschluss aus.

²³ Für nähere Informationen siehe Kapitel 3.3.1.

Allerdings wurden unter den negativ gescreenten Rufnummern lediglich 30 Kontakte mit einer Person und sieben Interviews erzielt. Darüber hinaus muss festgehalten werden, dass aufgrund der vergangenen Zeit zwischen Lookup und Feldphase durchaus zum Zeitpunkt der Überprüfung nicht-geschaltete Rufnummern neu geschaltet sein können.

Zur Überprüfung der Reliabilität und Validität solcher HLR-Lookup-Verfahren wurde daher von GESIS 2015 das Projekt *Verifizierung mobiler Informationen (Vermln)* durchgeführt. Zusätzlich wurde im Rahmen dieses Projektes die Dauer der Gültigkeit solcher Ergebnisse genauer begutachtet. Neben der Stichprobenziehung ist die Dauer der Gültigkeit vor allem unter dem Aspekt der Bereinigung der Auswahlgrundlage interessant. Sind Lookup-Ergebnisse über einen längeren Zeitraum gültig, so würde dies bspw. ermöglichen, die gesamte Auswahlgrundlage zu überprüfen, ohne dass dadurch hohe Kosten durch die Reevaluierung zuvor ausgefilterter Rufnummern entstehen. Die dadurch erreichte Reduktion des Overcoverage bzw. die Steigerung der Hitrate würde somit auch die Präzision der Schätzung verbessern. Eine kurze Darstellung des Projektaufbaus sowie der wichtigsten Ergebnisse sind Teil dieses Abschnittes. Zuvor wird der Ablauf eines solchen Verfahrens anhand des Aufbaus eines Mobilfunknetzes erläutert.

2.5.1 Aufbau von Mobilfunknetzen

Bei einem Mobilfunknetz wird über Knotenpunkte, sog. *Mobile Switching Center (MSC)s*, der gesamte Ablauf von Gesprächsanfragen gesteuert. Dabei sind das *Home Location Register (HLR)* ebenso wie das *Visitor Location Register (VLR)* als Komponenten des Mobilfunknetzes zu verstehen, die eng mit dem MSC in Verbindung stehen. Die beiden Datenbanken sind bspw. für die Erfassung von Rechnungsdaten und der Überprüfung der Zugehörigkeit eines Teilnehmers zu einem bestimmten Netz notwendig. Zur Ausfallsicherung erfolgt daher eine redundante Speicherung der relevanten Daten.

Das HLR speichert Kundendaten in einer zu Stammdatenblättern vergleichbaren Weise und vermerkt überdies, ob ein Kunde gerade im eigenen Netz eingebucht ist bzw. wo er zuletzt eingebucht war. Ausgehend von den Angaben des Mobilfunk-FAQs von Gajek (2015) wird das HLR oftmals anhand der ersten beiden Ziffern der Mobilfunkrufnummer (nach der Mobilfunkvorwahl) identifiziert. Es verwaltet darüber hinaus die *International Mobile Subscriber Identity (IMSI)*, die einer eindeutigen Kundenkennung entspricht, die *Mobile Station ISDN Number (MSISDN)*, also die Rufnummer des Teilnehmers sowie die Identifikation des aktuell für den Teilnehmer zuständigen VLRs.

Die Aufgabe des VLRs besteht unter anderem darin, festzustellen, ob ein Teilnehmer, der sich in ein Netz eingebucht hat, "netzfremd" ist, um so auch ohne Rückfrage an das HLR abgehende Gespräche einzuleiten. Die wichtigste Aufgabe der Datenbank liegt jedoch in der Unterstützung des MSCs beim Verbindungsaufbau. Hierzu werden vom VLR die Daten aller Mobilstationen,

die sich im Geltungsbereich eines MSCs befinden, gespeichert und die Nutzerdaten, die beim HLR abgefragt wurden, zur Authentifizierung bereitgestellt. Der eigentliche Aufbau einer solchen Netzwerkarchitektur kann anhand der schematischen Darstellung eines *Universal Mobile Telecommunication System* (UMTS)-Netzes in Abbildung 2 nachvollzogen werden.²⁴

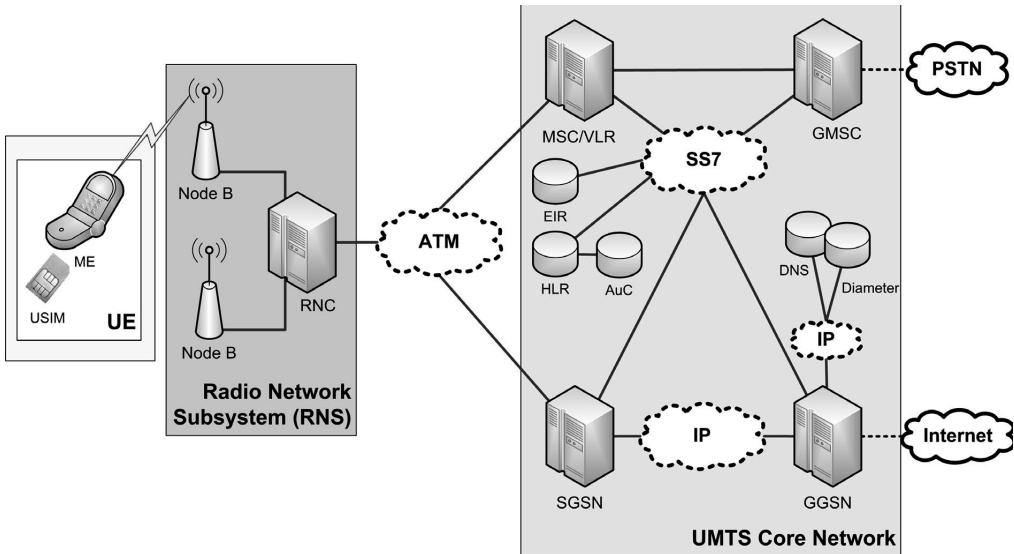


Abbildung 2 Vereinfachte Darstellung der UMTS-Netzwerkarchitektur

Quelle: Bundesamt für Sicherheit in der Informationstechnik (2008), S. 35.

Ein Mobilfunkgerät, das eine SIM-Karte enthält, wird durch Einschalten und Authentifizierung in ein Netzwerk eingebunden. Die Authentifizierung erfolgt dabei durch die auf der SIM gespeicherten Identifikationsinformationen des Inhabers und dem Austausch eines nicht auslesbaren Algorithmus zwischen Teilnehmer und Netzbetreiber (*Shared Secret*). Beim Rufaufbau erfolgt die Integration eines Endgerätes in mehreren Schritten bis hin zum eigentlichen Kernnetzwerk des Betreibers. Dabei gilt das MSC als Schnittstelle zwischen dem Betreiber Netzwerk und dem physischen System aus Sendemasten (Nodes) und Schaltstellen (*Radio Network Subsystem* (RNS)). Die Kommunikation innerhalb des Netzwerkes des Betreibers erfolgt hierbei über das SS7 (vgl. Bundesamt für Sicherheit in der Informationstechnik (2008), S. 35f.; Ericsson GmbH (2001), S. 16.). Dieses System ist eine Sammlung von Protokollen, die zur internen Kommunikation des Netzwerkbetreibers innerhalb des Mobilfunknetzes verwendet wird.

In einem solchen System erlaubt ein *HLR-Lookup* die Abfrage mehrerer HLRs nach einzelnen

²⁴ Auf eine genauere Beschreibung des *Long Term Evolution* (LTE)-Netzwerks, das bis 2018 Standard sein soll, wird aus Gründen der Übersichtlichkeit an dieser Stelle verzichtet. Dieses setzt jedoch auf eine vereinfachte Architektur und speichert Nutzerdaten nicht mehr im Rahmen eines HLRs, sondern auf einen *Home Subscriber Server*. Daten werden allerdings noch immer über das gleichnamige Protokoll *Signaling System 7* (SS7) ausgetauscht (vgl. Gutt (2010), S. 23ff.).

oder gebündelten Mobilfunkrufnummern und kann dadurch Auskunft darüber geben, ob eine Rufnummer tatsächlich zu einem geschalteten Anschluss führt und vergeben ist. Das Verfahren erfragt folglich den Status dieser Rufnummer. Darüber hinaus können Parameter wie die Länder- und Netzentifikation ermittelt werden, um so Informationen darüber zu erhalten, ob eine bestimmte Rufnummer von einem in ein anderes Mobilfunknetz portiert wurde und ob sich ein bestimmter Anschluss gerade im Roaming, also in einem fremden (bspw. ausländischen) Netz befindet. Die Anbieter dieser Verfahren wenden sich mit den entsprechenden Rufnummern an das Kernnetz und greifen über das SS7 direkt auf die entsprechenden HLRs zu (vgl. Ruto Messaging (2009), S. 4ff.; Infobib, S. 1.).

Bei einer solchen Abfrage wird zusätzlich die Identifikation des MSCs erfragt, die gewöhnlich jedoch aufgrund von datenschutzrechtlichen Bestimmungen nicht bereitgestellt wird. Heckel (2014) zeigt in einer Untersuchung, dass zumindest bei drei der (zum damaligen Zeitpunkt) vier Netzbetreiber die Kenntnis über das zuständige MSC eine rudimentäre geografische Einordnung des Mobilfunkanschlusses erlaubt. Eine wirklich trennscharfe Lokalisierung wird aber selbst mit der Kenntnis des Standortes nicht über alle (vier) Mobilfunknetze möglich, da sowohl der Standort, die Anzahl und der Zuständigkeitsbereich der MSCs über die Netzbetreiber hinweg variieren (vgl. Heckel (2014)). Darüber hinaus wäre die Frage zu klären, ob eine Sendestation stets das nächstgelegene MSC verwendet.

2.5.2 Reliabilität und Validität von Lookup-Verfahren

Im Rahmen des Projektes *VermIn* wurden zur Überprüfung der Reliabilität und Validität der Ergebnisse von HLR-Lookup-Verfahren ca. 30.000 Rufnummern zeitgleich durch zwei verschiedene Anbieter zu drei unterschiedlichen Zeitpunkten überprüft. Die Abfrage zu mehreren Zeitpunkten dient hierbei der Einschätzung der Volatilität der Ergebnisse einer solchen Abfrage. Von Interesse ist die Frage, wie alt ein solches Lookup sein kann, um noch seine Gültigkeit in der Feldphase zu behalten. Die Abfragen wurden daher in einem Abstand von ca. zwei Wochen und ca. vier Monaten zur ersten Abfrage wiederholt.

Zur Überprüfung der Reliabilität wurde die HLR-Lookup-Abfrage zeitgleich von zwei verschiedenen Anbietern²⁵ realisiert. Dabei wurde die Annahme zugrunde gelegt, dass beide Anbieter die gleiche technische Abfrage bei einer gegebenen Rufnummer durchführen. Demnach ist zu vermuten, dass auch die Ergebnisse beider Anbieter vergleichbar sind.

Zusätzlich wurde zur Überprüfung der Gültigkeit solcher Ergebnisse eine Substichprobe der 30.000 Rufnummern (RN) im Umfang von ca. 12.000 durch die *USUMA GmbH* im Rahmen einer Erhebung eingesetzt. Die zweite der drei HLR-Abfragen lag hierbei in der Feldzeit der Erhebung, während die erste Abfrage eine Woche vor Feldbeginn stattgefunden hat. Die Ergeb-

²⁵ Als Anbieter wurden dabei aufgrund ihres Angebotes *MyCoolSMS* und *txtnation* gewählt.

nisse der HLR-Abfragen waren USUMA in der Feldzeit unbekannt und die Substichprobe war eine Zufallsauswahl aus allen 17 Vorwahlbereichen. Der Vergleich zwischen den tatsächlichen Kontaktergebnissen und den Ergebnissen der Abfrage soll hierbei Aufschlüsse über die Validität von HLR-Lookups geben.

Der Vergleich beider Anbieter, getrennt nach Netz und Vorwahlbereich der Mobilfunknummern, liefert zum Zeitpunkt der ersten Abfrage die in Abbildung 3 dargestellten Ergebnisse.

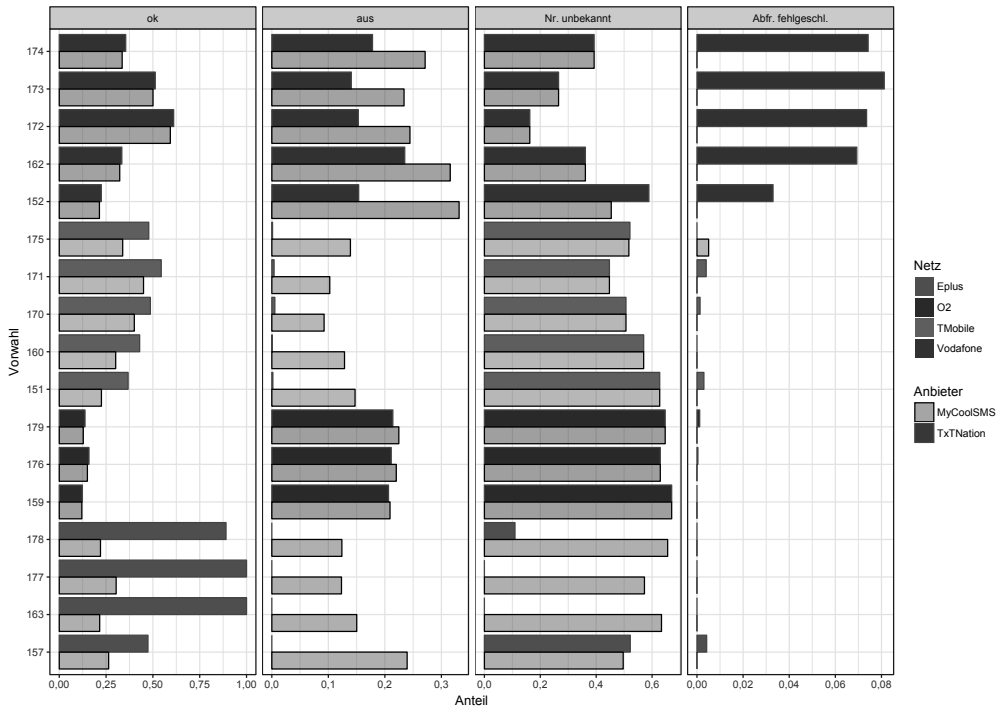


Abbildung 3 HLR-Lookup: Abfrage zum ersten Zeitpunkt nach Resultat und Netz

Abbildung 3 stellt die Ergebnisse der Abfrage getrennt nach Vorwahl, originärem Netz der Vorwahl sowie HLR-Anbieter gegenüber. Der Status *ok* gibt in diesem Kontext wieder, dass eine Rufnummer geschaltet ist und das Mobilfunkgerät empfangsbereit ist.²⁶ Der Status *aus* hingegen signalisiert, dass eine Rufnummer zwar vergeben, allerdings nicht empfangsbereit ist.²⁷ Die Ergebnisse des Anbieters *My-Cool-SMS* werden hier durch den helleren der beiden Balken angegeben und die jeweiligen, originären Netzbetreiber sind farblich unterschieden.

Ein Unterschied beim berichteten Status einer Mobilfunknummer lässt sich insbesondere in Abhängigkeit der Vorwahl und des Netzes feststellen. So fällt auf, dass einer der beiden Anbie-

26 Das Mobilfunkgerät ist eingeschaltet oder die SIM-Karte befindet sich in einem eingeschalteten Mobilfunkgerät
27 Das Mobilfunkgerät ist ausgeschaltet, bzw. die zugehörige SIM-Karte befindet sich nicht in einem empfangsbereiten Mobilfunkgerät.

ter den Status für Rufnummern aus dem Netz des ehemaligen Betreibers *Eplus* durchgehend als *vergeben und eingeschaltet* berichtet. Da ein solches Ergebnis nicht plausibel scheint und stark von den Ergebnissen des anderen Anbieters abweicht, kann davon ausgegangen werden, dass hier eine fehlerhafte Berichterstattung vorliegt. Bezüglich der Ergebnisse der Abfragen in anderen Netzen stellen sich nur geringfügige Unterschiede ein, deren Begründung oftmals in einer abweichenden Interpretation der Abfrage zu finden ist. So kann bspw. ein Großteil der abweichenden Ergebnisse im Netz von *T-Mobile* mit der Frage, ob ein Endgerät eingeschaltet ist, erklärt werden. Eine gröbere Einteilung zwischen *Anschluss ist geschaltet* und *Anschluss ist nicht geschaltet* würde jedoch ähnliche Resultate hervorbringen.

Interessant ist ebenfalls der Vergleich mit den Ergebnissen der dritten Abfrage, dargestellt in Abbildung 4.

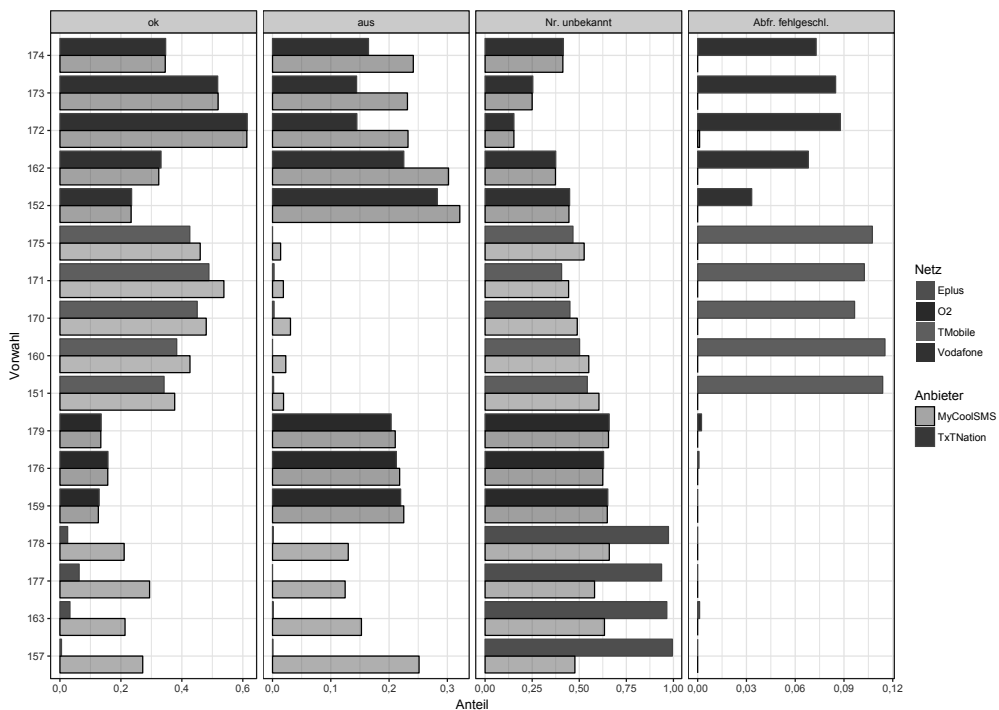


Abbildung 4 HLR-Lookup: Abfrage zum dritten Zeitpunkt nach Resultat und Netz

Hierbei fällt auf, dass im Zuge der dritten Abfrage die Ergebnisse im Netz von *Eplus* sich gerade zum anderen "Extrem" beim Anbieter *txtnation* gewandelt haben. Demnach kann davon ausgegangen werden, dass die Berichterstattung dieses Anbieters für das Netz von *Eplus* fehlerhaft ist.

Eine Bereinigung der Abfragen um diejenigen aus dem Netz von *Eplus* zeigt jedoch, dass für

den ersten Zeitpunkt ca. 20 % aller Ergebnisse und für den dritten Zeitpunkt ca. 14 % aller Ergebnisse voneinander abweichen. Diese Unterschiede sind hauptsächlich auf fehlgeschlagene Abfragen des einen Anbieters oder die Angaben, ob ein Mobilfunkgerät eingeschaltet ist, zurückzuführen. Es kommt daher ohne die Abfragen des *Eplus*-Netzes nicht dazu, dass einer der beiden Anbieter berichtet, eine Rufnummer sei geschaltet, während der andere das Gegenteil angibt.

Bei der Überprüfung der zeitlichen Variabilität von Lookup-Ergebnissen war insbesondere der Wechsel vom Status *Rufnummer nicht geschaltet* hin zu einem geschalteten Anschluss von Interesse. Eine Screening-Regel, die alle Rufnummern, deren Status durch ein Lookup als nicht geschaltet gemeldet wurde, ausschließt, könnte bei einer hohen Volatilität des Status einer Rufnummer oder einer langen Feldzeit dazu führen, dass Einheiten, die eigentlich zur Zielpopulation gehören, fälschlicherweise aus der Stichprobenauswahl ausgescreent werden. Trifft dies zu, führt das Screening dazu, dass potenzielle Zielpersonen über den getesteten Anschluss eine Wahrscheinlichkeit von Null aufweisen, um Teil der Stichprobe zu sein. Daraus ergibt sich die Gefahr potenziell verzerrter Schätzergebnisse.

Zur Überprüfung wurden die Rufnummern zum Zeitpunkt der ersten beiden Lookups in derselben Reihenfolge getestet. Da diese jedoch zwischen 18 und 22 Stunden dauerten und einige der Abfragen deshalb nachts stattfanden, wurde die Reihenfolge der Rufnummern zum dritten Zeitpunkt geändert. So wurde die Vermutung, dass einige Mobilfunkgeräte nachts nicht empfangsbereit sind, berücksichtigt. Die zeitliche Veränderung nach Anbieter kann anhand von Abbildung 5 nachvollzogen werden.

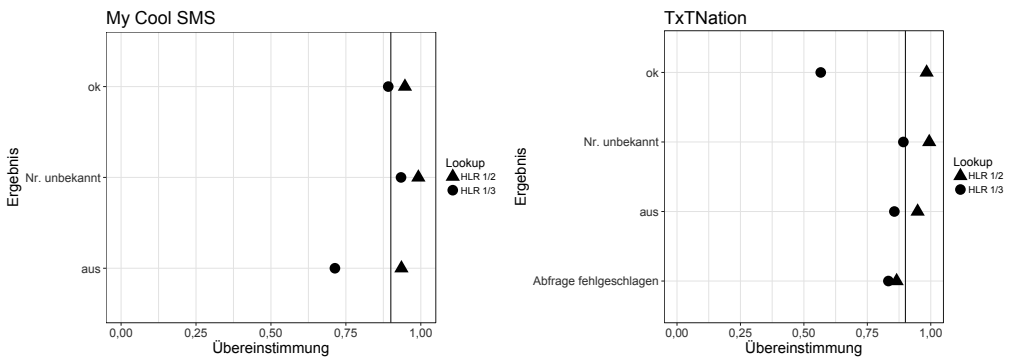


Abbildung 5 HLR-Lookup: Übereinstimmung des Status im Zeitverlauf

Bei beiden Anbietern zeigt sich von der ersten auf die zweite Abfrage nur wenig Veränderung des berichteten Status einer Rufnummer. Bei der dritten Abfrage lässt sich besonders bei dem Anbieter *txtnation* ein deutlicher Unterschied zur ersten Abfrage feststellen. Aufgrund der bereits zuvor festgestellten Probleme dieses Anbieters bei Abfragen im Netz von Eplus sowie

der geänderten Reihenfolge der Rufnummern bei der dritten Abfrage wurden anschließend nur diejenigen Resultate überprüft, bei denen ein vergleichbares Ergebnis beider Anbieter vorgelegen hat.²⁸ Dadurch sind lediglich ca. 22.000 Rufnummern Grundlage für die in Tabelle 4 dargestellten, abweichenden Ergebnisse.

Tabelle 4 HLR-Lookup: Ungleiche Resultate nach Lookup (für Substichprobe)

		Abfrage 3			Summe
		Ausgeschaltet	Nr. unbekannt	Eingeschaltet	
Abfrage 1	Ausgeschaltet	0	261	189	450
	Nr. unbekannt	184	0	434	618
	Eingeschaltet	275	142	0	417
Summe		459	403	623	1.485

Für lediglich 7 % der betrachteten Substichproben ließ sich generell ein Statuswechsel beobachten. Der hierbei am häufigsten Auftretende ist jedoch derjenige von einem nicht geschalteten zu einem geschalteten Anschluss. Dies betrifft in der betrachteten Substichprobe 618 Beobachtungen bzw. 3 % der Substichprobe. Gleichzeitig waren ca. 2 % der geschalteten Rufnummern der Substichprobe zum Zeitpunkt der dritten Abfrage nicht mehr geschaltet. Dazu lässt sich festhalten, dass 65 % dieser ausgebuchten Rufnummern zuvor als inaktiv berichtet wurden. Demnach liegt die Vermutung nahe, dass es sich bei einem Großteil dieser Nummern um nicht mehr verwendete Prepaid-Anschlüsse oder Karten für mobile Zweitgeräte handelt. Ohne nähere Angaben lässt sich dies allerdings nicht überprüfen. Der Wechsel zwischen inaktiven zu aktiven Anschlüssen (und umgekehrt) ist vermutlich der Permutation der Rufnummern zum Zeitpunkt der dritten Abfrage zuzuschreiben.

Innerhalb eines Zeitraumes von vier Monaten kann folglich generell ein Statuswechsel einzelner Anschlüsse beobachtet werden. Jedoch ist dies mit einem Anteil von ca. 7 % als vergleichsweise gering einzuschätzen. Weiterhin kann angenommen werden, dass die Feldzeit einer telefonischen Befragung zwar durchaus mehrere Wochen betragen kann, aber selten mehrere Monate umfasst. Daher kann davon ausgegangen werden, dass HLR-Ergebnisse, die aus einer möglichst nahe am Feldbeginn liegenden Abfrage stammen, für die Dauer der Erhebung gültig sind. Eine Screening-Regel, die ausschließlich nicht-geschaltete Anschlüsse aus der Erhebung ausschließt, ist dabei zu empfehlen, da beobachtet werden kann, dass durchaus ein Wechsel zwischen einem inaktiven hin zu einem aktiven Status erfolgt. Bei einer mehrmonatigen Feldzeit lohnt es sich darüber hinaus, die gesamte Bruttostichprobe nach einigen Wochen nochmals zu

28 Ein Großteil der Rufnummern des Eplus-Netzes fällt dadurch aus der Betrachtung heraus.

überprüfen, um potenziell Verzerrungen zu vermeiden.

Zur Überprüfung der Validität der Ergebnisse wurde eine Substichprobe der 30.000 Rufnummern im Umfang von ca. 12.000 in der RKI-Erhebung *Influenza 2* durch die *Usuma GmbH* eingesetzt. Die Ziehung der Substichprobe erfolgte dabei uneingeschränkt zufällig und beinhaltete Rufnummern aller 17 Vorwahlbereiche. Darüber hinaus war dem Umfrageinstitut das Ergebnis der Statusabfrage nicht bekannt. Durch die Verwendung systemfreier IDs konnte ein Vergleich der Ergebnisse hergestellt werden. Tabelle 5 stellt die Ergebnisse der zweiten Abfrage den Kontaktergebnissen aus der Überprüfung durch USUMA gegenüber.

Tabelle 5 HLR-Lookup: Vergleich der zweiten Abfrage und der finalen Kontaktergebnisse

		Finale Kontaktergebnisse			Summe
		RN geschaltet	nicht vergeben	Status unklar	
MyCoolSMS	Abfr. fehlgeschl.	0	0	0	0
	Ausgeschaltet	101	978	1.552	2.631
	Nr. unbekannt	8	6.516	61	6.585
	Eingeschaltet	1.572	335	1.006	2.913
txtnation	Abfr. fehlgeschl.	3	204	5	212
	Ausgeschaltet	40	352	841	1.233
	Nr. unbekannt	129	5.716	417	6.552
	Eingeschaltet	1.509	1.557	1.356	4.422
Summe (pro Lookup)		1.681	7.829	2.619	12.129

Auch die Gegenüberstellung zwischen der zweiten Abfrage und den finalen Kontaktergebnissen weist deutliche Unterschiede in Abhängigkeit des gewählten Anbieters auf. Von entscheidender Bedeutung ist dabei der Anteil der *falsch-negativen Selektion*. Also diejenigen Beobachtungen, die unter Verwendung der HLR-Ergebnisse dazu geführt hätten, dass eine bestimmte Rufnummer ausgescreent worden wäre, obwohl anhand der Kontaktergebnisse nachgewiesen werden kann, dass zumindest der Kontakt zu einer Person hergestellt worden ist. Ein Screening basierend auf einer Grundlage, die Falsch-Negative beinhaltet, kann dabei zu einer Verzerrung der Schätzwerte führen, da unter Umständen Einheiten, die grundsätzlich zur Zielpopulation gehören, eine Inklusionswahrscheinlichkeit von Null bekommen würden.

Unter Verwendung der Ergebnisse des Anbieters *MyCoolSMS* würde es bei der Verwendung des zweiten Lookups lediglich in acht Fällen bzw. 0,5 % eindeutig zu einer solchen (falsch-negativen)

Selektion kommen. Unter Verwendung des anderen Anbieters wären dies bereits 129 bzw. 7,7 %. In diesem Kontext ebenfalls interessant sind die Anrufketten der Kontaktergebnisse der falsch-negativen Fälle. Diese beginnen beim Anbieter *MyCool/SMS* alle mit dem Dialer-Ergebnis *„Kein Anschluss unter dieser Nummer“*. Daher kann angenommen werden, dass die Rufnummern erst im Verlauf der Feldzeit geschaltet wurden. Bestärkt wird diese Vermutung dadurch, dass im dritten Lookup sechs der acht Rufnummern als geschaltet ausgewiesen werden. Beim Anbieter *txtnation* liegen diese Ergebnisse lediglich für die selben acht Rufnummern vor. Die restlichen Anrufketten der 129 Rufnummern beginnen mit Kontaktergebnissen, die auf einen geschalteten Anschluss schließen lassen.

Eine *falsch-positive Selektion*, also der Ausweis einer Rufnummer als geschaltet, obwohl diese den Kontaktergebnissen zufolge nicht vergeben ist, lässt sich für beide Anbieter feststellen, jedoch insgesamt häufiger für den Anbieter *txtnation*. Allerdings kann für beide Anbieter gezeigt werden, dass ein Großteil der nicht geschalteten Rufnummern anhand einer solchen Abfrage erkannt werden kann. Daher kann der Aussage von Struminskaya et al. (2011) sowie Kunz und Fuchs (2011), dass HLR-Abfragen dazu beitragen können, Interviewkosten einzusparen, entsprochen werden. Durch den Anbieter *MyCool/SMS* würden dabei ca. 83 % aller vermeintlich nicht vergebenen Anschlüsse aus der vorliegenden Stichprobe ausgescreent. Beim Anbieter *txtnation* beläuft sich der Anteil voraussichtlich auf ca. 73 %.

Als Entscheidungsgrundlage zur vorherigen Bereinigung einer Stichprobe ist anhand der vorliegenden Ergebnisse daher einzig das Lookup des Anbieters *MyCool/SMS* dienlich, da der Anteil der falsch-negativen sowie der falsch-positiven Selektion bei *txtnation* zu hoch ist. Als Screening-Regel sollte ein Ansatz verwendet werden, der nur diejenigen Rufnummern aus der Stichprobe entfernt, die eindeutig als nicht geschaltet ausgewiesen werden. Eine Screening-Regel, die ebenfalls Anschlüsse aus der Stichprobe entfernt, deren Lookup kein eindeutiges Ergebnis hervorgebracht hat oder deren Status als inaktiv gemeldet wurde, läuft Gefahr, Einheiten der Zielpopulation eine Auswahlwahrscheinlichkeit von Null zuzuweisen. Dies lässt sich auch anhand des hohen Anteils von Rufnummern, deren wahrer Status nicht im Rahmen der Einsatzstichprobe ermittelt werden konnte, begründen. Grundsätzlich kann ein solches Lookup der Qualität der Erhebung zuträglich sein. Anhand der Ergebnisse zeigt sich allerdings, dass die Wahl des Anbieters eine entscheidende Rolle spielt. Daher empfiehlt es sich vor der Durchführung des Lookups der gesamten Stichprobe durch einen bestimmten Anbieter, eine Substichprobe zu verwenden und deren Ergebnisse genauer zu überprüfen.

Bei einer Feldzeit, die mehrere Monate beträgt, sollte darüber hinaus ein Lookup für die gesamte Stichprobenauswahl wiederholt werden, um sicherzustellen, dass keine Neuschaltung von Rufnummern, die zuvor ausgescreent wurden, erfolgt ist.

Unter dem Aspekt der späteren Schätzung kann ein HLR-Lookup genau dann als dienlich erachtet werden, wenn hierdurch die Anzahl der ungeklärten Kontaktergebnisse reduziert wird

und/oder ein solches Verfahren auf die gesamte Auswahlgrundlage angewendet werden kann und dadurch eine (grundsätzliche) Steigerung der Hitrate von Mobilfunkstichproben bewirkt wird. Gerade die Lookups des Anbieters *MyCool/SMS* scheinen zumindest die Anzahl der ungeklärten Kontaktergebnisse in der Stichprobe zu reduzieren. Dadurch kann ein Lookup zur Reduktion der Bruttostichprobe beitragen, was sich positiv auf die Schätzung auswirken kann. Bezüglich der Steigerung der Hitrate der Auswahlgrundlage gilt es neben der Validität, auch die Dauer der Gültigkeit und die Kosten einer solchen Abfrage zu berücksichtigen. Um Undercoverage zu vermeiden, sollten diejenigen Rufnummern, die bei einer Überprüfung des gesamten Auswahlrahmens als nicht-vergeben identifiziert wurden, nochmals bei einer Aktualisierung des Auswahlrahmens überprüft werden. Die Häufigkeit dieser erneuten Abfrage wird durch die Dauer der Gültigkeit der Abfrage determiniert. Da die Ergebnisse der Untersuchung Vermln gezeigt haben, dass sich nach einer dreimonatigen Frist bereits der Status einer Rufnummer von nicht-vergeben zu vergeben ändern kann, müsste eine solche Abfrage vergleichsweise häufig erfolgen. Hinzu kommt die Abfrage neuer Rufnummernblöcke, die der Auswahlgrundlage hinzugefügt werden. Bei einer initialen Überprüfung der Auswahlgrundlage von GESIS für Herbst 2016 mit einem Nummernuniversum von ca. 328 Mio. potenziellen Rufnummern entstünden dadurch erstmalig Kosten in Höhe von ca. 1,6 Mio. €.²⁹ Hinzu kämen noch die weiteren Abfragen bereits überprüfter und "aussortierter" Rufnummern. Da ca. 55 % der Rufnummern als nicht-geschaltet ausgewiesen werden, kann davon ausgegangen werden, dass die initialen Kosten sowie die Kosten der Pflege dieser Grundlage (deutlich) zu hoch sind, um die Auswahlgrundlage für Mobilfunkstichproben dauerhaft durch HLR-Abfragen zu pflegen. Eine grundsätzliche Steigerung der Hitrate der Auswahlgrundlage kann bei den aktuellen Kosten der Abfrage demnach ausgeschlossen werden.

29 Unter der Berücksichtigung der Rabattierung großer Losgrößen.

3 Gewichtungsverfahren bei Dual-Frame-Erhebungen

Ziel dieses Kapitels ist es, den aktuellen Stand der Forschung bei der Gewichtung von Dual-Frame-Erhebungen umfassend darzustellen. In diesem Kontext werden eingehend die Begriffe Inklusionswahrscheinlichkeit, Designgewichtung und Kalibrierung (auch zur Berücksichtigung von Unit-Nonresponse) erläutert, um anschließend Verfahren zur Gewichtung von Dual-Frame-Erhebungen theoretisch/formal zu beschreiben. Darüber hinaus werden die Schätzer der hier vorgestellten Gewichtungsverfahren anhand von geeigneten Beispielen mit ungewichteten Stichprobenergebnissen verglichen, um einen Eindruck über die Auswirkung von Designgewichtung zu vermitteln. Der Vergleich dieser Gewichtungsverfahren wurde eigens für die vorliegende Arbeit unter der Zuhilfenahme eines synthetischen Datensatzes sowie unter der Verwendung der in Kapitel 2.4.4 und 2.4.5 beschriebenen Erhebungen ausgearbeitet. Er dient als Grundlage zur späteren Diskussion über die Eignung verschiedener Vorgehensweisen zur Ermittlung von Schätzwerten für eine zu beschreibende Zielpopulation. Zusätzlich zu den hier angeführten Erläuterungen und Untersuchungen unterschiedlicher Gewichtungsverfahren, befasst sich Kapitel 3.4.4 als Exkurs mit der Klumpung von Interviews im Rahmen der Dissimilation der Schätzergebnisse und stellt eine Überlegung zu deren Einbeziehung innerhalb der Gewichtungsprozedur von Erhebungsdaten an.

3.1 Ziel der Gewichtung von Erhebungsdaten

Bei der Auswertung von Erhebungsdaten steht jeder Forscher vor der Frage, ob diese gewichtet oder ungewichtet erfolgen soll. Bei der für die Planung und/oder Durchführung einer Erhebung verantwortlichen Person sollte dabei die Kenntnis vorhanden sein, dass eine mangelhaft geplante oder durchgeführte Erhebung durch eine anschließende Gewichtung nicht ausgeglichen wird. Gewichtung im Allgemeinen dient daher nicht dem Ausgleich mangelhafter Stichprobenpraxis. Vielmehr ist der Zweck einer Gewichtung das Ziehungsverfahren, das zur Generierung der Stichprobe verwendet wurde, sowie besondere Gegebenheiten, die sich während des Erhebungsprozesses eingestellt haben (bspw. Nonresponse) zu berücksichtigen.

Valliant et al. (2013) formulieren, dass das generelle Ziel der Gewichtung von Erhebungsdaten darin besteht, eine Menge an Gewichten w_i zu ermitteln, die für jegliche Analysen eines Datensatzes herangezogen werden können. Das Ziel der Gewichtung von Erhebungsdaten ist somit, geeignete Schätzwerte für die zugrunde liegende Zielpopulation der Erhebung zu bestimmen (vgl. Valliant et al. (2013), S. 308.). Indes sind nach Gabler et al. (2011) die Hauptgründe der Verwendung von Gewichten, bei Erhebungsdaten ungleiche Auswahlwahrscheinlichkeiten zu berücksichtigen, potenzielle Verzerrungen, bspw. aufgrund von Nonresponse, zu reduzieren sowie

der Anpassung der Erhebungsdaten an die in der Zielpopulation vorherrschenden Verteilung (vgl. Gabler et al. (2011), S. 147.). Bei einer imperfekten Auswahlgrundlage kann Gewichtung dabei helfen, die Lücke zwischen der Population der Auswahlgrundlage und der der Zielpopulation zu schließen oder zu verkleinern. Daher ist die Gewichtung zur Vermeidung von Fehlerquellen von zentraler Bedeutung (vgl. Lundström und Särndal (2001), S. 17ff.). Somit besteht das Ziel der Gewichtung von Erhebungsdaten darin, das Erhebungsdesign bei der Schätzung von Populationswerten zu berücksichtigen sowie Verzerrungen (aufgrund von Nonresponse) zu vermindern (vgl. Sand und Gabler (2018), S. 4f.). Um dies zu veranschaulichen, stellt Abbildung 6 nochmals die potenziellen Fehlerquellen einer Erhebung dar.

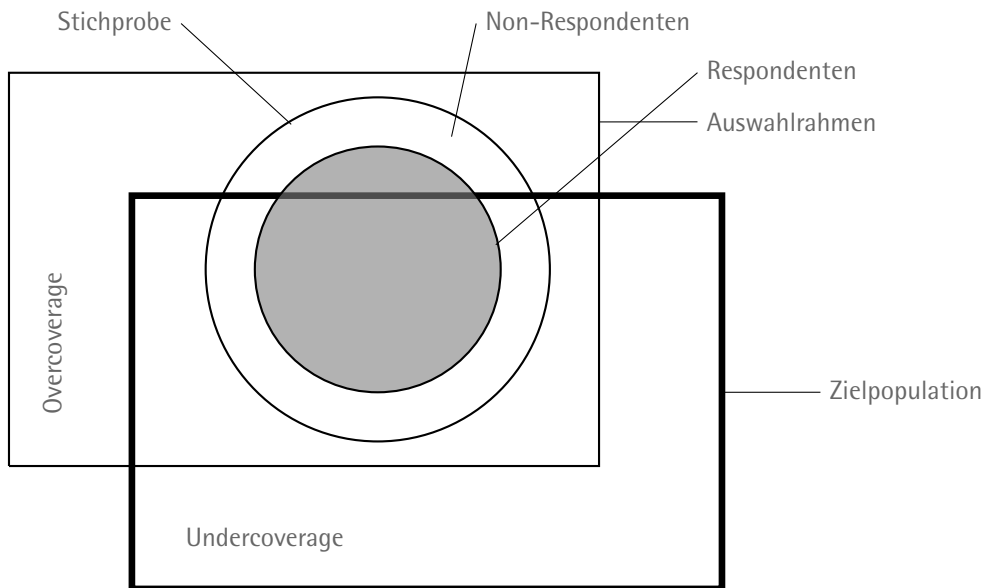


Abbildung 6 Potenzielle Fehlerquellen einer Stichprobe

Bei der Ziehung aus einer Auswahlgrundlage zählt eine Einheit, die zwar Teil der Auswahlgrundlage, aber nicht der Zielpopulation, ist, als Karteileiche und fällt in den Bereich des Overcoverage. Der Auswahlrahmen kann dadurch Einheiten beinhalten, die generell oder für eine gegenwärtige Erhebung nicht von Interesse sind. Am Beispiel von Telefonstichproben könnte dies bspw. eintreten, wenn eine Auswahlgrundlage auch geschäftliche Anschlüsse beinhaltet, eine Erhebung jedoch Personen in privaten Haushalten als Zielpopulation hat. Andererseits kann es vorkommen, dass die Auswahlgrundlage nicht alle Einheiten der Zielpopulation beinhaltet und es dadurch zu Fehlbeständen bzw. Undercoverage kommt. Ein deutliches Beispiel für diese Problematik stellen die Mobile-Only-Haushalte bei alleiniger Verwendung von Festnetzstichproben dar. Gewöhnlich wird davon ausgegangen, dass keine großen Unterschiede zwischen Auswahl- und Zielpopulation

liegen und dass der Anteil der Karteileichen zu vernachlässigen ist. Dennoch zeigt sich am Beispiel von Festnetzstichproben, dass diese Annahmen nicht immer zutreffen. Des Weiteren kann eine Gewichtung sinnvoll sein, um unverzerrte Schätzwerte für einen Parameter der Zielpopulation zu erhalten.

Grundsätzlich ist im Zuge der Gewichtung zwischen der Verwendung von Designgewichten und der Kalibrierung bzw. der Verwendung von Anpassungsgewichten zu unterscheiden. Die Darstellung unterschiedlicher Kalibrierungsverfahren erfolgt in Abschnitt 3.4. Eine grundlegende Darstellung der Designgewichtung sowie die Anwendung einer solchen in Dual-Frame-Erhebungen geschieht in den folgenden beiden Abschnitten.

3.2 Designgewichtung im Allgemeinen

Im Falle von *Frame-Imperfections*, wie es u. a. bei telefonischen Befragungen oftmals zu beobachten ist, kann es demnach zu Under- und Overcoverage kommen. Maßnahmen, um mögliche Ursachen einer Verzerrung zu vermeiden oder zu minimieren, wie bspw. die Verwendung von Dual-Frame-Erhebungen, sind oftmals Teil der Stichprobenziehung. Diese Maßnahmen sollten jedoch ebenfalls in der Designgewichtung Berücksichtigung finden, um unverzerrte Schätzwerte für Parameter der Population zu erhalten.

Designgewichte d_i (mit $i = 1, \dots, N$; N entspricht der Anzahl der Einheiten der Zielpopulation), auch *Base Weights* genannt, werden beim Vorliegen einer Zufallsstichprobe S vom Umfang n als Inverse der *Inklusions-* bzw. Auswahlwahrscheinlichkeit π_i berechnet.³⁰ Durch die Verwendung dieser Gewichte können Schätzwerte aus der zugrunde liegenden Stichprobe für die gesamte, endliche Population bestimmt werden. Dieser Schätzwert ist genau dann unverzerrt, wenn sein (Design-) Erwartungswert gleich dem zu schätzenden Parameter der Zielpopulation ist. Ein Schätzwert für einen Totalwert Y einer Zielpopulation lässt sich bspw. anhand des Horvitz-Thompson-Schätzers durch

$$\hat{Y} = \sum_{i \in S} \frac{1}{\pi_i} \cdot y_i = \sum_{i \in S} d_i \cdot y_i \quad (3.1)$$

bestimmen (vgl. S. L. Lohr (2009), S. 240f.). Bei diesem liegt die Annahme zugrunde, dass eine (nahezu) perfekte Auswahlgrundlage besteht, die alle Einheiten der Zielpopulation beinhaltet, während gleichzeitig der Nonresponse vernachlässigbar ist (vgl. Sand und Gabler (2018), vgl. Valliant et al. (2013), S. 307). Nach Gabler et al. (2011) ist eine Designgewichtung immer dann vonnöten, wenn ein unverzerrter Schätzwert für eine (Teil-) Population bestimmt werden soll und unterschiedliche Inklusionswahrscheinlichkeiten der jeweiligen Erhebungseinheiten zugrunde

³⁰ In der vorliegenden Arbeit wird grundsätzlich von der Verwendung des Horvitz-Thompson-Schätzers bei der Verwendung eines designgewichteten Schätzwertes ausgegangen.

liegen. Bei Telefonstichproben trifft dies bspw. immer dann zu, wenn Rufnummern zwar mit gleichen Wahrscheinlichkeiten aus der Auswahlgrundlage gezogen werden, eine Einheit jedoch über mehrere Rufnummern kontaktiert werden kann (vgl. Gabler et al. (2011), S. 147.).

Im Rahmen von Festnetzstichproben gilt überdies zu beachten, dass hier in einem ersten Schritt lediglich ein Haushalt kontaktiert wird. Besteht die Zielpopulation der Erhebung nun aus Personen in privaten Haushalten, so sollte anhand eines sog. *Kish-Grids* oder eines (randomisierten) Geburtstagsverfahrens eine Zielpersonenauswahl im Haushalt getroffen werden. Bei Mehrpersonenhaushalten ist daher die Auswahlwahrscheinlichkeit einer einzelnen Einheit ebenfalls abhängig von der *erhebungsrelevanten Haushaltsgröße*. Die erhebungsrelevante Haushaltsgröße misst hierbei all diejenigen Einheiten eines Haushaltes, die Teil der Zielpopulation sind.

Bei Dual-Frame-Erhebungen, denen kein *Screening-Ansatz* zugrunde liegt, gilt es darüber hinaus zu beachten, dass die Auswahlwahrscheinlichkeit eines Dual-Users sowohl über das Festnetz als auch über das Mobilfunknetz in der Bestimmung der Designgewichte berücksichtigt werden muss (vgl. S. Häder et al. (2009b), S. 25.).

3.3 Designgewichtung in Dual-Frame-Erhebungen

Der Fragestellung, ob im Zuge von Dual-Frame-Erhebungen eine besondere oder gesonderte Art der Designgewichtung vonnöten ist, wurde sowohl in der amerikanischen als auch in der deutschsprachigen Literatur ausführlich diskutiert. Die von der AAPOR gegründete *Cell Phone Task-Force* hält diesbezüglich 2010 fest, dass eine generelle Designgewichtung immer dann notwendig ist, wenn unterschiedliche Inklusionswahrscheinlichkeiten einzelner Erhebungseinheiten vorliegen und generalisierbare, inferenzielle Aussagen über die Zielpopulation getroffen werden sollen. Nach Auffassung dieser Task-Force kann daher nur in den seltensten Fällen von einer Designgewichtung abgesehen werden (vgl. AAPOR Cell Phone Task Force (2010), S. 62ff.). So könnte bspw. genau dann von einer Designgewichtung abgesehen werden, wenn die Auswahl anhand von einer vollständigen Liste oder eines Registers uneingeschränkt zufällig getroffen wird. Denkbar wäre in diesem Kontext, dass innerhalb eines zentralen Melderegisters Festnetz- und/oder Mobilfunknummern hinterlegt wären, über die dann eine uneingeschränkt zufällige Stichprobenauswahl getroffen wird.

Die Designgewichtung von Dual-Frame-Erhebungen wird weitergehend dadurch erschwert, dass sich das Verständnis der Stichprobenauswahl in beiden Stichproben unterscheidet. Wie bereits in Abschnitt 2.4 beschrieben, wird eine Festnetzstichprobe als Haushaltsstichprobe erfasst, bei der im Zuge einer Personenerhebung eine Transformationsgewichtung auf Personenebene erfolgen muss, während eine Mobilfunkstichprobe i. d. R. als Personenstichprobe wahrgenommen wird (vgl. M. Häder (2009b), S. 18.).

Weiterhin gilt es, bei der Designgewichtung zwischen der Verwendung eines Screening-

Ansatzes sowie eines Overlap-Ansatzes zu unterscheiden. Bei einem Screening-Ansatz ist die Designgewichtung vergleichsweise einfach zu behandeln, da die drei disjunkten Bereiche (Landline-Only-Haushalte, Dual-User und Mobile-Only-Haushalte) aufgrund des Screenings nach einem bestimmten Only-Haushalt in der entsprechenden Stichprobe wie eine Schichtung behandelt werden können. Da die Dual-User nur über eines der beiden Telekommunikationsmedien befragt werden, ergibt sich für die Designgewichtung, dass sie separat nach Domain erfolgen kann (vgl. AAPOR Cell Phone Task Force (2010), S. 62ff.). Somit lässt sich ein Schätzwert $\hat{Y}$ für den Totalwert einer interessierenden Variable Y als Summe der Schätzwerte der einzelnen Domains wie folgt berechnen:

$$\begin{aligned}\hat{Y} &= \sum_{i \in S^F} \frac{1}{\pi_i^F} y_i + \sum_{i \in S^{F \cap C}} \frac{1}{\pi_i^F} y_i + \sum_{i \in S^C} \frac{1}{\pi_i^C} y_i = \sum_{i \in S^F} d_i^F y_i + \sum_{i \in S^{F \cap C}} d_i^F y_i + \sum_{i \in S^C} d_i^C y_i \\ &= \hat{Y}^F + \hat{Y}^{F \cap C} + \hat{Y}^C.\end{aligned}$$

In der vorliegenden Darstellung wird davon ausgegangen, dass nach Mobile-Only-Haushalten gescreent wird. Weiterhin besteht die Annahme, dass alle Dual-User aufgrund des Screenings lediglich über die Festnetzstichprobe befragt werden. π_i^F und π_i^C bezeichnen die Inklusionswahrscheinlichkeiten einer Zielperson i in einer Festnetz- oder Mobilfunkstichprobe. d_i^F und d_i^C stellen analog die jeweiligen Designgewichte dar. Die Menge der Landline-Only-Haushalte der Stichprobe wird anhand von S^F angegeben. Entsprechend beschreiben $S^{F \cap C}$ und S^C die Menge der Elemente der anderen beiden Domains (Dual-User und Mobile-Onlys) in der Stichprobe. Die domain-spezifischen Schätzwerte des Totalwertes einer interessierenden Variable werden schließlich durch $\hat{Y}^F$, $\hat{Y}^{F \cap C}$ und $\hat{Y}^C$ dargestellt.

Weitaus schwieriger stellt sich die Gewichtung bei der Verwendung eines Overlap-Ansatzes dar. Wird ein solcher Ansatz für eine Dual-Frame-Erhebung gewählt und der Überlappungsbereich beider Auswahlrahmen ist nicht leer, so besteht im Gegenzug zum Screening-Ansatz für einige Einheiten die Möglichkeit, entweder über den ersten oder den zweiten Auswahlrahmen in die Erhebung zu gelangen. Einzelnen betrachtet lassen sich für die beiden Stichproben im Rahmen einer Dual-Frame-Telefonerhebung daher folgende Erwartungswerte ($E(\bullet)$) des Totalwertes Y der einzelnen Teilpopulationen berechnen:

$$E\left(\sum_{i \in S^F} d_i^F \cdot y_i\right) = Y^F + Y^{F \cap C}$$

und

$$E\left(\sum_{i \in S^C} d_i^C \cdot y_i\right) = Y^C + Y^{F \cap C} \quad .^{31}$$

Würde nun ein gemeinsamer Schätzwert für Y durch die Addition der einzelnen, aus den jeweiligen Stichproben gewonnenen, Schätzwerten nach

$$\hat{Y} = \sum_{i \in S^F} d_i^F \cdot y_i + \sum_{i \in S^C} d_i^C \cdot y_i$$

berechnet, so ergäbe sich aufgrund der doppelten Anführung des Überlappungsbereiches, dass $\hat{Y}$ ein verzerrter Schätzer für Y wäre (vgl. S. L. Lohr (2011), S. 200.). Daher wurden mehrere Lösungsansätze in der englisch- und deutschsprachigen Literatur erarbeitet, die sich mit der Designgewichtung bei gerade solchen Overlap-Dual-Frame-Erhebungen befassen. Einige dieser Gewichtungsansätze werden in den folgenden Abschnitten näher diskutiert.

3.3.1 Single-Frame-Gewichte

Eine Möglichkeit, in einem solchen Ansatz eine Gewichtung durchzuführen, besteht in der Überlegung, die Beobachtungen beider Stichproben im Überlappungsbereich als eine Art *Super-Stratum* zu verstehen. Erforderlich dafür ist, dass die Inklusionswahrscheinlichkeiten der Einheiten des Überlappungsbereiches so angepasst werden, dass sie die Auswahlwahrscheinlichkeiten beider Stichproben reflektieren. Um dies zu ermöglichen, sind jedoch Angaben für die Auswahlwahrscheinlichkeit aus beiden Stichproben erforderlich (vgl. S. Lohr und J. N. K Rao (2006), S. 1023.). Im Rahmen einer Dual-Frame-Erhebung müssten dabei unabhängig von der zugrunde liegenden Stichprobe mehrere Parameter erfasst werden. Diese sollten dabei erlauben, die Wahrscheinlichkeit der Auswahl einer Einheit für beide Stichproben zu ermitteln. Zusätzlich müssen diejenigen Parameter erfasst werden, die die Wahrscheinlichkeit einer Einheit angeben, im Rahmen der anderen Stichprobe in die Erhebung zu gelangen. Bei Dual-Frame-Telefonstichproben (im Overlap-Ansatz) müsste für Dual-User sowohl erfasst werden, wie wahrscheinlich es ist, dass diese über den Auswahlrahmen Mobilfunk in die Erhebung gelangen als auch die Wahrscheinlichkeit dafür, dass sie über den Auswahlrahmen Festnetz gezogen werden. Dies muss unabhängig davon, über welche der beiden Stichproben eine Einheit tatsächlich in die Erhebung gelangt ist, erhoben werden. Erste Überlegungen zu diesem Ansatz wurden 1986 durch Karlton und Anderson sowie Bankier unternommen. Die Inklusionswahrscheinlichkeit kann somit nach

$$\pi_i = \pi_i^F + \pi_i^C - \pi_i^{F \cap C} \quad \forall i \in U \quad {}^{32} \quad (3.2)$$

31 Würde keiner der beiden Auswahlrahmen unter Undercoverage leiden, wären die Schätzwerte der jeweiligen Stichproben (bei Ausbleiben anderer Fehlerquellen) erwartungstreu für Y .

Wegen der Unabhängigkeit der beiden Stichproben folgt

$$\pi_i^{F \cap C} = \pi_i^F \cdot \pi_i^C.$$

bestimmt werden. Die Inklusionswahrscheinlichkeit einer Einheit i aus der Zielpopulation U ist demnach gleich der Summe der Inklusionswahrscheinlichkeiten, durch eines der beiden Medien in die Erhebung zu gelangen minus dem Produkt beider einzelnen Wahrscheinlichkeiten. Dieses gibt die Wahrscheinlichkeit an, über beide Telekommunikationsmedien gleichzeitig gezogen zu werden. Das Designgewicht berechnet sich dann nach

$$d_i = \begin{cases} (\pi_i^F)^{-1} & \text{wenn } i \in U^F \setminus C \\ (\pi_i^F + \pi_i^C - \pi_i^F \cdot \pi_i^C)^{-1} & \text{wenn } i \in U^F \cap C \\ (\pi_i^C)^{-1} & \text{wenn } i \in U^C \setminus F. \end{cases} \quad (3.3)$$

U^F und U^C beschreiben dabei die Menge der Einheiten, die über die Auswahlgrundlage Festnetz (U^F) und die Auswahlgrundlage Mobilfunk (U^C) erreicht werden können. $U^F \cap C$ entspricht der Menge der Dual-User der Zielpopulation, während $U^F \setminus C$ und $U^C \setminus F$ die Einheiten, die ausschließlich über eine Auswahlgrundlage (*Onlys*) erreichbar sind, beschreiben. Vorteilhaft an dieser Vorgehensweise ist Rao und Lohr (2000) zufolge, dass bei der Bestimmung von Schätzwerten keine Kenntnisse über die tatsächlichen Populationsgrößen derjenigen Einheiten, die über Festnetz (N^F) oder Mobilfunk (N^C) erreicht werden können, benötigt werden (vgl. S. L. Lohr und J. N. K. Rao (2000), S. 272.).

Eine ähnliche Vorgehensweise wurde während der Designgewichtung der beiden CELLA-Erhebungen gewählt. Zur Bestimmung der Inklusionswahrscheinlichkeit der einzelnen Einheiten sind nach Gabler und Ayhan (2007) oder S. Häder et al. (2009a) die Kenntnis über die Größe des Auswahlrahmens für Festnetzstichproben M^F sowie derer für Mobilfunkstichproben M^C von Belang. Weiterhin bedarf es der Bruttostichprobengröße³³ der beiden Stichproben m^F und m^C , der Anzahl der Rufnummern, über die eine Einheit über Festnetz (k_i^F) und Mobilfunk (k_i^C) erreicht werden kann sowie die Anzahl der erhebungsrelevanten Personen (z_i) im Haushalt der befragten Einheit. Erhebungsrelevant sind dabei jene Einheiten, die zur Zielpopulation gehören. Für nähere Informationen, wie und in welcher Art diese Parameter zu erheben sind, siehe Häder et al. (2009) bzw. Sand und Gabler (2018). Die zur Bestimmung der Auswahlwahrscheinlichkeit und anschließenden Designgewichtung erforderlichen Parameter werden nochmals in Tabelle 6 dargestellt.

32 Unter der Annahme, dass beide Auswahlrahmen gemeinsam die gesamte Zielpopulation abdecken.

33 Die Bruttostichprobengröße entspricht der Anzahl der Rufnummern, die aus dem jeweiligen Auswahlrahmen gezogen wurde, um die jeweilige Erhebung zu realisieren. Diese entspricht i. d. R. nicht der Anzahl der realisierten Befragungen.

Gerade die Abfrage der Anzahl der Festnetznummern (und in einem geringeren Maße diejenige der Mobilfunknummern) gestaltet sich in der Praxis durchaus schwierig. In der Erhebung CELLA 1 wurde daher lediglich nach der Anschlussart (Analog- oder ISDN-Anschluss) gefragt und ein entsprechender Durchschnittswert angenommen. Jedoch konnte ein nicht vernachlässigbarer Anteil keine Auskunft über die Art ihres Festnetzanschlusses geben. Weiterhin zeigte sich im Rahmen der anschließenden Erhebung CELLA 2, dass eine solche vereinfachte Annahme keine ausreichende Belastbarkeit liefert. Im Rahmen der zweiten CELLA-Erhebung ließ sich beobachten, dass die Annahme, ein Haushalt mit analogem Anschluss verfüge über einen Anschluss und ISDN-Haushalte (durchschnittlich) über 2,5, nach der Abfrage deutlich von den berechneten Stichprobenmittelwerten abweicht (vgl. Gabler et al. (2011), S. 153f.; Gabler und S. Häder (2009b), S. 52f.). Eine Untersuchung von Meier (2007), die sich mit dem Vergleich der angegebenen Anzahl von Anschlüssen, die durch die Befragung einer Einheit verfügbar ist, und der tatsächlichen Anzahl der vorhandenen Anschlüsse anhand einer persönlich mündlichen Befragung beobachten lässt, zeigt überdies Differenzen zwischen diesen beiden Werten auf. So stimmte lediglich für 75 % aller Befragten die angegebene Anzahl mit der beobachteten überein (vgl. Meier (2007), S. 94f.). Erschwerend kommt durch ISDN- und IP-Telefonie hinzu, dass ein Auftragnehmer eines Festnetzangebotes³⁴ i. d. R. eine bestimmte Anzahl von Rufnummern zugeteilt bekommt, unabhängig davon, ob diese genutzt werden oder nicht. Bei ungenutzten, aber zugeteilten Anschlüssen besteht weiterhin die Möglichkeit eines erfolgreichen Kontaktes. Da diese Rufnummern jedoch nicht aktiv genutzt werden, besteht die Gefahr, dass eine befragte Person diese nicht bei der Befragung angibt. Die Anzahl der Rufnummern wird in diesem Fall untererfasst, was zu einer Unterschätzung der Inklusionswahrscheinlichkeit einer Einheit führt.

Tabelle 6 Parameter zur Bestimmung der Auswahlwahrscheinlichkeit bei Dual-Frame-Telefonstichproben

π_i^F :	Inklusionswahrscheinlichkeit für i aus Festnetzstichprobe	π_i^C :	Inklusionswahrscheinlichkeit für i aus Mobilfunkstichprobe
m^F :	Bruttostichprobengröße Festnetz	m^C :	Bruttostichprobengröße Mobilfunk
M^F :	Anzahl Rufnummern im Auswahlrahmen Festnetz	M^C :	Anzahl Rufnummern im Auswahlrahmen Mobilfunk
k_i^F :	Anzahl Festnetzzufnummern der Einheit i	k_i^C :	Anzahl Mobilfunkrufnummern der Einheit i
z_i :	Anzahl erhebungsrelevanter Personen im Haushalt der Einheit i		

Die Inklusionswahrscheinlichkeit einer Einheit i aus einer Festnetzstichprobe (π_i^F) wird dabei

34 Die Person oder der Haushalt, der über eine bestimmte Anschlussart im Festnetz verfügt.

durch

$$\pi_i^F = \frac{m^F}{M^F} \cdot \frac{k_i^F}{z_i} \quad (3.4)$$

angegeben. Hieraus ersichtlich wird auch das Verständnis einer Festnetzstichprobe als zweistufiges Verfahren, bei dem zuerst eine Rufnummer mit gleichen Wahrscheinlichkeiten aus einer entsprechenden Auswahlgrundlage gezogen wird. Anschließend findet, bei erfolgreichem Kontakt, eine Zielpersonenauswahl im Haushalt statt.³⁵ Für eine zugrundeliegende Mobilfunkstichprobe lässt sich die Auswahlwahrscheinlichkeit (π_i^C) durch

$$\pi_i^C = \frac{m^C}{M^C} \cdot k_i^C \quad (3.5)$$

bestimmen. Wie bereits zuvor dargestellt, wird davon ausgegangen, dass es sich bei einem Mobiltelefon um einen personalisierten Gegenstand handelt. Eine Mobilfunkstichprobe ist daher als Personenstichprobe zu verstehen, bei der die Kontaktperson der Zielperson entspricht. Die Berücksichtigung der erhebungsrelevanten Haushaltsgröße ist folglich in einer einzelnen Mobilfunkstichprobe nicht erforderlich. In einer Dual-Frame-Erhebung, die einen Overlap-Ansatz verwendet, lässt sich Gabler und Ayhan (2007) zufolge die gemeinsame Inklusionswahrscheinlichkeit (π_i) einer Einheit i als Summe der einzelnen Inklusionswahrscheinlichkeiten der beiden Stichproben berechnen. Somit wird π_i bestimmt durch

$$\pi_i \approx \frac{m^F}{M^F} \cdot \frac{k_i^F}{z_i} + \frac{m^C}{M^C} \cdot k_i^C. \quad (3.6)$$

Darüber hinaus wird angenommen, dass

$$\pi_i^F \cdot \pi_i^C \approx 0.$$

Die Wahrscheinlichkeit, eine Person des Haushaltes sowohl über die Festnetz- als auch über die Mobilfunkstichprobe in der gleichen Erhebung zu ziehen, ist demnach vernachlässigbar (vgl. Gabler und Ayhan (2007), S.42ff.; vgl. S. Häder et al. (2009a), S. 23ff.; vgl. Gabler et al. (2011), S. 149ff.). Die Annahme beruht dabei auf der Größe der jeweiligen Auswahlrahmen, die oftmals mehrere Hundertmillionen Rufnummern beinhalten. Für nähere Angaben zur Gestaltung von Auswahlrahmen siehe bspw. Sand (2015b). Zur Herleitung der Bestimmung der Inklusionswahrscheinlichkeit siehe Gabler und Ayhan (2007). Als geeigneter Schätzer für den Totalwert Y einer Zielpopulation bietet sich hierbei, wie bereits dargestellt, der Horvitz-Thompson-Schätzer an, der mittels der Inversen der gemeinsamen Inklusionswahrscheinlichkeit über beide Stichproben

35 Ferner wird davon ausgegangen, dass jede erhebungsrelevante Einheit des Haushaltes die gleiche Wahrscheinlichkeit hat, ausgewählt zu werden. Ein Diskurs darüber, welche Verfahren dies ermöglichen, findet sich bspw. in Gabler et al. (2011).

nach

$$\hat{Y}_{SF} = \sum_{i \in S^F \cup S^C} \frac{y_i}{\pi_i}$$

bestimmt wird.³⁶ Im Rahmen von Dual-Frame-Erhebungen in Deutschland ist dieses Vorgehen in der geschilderten oder einer modifizierten Art gängige Praxis. Eine Modifikation dieses Vorgehens wird bspw. durch den ADM unternommen. Bei Dual-Frame-Erhebungen mit ungleichen Zusammensetzungen der abgeschlossenen Interviews nach Telekommunikationsmedium schlägt dieser zusätzlich eine Proportionalisierung der beiden Stichproben vor. Näheres wird bspw. durch Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute ADM (2012) erläutert.

Einen weiteren Schätzer dieser Kategorie stellt der sog. *Pseudo Single-Frame Schätzer* dar. Ähnlich wie bei dem vorherigen Schätzer werden dabei die Einheiten beider Stichproben zusammengefasst und eine gemeinsame Inklusionswahrscheinlichkeit bestimmt. Das Gewichtungsverfahren geht jedoch mit dem Überlappungsbereich in einer Art und Weise um, die dem eines Screening-Ansatzes gleichzusetzen ist. So wird ein gemeinsames Designgewicht d_i^{SF*} bestimmt nach

$$d_i^{SF*} = \begin{cases} (\pi_i^F)^{-1} & \text{wenn } i \in U^F \setminus C \cup (S^F \cap U^{F \cap C}) \\ 0 & \text{wenn } i \in S^C \setminus F \cap U^{F \cap C} \\ (\pi_i^C)^{-1} & \text{wenn } i \in U^C \setminus F. \end{cases}$$

Ein Nachteil dieses Gewichtungsansatzes stellt die Tatsache dar, dass bei einer solchen Vorgehensweise bewusst ein Teil des Überlappungsbereiches ausgelassen wird. Der Verlust von Daten kann diesbezüglich die Effektivität des Schätzers sowie dessen Verzerrung negativ beeinflussen (vgl. Winglee et al. (2007), S. 3196ff.). Weitere Variationen der Vorgehensweise werden, da die Ansätze aus 3.3 und 3.6 mittlerweile gängige Praxis sind, an dieser Stelle nicht weiter diskutiert. Diese werden jedoch in Winglee et al. (2007) sowie Sand (2014) ausführlich beschrieben.

3.3.2 Composite-Gewichte

Auch wenn der zuvor beschriebene Single-Frame-Gewichtungansatz in Deutschland gängige Praxis ist, findet sich gerade in der amerikanischen Literatur oftmals die Verwendung sog. Composite-Weighting-Ansätze. Im Gegensatz zum Vorgehen bei einem Single-Frame-Ansatz wird hier nicht die gemeinsame Inklusionswahrscheinlichkeit $(\pi_i^{F \cup C})$ einer Einheit i berechnet. Vielmehr werden die beiden Stichproben separat behandelt und mittels Konvexkombination ein gemeinsamer Schätzwert für die gesamte Dual-Frame-Erhebung berechnet. Die ersten Ansätze dieser Art stammen von Hartley (1962) sowie von Fuller und Burmeister (1972) und empfehlen

³⁶ Andere gängige Schätzer, wie bspw. der GREG-Schätzer, werden in den folgenden Kapiteln erläutert.

zur Berechnung eines gemeinsamen Schätzers aus beiden Stichproben die Einführung der beiden Parameter $c_{i,\phi}^F$ sowie $c_{i,\phi}^C$ für die jeweiligen Stichproben. Diese nehmen den Wert

$$c_{i,\phi}^F = \begin{cases} 1 & \forall i \in S^F \cap U^{F \setminus C} \\ \phi & \forall i \in S^F \cap U^{F \cap C} \end{cases} \quad (3.7)$$

bzw.

$$c_{i,\phi}^C = \begin{cases} 1 & \forall i \in S^C \cap U^{C \setminus F} \\ 1 - \phi & \forall i \in S^C \cap U^{F \cap C} \end{cases} \quad (3.8)$$

an, mit $\phi \in [0,1]$. Ein gemeinsamer Schätzer für den Totalwert einer interessierenden Variable y berechnet sich dann nach:

$$\begin{aligned} \hat{Y}_{CW} &= \sum_{i \in S(F)} c_{i,\phi}^F \cdot d_i^F \cdot y_i + \sum_{i \in S(C)} c_{i,\phi}^C \cdot d_i^C \cdot y_i \\ &= \hat{Y}_{S^F \cap U^{F \setminus C}} + \phi \cdot \hat{Y}_{S^F \cap U^{F \cap C}} + (1 - \phi) \cdot \hat{Y}_{S^C \cap U^{F \cap C}} + \hat{Y}_{S^C \cap U^{C \setminus F}}. \end{aligned} \quad (3.9)$$

Somit ist ein geeigneter Schätzwert nach diesem Vorgehen gleich der Summe der beiden Schätzwerte der Only-Domains aus beiden Stichproben ($\hat{Y}_{S^F \cap U^{F \setminus C}}, \hat{Y}_{S^C \cap U^{C \setminus F}}$) sowie der gewichteten Summe der beiden Schätzwerte des Überlappungsbereiches ($\hat{Y}_{S^F \cap U^{F \cap C}}, \hat{Y}_{S^C \cap U^{F \cap C}}$). Ein solcher Schätzer ist genau dann approximativ unverzerrt, wenn für eine Einheit i gilt, dass $c_{i,\phi}^F \approx 1$, falls $i \in S^F \cap U^{F \setminus C}$, $c_{i,\phi}^C \approx 1$, falls $i \in S^C \cap U^{C \setminus F}$ sowie $c_{i,\phi}^F + c_{i,\phi}^C \approx 1$, falls $i \in (S^F \cap U^{F \cap C}) \cup (S^C \cap U^{F \cap C})$ entspricht (vgl. S. L. Lohr (2011), S. 200.; vgl. Brick et al. (2006), S. 786.).

Die Wahl eines Wertes von 0 bzw. 1 für ϕ führt in diesem Ansatz zum Ausschluss der Menge an Elementen *einer* der beiden Stichproben im Überlappungsbereich und wäre daher mit den Ergebnissen eines Screening-Ansatzes vergleichbar. Bezüglich der Wahl des Konvexgewichtes ϕ (auch *Composite-Factor* genannt) empfiehlt Hartley (1962) ein vergleichsweise simples Vorgehen. Das *Fixed Weight Adjustment* verwendet dabei ein bereits zuvor festgelegtes Gewicht zur Kombination der beiden Schätzwerte des Überlappungsbereiches. Vorteilhaft an diesem Vorgehen ist die erreichte interne Konsistenz der Gewichte, da für alle Variablen einer Erhebung die gleiche Menge an angepassten Gewichten verwendet wird. Zusätzlich ist die Implementierung eines solchen Verfahrens verhältnismäßig unkompliziert. Bezüglich der Effizienz eines solchen Schätzers gibt S. L. Lohr (2011) zu bedenken, dass diese stark von den zuvor gewählten Gewichten abhängig ist.

Nach Brick et al. (2006) besteht das einfachste Vorgehen darin, für ϕ einen Wert von $\phi = 0,5$ anzunehmen. Dieses Vorgehen beschränkt die originären Designgewichte einer Erhebungseinheit i aus dem Überlappungsbereich auf die Hälfte, unter der Annahme, dass dies die "doppelte

Erhebung" diese Bereiche ausgleichen kann. In ihrer Untersuchung des amerikanischen *Current Population Surveys (CPS)* schätzten die Autoren einen optimalen Wert für ϕ in Höhe von 0,42. Die geringfügige Abweichung hat den Autoren zufolge nur wenig Einfluss auf die Effizienz des Schätzers (vgl. Brick et al. (2006), S. 786). Andere Untersuchungen, wie bspw. diejenige von Xia et al. (2010), warnen davor, dass gerade dieser, vergleichsweise einfache Ansatz zu ineffizienten Schätzungen führen kann, wenn sich die Auswahlwahrscheinlichkeiten einer Einheit in beiden Auswahlrahmen stark unterscheiden (vgl. Xia et al. (2010), S. 3914ff.).

Eine weitere Möglichkeit zur Bestimmung dieser *Composite-Faktoren* sehen die Autoren daher darin, ϕ proportional zu der Zusammensetzung der Erhebung nach beiden Stichproben zu wählen. Wird bei einer Dual-Frame-Erhebung eine Festnetzstichprobe S^F vom Umfang n^F und eine Mobilfunkstichprobe S^C vom Umfang n^C gezogen, so sollte nach Xia et al. (2010) ϕ durch

$$\phi = \frac{n^F}{n^F + n^C} \quad (3.10)$$

bestimmt werden. Die zugrunde liegende Überlegung an diesem Ansatz ist, dass beide Stichproben versuchen, dieselbe Zielpopulation abzubilden. Ein Vorzug dieses Vorgehens besteht darin, dass größere Gewichte von Einheiten im Überlappungsbereich, die aus kleineren Stichproben kommen, stärker reduziert werden als kleine Gewichte aus größeren Stichproben. Bei gleich großen Stichprobenumfängen entspricht Gleichung 3.10 dem zuvor dargestellten Ansatz.

Alternativ besteht die Möglichkeit, anstelle des Stichprobenumfanges der jeweiligen Stichproben den *effektiven Stichprobenumfang* zu verwenden. Der effektive Stichprobenumfang n_{eff} gibt die Anzahl der Einheiten einer Stichprobe wieder, die unter einer uneingeschränkten Zufallsauswahl benötigt würde, um einen Schätzwert der gleichen Präzision hervorzubringen wie im gegebenen, komplexen Ziehungsverfahren. Dieser Stichprobenumfang lässt sich als das Verhältnis zwischen dem tatsächlich realisierten Stichprobenumfang n und dem *Designeffekt* def der komplexen Stichprobe durch

$$n_{eff} = \frac{n}{def} \quad (3.11)$$

berechnen. Der Designeffekt wiederum gibt die Divergenzen zwischen den Ergebnissen, die unter einer realisierten Stichprobenerhebung vorliegen und denjenigen einer hypothetischen uneingeschränkten Zufallsauswahl wieder. Im Idealfall lässt sich dieser, bezogen auf eine interessierende Variable y , bzw. deren (geschätzten) Totalwert $\hat{Y}$ bestimmen durch das Verhältnis der Varianz dieser Variable unter dem komplexen Stichprobendesign ($Var(\hat{Y})_c$) und demjenigen unter uneingeschränkter Zufallsauswahl ($Var(\hat{Y})_{srs}$). Im besten Fall kann dieser daher nach

$$def = \frac{Var(\hat{Y})_c}{Var(\hat{Y})_{srs}}$$

berechnet werden (vgl. Ganninger (2009), S. 36.). Allerdings ist die Varianz die sich unter einer uneingeschränkten Zufallsstichprobe ergeben würde, bei Daten, die aus einem komplexen Design stammen, oftmals nur schwer zu schätzen. Ähnlich verhält es sich mit der Varianzschätzung innerhalb des komplexen Designs einer Dual-Frame-Erhebung. Erschwerend kommt hinzu, dass unter dem gegebenen Vorgehen ϕ für jede der Variablen unterschiedlich ausfallen würde. Zur vereinfachten Handhabung wird im Rahmen der vorliegenden Arbeit daher empfohlen, bei diesem Vorgehen eine modellbasierte Schätzung des Designeffektes durchzuführen. Grundlage dieser Empfehlung ist einerseits, dass die Varianz der jeweiligen Schätzer nur sehr aufwendig zu bestimmen sind (siehe: Kapitel 8). Andererseits ist die interne Konsistenz von Designgewichten durchaus erstrebenswert für die spätere Schätzung von Erhebungsdaten. Würde man den Designeffekt auf der Basis der Varianz bestimmen, so würde sich dieser in Abhängigkeit einer Variable unterscheiden. Die Konsequenz daraus wären unterschiedliche Designgewichte, die abhängig davon sind, welche Variable man gerade schätzen möchte. Dies würde bei zunehmender Anzahl von Erhebungsvariablen auch die Komplexität der Schätzung und der späteren Analyse der Erhebung steigern. Der modellbasierte Designeffekt hingegen ist, da er lediglich von den Designgewichten der einzelnen Stichproben und deren Klumpung abhängig ist, für alle Variablen dieser Stichprobe konstant³⁷ und führt daher auch zu Composite-Gewichten, die intern konsistent sind.

Für eine ungeklumpte Stichprobe kann dieser modellbasierte Designeffekt $def f^*$ dann durch

$$def f^* = n \cdot \frac{\sum_{i \in S} d_i^2}{(\sum_{i \in S} d_i)^2} \quad (3.12)$$

bestimmt werden.³⁸ In Anlehnung an 3.11 berechnet sich ϕ nach

$$\phi = \frac{n_{eff}^F}{n_{eff}^F + n_{eff}^C}. \quad (3.13)$$

Gerade bei stark variierenden Designgewichten kann dieses Vorgehen neben der reinen Stichprobengröße auch weitere Stichprobeninformationen berücksichtigen. Weiterhin besteht die Möglichkeit, ϕ direkt über das Verhältnis der summierten Designgewichte (der Einheiten des Überlappungsbereiches) zu bestimmen. Xia et al. (2010) merken bei ihrer Untersuchung unterschiedlichster Vorgehensweisen zur Bestimmung von ϕ an, dass die Varianz der Schätzwerte eines Totalwertes Y stark von der Variation der Designgewichte abhängt. Zudem ergibt ihre Untersuchung, dass der Ansatz, der lediglich die Stichprobengrößen verwendet, als unverzerrter

³⁷ Eine Klumpung wird im Zuge der Schätzung auf der Basis von Telefonstichproben i. d. R. ausgeschlossen.

³⁸ Hier wird davon ausgegangen, dass lediglich die unterschiedlichen Auswahlwahrscheinlichkeiten der einzelnen Einheiten varianztreibend sind und kein Klumpeneffekt vorliegt. Bei einer Klumpenstichprobe ist dieser Ansatz deutlich komplexer. Eine ausführliche Erläuterung zu diesem Thema findet sich in Ganninger (2009).

Schätzer herangezogen werden kann (vgl. Xia et al. (2010), S. 3914ff.).

Hartley ([1962, 1974] führt weitergehend an, dass die Berechnung des Composite-Faktors ebenso unter der Berücksichtigung der Varianzminimierung erfolgen kann. Unter der Annahme, dass die Ziehung beider Stichproben einer Dual-Frame-Erhebung unabhängig von einander erfolgt, lässt sich die Varianz des geschätzten Totalwertes $\hat{Y}$ ausdrücken als

$$Var(\hat{Y}) = Var(\hat{Y}_{SF \cap UF \setminus C} + \phi \cdot \hat{Y}_{SF \cap UF \cap C}) + Var(\hat{Y}_{SC \cap UC \setminus F} + (1 - \phi) \cdot \hat{Y}_{SC \cap UF \cap C}).$$

Bei einem *optimalen Schätzer* wird ϕ dabei so gewählt, dass

$$\phi_{opt} = \frac{Var(\hat{Y}_{SC \cap UF \cap C}) + Cov(\hat{Y}_{SC \cap UC \setminus F}, \hat{Y}_{SC \cap UF \cap C}) - Cov(\hat{Y}_{SF \cap UF \setminus C}, \hat{Y}_{SF \cap UF \cap C})}{Var(\hat{Y}_{SC \cap UF \cap C}) + Var(\hat{Y}_{SF \cap UF \cap C})}. \quad (3.14)$$

Da die dafür notwendigen Varianzen und Kovarianzen i. d. R. bei einer derartigen Erhebung unbekannt sind, empfiehlt es sich zur Berechnung von ϕ_{opt} geeignete Schätzwerte für die jeweiligen Parameter zu verwenden. Die in diesem Vorgehen verwendeten Designgewichte sind jedoch als Funktionen der Varianzen und Kovarianzen der Bereichstotalwerte zu bewerten und somit von den entsprechenden Zufallsvariablen abhängig. Weiterhin sind die Designgewichte nicht konstant für alle Variablen. Daher muss deren Variabilität in der Berechnung der Standardfehler der Schätzwerte berücksichtigt werden. Zusätzlich ergibt sich nach diesem Vorgehen für jede betrachtete Variable ein neues Designgewicht. Dies kann zu Inkonsistenzen in der Schätzung führen, da sich die optimalen Gewichte für jede betrachtete Variable unterscheiden (vgl. S. L. Lohr (2011), S. 200.; vgl. S. L. Lohr (2007), S. 3259.).

Eine Modifikation des von Hartley angeführten Schätzers wird durch Fuller und Burmeister (1972) angeführt. Diese sieht vor, neben der Konvexkombination des Überlappungsbereiches noch die Schätzwerte der Größe der Zielpopulation im Überlappungsbereich bei der Schätzung zu inkludieren. Hierfür muss in einem ersten Schritt die Anzahl derjenigen Einheiten der Zielpopulation, die sich in diesem Überlappungsbereich befinden, anhand der Festnetz- ($\hat{N}_{SF \cap U^F C}$) und Mobilfunkstichprobe ($\hat{N}_{SC \cap U^F C}$) geschätzt werden. Deren Differenz wird dann unter Verwendung eines weiteren Gewichtungsfaktors (ψ) in der Schätzung zusätzlich verwendet. Ein entsprechender Schätzwert des Totalwertes einer interessierenden Variable ($\hat{Y}_{FB}$) lässt sich dann durch

$$\begin{aligned} \hat{Y}_{FB} = & \hat{Y}_{SF \cap UF \setminus C} + \phi \cdot \hat{Y}_{SF \cap UF \cap C} + (1 - \phi) \cdot \hat{Y}_{SC \cap UF \cap C} \\ & + \hat{Y}_{SC \cap UC \setminus F} + \psi \cdot (\hat{N}_{SF \cap U^F C} - \hat{N}_{SC \cap U^F C}) \end{aligned} \quad (3.15)$$

errechnen. S. L. Lohr (2007) führt dazu an, dass auch für diesen Schätzer eine Wahl der Composite-Faktoren so gestaltet werden kann, dass dieser Schätzer varianzminimierend wird. Allerdings

gelten hierbei ähnliche Probleme wie bei dem vorherigen optimalen Schätzer. Demnach müssen auch hierfür, da die notwendigen Varianzen und Kovarianzen nicht bekannt sind, entsprechende Werte aus den Stichproben geschätzt werden (vgl. S. L. Lohr (2007), S. 3260.).

3.3.3 Weitere Designgewichtungsverfahren für Dual-Frame-Erhebungen

In der vorherigen Darstellung zur Bestimmung zur Wahl eines optimalen Composite-Faktors ist grundsätzlich davon ausgegangen worden, dass die Allokation der abgeschlossenen Interviews in einer Dual-Frame-Erhebung extern fixiert wurde. Unter der Berücksichtigung der Kosten eines abgeschlossenen Interviews wird jedoch diese Allokation (und damit verbunden auch die Wahl eines Composite-Faktors) genau dann interessant, wenn abgeschlossene Interviews aus einem Auswahlrahmen höhere Kosten verursachen als diejenigen des anderen. Gerade unter dem Aspekt der Budgetrestriktion kamen zu Beginn der 1960er-Jahre erste Überlegungen zur Durchführung von Dual-Frame-Erhebungen auf. Wie bereits in Kapitel 2.3 beschrieben, war bei den ersten Untersuchungen solcher Erhebungen die Zielpopulation einer Auswahlgrundlage oftmals Teilmenge einer anderen, deren Erhebungskosten jedoch geringer waren.

Geht man nun von einem fest vorgegebenem Budget T aus, so setzt sich dieses nach

$$T \geq t^0 + t^F \cdot n^F + t^C \cdot n^C$$

zusammen. t^0 bezeichnet hierbei die fixen Kosten einer Erhebung, die unabhängig von der Allokation der jeweiligen Nettostichprobengrößen in einer Erhebung entstehen. Da diese ohnehin anfallen und für die Allokation nicht weiter relevant sind, wird im Rahmen der folgenden Betrachtung davon ausgegangen, dass sie bereits vom zur Verfügung stehenden Budget abgezogen wurden. t^0 nimmt daher einen Wert von Null an. t^F und t^C geben die durchschnittlichen Kosten pro abgeschlossenem Interview wieder, während n^F und n^C der (geplanten) Anzahl der abgeschlossenen Interviews entsprechen. Die Kosten eines realisierten Mobilfunkinterviews können in Relation zu einem abgeschlossenen Festnetzinterview ausgedrückt werden. In der tatsächlichen Umfragepraxis können die Kosten eines abgeschlossenen Mobilfunkinterviews aufgrund unterschiedlicher Längen der Befragung oder einer unterschiedlich gearteten Salienz bezüglich des Befragungsthemas nach Modus variieren. Da die durchschnittlichen Kosten pro abgeschlossenem Interview jedoch (innerhalb einer Erhebung) als konstant erachtet werden können, gilt für die Kostenrelation v

$$v = \frac{t^C}{t^F}.$$

Aus mehreren Gründen, wie bspw. der geringeren Hitrate bei Mobilfunkstichproben oder höherer Anrufrufen, wird dabei angenommen, dass ein abgeschlossenes Interview im Mobilfunk teurer

ist, als eines im Festnetz. Im Rahmen einer solchen Allokation kann entweder versucht werden, bei einer vorgegebenen Varianz eines Schätzwertes das Budget zu optimieren oder bei gegebenem Budget, die Varianz zu minimieren. Da i. d. R. das Budget als (extern) vorgegeben betrachtet werden kann, wird im Folgenden die kostenoptimale Allokation dargestellt, welche die Varianz minimieren soll. Weiterhin wird von einem Overlap-Ansatz ausgegangen.³⁹

Zur Bestimmung einer kostenoptimalen Allokation stellen Callegaro et al. (2011) einen Ansatz unter Verwendung der Daten der CELLA 1-Erhebung vor. Im Gegensatz zum ursprünglichen Gewichtungsansatz dieser Erhebung wird dabei jedoch ein Composite-Ansatz zur Ermittlung der Designgewichte verwendet.

Da im Vorfeld i. d. R. unbekannt ist, ob eine generierte Rufnummer zu einem geschalteten Anschluss führt und sich das Teilnahmeverhalten in beiden Stichproben unterscheiden kann, wird in diesem Ansatz nach einer Allokation der Bruttostichprobe anstelle der Nettostichprobe gesucht. Diese lässt sich als Verhältnis der Kosten pro abgeschlossenem Interview (v) sowie der Relation der jeweiligen Varianzen in beiden Stichproben ausdrücken (τ). Das Verhältnis der beiden Bruttostichproben berechnet sich dann nach

$$\frac{m^F}{m^C} = \sqrt{\tau \cdot v}, \quad (3.16)$$

mit

$$\tau = \frac{Var^F(\kappa^F)}{Var^C(\kappa^C)} = \frac{\sum_{i \in U^F} \kappa_i^F \left(\frac{\theta_i^F}{\kappa_i^F} - Y^F \right)^2}{\sum_{i \in U^C} \kappa_i^C \left(\frac{\theta_i^C}{\kappa_i^C} - Y^C \right)^2},$$

$$\kappa_i^F = \frac{\pi_i^F}{m^F} \forall i \in U^F; \quad \kappa_i^C = \frac{\pi_i^C}{m^C} \forall i \in U^C$$

und

$$\theta_i^F = c_i^F \cdot y_i \forall i \in U^F; \quad \theta_i^C = c_i^C \cdot y_i \forall i \in U^C.$$

c_i^F und c_i^C entsprechen hierbei den Composite-Faktoren wie sie bspw. in Kapitel 3.3.2 beschrieben wurden. Bei ihrer Untersuchung am Beispiel der Erhebung CELLA 1 bemerken die Autoren, dass unter dem Gesichtspunkt der Budgetrestriktion wohl eine andere Allokation (der Nettointerviews) zu bevorzugen gewesen wäre als 50% über das jeweilige Medium (vgl. Callegaro et al. (2011), S. 17f.). Eine ähnliche Empfehlung wird seitens des ADMs ebenfalls geäußert. Dieser empfiehlt nach kostenoptimaler Allokation einen Mobilfunkanteil von 30% bis 40% (vgl. Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute ADM (2012), S. 29ff.). Weiterhin von Interesse ist in diesem Kontext jedoch nicht nur die Bruttostichprobengröße oder die angestrebte Anzahl von Nettointerviews, sondern die *effektive Stichprobengröße*. Ein solcher Ansatz, der

³⁹ Levine und Harter (2014) sowie Wolter et al. (2015) diskutieren sowohl die Aufteilung bei vorgegebener Varianz als auch unter der Verwendung eines Screening-Ansatzes.

diese unter Budgetrestriktionen berücksichtigt, findet sich bspw. bei Levine und Harter (2014).

Um die Probleme eines optimalen Schätzers nach Fuller und Burmeister zu umgehen, beschreiben Skinner und J. N. K. Rao (1996) ein Vorgehen zur Bestimmung von Composite-Faktoren, das im Rahmen einer Dual-Frame-Erhebung für alle Variablen die gleichen Gewichte erzeugt. Demnach besteht in diesem Ansatz nicht mehr das zuvor geschilderte Problem der internen Inkonsistenz. Im Gegensatz zu dem Ansatz von Fuller und Burmeister ist es darüber hinaus möglich, einen solchen Gewichtungsansatz auch auf komplexe Stichproben anzuwenden. Demnach muss im Rahmen dieses sog. **Pseudo-Maximum Likelihood Schätzers** nicht mehr eine uneingeschränkte Zufallsauswahl zugrunde liegen. Ein Schätzer für den Totalwert Y einer interessierenden Variable berechnet sich in diesem Ansatz nach

$$\hat{Y}_{PML} = (N^F - \hat{N}_{PML}^{F \cap C})E(y^{F \setminus C}) + (N^C - \hat{N}_{PML}^{F \cap C})E(y^{C \setminus F}) + \hat{N}^{F \cap C}E(y^{F \cap C}), \quad (3.17)$$

mit

$$\begin{aligned} E(y^{F \setminus C}) &= \sum_{i \in S^F \cap U^{F \setminus C}} \frac{y_i}{n^{S^F \cap U^{F \setminus C}}}; \quad E(y^{C \setminus F}) = \sum_{i \in S^C \cap U^{C \setminus F}} \frac{y_i}{n^{S^C \cap U^{C \setminus F}}}; \\ E(y^{F \cap C}) &= \frac{\frac{n^F}{N^F} \cdot \hat{N}^{S^F \cap U^{F \cap C}} \cdot E(y^{S^F \cap U^{F \cap C}}) + \frac{n^C}{N^C} \cdot \hat{N}^{S^C \cap U^{F \cap C}} \cdot E(y^{S^C \cap U^{F \cap C}})}{\frac{n^F}{N^F} \cdot \hat{N}^{S^F \cap U^{F \cap C}} + \frac{n^C}{N^C} \cdot \hat{N}^{S^C \cap U^{F \cap C}}}. \end{aligned}$$

Zur Bestimmung von $\hat{N}_{PML}^{F \cap C}$ wird die kleinere Wurzel der Gleichung

$$px^2 - qx + h = 0$$

mit

$$p = n^F + n^C$$

$$q = n^F N^F + n^C N^C + n^F \hat{N}^{S^F \cap U^{F \cap C}} + n^C \hat{N}^{S^C \cap U^{F \cap C}}$$

$$h = n^F \hat{N}^{S^F \cap U^{F \cap C}} N^C + n^C \hat{N}^{S^C \cap U^{F \cap C}} N^F$$

verwendet (vgl. S. L. Lohr und J. N. K. Rao (2000), S. 272.).

Unter der Verwendung eines Composite-Ansatzes lassen sich im Kontext dieses Gewichtungsver-

fahrens die Gewichte wie folgt bestimmen:

$$c_{i,PML}^F = \begin{cases} \frac{N^F - \hat{N}_{PML}^{F \cap C}}{\hat{N}_{S^F \cap U^F \setminus C}^F} & \forall i \in S^F \cap U^F \setminus C \\ \frac{\hat{N}_{PML}^{F \cap C}}{\gamma \cdot \hat{N}_{S^F \cap U^F \cap C}^F + (1-\gamma) \cdot \hat{N}_{S^C \cap U^F \cap C}^C} \cdot \gamma & \forall i \in S^F \cap U^F \cap C \end{cases} \quad (3.18)$$

und

$$c_{i,PML}^C = \begin{cases} \frac{N^C - \hat{N}_{PML}^{C \cap F}}{\hat{N}_{S^C \cap U^C \setminus F}^C} & \forall i \in S^C \cap U^C \setminus F \\ \frac{\hat{N}_{PML}^{C \cap F}}{\gamma \cdot \hat{N}_{S^F \cap U^F \cap C}^F + (1-\gamma) \cdot \hat{N}_{S^C \cap U^F \cap C}^C} \cdot (1-\gamma) & \forall i \in S^C \cap U^F \cap C \end{cases} \quad (3.19)$$

N^F bzw. N^C stellen hierbei die Anzahl der Einheiten der Zielpopulation dar, die über die entsprechende Auswahlgrundlage erreichbar sind. Die Schätzwerte $\hat{N}_{S^F \cap U^F \setminus C}^F$, $\hat{N}_{S^C \cap U^C \setminus F}^C$, $\hat{N}_{S^F \cap U^F \cap C}^F$ und $\hat{N}_{S^C \cap U^F \cap C}^C$ entsprechen denjenigen, die für die jeweiligen Teilpopulationen (Onlys und Dual-User) aus der Festnetz- und Mobilfunkstichprobe gewonnen wurden. γ wird in diesem Ansatz durch

$$\gamma = \frac{N^F \setminus C \cdot N^C \cdot Var(\hat{N}_{S^C \cap U^F \cap C}^C)}{N^F \setminus C \cdot N^C \cdot Var(\hat{N}_{S^C \cap U^F \cap C}^C) + N^C \setminus F \cdot N^F \cdot Var(\hat{N}_{S^F \cap U^F \cap C}^F)} \quad (3.20)$$

bestimmt. Zur Bestimmung von γ ist die Kenntnis von Parametern notwendig, deren Größen in der Zielpopulation oftmals unbekannt sind. Bspw. gibt es im Rahmen von Dual-Frame-Erhebungen lediglich Schätzwerte über die tatsächliche Anzahl der Einheiten, die über ein bestimmtes Medium erreicht werden können. Daher empfiehlt S. L. Lohr (2011) anstelle von γ den Schätzwert $\hat{\gamma}$ zu verwenden. Die Autorin merkt an, dass bei der Übereinstimmung der Zielpopulationen, oder wenn eine Zielpopulation Teilmenge der anderen ist, $\hat{\gamma}$ als durchschnittlicher Designeffekt beider Stichproben geschätzt werden kann. Allgemein kann dieser Schätzer als effizient erachtet werden, auch wenn angemerkt wird, dass dieser gelegentlich nicht den geringsten *Mean Squared Error* (MSE) liefert (vgl. S. L. Lohr (2011), S. 201.).

Neben den hier angeführten Schätzern gibt es für Dual-Frame-Erhebungen noch weitere Ansätze, die jedoch im Rahmen dieser Arbeit nicht weiter erläutert werden. Eine Übersicht über jene Verfahren findet sich bspw. in S. L. Lohr (2011), Winglee et al. (2007), Mecatti (2007) oder Sand (2014).

3.4 Kalibrierung und Anpassungsgewichtung

Der durch eine Designgewichtung bestimmte Schätzer (bspw. der Horvitz-Thompson-Schätzer) für einen Parameter der Zielpopulation gilt immer dann als unverzerrt für den wahren Wert in der

Population, wenn alle Einheiten eine positive Auswahlwahrscheinlichkeit aufweisen und zusätzliche Fehlerquellen, wie bspw. *Undercoverage* oder *Nonresponse*, ausgeschlossen werden können. Allerdings sind die meisten Erhebungen von Nonresponse betroffen (vgl. Gabler et al. (2011), S. 162.). Darüber hinaus kann ein systematischer Unterschied zwischen teilnehmenden und nicht teilnehmenden Zielpersonen bestehen. Um dennoch die Lücke zwischen dem geschätzten Wert, basierend auf einer Erhebung, und dem wahren Wert der Zielpopulation zu schließen, bzw. zumindest die Verzerrung zu verringern, wird oftmals, im Anschluss an die Designgewichtung, eine Anpassungsgewichtung durchgeführt.⁴⁰

Die Menge der Stichprobenelemente bereinigt um die Nicht-Respondenten ergibt dabei die Menge der Respondenten ($R = \{r_1, \dots, r_i, \dots, r_n\} \in (0,1)^n$), mit

$$r_i = \begin{cases} 1 & \text{wenn Element } i \text{ Respondent ist} \\ 0 & \text{wenn Element } i \text{ Nicht-Respondent ist} \end{cases}, \forall i \in S.$$

Systematische Unterschiede zwischen der Menge der Respondenten und der Nicht-Respondenten können dabei zu verzerrten Schätzergebnissen führen. In diesem Zusammenhang gilt es, die Ursache der Antwortverweigerung genauer zu untersuchen. Die Wahrscheinlichkeit, dass eine Einheit der Stichprobe antwortet, oder die *Propensity Score*, wird dabei durch

$$\eta_i = Pr(r_i = 1)$$

beschrieben. Ein entsprechender Schätzer für eine Variable y_i lässt sich dabei nach

$$\hat{Y} = \sum_{i \in S} \frac{y_i}{\pi_i \cdot \eta_i}$$

bestimmen. Bei der Berücksichtigung der (Unit-) Nonresponse anhand einer Anpassungsgewichtung merkt S. L. Lohr (2009) an, dass die zugrunde liegenden Modelle starke Annahmen bezüglich des Ausfallmechanismus treffen, die jedoch in den seltensten Fällen den wahren Sachverhalt vollständig widerspiegeln (vgl. S. L. Lohr (2009), S. 338ff.). Diesbezüglich werden i. d. R. Modelle betrachtet, die sich auf den Zusammenhang einer Untersuchungsvariablen y bzw. der Ausprägung dieser bei einer Einheit i (y_i) und einem Vektor mit zusätzlichen Informationen (bspw. soziodemografische Charakteristika) der Einheit i ($\mathbf{x}_i$) beziehen. Wenn die Wahrscheinlichkeit der Teilnahme η_i weder von y_i oder $\mathbf{x}_i$, noch von dem Erhebungsdesign abhängig ist, so spricht man von *Missing Completely at Random* (MCAR). Bei Nonresponse aufgrund von *Missing at Random* (MAR) ist η_i von $\mathbf{x}_i$, aber nicht von y_i abhängig. Bei einem Ausfall nach *Not Missing at Random* (NMAR) hingegen besteht eine Abhängigkeit zwischen

⁴⁰ Die folgenden Abschnitte beschreiben Kalibrierungsverfahren im Allgemeinen und gelten deshalb nicht bloß für telefonische Befragungen.

η_i und y_i und der Ausfallmechanismus kann nicht (vollständig) anhand der Hilfsvariablen (x_i) erklärt werden (vgl. S. L. Lohr (2009), S. 17f.; S. 338f.).

In der Praxis wird bei Vorliegen von Unit-Nonresponse oftmals MAR als Ursache angenommen. Dabei zeigen Erfahrungen aus verschiedensten Nonresponseanalysen, dass bei Erhebungen von Individuen oftmals eine geringere *Response rate* (RR) zum Beispiel für Personen aus Großstädten, Alleinstehende oder Personen mit geringem Bildungsabschluss zu beobachten ist (für weitere Faktoren siehe bspw. Lundström und Särndal (2001), S. 23f.; Montaquila et al. (2008), S. 583.; Schneiderat und Schlinzig (2009c), S. 138f.; Schneiderat und Schlinzig (2011), S. 126ff.).

Die Art des zugrunde liegenden Nonresponse ist also von entscheidender Bedeutung, um das Ausmaß der dadurch verursachten Verzerrung (*Nonresponse Bias*) zu bewerten. Allgemein kann Nonresponse zu einer Reduktion der Genauigkeit der Schätzergebnisse aufgrund einer reduzierten effektiven Stichprobengröße sowie einer steigenden Variation der Erhebungsgewichte führen. Der Verzerrung durch systematische Unterschiede zwischen teilnehmenden und nicht-teilnehmenden Einheiten wird jedoch i. d. R. eine dominante Rolle in der Begutachtung der Fehlerquellen zugesprochen (vgl. Brick (2013), S. 330.).

Aufgrund der starken Annahmen, die durch ein solches Modell angestellt werden, ist anzumerken, dass die Kalibrierung nicht als Allheilmittel gegen die Verzerrung verstanden werden darf, die durch Nonresponse bewirkt wird. Vielmehr besteht die Aufgabe einer solchen Gewichtung darin, die Verzerrung, die aufgrund der (systematischen) Teilnahmeverweigerung entstehen kann, zu verringern. Hierbei gilt nochmals hervorzuheben, dass dies umso besser gelingt, je mehr das zugrunde liegende Modell des Ausfalls einer erhobenen Einheit dem tatsächlichen Grund für das Zustandekommen fehlender Werte entspricht.

Neben der Reduktion des Nonresponse Bias kann die Kalibrierung jedoch auch eingesetzt werden, um die Präzision der Schätzergebnisse zu steigern. Dabei kann es bspw. auch ohne das Aufkommen von Nonresponse im Rahmen einer Erhebung erstrebenswert sein, eine Kalibrierung der Ergebnisse durchzuführen. Ein Beispiel bei einer uneingeschränkten Zufallsauswahl wäre die Kalibrierung mittels Regressions- oder Verhältnisschätzer, wenn zwischen den Untersuchungsvariablen und einer (oder mehrerer) Hilfsvariable(n) ein linearer Zusammenhang besteht.

Zusätzlich ist es gerade im Kontext von Telefonstichproben denkbar, dass Informationen über die Auswahlgrundlage vorliegen, diese allerdings für eine Schichtung vor der Stichprobenziehung nicht geeignet sind. Bei Telefonerhebungen besteht die Auswahlgrundlage oftmals lediglich aus Rufnummern, die zwar bestimmte Informationen, wie Netzanbieter oder geografische Lage beinhalten können, jedoch keine Schichtung nach soziodemografischen Charakteristika erlauben. Eine geschichtete Zufallsauswahl anhand von (soziodemografischen) Schichtinformationen ist daher im Zuge von Telefonstichproben nicht möglich. Die nachträgliche Schichtung (die durch die Kalibrierung bewirkt wird,) erlaubt hierbei, die im Rahmen der Erhebung gewonnenen Schichtinformationen in den (finalen) Schätzwert eines interessierenden Parameters

einzu beziehen.

Somit kann das Ziel der Kalibrierung bzw. der Anpassungsgewichtung einmal in der **Steigerung der Präzision der Schätzergebnisse** liegen und darüber hinaus in der **Verringerung der Verzerrung durch (systematischen) Nonresponse**.

Die Anpassungsgewichtung aufgrund von Nonresponse kann infolgedessen als Sonderfall der Kalibrierung betrachtet werden. Für die Hilfsvariablen, die für die Anpassungsgewichtung verwendet werden, stimmen die geschätzten Werte dann mit den Werten der Zielpopulation überein. Entscheidend für eine sachgerechte Anpassungsgewichtung ist, wie aus der MAR-Annahme implizit hervorgeht, dass die Hilfsvariablen mit der Untersuchungsvariablen stark korrelieren und dadurch eine Verbesserung der Präzision erwirkt wird. Daher ist die Wahl der Hilfsvariablen von großer Bedeutung (vgl. Lumley (2010), S. 135ff.). Für Mehrthemenbefragungen (wie bspw. die *Allgemeine Bevölkerungsumfrage für Sozialwissenschaften (ALLBUS)* oder dem *European Social Survey (ESS)*) gestaltet sich die Wahl der Hilfsvariablen erheblich schwieriger. Da bei diesen Umfragen der Fokus nicht auf einer einzigen Variable liegt, für die die Anpassung eine möglichst große Annäherung zur tatsächlichen Verteilung erreicht werden soll, sondern mehrere Untersuchungsvariablen von Interesse sind, werden höhere Anforderungen an die Kalibrierung gestellt. Dies ergibt sich daraus, dass es i. d. R. wünschenswert ist, möglichst nur ein Kalibrierungsvariable für die gesamte Erhebung zu haben. Daher wird im Zuge einer solchen Kalibrierung oftmals versucht, Hilfsvariablen zu finden, die mit dem Antwortverhalten zusammenhängen, um so auf diese anzupassen. Da das Ziel dieser Kalibrierung gewöhnlich nicht darin besteht, unverzerrte Schätzergebnisse für eine einzige Variable zu liefern, sondern für alle in der Erhebung zu findenden Variablen möglichst unverzerrt zu schätzen, ist die Homogenität im Antwortverhalten von deutlich höherer Relevanz. Dies bedeutet, dass nach der Kalibrierung das Antwortverhalten nicht mehr durch bestimmte soziodemografische Charakteristika moderiert wird, bzw. eine Homogenität des Antwortverhalten innerhalb der Hilfsvariablen erreicht wird. Lumley (2010) bemerkt diesbezüglich, dass dies zwar kein realistisches Vorhaben ist, dass jedoch die Schätzergebnisse nach einer Kalibrierung weniger verzerrt sind als zuvor (Lumley (2010), S. 153.).

Unabhängig davon, wofür eine Kalibrierung eingesetzt wird, ist eine Voraussetzung für die Hilfsvariablen, dass sowohl deren Stichproben- als auch Populationswerte bekannt sind.

Weiterhin führen Särndal und Lundström (2005) folgende Eigenschaften an, die für die Hilfsvariablen wünschenswert sind:

- Zwischen den gewählten Hilfsvariablen und dem Antwortverhalten sollte ein eindeutiger Zusammenhang bestehen.
- Ein explanatorischer Zusammenhang zwischen den Hilfsvariablen und den Hauptuntersuchungsmerkmalen sollte gegeben sein.

- Die bedeutendsten Wertebereiche/Domains der Erhebung werden durch die Hilfsvariablen (näherungsweise) identifiziert. Des Weiteren beinhaltet der Vektor der Hilfsvariablen alle Merkmale, die für eine spätere Betrachtung nach Subgruppen herangezogen werden.

Diese Eigenschaften decken sich weitestgehend mit der MAR-Annahme (vgl. Särndal und Lundström (2005), S. 22.).

Brick et al. (2005) sowie Kennedy (2010) führen als mögliche Grundlagen zur Begründung von Nonresponse, die gerade im Rahmen von Dual-Frame-Telefonbefragungen von Relevanz sind, neben *soziodemografischen Charakteristika* auch *Topic Salience* an.

- Ausfälle aufgrund von **soziodemografischen Charakteristika** treten für gewöhnlich immer dann auf, wenn bestimmte Eigenschaften oder Merkmale einer Person oder eines Haushaltes einen direkten Einfluss auf das Teilnahme- und/oder Antwortverhalten bei einer Erhebung haben. Typischerweise werden solche Ursachen bspw. im Rahmen der MAR-Annahme, die in Kapitel 3 beschrieben wurde, erfasst. Im Zuge der CELLA 2 - Erhebung wurden solche Ursachen durch die Kalibrierung berücksichtigt. Exemplarisch ist in diesem Zusammenhang der sog. Bildungsbias zu nennen, der das Phänomen beschreibt, dass gerade Einheiten mit einem höheren, formalen Bildungsabschluss eher dazu neigen, an einer (telefonischen) Erhebung teilzunehmen.
- **Topic Salience** hingegen lässt sich unter anderem dann beobachten, wenn das Erhebungsinstrument selbst mit dem Gegenstand der Erhebung in Verbindung steht. Eine solche Ursache kann bspw. bei der Untersuchung des *Practicum Surveys* vermutet werden. In diesem wurde mitgeteilt, dass sie im Rahmen einer Erhebung, die sich mit neuen Technologien und mobiler Kommunikation beschäftigt, ausgewählt wurden. Innerhalb dieser Erhebung war zu beobachten, dass sich das Antwortverhalten der über das Mobilfunknetz Befragten von demjenigen der Festnetzbefragten unterscheidet. Die Verweigerungsrate derjenigen, die über das Mobilfunknetz kontaktiert wurden, war geringer. Dabei besteht die Vermutung, dass diejenigen, die über das Mobilfunknetz zu diesem Thema befragt wurden, auch ein größeres Interesse an der thematischen Ausrichtung der Erhebung hatten (vgl. Brick et al. (2005), S. 2794ff.; Kennedy (2010), S. 18ff.). Auch im Zuge der CELLA 1-Erhebung ließen sich ähnliche Beobachtungen machen. Diese verwendete für ca. 70 % der Rufnummern in der Mobilfunkstichprobe eine Ankündigungs-SMS, die potentielle Teilnehmer darauf hinwies, dass sie für eine Befragung ausgewählt wurden, die sich mit dem Thema *Mobilfunk* auseinandersetzt. Die Ausschöpfungsrate mit Ankündigung lag dabei ca. 7 % höher als ohne diese. Schneiderat und Schlinzig (2009) begründen den Unterschied mit dem Effekt, der sich zwischen sog. *kalten* und *warmen* Kontakten einstellt. Jedoch können die Resultate ebenfalls vermuten lassen, dass durch die thematische

Ankündigung gerade diejenigen, die an diesem Thema interessiert sind, eher teilnehmen (vgl. Schneiderat und Schlinzig (2009d), S. 95ff.). Somit lässt sich nicht ausschließen, dass Einheiten, die eine Affinität zu diesem Thema besitzen, auch eher positiv auf eine solche Ankündigung (gerade über dieses Telekommunikationsmedium) reagieren.

Die Berechnung eines Anpassungsgewichtes für eine Erhebungseinheit i (g_i) kann hierbei anhand von mehreren Verfahren durchgeführt werden. So kann bspw. eine *Poststratifizierung* oder ein *Raking*-Verfahren durchgeführt werden. Eine Besonderheit stellt in diesem Kontext die Kombination von Design- und Anpassungsgewichtung mittels *Generalized Regression Estimator* (GREG-Schätzer) dar. Die drei unterschiedlichen Vorgehensweisen werden folgend kurz beschrieben und abschließend verglichen.

3.4.1 Poststratifizierung (nachträgliche Schichtung)

Ein gängiges Verfahren bei Personenerhebungen ist die Poststratifizierung. Als Hilfsvariablen werden dabei oftmals Indikatoren soziodemografischer Charakteristika wie bspw. *Alter*, *Geschlecht* und *Bildungsabschluss* verwendet. Erforderlich für die Verwendung bestimmter Hilfsvariablen ist, dass alle Merkmalskombinationen in der Zielpopulation vorhanden sind (vgl. Valliant et al. (2013), S. 353.). Zur Anpassungsgewichtung werden dann die jeweiligen Poststrata ($h = 1, \dots, H$) der Merkmalskombinationen verwendet. Bei einer Anpassungsgewichtung nach den obigen Charakteristika muss somit die Kontingenztafel *Alter X Geschlecht X Bildungsabschluss* für die Zielpopulation verfügbar sein. Die Schichten werden in diesem Kontext auch oftmals als Gewichtungszellen bezeichnet. Die eigentliche Gewichtung erfolgt dann nach dem *Soll-durch-Ist*-Prinzip. Weiterhin ist es erforderlich, dass alle Gewichtungszellen (v. a. in den Erhebungsdaten) besetzt sind (vgl. Sand und Gabler (2018), S. 11.).

Das Ziel einer Poststratifizierung besteht bei einer gegebenen Stichprobe S darin, für eine Einheit i in einer Schicht h Gewichte g_i^{POST} zu entwickeln, für die gilt:

$$\sum_{i \in S} g_i^{POST} \cdot d_i \cdot v_{hi} = \sum_{i \in R_h} g_i^{POST} \cdot d_i = N_h. \quad (3.21)$$

R_h entspricht hier der Menge der Respondenten in einer Schicht. Diese sind eine Teilmenge der Menge der Einheiten in der gegebenen Stichprobe S . Die binäre Variable v_{hi} nimmt bei dieser Betrachtung den Wert 1 an, wenn eine Stichprobeneinheit i antwortet und 0, wenn diese die Teilnahme verweigert. Die Summe der Variable v_{hi} ergibt demnach die Anzahl der Respondenten innerhalb einer bestimmten Schicht h und kann als Responseindikator verstanden werden. N_h gibt dabei die Anzahl der Einheiten der Schicht h in der Zielpopulation U an (vgl. S. L. Lohr (2009), S. 343). Bei einer vorliegenden Erhebung kann die Anzahl der Einheiten in einer Schicht anhand des designgewichteten Schätzwertes ($\hat{N}_h$) als Ausgangswert herangezogen

werden. Dieser ermittelt sich durch

$$\hat{N}_h = \sum_{i \in R_h} d_i.$$

Das Anpassungsgewicht g_i^{POST} für eine Einheit i in Zelle h bestimmt sich dann anhand von

$$g_i^{POST} = \frac{N_h}{\hat{N}_h} \text{ (vgl. Valliant et al. (2013), S. 353.)} \quad (3.22)$$

Im Zähler dieser Gleichung steht die Anzahl der Einheiten der Zielpopulation in Schicht h . Der Nenner ist der entsprechende Schätzwert, der aus der Erhebung ermittelt wurde. Unter der Verwendung eines Schichtungstableaus entspricht N_h der Anzahl der Einheiten der Zielpopulation, die eine bestimmte Merkmalskombination aufweisen. So könnte, bspw. bei der Kalibrierung auf Alter, Geschlecht und Bildungsabschluss, eine Zelle des Schichtungstableaus die Anzahl derjenigen Einheiten der Zielpopulation angeben, die männlich und zwischen 20 und 30 Jahren sind und einen mittleren Bildungsabschluss aufweisen. $\hat{N}_h$ gibt den Schätzwert für die Anzahl der Einheiten der Zielpopulation innerhalb einer Zelle an. Dieser wird aus der Stichprobe geschätzt. Der Quotient zwischen tatsächlicher und geschätzter Anzahl der Einheiten entspricht dann dem Kalibrierungsgewicht. Für jede Einheit der Stichprobe wird so, in Abhängigkeit der zugehörigen Schicht der Einheit, ein Kalibrierungsgewicht bestimmt. Dieser Ansatz liefert daher (maximal) H unterschiedliche Kalibrierungsgewichte.

3.4.2 Raking

Im Zuge einer (geplanten) Poststratifizierung tritt oftmals das Problem auf, dass nicht alle Merkmalskombinationen, nach denen gewichtet werden soll, in der Zielpopulation verfügbar sind oder die jeweiligen Zellen in den Erhebungsdaten nicht (ausreichend) besetzt sind. Die Verwendung vieler Anpassungsvariablen kann dabei mitunter dazu führen, dass mehr Gewichtungszellen gebildet werden müssen, als Einheiten in der Erhebung vorhanden sind. Somit kann es vorkommen, dass die Zellbesetzung der Erhebungsdaten Null entspricht, obwohl Einheiten einer entsprechenden Merkmalskombination in der Zielpopulation vorhanden sind.⁴¹ Weiterhin können auch für die Zielpopulation die Zellkombinationen fehlen. Dadurch können lediglich die Randverteilungen der einzelnen Merkmale und nicht die Verteilung von Merkmalskombinationen vorhanden sein. Beide Voraussetzungen verhindern die Durchführung einer Poststratifizierung.

In solchen Fällen kann ein *Raking*-Verfahren, wie bspw. der *Iterative Proportional Fitting (IPF)*-Algorithmus, dazu dienen, dass zumindest die Verteilung der Ränder der Erhebungsdaten mit denen der Zielpopulation übereinstimmt. Dies erlaubt eine Schichtung nach mehreren Merkmalen, ohne dass die Häufigkeiten der einzelnen Zellen der Zielpopulation bekannt oder alle

⁴¹ Allgemein unterscheidet man hierbei zwischen einer strukturellen Null und einer Stichprobennull.

Zellen der Erhebungsdaten besetzt sind. Das Vorgehen gleicht dem der Poststratifizierung, jedoch werden die Randverteilungen der einzelnen Variablen der Erhebungsdaten in jedem separaten Schritt an diejenigen der Zielpopulation angepasst. Dieses Prozedere wird so lange wiederholt, bis sich die Anpassungsgewichte nicht mehr (oder kaum noch) verändern und die kalibrierten Randverteilungen der Hilfsvariablen der Erhebung denjenigen der Zielpopulation entsprechen (vgl. Lumley (2010), S. 131).

Exemplarisch kann dies durch die Anpassung an die beiden Hilfsvariablen *Alter* und *Bildungsabschluss* dargestellt werden, für die angenommen wird, dass lediglich die Randverteilungen in der Zielpopulation verfügbar sind. Die Anpassungsgewichte werden dann durch den nachfolgenden Algorithmus ermittelt:

- **Schritt 1:** Zeilenweise wird die Zellhäufigkeit durch deren Summe dividiert und anschließend mit den Anpassungshäufigkeiten der Randverteilung der ersten Anpassungsvariable (bspw. Alter) in der entsprechenden Zeile multipliziert.
- **Schritt 2:** Spaltenweise wird die Zellhäufigkeit durch deren Summe dividiert und darauf folgend mit den Anpassungshäufigkeiten der Randverteilung der zweiten Anpassungsvariable (bspw. Bildungsabschluss) in der entsprechenden Spalte multipliziert.
- **Schritt 3:** Die ersten beiden Schritte werden so lange wiederholt, bis sich keine sichtbare Veränderung der Gewichtungsfaktoren mehr einstellt (vgl. Sand und Gabler (2018), S. 13f.).

3.4.3 Der GREG-Schätzer und die Kombination von Design- und Anpassungsgewichtung

Wie bereits zuvor beschrieben, ist das Ziel einer Gewichtung von Erhebungsdaten, ein einziges Gewicht w_i für eine Erhebungseinheit i zu erstellen, das für jegliche Analysen des Datensatzes der Erhebung herangezogen werden kann. Dieses Gewicht lässt sich als Kombination von Design- und Anpassungsgewichtung verstehen. Im einfachsten Fall kann es durch die Multiplikation der beiden einzelnen Gewichte berechnet werden. w_i entspricht dann

$$w_i = d_i \cdot g_i \quad \forall i \in R \quad (\text{vgl. S. L. Lohr (2009), S. 132.}).$$

Der Totalwert einer Variable y kann demnach anhand von

$$\hat{Y} = \sum_{i \in R} w_i \cdot y_i = \sum_{i \in R} d_i \cdot g_i \cdot y_i$$

geschätzt werden. Eine weitere Möglichkeit zur Bestimmung von Gesamtgewichten stellt in diesem Zusammenhang der *verallgemeinerte Regressionsschätzer* (GREG-Schätzer) dar. Aufgrund der positiven Eigenschaften dieses Schätzers wird er u. a. durch Särndal und Lundström

(2005) empfohlen. Auch im Zuge der Gewichtung der Daten der CELLA 2-Erhebung wurde dieses Verfahren verwendet und seine positiven Eigenschaften hervorgehoben. Im Grundsatz verwendet dieser Schätzer ebenso wie andere Kalibrierungsansätze Hilfsvariablen, anhand derer eine Anpassung an die Verteilung innerhalb der Zielpopulation erfolgt. Das Design wird jedoch im Zuge der Schätzung direkt in die Kalibrierung eingebunden. Verwendet wird dafür der Vektor der Anpassungsvariablen $\mathbf{x}$, für die sowohl in der Erhebung als auch in der Zielpopulation Angaben vorliegen. Unabhängig von der Existenz von Nonresponse kann durch die Verwendung des GREG-Schätzers eine Reduktion der Varianz für den Schätzer des Totalwertes einer interessierenden Variable ($\hat{Y}$) herbeigeführt werden. Das Ausmaß dieser Varianzreduktion ist dabei abhängig von der Stärke des (linearen) Zusammenhanges der interessierenden Variable y und dem Vektor der Hilfsvariablen $\mathbf{x}$ (vgl. Särndal und Lundström (2005), S. 34.; vgl. Gabler et al. (1998), S. 162f.).

Weiterhin als positive Eigenschaft zu nennen ist, dass diese Kalibrierungsmethode in der Berechnung der Anpassungsgewichte direkt die Inklusionswahrscheinlichkeiten der jeweiligen Einheiten berücksichtigt. Särndal und Lundström (2005) beschreiben im Zuge dessen, dass die Kalibrierung in zwei Phasen erfolgen kann. An die Designgewichtung anknüpfend kann in einem ersten Schritt eine Anpassung des Nonresponse etwa anhand der zuvor geschätzten Antwortwahrscheinlichkeiten erfolgen. Anschließend geschieht die Kalibrierung unter Verwendung des GREG-Schätzers.

Der Notation von Gabler et al. (2011) folgend, die diesen Schätzer bei der Gewichtung der Daten der CELLA 2-Erhebung verwendet haben, lässt sich für eine Dual-Frame-Erhebung ein entsprechender Schätzer für einen Totalwert ($\hat{Y}^{GREG}$) anhand von

$$\hat{Y}^{GREG} = \sum_{i \in R} d_i \cdot g_i^{GREG} \cdot y_i \quad (3.23)$$

bestimmen. Das Kalibrierungsgewicht g_i^{GREG} wird hierbei durch

$$g_i^{GREG} = \frac{1}{q_i} \left(1 + c_i \left(\sum_{i \in U} \mathbf{x}_i - \sum_{i \in R} \frac{1}{\pi_i q_i} \mathbf{x}_i \right)' \left(\sum_{i \in R} \frac{c_i}{\pi_i q_i} \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \mathbf{x}_i \right) \quad (3.24)$$

angegeben. q_i entspricht dabei der (geschätzten) Antwortwahrscheinlichkeit der Einheit i , während c_i Gewichte sind, die der Forscher selbst (frei) wählt. i. d. R. wird für c_i der Wert 1 angegeben, sofern der Forscher keine begründete Wahl für einen anderen Wert angibt. Der Term $1/q_i$ beschreibt die Berücksichtigung der Antwortwahrscheinlichkeit im gesamten Gewichtungsprozedere des Zwei-Phasen-GREG-Schätzers nach Särndal und Lundström (2005) (vgl. Gabler et al. (2011), S. 163, Särndal und Lundström (2005), S. 52 & S. 64). Die Einbeziehung der Antwortwahrscheinlichkeit in die Schätzung wird durch Little und Rubin (1987) dahin gehend kritisiert, dass dies zwar das Potenzial besitzt, eine mögliche Verzerrung aufgrund von Nonresponse zu

negieren, jedoch Gefahr läuft, eine hohe Varianz aufzuweisen. Sand und Gabler (2018) diskutieren dazu, dass durch die Einbeziehung eines Nonresponsegewichtes $1/\hat{q}_i$, Respondenten, die aufgrund des Modells eine niedrige Teilnahmewahrscheinlichkeit erhalten, ein höheres Gewicht erhalten. Dadurch können diese die geschätzten Ergebnisse übermäßig beeinflussen (vgl. Sand und Gabler (2018), S. 16.). Zusätzlich legt ein solches Modell ein (zu) hohes Vertrauen in das zugrunde liegende (Regressions-) Modell, das zur Schätzung der Antwortwahrscheinlichkeiten einer beliebigen Einheit i verwendet wird (vgl. Little und Rubin (1987), S. 58.). Da zusätzlich i. d. R. die MAR-Annahme zugrunde liegt, wird oftmals von einer separaten Nonresponsegewichtung abgesehen und für q_i ebenfalls der Wert 1 angenommen. Daher reduziert sich das obige Kalibrierungsgewicht $g_i^{GREG^*}$ auf

$$g_i^{GREG^*} = 1 + \left(\sum_{i \in U} \mathbf{x}_i - \sum_{i \in R} \frac{1}{\pi_i} \mathbf{x}_i \right)' \left(\sum_{i \in R} \frac{1}{\pi_i} \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \mathbf{x}_i. \quad (3.25)$$

3.4.4 Exkurs: Berücksichtigung der "Klumpung" über die Interviewer bei telefonischen Befragungen

Wie bereits in Kapitel 2 diskutiert, liegt einer der Vorteile der telefonischen Befragung darin, dass es möglich ist, regional ungeklumpte Stichproben zu ziehen. Jedoch ist dem Interviewer sowie dessen Fähigkeit, einen potenziellen Befragten zu rekrutieren, eine besondere Aufmerksamkeit zu schenken.

Während bei Face-to-Face-Befragungen aus Kostengründen oftmals eine primäre Erhebungseinheit (PSU) (als Klumpen), wie bspw. eine Gemeinde, sowie ein Interviewer zusammenfallen, ist dies bei telefonischen Befragungen nicht die Regel. Hier werden Befragte aus allen Erhebungsgebieten gewöhnlich durch alle vorhandenen Interviewer abgedeckt. Demnach kann die Häufigkeit des Einsatzes eines bestimmten Interviewers sowie dessen Fähigkeit einen befragten zu rekrutieren, eine Klumpung der Interviews über bestimmte Interviewer bewirken. In Abhängigkeit davon, wie stark sich ein bestimmter Interviewer an die Vorgaben des zugrunde liegenden Instrumentes (zum Beispiel den Fragebogen) hält, besteht die Möglichkeit durch diese Klumpung, dass durch eine bewusste und systematische Veränderung bspw. der Antwortreihenfolge oder durch das Auslassen bestimmter Antwortkategorien, eine systematische Verzerrung der Erhebungsergebnisse eintreten kann. Solche Verzerrungen können sich ebenfalls durch die Art und Weise der Beantwortung von Rückfragen durch den Interviewer sowie durch dessen Fähigkeit zur Rekrutierung und die Einleitung der Befragung durch den Einleitungstext ergeben.⁴² Zuletzt

42 Ein Großteil der Interviewer in Deutschland ist als freier Mitarbeiter beschäftigt. Freien Mitarbeitern muss in der Ausübung ihrer Tätigkeit eine gewisse Freizügigkeit eingeräumt werden. Dies steht jedoch im starken Kontrast zu standardisierten Fragebögen, die häufig bei Befragungen eingesetzt werden. Um rechtliche Probleme zu umgehen, wird die Freizügigkeit der Arbeit damit begründet, dass dem Interviewer freigestellt wird, wie er eine Befragung einleitet. Im Gegensatz zu fest angestellten Interviewern, die dazu verpflichtet werden können nach genauen

besteht die Möglichkeit, dass bestimmte Charakteristika des Interviewers (die aural durch den Befragten wahrgenommen werden können) Einfluss auf das Antwortverhalten ausüben können. Die Mechanismen, die im Zuge einer Telefonerhebung dazu beitragen können, sind im Rahmen dieser Arbeit nicht weiter im Fokus. Vielmehr ist hier von Interesse, in welchem Umfang eine solche Klumpung auftreten kann und wie diese bspw. durch eine Kalibrierung berücksichtigt werden könnte. Nähere Ausführungen über mögliche Ursachen der unterschiedlichen Rekrutierungsraten finden sich bspw. bei O'Muircheartaigh und Campanelli (1998), Conrad et al. (2012), West und Olson (2010) oder Meier (2011).

Untersuchung der Klumpung von Interviews

Zur Untersuchung der Klumpung von Interviews im Rahmen von Dual-Frame- Telefonstichproben wurde in der vorliegenden Arbeit die Erhebung Influenza 2014 herangezogen, da hierfür eine Identifizierung der Interviewer möglich ist und sich deren Anzahl in einem überschaubaren Maß aufhält. Die hier präsentierten Ergebnisse wurden eigens für die vorliegende Arbeit neu berechnet. In der Auseinandersetzung mit dem Kontaktdatensatz der Bruttostichprobe konnten von den ca. 22.000 vorhandenen Kontaktversuchen ca. 17.000 nicht direkt einem Interviewer zugeordnet werden, da entweder kein Personenkontakt erfolgt ist, die Rufnummer bereits im Vorfeld durch den Predictive Dialer aussortiert wurde oder die Interviewer-ID nicht erfasst wurde. Unter diesen 17.000 Kontaktversuchen wurden jedoch lediglich 13 Interviews geführt. Weiterhin beinhaltet der Datensatz fünf Interviewer-IDs, unter denen kein Interview geführt wurden. Jedoch zeigt eine nähere Untersuchung dieser IDs, dass bei diesen auch nur wenige Kontaktversuche erfolgt sind. Demnach besteht die Vermutung, dass es sich hierbei entweder um Interviewer handelt, die bereits zu einem frühen Zeitpunkt aus der Erhebung ausgeschieden sind oder um Kontaktversuche, die zu Testzwecken unternommen wurden.

Für die Bewertung der Klumpung von Interviews über bestimmte Interviewer sind beginnend mehrere Faktoren relevant. Dazu gehören die Anzahl der kontaktierten Rufnummern pro Interviewer, die Anzahl der abgeschlossenen Befragungen pro Interviewer, sowie das Verhältnis dieser beiden Zahlen, um zu beurteilen, ob es, aus welchen Gründen auch immer, Unterschiede in der Rekrutierungsrate zwischen den verschiedenen Interviewern gibt. Die dafür neu berechneten Kennwerte aus der Erhebung Influenza 2014 werden daher in Abbildung 7 dargestellt.

Vorgaben eine Einleitung des Interviews vorzunehmen (bspw. das wörtliche Vorlesen eines Einleitungstextes), wird freien Interviewern oftmals lediglich diejenigen Bestandteile vorgegeben, die in der Anbahnung eines Interviews genannt werden müssen. Diese umfassen bspw. die Freiwilligkeit der Teilnahme und das erhebende Institut. Die restliche Anbahnung des Interviews obliegt der Freizügigkeit des freien Mitarbeiters.

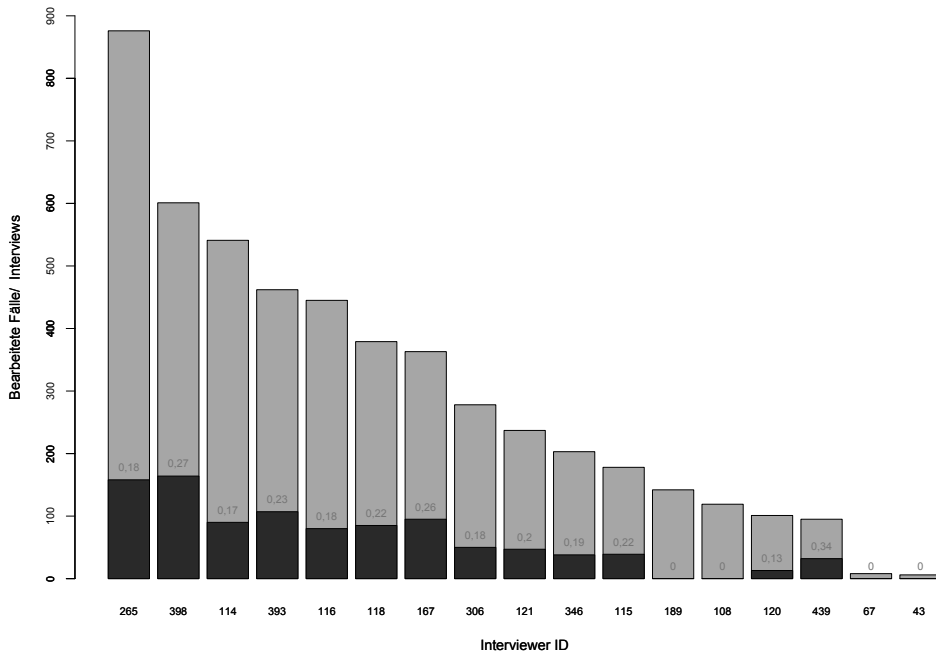


Abbildung 7 Influenza 2014: Bearbeitete Fälle pro Interviewer

Die Anzahl der kontaktierten Rufnummern sind in dieser Abbildung grau und die Anzahl der abgeschlossenen Interviews blau dargestellt. Das Verhältnis der beiden Größen wird durch die Zahl oberhalb der Anzahl der abgeschlossenen Interviews wiedergegeben. Zu bemerken ist indes, dass eine Rufnummer durch mehrere Interviewer kontaktiert worden sein kann.

Unter der Berücksichtigung von Abbildung 7 zeigt sich deutlich, dass es eine unterschiedliche Arbeitslast der Interviewer im Rahmen der Erhebung gegeben hat. Damit verbunden ist, dass die Anzahl der kontaktierten Rufnummern, aber auch die Anzahl der abgeschlossenen Interviews (absolut) variiert. Weiterhin zeigt sich, dass das Verhältnis zwischen kontaktierten Rufnummern und abgeschlossenen Interviews zwischen den einzelnen Interviewern stark abweicht. Allerdings sollte dieses Verhältnis nicht unbedingt als "Erfolgsrate" eines Interviewers verstanden werden, da auch unterschiedliche Aspekte der Feldsteuerung dieses Verhältnis beeinflussen können. So ist es bspw. denkbar, dass bestimmte Interviewer häufiger an Tagen eingesetzt wurden, an denen eine höhere Anzahl von Rufnummern kontaktiert wurden, die bereits mehrfach einen erfolglosen Kontaktversuch aufweisen. Auf der anderen Seite besteht die Möglichkeit, dass bestimmte Interviewer häufiger zu Zeitpunkten eingesetzt wurden, bei denen eine "frische" Tranche von noch nicht kontaktierten Rufnummern ins Feld ging. Diese und weitere Faktoren können neben der Fähigkeit des Interviewers zur Rekrutierung von Befragten jenes Verhältnis

beeinflussen. Dennoch kann anhand der vorliegenden Auswertung nicht ausgeschlossen werden, dass eine Klumpung von Interviews über die Interviewer vorliegt. Jedoch ist unter diesem Aspekt vorwiegend relevant, ob es systematische Unterschiede zwischen den jeweiligen Befragten nach Interviewer gibt. Daher sollte ein gesteigerter Fokus auf der Homogenität bzw. Heterogenität der Befragten nach Interviewer liegen.

Da der Einsatz der Interviewer in Abhängigkeit der Maßnahmen der Feldsteuerung schwer nachzuvollziehen ist und nach Abschluss der Erhebung nicht oder nur sehr aufwendig eruiert werden kann, wann eine neue Tranche von Rufnummern eingesetzt wurde, welche Interviewer zu diesem Zeitpunkt im Einsatz waren und wie groß der Anteil der Rufnummern mit mehreren erfolglosen Kontaktversuchen zu bestimmten Zeitpunkten gewesen ist, wurde in einem nächsten Schritt lediglich die durchschnittliche Anzahl der Kontaktversuche pro Rufnummer pro Interviewer berechnet. Diese wurde dann dem zuvor beschriebenen Verhältnis aus der Anzahl der kontaktierten Rufnummern und abgeschlossenen Interviews gegenübergestellt. Das Ergebnis dieser Gegenüberstellung kann anhand von Abbildung 8 nachvollzogen werden.

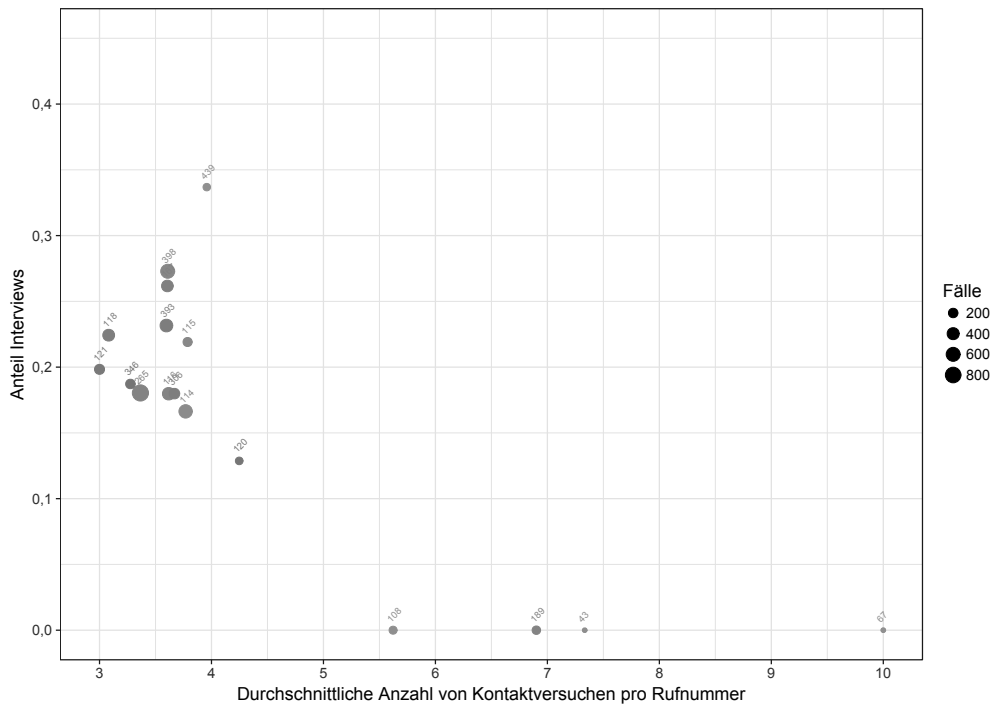


Abbildung 8 Influenza 2014: Durchschnittliche Anzahl der Kontaktversuche nach Interviewer und Anteil der Interviews

Auf der Abszisse ist hier die durchschnittliche Anzahl der Kontaktversuche bis zum letzten Kontakt der Rufnummer nach Interviewer angegeben. Die Ordinate trägt den Anteil der abge-

Tabelle 7 Median der Kontaktversuche nach Interviewer

Interviewer-ID	108	114	115	116	118	120	121	167		
Median Kontaktversuche	7,0	3,0	3,0	3,0	2,0	4,0	2,0	3,0		
Interviewer-ID	189	265	306	346	393	398	43	439	67	
Median Kontaktversuche	7,0	2,0	3,0	2,0	2,0	2,0	7,5	3,0	10,0	

schlossenen Interviews an der gesamten Anzahl der Rufnummern, die durch einen Interviewer bearbeitet wurden, ab. Die Größe der Punkte verweist auf die Anzahl der Rufnummern, die durch die Interviewer bearbeitet wurden. Abgesehen von denjenigen Interviewer-IDs, unter denen keine Interviews geführt wurden, fällt auf, dass die durchschnittliche Anzahl der Interviews, unabhängig des Anteils der abgeschlossenen Interviews, vergleichsweise nahe beieinander zwischen einem Wert von drei und vier Kontaktversuchen liegt. Die einzige Ausnahme stellt hierbei die ID 120 dar, deren Kontaktversuche oberhalb des Wertes vier zu finden sind. Auch der Vergleich mit der mittleren Anzahl der Kontaktversuche, die die Rufnummern nach Interviewer aufweisen, zeigt eine ähnliche Tendenz.

So liegt diese, unter Ausschluss derjenigen IDs, unter denen keine Interviews geführt wurden, grundsätzlich unterhalb des Durchschnittswertes, was den Einfluss gerade derjenigen Rufnummern, die bis zur maximalen Anzahl der Anrufversuche kontaktiert wurden, widerspiegelt. Weiterhin zeigt sich, dass sich der Abstand zwischen dem Durchschnittswert und dem Median gerade dann vergrößert, wenn unter einer ID mehr Interviews geführt wurden. Allgemein lässt sich anhand dieser Auswertung jedoch nicht vermuten, dass bestimmte Interviewer vorwiegend oder ausschließlich Rufnummern bearbeitet haben, die bereits länger im Feld gewesen sind und mehrfach erfolglos kontaktiert wurden, während unter anderen IDs vorwiegend "frische" Rufnummern bearbeitet wurden. Zur Überprüfung dieser Annahme wurde, im Zuge dieser Arbeit, in einem weiteren Schritt die Differenz zwischen dem Datum, an dem eine Rufnummer zum ersten Mal ins Feld gegangen ist und demjenigen, an dem die Rufnummer zum letzten Mal kontaktiert wurde, gebildet. Diese Differenz gibt Ausschluss darüber, über welchen Zeitraum eine Rufnummer bearbeitet wurde. Aufgeschlüsselt nach Interviewer-ID kann so ermittelt werden, ob Interviewer einen höheren Anteil an "frischen" Rufnummern bearbeitet haben. Vergleicht man dies mit den in Abbildung 7 dargestellten Verhältnis zwischen kontaktierten Rufnummern und abgeschlossenen Interviews, so kann dies Aufschlüsse darüber geben, ob bestimmte Interviewer lediglich davon profitiert haben, dass diese hauptsächlich Rufnummern kontaktiert haben, die noch nicht lange eingesetzt wurden. Tabelle 8 stellt diese Ergebnisse nochmals dar.

Tabelle 8 Influenza 2014: Tage einer Rufnummer im Feld nach Interviewer-ID (in %)

Tage im Feld/ID	439	398	167	121	265	114	120
0	33,7	38,3	31,7	43,5	37,1	34,8	25,7
1 - 3	5,3	9,5	10,2	6,3	10,3	12,4	13,9
4 - 6	21,1	8,7	17,6	14,3	12,1	12,9	25,7
7 - 10	15,8	27,0	21,5	25,3	27,3	19,0	19,8
11+	24,2	16,6	19,0	10,5	13,2	20,9	14,9
AZ Interviews/AZ Kontakte	0,34	0,27	0,26	0,2	0,18	0,17	0,13

Die in Tabelle 8 dargestellten Interviewer sind lediglich eine Auswahl der gesamten, im Datensatz enthaltenen IDs, die jedoch einen Einblick darin gewähren, ob Interviewer, deren Verhältnis von abgeschlossenen Interviews zu kontaktierten Rufnummern höher gewesen ist, vornehmlich Rufnummern kontaktiert haben, die noch nicht lange im Feld eingesetzt wurden. Dazu gibt der Wert 0 wieder, dass eine Rufnummer noch am Tag des Einsatzes vollständig bearbeitet wurde. Die restlichen Differenzen wurden als Zeiträume angegeben. Die letzte Zeile gibt nochmals das Verhältnis zwischen der Anzahl abgeschlossenen Interviews und der Kontakte der jeweiligen Interviewer wieder. Die hier dargestellte Auswahl beinhalten diejenigen drei Interviewer mit dem höchsten und die drei Interviewer mit dem niedrigsten Verhältnis sowie eine Interviewer-ID "im mittleren Bereich".

Hier zeigt sich, dass es durchaus Unterschiede im Anteil des jeweiligen Anteils des Alters einer Rufnummer zwischen den unterschiedlichen Interviewern gegeben hat. Allerdings kann anhand dieser Angaben nicht davon ausgegangen werden, dass Interviewer, deren Verhältnis zwischen der Anzahl an abgeschlossenen Interviews und kontaktierten Rufnummern höher war, mehr Rufnummern kontaktiert haben, die erst kürzlich eingesetzt wurden. Dies begründet sich darin, dass auch Interviewer-IDs mit geringem Verhältnis (wie bspw. ID 265) eine ähnliche Aufteilung des "Alters" einer Rufnummer nachweisen. Der Umkehrschluss, dass Interviewer mit einem entsprechend niedrigen Verhältnis dieser beiden Anzahlen vermehrt ältere Rufnummern kontaktiert haben, lässt sich jedoch nicht eindeutig ablehnen. So kann zumindest für den Interviewer mit der ID 120, dessen Verhältnis am geringsten ist, aufgezeigt werden, dass dieser vermehrt Rufnummern kontaktiert hat, die bereits für eine längere Zeit im Einsatz gewesen sind. Insgesamt kann jedoch davon ausgegangen werden, dass trotz bestehender Unterschiede in der Verteilung des Alters von Rufnummern über die jeweiligen Interviewer, diese Unterschiede (alleinig) für die Differenzen im Verhältnis der Anzahl der Interviews und der kontaktierten Nummern verantwortlich sind.

Eine weitere Möglichkeit zur Erklärung der Unterschiede zwischen den jeweiligen Interview-

ern, die teilweise unabhängig von der Fähigkeit des Interviewers zur Rekrutierung sein kann, ist die Dauer, die ein bestimmter Interviewer durchschnittlich oder im Median zur Durchführung einer Befragung benötigt. Eine weitere Vermutung, weswegen unterschiedliche Interviewer verschiedene Rekrutierungsraten aufweisen, könnte darin bestehen, dass diese bspw. dadurch, dass Erhebungsinstrumente verkürzt werden oder weniger Aufwand in der Erläuterung der unterschiedlichen Items (bei Rückfragen) erfolgt, eine Befragung schneller bearbeiten und dadurch durchschnittlich mehr Rufnummern in einer gegebenen Zeit bearbeiten, als dies andere Interviewer tun. In der Regel werden solche Vorgehensweisen durch das Feldmonitoring während der Erhebung unterbunden. Um dennoch zu überprüfen, ob ein geringerer Zeitaufwand der Interviewer zu einer höheren Zahl von Interviews führt, wird dies anhand von Abbildung 9 genauer betrachtet. Eine erste Untersuchung dieser beiden Größen zeigt bereits, dass es wenig Unterschiede zwischen der mittleren und der durchschnittlichen Dauer der Befragungen nach Interviewer gibt. Daher wird in Abbildung 9 die mittlere Interviewdauer zu Vergleichszwecken herangezogen.

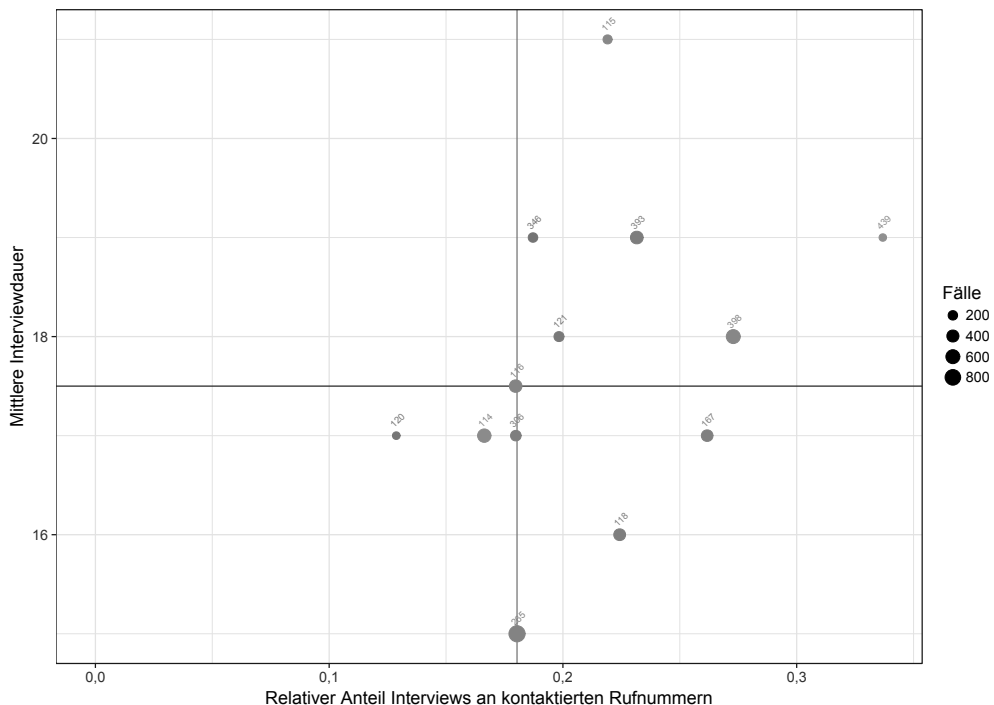


Abbildung 9 Influenza 2014: Mittlere Dauer der Befragung nach Interviewer und Anteil der Interviews

An der Abszisse der Darstellung wurde hier der Anteil der abgeschlossenen Interviews abgetragen. Die Ordinate gibt die mittlere Dauer der geführten Interviews pro Interviewer wieder. Die Lage

der beiden Geraden der Grafik ergibt sich durch den Median der jeweiligen Variable über alle Interviewer. Analog zu Abbildung 8 ist die Größe der Punkte abhängig von der Anzahl der Rufnummern, die durch einen Interviewer bearbeitet wurden.

Anhand der Abbildung zeigt sich, dass der zuvor geäußerten Vermutung nicht (eindeutig) entsprochen werden kann. Zwar liegt die mittlere Dauer der Interviews, die unter der ID, welche die meisten Rufnummern kontaktiert hat, deutlich unterhalb des Medians aller IDs, die Ursache lässt sich jedoch nicht direkt auf ein bestimmtes, anderes Verhalten in der Bearbeitung der Befragung zurückführen. Vielmehr könnte die niedrigere Dauer auch durch einen Größeneffekt hervorgerufen worden sein. Weiterhin finden sich im Gegenzug auch einige IDs mit einer höheren Rekrutierungsrate, deren Anzahl der Fälle ähnlich hoch sind und deren mittlere Dauer der Befragung ebenso oberhalb des Medians der Befragungsdauer liegt. Tendenziell lässt sich anhand der Abbildung kein direkter Zusammenhang zwischen der Dauer der Befragung und der Rekrutierungsrate nachweisen.

Die Homogenität der Ergebnisse nach Interviewer

Unabhängig der hier dargestellten Untersuchung möglicher Ursachen des Unterschiedes im Anteil der Interviews an kontaktierten Rufnummern nach Interviewer zeigt sich (anhand von Abbildung 7), dass die absolute Zahl der Interviews nach Interviewer stark variiert. Ob diese Klumpung von Interviews über die Interviewer sich negativ auf die Schätzergebnisse der Erhebung auswirkt und somit zu Ergebnissen führen kann, die weniger präzise sind als bei einer ungeklumpten Erhebung, hängt jedoch von der Homogenität der Ergebnisse der Befragung pro Klumpen⁴³ ab. Dabei gilt zu beachten, dass eine Homogenität der Ergebnisse innerhalb der Klumpen tendenziell die Varianz erhöht und sich somit negativ auf die Präzision der Schätzergebnisse auswirkt. Demnach gilt, je heterogener die Verteilung der Ergebnisse der Befragung eines Interviewers sind, desto geringer ist der (negative) Effekt der Klumpung auf die Gesamtvarianz des Schätzers. Die Annahme(n) beziehen sich jeweils auf den Vergleich der Verteilung in der Zielpopulation.

O'Muircheartaigh und Campanelli (1998) beschreiben, dass im Rahmen von Face-to-Face-Befragungen Klumpen- und Interviewereffekt oftmals schwer voneinander zu trennen sind, da Interviewer aus Kostengründen häufig ein bestimmtes Erhebungsgebiet vollständig und alleinig bearbeiten. Da bei telefonischen Befragungen im Gegensatz hierzu jedoch die Möglichkeit besteht, diese regional ungeklumpt durchzuführen, handelt es sich um einen interpenetrierenden Erhebungsansatz. Unabhängig von der Region, aus der die Erhebungseinheit stammt, sollte es demnach für jeden Interviewer möglich sein, diese zu befragen. Daher ist zu vermuten, dass die Klumpung über die Interviewer einzig mit bestimmten Eigenschaften, Vorgehensweisen und der Art und Häufigkeit des Einsatzes desselben während der Feldarbeit verbunden ist (vgl.

⁴³ In diesem Fall entspricht ein Klumpen einem Interviewer.

O'Muircheartaigh und Campanelli (1998), S. 64 f.). Im Gegensatz zu einer Homogenität der Ergebnisse aufgrund der Klumpung der Stichprobe, die i. d. R. varianztreibend ist und damit die Präzision der Schätzergebnisse einschränkt, kann ein solcher Interviewereffekt systematische Unterschiede hervorrufen und somit verzerrend auf die Schätzergebnisse wirken. Dies kann bspw. dann geschehen, wenn Befragte aufgrund spezieller Eigenschaften des Interviewers dazu geneigt sind, in einer bestimmten Art und Weise zu antworten (etwa durch ein Antwortverhalten nach sozialer Erwünschtheit) oder, wenn Interviewer bewusst Einfluss auf die (ursprünglich standardisierte) Befragung nehmen (etwa durch das Fälschen von Ergebnissen oder dem Auslassen bestimmter Antwortkategorien).

Ob und in welchem Ausmaß es im Rahmen der Erhebung Influenza 2014 zu systematischen Verzerrungen aufgrund der Klumpung über die Interviewer gekommen ist und was diese, falls vorliegend, genau bewirkt hat, ist nach der Erhebung nur schwer zu determinieren. Da bei Telefonstichproben eine Erhebungseinheit i. d. R. zufällig einem Interviewer zugeordnet wird, sollte jedoch die Klumpung bei Ausbleiben eines Interviewereffektes lediglich durch die Häufigkeit des Einsatzes eines Interviewers entstehen. Bei Unabhängigkeit des Merkmals mit dem Interviewer sollten die Ergebnisse der Befragung einer Variable bei der Betrachtung eines Interviewers nicht miteinander korreliert sein. Die Intraklassen-Korrelation der betrachteten Merkmale nach Interviewer (ρ_{Int}) lässt sich dabei anhand des Verhältnisses der Varianz zwischen den jeweiligen Interviewern (σ_b^2) an der gesamten Varianz beschreiben. Diese berechnet sich aus der Summe der Varianz zwischen den Interviewern und derjenigen innerhalb der Interviewer (σ_w^2) bzw. Klumpen. ρ_{Int} berechnet nach Kish (1962) sich daher durch

$$\rho_{Int} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}.$$

In der ursprünglichen Untersuchung der Interviewervarianz (σ_w^2) wird davon ausgegangen, dass eine Anzahl von h Interviewern aus der Menge aller potentiellen Interviewer ausgewählt wird. Jeder dieser h ausgewählten Interviewer hat hierbei einen eigenen durchschnittlichen Interviewer-Bias über die Anzahl der durch diesen abgeschlossenen Interviews. Weiterhin wird davon ausgegangen, dass der Effekt dieser Fehler auf die Varianz eines Schätzwertes im Verhältnis zu der Anzahl der eingesetzten Interviewer steht.⁴⁴ Demnach wird vermutet, dass mit steigender Anzahl der Interviewer der Effekt des Interviewer-Bias auf die Varianz abnimmt (vgl. Kish (1962), S. 94.).

Da es sich hierbei um die (wahren) Populationswerte der Varianzen handelt, die anhand einer Stichprobe lediglich geschätzt werden können, kann für die Intraklassen-Korrelation oftmals auch nur ein geeigneter Schätzwert ermittelt werden. Dieser lässt sich nach Ganninger (2009) anhand von

⁴⁴ Es wird angenommen, dass Interviewer zufällig ausgewählt wurden.

$$\hat{\rho}_{Int} = \frac{MSB - MSW}{MSB + (K - 1)MSW}$$

beschreiben, mit

$$MSB = \frac{SSB}{h-1}, \quad MSW = \frac{SSW}{n-h},$$

$$K = \frac{1}{h-1} \left(n - \sum_{k \in \mathcal{H}} \frac{n_k^2}{n} \right),$$

$$SSB = \sum_{k \in \mathcal{H}} n_k (\bar{y}_k - \bar{y})^2 \text{ und } SSW = \sum_{k \in \mathcal{H}} \sum_{i \in S^k} (y_{ik} - \bar{y}_k)^2.$$

SSW und SSB beschreiben hier die summierte, quadrierte Abweichung innerhalb und zwischen den Klumpen. MSW und MSB ist demnach die mittlere quadrierte Abweichung. h entspricht der Anzahl der Klumpen/ Interviewer, $\mathcal{H}$ ist die Menge der Interviewer (der Erhebung), n_k ist die Anzahl der Elemente innerhalb eines Klumpens, S^k ist Menge der Elemente der Sub-Stichprobe eines bestimmten Klumpens k und n entsprechend die Anzahl der Elemente innerhalb der Stichprobe (vgl. Ganninger (2009), S. 45.).

$\hat{\rho}_{Int}$ kann dabei in einem Wertebereich von $-\frac{1}{K-1} \leq \hat{\rho}_{Int} \leq 1$. Bei Erhebungen mit einer entsprechend ausreichend großen Zielpopulation wird vereinfacht angenommen, dass $\frac{1}{K-1} \approx 1$. Ein negativer Wert von $\hat{\rho}_{Int}$ spricht somit dafür, dass ein Großteil oder die gesamte Variation der Variable durch die Variation innerhalb der Klumpen erklärt werden kann. Demnach besteht in einem solchen Fall nur ein geringer oder kein Interviewereffekt.

Die Auswirkung der Klumpung über die Interviewer kann anschließend anhand des (modellbasierten) Designeffektes, der durch die Klumpung hervorgerufen wird ($deff_c$), bestimmt werden. Dieser vergleicht die Varianz unter der gegebenen Klumpung mit derjenigen, die sich durch eine uneingeschränkte Zufallsauswahl ergeben würde. Ein Schätzwert dieses (Teiles des) Designeffektes lässt sich durch

$$\widehat{deff}_c = 1 + \hat{\rho}_{Int}(\bar{n}_k - 1)$$

bestimmen. $\bar{n}_k$ meint hierbei die durchschnittliche Anzahl von Elementen innerhalb der Klumpen.

Berechnung des Interviewer-Designeffektes anhand der Daten aus Influenza 2014

Beispielhaft soll gerade dieser Klumpeneffekt anhand von ausgewählten Variablen der Erhebung Influenza 2014 veranschaulicht und untersucht werden. Dafür wurde im Zuge der vorliegenden Erhebung der $\widehat{deff}_c$ für ausgewählte Variablen geschätzt. Dies dient der Grundlage zur

Diskussion, ob und wie der Klumpeneffekt der Interviewer im Zuge der Gewichtung eines Erhebungsdatensatzes stärker berücksichtigt werden soll. Diese Diskussion schließt den vorliegenden Exkurs ab. Die entsprechenden Schätzwerte des modellbasierten Designeffektes der Klumpung werden anhand von Abbildung 10 dargestellt.

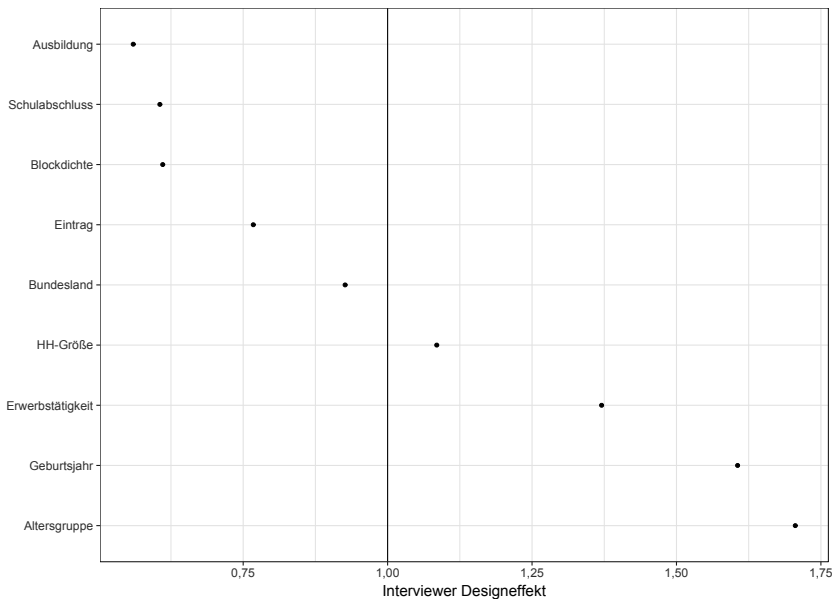


Abbildung 10 Influenza 2014: Designeffekt durch Klumpung

Die von der Abszisse ausgehende Gerade bei dem Wert 1 stellt in der Abbildung denjenigen Designeffekt dar, der sich bei einer uneingeschränkten Zufallsauswahl einstellen würde. Hierbei zeigt sich, dass einige der betrachteten Variablen unterhalb des Designeffektes von 1 liegen. Im Falle der Variablen *Blockdichte* und *Eintrag*, die beide eher als technisches Datum zu verstehen sind und auf die der Interviewer ohnehin keinen Einfluss nehmen kann, ist dies nicht weiter verwunderlich, da hier, wie bereits zuvor geschildert, die Variation einzig innerhalb der betrachteten Klumpen und nicht zwischen den jeweiligen Interviewern besteht. Ebenso plausibel wirkt daher der Wert der Variable *Land*, die die Bundeslandkennung beinhaltet, wenn man davon ausgeht, dass ein Vorzug telefonischer Befragungen gerade darin besteht, dass diese regional ungeklumpt sind. Die Begründung ist daher vergleichbar mit derjenigen der technischen Variablen. Bei einer Telefonstichprobe handelt es sich um einen interpenetrierenden Ansatz, somit hat der Interviewer selbst keinen (oder nur einen sehr geringen) Einfluss auf diese Variable, weshalb die Variation dieser nur innerhalb der Befragungen nach Interviewer und nicht zwischen diesen besteht.

Die Betrachtung der soziodemografischen Charakteristika führt indes zu unterschiedlichen Ergebnissen in Abhängigkeit der betrachteten Variablen. So zeigt sich bspw. für die beiden

Variablen, welche den Schul- und Ausbildungsabschluss messen ein Designeffekt, der kleiner eins ist, was durch ein $\hat{\rho}_{Int}$ kleiner Null erklärt werden kann. Demnach ist bei diesen beiden Variablen keine Verzerrung aufgrund der Klumpung über die Interviewer zu vermuten. Jedoch muss dazu angemerkt werden, dass gerade bei der Bildungsvariable bereits eine nicht zu vernachlässigende Verzerrung (in den designgewichteten Erhebungsdaten) aufgrund von bspw. gehäufte Antwortverweigerung gerade von Personen mit geringerem Bildungsabschluss besteht. Dieser sog. Bildungsbias ist für gewöhnlich in telefonischen Erhebungen besonders ausgeprägt. Daher ist es zumindest interessant hervorzuheben, dass scheinbar alle Interviewer gleichermaßen durch diesen Bildungsbias betroffen sind.

Dennoch können in dem vorliegenden Datensatz auch Variablen identifiziert werden, bei denen die Klumpung über die Interviewer mit einer Homogenität der Ausprägungen einer Variable einhergeht. So zeigen die Variablen *Haushaltsgröße*, *Erwerbstätigkeit* sowie die beiden *Altersvariablen* einen Designeffekt, der deutlich über 1 liegt. Damit zusammenhängend ist ein höherer Intraklassenkorrelationskoeffizient. Die damit verbundene Homogenität innerhalb der Variablen bewirkt daher eine höhere Varianz dieser Variablen verglichen zu einer uneingeschränkten Zufallsauswahl. Damit verbunden ist ebenfalls eine geringere Präzision der Schätzwerte gerade dieser Variablen. Hierbei ist es wichtig zu erwähnen, dass diese Ergebnisse nicht als Indiz für verzerrte Schätzergebnisse interpretiert werden dürfen. Weiterhin ist die Ursache der Homogenität innerhalb der unterschiedlichen Interviewer-IDs nicht geklärt. So ist es zwar potenziell möglich, dass bestimmte Charakteristika des Interviewers oder dessen Bearbeitungsweisen des Fragebogens Einfluss auf die Variabilität der Ergebnisse haben, andererseits können diese Klumpungen auch durch Maßnahmen der Feldsteuerung oder Arbeitszeiten bestimmter Interviewer hervorgerufen werden. So ist es bspw. denkbar, dass bestimmte Interviewer lediglich morgens eingesetzt werden. Da gerade zu diesen Zeiten insbesondere in Festnetzstichproben eher Personen erreichbar sind, die nicht vollzeit-erwerbstätig sind oder, die einer bestimmten Altersstruktur folgen, könnte dies ebenfalls einen Teil der Homogenität innerhalb der Variablen der Interviewer begründen.

Berücksichtigung der Klumpung nach Interviewern in der Kalibrierung

Grundsätzlich ist die Berücksichtigung der Klumpung über die Interviewer im Rahmen der Anpassungsgewichtung⁴⁵ möglich und denkbar. Die Voraussetzung dafür ist jedoch das Vorhandensein

⁴⁵ Ebenfalls denkbar wäre, eine entsprechende Berücksichtigung bereits im Zuge der Designgewichtung durchzuführen. Dazu müsste jedoch eine Telefonstichprobe direkt als zweistufiges Verfahren interpretiert werden, bei der auf der ersten Stufe ein Klumpen (Interviewer) und im Anschluss dessen erst der Befragte innerhalb des Klumpens ausgewählt wird. Da es sich jedoch bei Telefonstichproben um interpenetrierende Verfahren handelt und darüber hinaus die Inklusionswahrscheinlichkeit eines Interviewers von einer Vielzahl von Faktoren abhängig ist, die schwer zu erfassen sind (bspw. Arbeits- und Pausenzeiten oder ob der Interviewer zu dem Zeitpunkt, zu dem ein potenzieller Befragter ausgewählt wurde, gerade eine Befragung durchführt und deshalb nicht ausgewählt werden kann), ist eine solche Berücksichtigung weder praktikabel noch zielführend.

geeigneter Paradata, die zumindest die Identifikation des Interviewers erlauben. Wünschenswert sind weiterhin Angaben bezüglich soziodemografischer Charakteristika der Interviewer sowie eine Einschätzung der Erfahrung dieser. Gegeben dieser Angaben lässt sich nach Hox (1994) die zusätzliche Varianz, die durch die Klumpung über Interviewer entstehen kann, anhand von Mehrebenenmodellen darstellen. Dabei ist die Angabe der Erhebungseinheit innerhalb der Klumpen des Interviewers *genested* (vgl. Hox (1994), S. 305ff.).

Die Berücksichtigung im Zuge der Anpassungsgewichtung könnte, gegeben dem Vorhandensein entsprechender Paradata, im Rahmen der Verwendung des GREG-Schätzers angewendet werden. Zu überlegen wäre dabei, einem Interviewer, der, bezogen auf eine Untersuchungsvariable, eine geringere Variabilität verursacht, als dies über die gesamte Variable zu beobachten ist, ein geringeres Gewicht zu geben und im Gegenzug dazu denjenigen Interviewern, deren Angaben bezüglich einer Variablen heterogener sind, ein entsprechend höheres Gewicht zuzuteilen. Dies könnte bspw. durch die Verwendung des Parameters c_i (bzw. in diesem Fall c_{ik}) vorgenommen werden. So könnte bspw. ein Interviewer mit besonders homogenen Angaben einen Wert für diesen Parameter bekommen, der kleiner ist als 1 und diejenigen Interviewer mit vergleichsweise heterogenen Angaben einen Wert, der größer 1 ist. Zu berücksichtigen ist hierbei jedoch, dass das Gewichtungungsverfahren weiterhin erwartungstreue Schätzergebnisse hervorbringt. Damit darf durch die Einführung dieser Art von Gewichten die Schätzung bspw. des Totalwertes nicht (weiter) verzerrt werden, weswegen die Ausprägungen von c_{ik} so zu wählen sind, dass der geschätzte Totalwert unverändert bleibt. Da Särndal und Lundström (2005) jedoch darlegen, dass der GREG-Schätzer für einen beliebigen Wert von c_i (asymptotisch) erwartungstreu ist, dürfte die Einführung eines solchen, weiteren Parameters weniger problematisch sein. Jedoch gilt bei diesem Vorgehen zu bedenken, dass die Gewichte entsprechend für jede Variable einzeln bestimmt werden müssten, da, wie Abbildung 10 dargestellt, die Heterogenität nach Interviewer abhängig von der betrachteten Variablen ist. Ein solches Vorgehen kann daher zu Inkonsistenzen innerhalb der Gewichtung aufgrund der sich ändernden Gewichte in Abhängigkeit der betrachteten Variable führen.

Alternativ bestünde die Möglichkeit ein fixes Gewicht für einen gegebenen Interviewer einzuführen, das sich bspw. anhand der gemittelten Abweichung der Variabilität innerhalb dieses Interviewers berechnet. Betrachtet man jedoch die deutlichen Unterschiede der Designeffekte über die Variablen hinweg, läuft ein solcher Ansatz Gefahr, weitere Quellen der Verzerrung in die Schätzung mit einzubringen oder die Präzision bestimmter Schätzer zu reduzieren.

Demnach wird anhand der hier angeführten Diskussion ersichtlich, dass es bezogen auf bestimmte Variablen durchaus zu einer Klumpung über Interviewer kommen kann, die die Präzision der Schätzergebnisse negativ beeinflusst, dass eine Berücksichtigung dieser jedoch nur wohlüberlegt und evtl. nur für bestimmte Variablen in gesonderten Fällen durchgeführt werden sollte.

3.5 Vergleich der Ansätze zur Designgewichtung und Kalibrierung

Nachdem die unterschiedlichen Ansätze zur Designgewichtung in Dual-Frame-Erhebungen und Kalibrierung anhand der einschlägigen Literatur erläutert wurden, liegt das Interesse vor allem darin darzustellen, wie sich diese auf eine Erhebung anwenden lassen und wie sie sich in ihrer Schätzung von Parametern unterscheiden. Daher wurden einige der zuvor beschriebenen Ansätze näher untersucht. Für die Untersuchung der unterschiedlichen Designgewichte wurde dies anhand eines synthetischen Datensatzes unternommen. Die Wahl des Vorgehens liegt darin begründet, dass einerseits alle relevanten Parameter der Zielpopulation des synthetischen Datensatzes bekannt sind, andererseits ist die Stichprobenziehung unter vollständiger Kontrolle. Somit können Prozesse, die nicht direkt durch die Stichprobenziehung und der Auswahlwahrscheinlichkeit einer Erhebungseinheit in Verbindung stehen wie bspw. Interviewereffekte oder Messfehler unterbunden werden. Die Evaluation der Designgewichtungsverfahren erfolgt anhand des Vergleiches der Punktschätzer zu den wahren Werten der Zielpopulation sowie unter der Verwendung des modellbasierten Designeffektes aus Kapitel 3.5.1. Weiterhin legt dieses Kapitel die Erzeugung der synthetischen Grundgesamtheit sowie deren relevante Parameter dar. Die synthetische Grundgesamtheit wird an späterer Stelle im Rahmen der eigentlichen Forschungsfrage, die sich der Verwendung von Nutzungsgewohnheiten und der unterschiedlichen Erreichbarkeit in Abhängigkeit des gewählten Telekommunikationsmediums widmet, als Hilfsmittel zu Begutachtung unterschiedlicher Schätzer verwendet.

Da die Kalibrierung einerseits zur Reduktion der Verzerrung aufgrund von Fehlerquellen im Rahmen einer Erhebung bspw. durch Nonresponse eingesetzt wird und andererseits in einer nachträglichen Schichtung Verwendung findet, empfiehlt es sich, diese Verfahren an Daten, die aus einer tatsächlichen Erhebung stammen, zu untersuchen. Der Grund dafür liegt darin, dass bspw. ein Verweigerungsprozess sowie die damit verbundenen soziodemografischen Charakteristika, die zu dessen Erklärung herangezogen werden, nur schwer in allen Aspekten nachgebildet werden können. Daher verwendet die Untersuchung der unterschiedlichen Kalibrierungsverfahren die Daten der Erhebung Influenza 2014. Hierbei werden Punktschätzer für soziodemografische Charakteristika erzeugt, die einen direkten Vergleich mit Angaben der amtlichen Statistik ermöglichen. Da jedoch gerade der Effekt einer Kalibrierung auf die Ergebnisse einer Erhebung von Interesse sind, die sich nicht als soziodemografische Charakteristika beschreiben lassen, sondern bspw. durch diese beeinflusst werden, wie u. a. Meinungen und Prävalenzen, werden auch Schätzergebnisse zu diesen Variablen in Kapitel 3.5.2 betrachtet. Die nachfolgenden Schätzungen wurden eigens anhand der verfügbaren Datengrundlage für die vorliegende Arbeit durchgeführt, um die Eignung der jeweiligen Gewichtungsansätze adäquat diskutieren zu können.

3.5.1 Vergleich unterschiedlicher Designgewichtungsverfahren

Zur Beurteilung der unterschiedlichen Designgewichtungsverfahren wurden einige der soeben vorgestellten Ansätze im Rahmen einer Simulation auf ihre Unverzerrtheit untersucht. Hierzu wurde im ersten Schritt eine synthetische Grundgesamtheit im Umfang von $N = 1.000.000$ Einheiten unter der Verwendung des Programmes *R, Version 3.1.3* erzeugt. Die Zusammensetzung dieser synthetischen Grundgesamtheit wurde so gewählt, dass diese sowohl Landline- und Mobile-Onlys als auch Dual-User beinhaltet, die über eine unterschiedliche Anzahl von Anschlüssen der jeweiligen Netzart verfügen und deren Haushaltsgröße voneinander abweichend ist. Die Verteilung dieser Merkmale wird in Tabelle 9 zusammengefasst.

Tabelle 9 Merkmale der synthetischen Grundgesamtheit

Anzahl	Festnetz (Anteil)	Mobilfunk (Anteil)	HH-Größe (Anteil)
0	0,1	0,14	••
1	0,6	0,68	0,3
2	0,2	0,18	0,5
3	0,1	••	0,15
4	••	••	0,05

Die Tabelle gibt den Anteil bestimmter Merkmale (einzeln) an der synthetischen Grundgesamtheit wieder. Spalte zwei und drei gibt dabei die Verteilung der Anzahl der Festnetz- und Mobilfunkanschlüsse in der Population wieder. Spalte vier stellt die Verteilung der Haushaltsgröße in der Population dar. So entstammen bspw. 30 % der darin enthaltenen Einheiten aus Einpersonenhaushalten, während 5 % aus Vierpersonenhaushalten stammen. Weiterhin haben 10 % der Einheiten der Grundgesamtheit keinen und 60% einen Festnetzanschluss. Der Anteil der Einheiten, die zwei oder drei Festnetzanschlüsse besitzen liegt bei 20 % bzw. 10 %. Ebenfalls enthält diese Grundgesamtheit 14% Landline-Onlys. Der Anteil der Einheiten, die einen oder zwei Mobilfunkanschlüsse aufweisen, liegt entsprechend bei 68 % bzw. 18 %. Einheiten mit drei oder mehr Mobilfunkanschlüssen kommen in dieser Simulation nicht vor.

Diesbezüglich gilt festzuhalten, dass keine der Einheiten weder über Festnetz noch über Mobilfunk erreichbar ist. Die Coverage beider Auswahlgrundlagen gemeinsam ist demnach vollständig.⁴⁶ Es wird davon ausgegangen, dass die jeweilige Summe aller Festnetz- und Mobilfunknummern das entsprechende Nummernuniversum vollständig abdeckt und alle Rufnummern in den zugehörigen Auswahlgrundlagen enthalten sind. Dabei ergibt sich eine Auswahlgrundlage

⁴⁶ Dies wurde dadurch erreicht, dass alle Einheiten, die keine Festnetznummern aufweisen, mindestens eine Mobilfunknummer zugewiesen bekommen haben.

für Festnetzstichproben vom Umfang $M^F = 1.300.330$ und $M^C = 1.044.651$. Weiterhin wird in diesem stark vereinfachten Modell angenommen, dass keine Rufnummern existieren, die nicht zu einem geschalteten Anschluss einer Einheit der Zielpopulation führt.

Anhand der angeführten Parameter wurden die Inklusionswahrscheinlichkeiten für eine Festnetz- und eine Mobilfunkstichprobe der einzelnen Einheiten berechnet und zwei unabhängige Stichproben vom Umfang $m^F = 6.000$ und $m^C = 4.000$ gezogen.⁴⁷ Daran anschließend wurden anhand der beiden Stichproben die Designgewichte mittels *Single-Frame-Ansatz* (SF-Schätzer), *Composite-Ansatz* unter Verwendung von $3 \cdot 10^{48}$ (CW-Schätzer) sowie nach $3 \cdot 13^{49}$ (CW*-Schätzer) und nach der *Pseudo-Maximum-Likelihood-Methode* (PML-Schätzer) berechnet. Anhand der berechneten Designgewichte wurde dann die Schätzung der Anteile für die vier Ausprägungen einer Variable der Zielpopulation durchgeführt. Die Ergebnisse dieser Schätzung können in Tabelle 10 abgelesen werden.

Tabelle 10 Simulationsergebnisse nach Schätzverfahren

Variable	SF-Schätzer	CW-Schätzer	CW*-Schätzer	PML-Schätzer	Population
1	0,253	0,256	0,255	0,250	0,25
2	0,147	0,150	0,149	0,147	0,15
3	0,499	0,495	0,496	0,503	0,5
4	0,101	0,099	0,100	0,100	0,1
Designeffekt	1,33	1,50	1,45	1,50	●●
n_{eff}	7.540	6.657	6.882	6.886	●●

Aufgrund der Verwendung einer synthetischen Zielpopulation sind sämtliche für die Schätzung nach den jeweiligen Vorgehensweisen erforderlichen Parameter bekannt. Es zeigt sich, dass alle vier Vorgehensweisen die Variable der Zielpopulation (in der Punktschätzung) sehr genau abbilden und sich lediglich in der dritten Nachkommastelle unterscheiden. Der errechnete Bias nach den unterschiedlichen Verfahren ist für alle daher annähernd Null. Weiterhin wurde der modellbasierte Designeffekt der jeweiligen Schätzer nach 3.12 für die betrachtete Variable berechnet.⁵⁰ Generell kann der hier betrachtete Designeffekt für die vier Vorgehensweisen als gering bis moderat betrachtet werden, während der SF-Schätzer scheinbar dasjenige Verfahren ist, welches den geringsten Designeffekt mit einem Wert von ca. 1,33 hervorbringt. Die restlichen

47 Um das Modell möglichst einfach zu halten, wurde nicht nach Doppelziehung einer Einheit über beide Stichproben kontrolliert. Daher stimmen ca. 0,27 % der Einheiten in beiden Stichproben überein. Dieser Anteil kann jedoch als vernachlässigbar angesehen werden.

48 ϕ ist proportional zu den Stichprobengrößen.

49 ϕ ist proportional zu den effektiven Stichprobengrößen.

50 Da es sich um den modellbasierten Ansatz handelt und keine Klumpung vorherrscht, ist dieser Designeffekt bei einem gegebenem Schätzverfahren gleich für alle betrachteten Variablen des Datensatzes.

Designeffekte unterscheiden sich nur unwesentlich und liegen alle bei einem Wert von ca. 1,5. Da sich der hier betrachtete, modellbasierte Designeffekt bei der vorliegenden Vorgehensweise zu dessen Berechnung lediglich auf die Variation der Designgewichte bezieht, zeigt sich ferner, dass diese beim SF-Ansatz am geringsten ist bzw. dass die Designgewichte in diesem Ansatz näher zusammenliegen. Jedoch besteht nur ein geringer Unterschied zu den anderen Verfahren.

Die effektive Stichprobengröße (n_{eff}), also diejenige Stichprobengröße, die bei einer uneingeschränkten Zufallsauswahl notwendig wäre, um einen Schätzwert mit vergleichbarer Präzision zu erhalten, reduziert sich in allen Fällen um mehr als 25 %. Nicht weiter verwunderlich ist dabei, dass der SF-Schätzer noch die höchste effektive Stichprobengröße aufweist, bedenkt man, dass diese auch den geringsten Schätzwert des Designeffektes aufweist.

Der hier betrachtete, synthetische Datensatz lässt einige der Schwierigkeiten, die sich in einer Erhebungssituation einstellen, außer Acht. So hat eine reale Erhebung bspw. mit Antwortverweigerung umzugehen. Ferner führen bei einer realen Erhebung nicht alle Rufnummern zu einem geschalteten Anschluss einer Einheit der Zielpopulation und die Auswahlgrundlagen müssen auch nicht die gesamte Zielpopulation abdecken (wie bspw. aus Abbildung 1 ersichtlich wurde). Zuletzt sind oftmals nicht alle Parameter der Zielpopulation bekannt, die zur Bestimmung eines speziellen Schätzers notwendig sind. So ist bspw. die Anzahl der Einheiten in einer bestimmten Zielpopulation, die über einen bestimmten Auswahlrahmen erreicht werden kann, oftmals unbekannt. Dies liegt daran, dass beiden Auswahlrahmen lediglich ein *Universum an Rufnummern* zugrunde liegt, während Einheiten der Zielpopulation über mehrere dieser Rufnummern erreichbar sind. Die genaue Anzahl dieser Rufnummern stellt sich aber erst in der Befragungssituation heraus.

Folgend werden nur die ersten drei Verfahren auf die Ergebnisse der CELLA 2-Erhebung angewendet. Tabelle 11 vergleicht dazu die Ergebnisse der Erhebung ohne eine Designgewichtung sowie den SF-Schätzer, den CW-Schätzer und den CW*-Schätzer mit den Ergebnissen des Mikrozensus 2010 (Destatis (2016)). Zur Untersuchung der jeweiligen Gewichtungsverfahren wurde in einem ersten Schritt eine Imputation für fehlende Werte des Datensatzes unter der Verwendung der Software *R, Version 3.1.3* und des Paketes *mice* berechnet, um mit einer vollständigen Datenmatrix arbeiten zu können. Die Bildungsvariable wurde weiterhin dahingehend angepasst, dass deren Kategorien denjenigen des Mikrozensus 2010 entsprechen.

Tabelle 11 **CELLA 2: Höchster allgemeiner Schulabschluss nach Schätzverfahren (in %)**

Bildungsabschluss	Keine Gew.	SF-Schätzer	CW-Schätzer	CW*-Schätzer	MZ 2010
Ohne/ noch kein	3,5	4,0	4,0	3,9	7,6
Haupt-/ Volksschulabschl.	18,2	20,8	20,6	20,2	37,0
Polytechn. OS	6,0	6,3	6,0	5,9	7,1
Mittlere Reife	27,0	26,6	26,5	25,7	21,7
FH/ HS – Reife	43,6	40,8	41,3	40,1	25,8
Andere	1,7	1,6	1,7	1,6	0,4
Designeffekt	●● ⁵¹	1,38	1,51	1,45	●●

Im Gegensatz zu der Betrachtung des synthetischen Datensatzes lassen sich deutliche Unterschiede bei der Betrachtung der realen Erhebungsdaten aus CELLA 2 feststellen. So ist die Übereinstimmung zwischen der tatsächlichen Verteilung in der Population und den ermittelten Werten nicht mehr so eindeutig. Gründe hierfür können die (selektive/systematische) Antwortverweigerung bestimmter Zielpersonen, Mängel in der Auswahlgrundlage sowie diverse Fehlerquellen in der Messung bestimmter Erhebungsgegenstände sein. Gerade bei Dual-Frame-Erhebungen kann es, wie bereits Sand (2015a) berichtete, zu einer abweichenden Antwortneigung der erhobenen Einheit in Abhängigkeit des gewählten Mediums kommen. Genau dieser Aspekt wird in Kapitel 5 näher betrachtet.

Ein Vergleich mit den ungewichteten Ergebnissen der Erhebung zeigt jedoch auch hier, dass die Designgewichtung bereits zu einer Annäherung der geschätzten Werte an diejenigen der Zielpopulation führt. Dennoch sollte gerade in einer solchen Erhebung nicht auf eine anschließende Kalibrierung verzichtet werden. Die dafür geeigneten Verfahren werden folgend in Abschnitt 3.4 weiter erläutert.

Der Vergleich der einzelnen Designgewichtungsverfahren zeigt nur geringfügige Unterschiede der einzelnen Schätzwerte. Dennoch schneidet der CW-Schätzer schlechter ab als die anderen beiden Vorgehensweisen, die ähnlich genaue Werte liefern. Dies lässt sich insbesondere an einer höheren Abweichung der geschätzten Anteile dieser Variablen von denjenigen der amtlichen Statistik sowie anhand eines höheren Schätzwertes des Designeffektes begründen. Der Vergleich des modellbasierten Designeffektes zwischen dem SF- und dem CW*-Schätzer ist in diesem Beispiel ebenfalls weniger eindeutig. Auch hier wird deutlich, dass die Designgewichte des SF-Schätzers weniger stark variieren bzw. eine geringere Spannweite aufweisen.

51 Da hier eine modellbasierte Schätzung des Designeffektes unter der Berücksichtigung der Designgewichte vorgenommen wird, kann hier kein Designeffekt auf Basis dieser Schätzung angegeben werden. Dies ist jedoch nicht gleichbedeutend damit, dass kein Designeffekt besteht.

3.5.2 Vergleich der Kalibrierungsverfahren

Um eine bessere Einschätzung der drei vorgestellten Kalibrierungsverfahren zu erhalten, werden diese auf die Daten der Erhebung *Influenza 2014* angewendet und mit den Werten des Zensus 2011 sowie des Mikrozensus 2012 verglichen. Die vorgelagerte Designgewichtung wurde nach Gleichung 3.6 durchgeführt und nach der in Kapitel 3.3.1 beschriebenen Art und Weise angewendet. Weiterhin erfolgte, um mit einer vollständigen Datenmatrix zu arbeiten, eine Imputation fehlender Werte unter der Verwendung des Paketes *mice* der Software *R*, Version 3.1.3. Dabei wurde auf die Daten der gesamten Erhebung inklusive des Oversamples der mindestens 60-Jährigen zurückgegriffen. Als Anpassungsvariablen wurden *Alter*, *Bildungsabschluss* und *Geschlecht* verwendet.

Zur Berechnung der Poststratifizierungsgewichte wurde das Verhältnis der jeweiligen Zellen in den Angaben der offiziellen Statistik und innerhalb der Erhebung berechnet. Für die Anwendung des Raking-Verfahrens bzw. den Gewichten, die nach dem *Iterative Proportional Fitting (IPF)*-Algorithmus bestimmt wurden, mussten zusätzlich die jeweiligen Randverteilungen *AlterXGeschlecht*, *AlterXBildungsabschluss* und *GeschlechtXBildungsabschluss* aggregiert werden. Anschließend wurde der Algorithmus bis zur Konvergenz der Randverteilungen durchgeführt. Für die Berechnung der Gewichte des GREG-Schätzers mussten lediglich die Totalwerte der Population der einzelnen Ausprägungen bekannt sein. Diese wurden in der Anpassungsgrundlage durch Summieren gewonnen. Es zeigt sich, dass sich die Anforderungen an die Anpassungsvariablen, die zur Kalibrierung verwendet werden, deutlich unterscheiden.

Die Auswirkung der jeweiligen Kalibrierungsverfahren auf die Schätzwerte für die Anpassungsvariablen, die aus der Erhebung stammen, wird anhand von Tabelle 12 verdeutlicht.

Tabelle 12 Influenza 2014: Alter nach Art der Kalibrierung (in %)

Alter	Ungewichtet	Designgew.	Poststrat	IPF	GREG	Zensus 2011 ⁵²
18 - 29	9,4	15,4	14,9	14,9	14,9	14,8
30 - 49	22,7	30,4	35,0	35,0	35,0	35,0
50 - 59	14,2	19,0	17,4	17,3	17,5	17,8
60 - 69	27,5	17,9	13,7	13,7	13,7	13,6
70+	26,3	17,4	19,0	19,0	19,0	18,8

Hier zeigt sich, dass die Verwendung der reinen Stichprobenergebnisse in verzerrten Schätzergebnissen resultiert. Die Designgewichtung führt erstmals eine Reduktion dieser Verzerrung herbei.

⁵² Ergebnisse nach Trimmen der Gewichte; zuvor stimmen die Ergebnisse mit den Anpassungsvariablen (annähernd) überein.

Aufgrund von (selektivem) Nonresponse besteht jedoch weiterhin eine Diskrepanz zwischen den Ergebnissen der Erhebung und der Verteilung der amtlichen Statistik. Erst durch die anschließende Kalibrierung stimmen die Schätzwerte mit den Daten der amtlichen Statistik (annähernd) überein. Der geringfügige Unterschied ist an dieser Stelle nicht weiter verwunderlich, da die Altersvariable auch zur Kalibrierung verwendet wurde. Die noch bestehenden Abweichungen erklären sich durch die Verwendung von getrimmten Gewichten.

Die Betrachtung der Schätzwerte weiterer Variablen, die nicht bei der Kalibrierung herangezogen wurden, ist von höherem Informationsgehalt und bietet eine bessere Möglichkeit die Kalibrierungsverfahren zu beurteilen. Tabelle 13 vergleicht die Schätzwerte der Haushaltsgröße getrennt nach Kalibrierungsverfahren mit dem Mikrozensus 2012. Dieser wurde gewählt, da entsprechende Angaben im Zensus 2011 nicht verfügbar waren.

Tabelle 13 Influenza 2014: Haushaltsgröße nach Art der Kalibrierung (in %)

Haushaltsgröße	Ungewichtet	Designgew.	Poststrat	IPF	GREG	MZ 2012
1 Person	48,0	39,4	42,4	42,2	42,2	40,5
2 Personen	36,6	36,9	34,3	34,3	34,2	34,5
3 Personen	7,8	11,1	10,6	10,7	10,9	12,5
4 Personen	4,4	6,2	6,4	6,3	6,4	9,2
5+ Personen	1,4	2,3	2,6	2,6	2,5	3,3

Hier wird, ebenso wie bei dem dargestellten Vergleich in Tabelle 12, deutlich, dass erst durch die Designgewichtung sowie eine anschließende Kalibrierung, eine Annäherung an die Daten der amtlichen Statistik erfolgt und die Verwendung der reinen Stichprobenergebnisse zu verzerrten Schätzergebnissen führt.

Der Vergleich der Verfahren zur Anpassungsgewichtung betont, dass alle drei (zumindest in der Punktschätzung) keine stark voneinander abweichenden Ergebnisse liefern. Jedoch sind sowohl die Anforderungen an die Anpassungsvariablen als auch der Aufwand zur Bestimmung der jeweiligen Gewichte unterschiedlich. Zur Durchführung einer Poststratifizierung ist die gemeinsame Verteilung aller Anpassungsvariablen notwendig, die oftmals durch die amtliche Statistik nicht zur Verfügung steht. Für die Anwendung des IPF-Algorithmus sind lediglich die Randverteilungen der jeweiligen Variablen erforderlich. Diese sind oftmals verfügbar. Der GREG-Schätzer benötigt "nur" die Totalwerte der Population der entsprechenden Ausprägungen. Darüber hinaus hat der GREG-Schätzer den Vorzug, dass die Inklusionswahrscheinlichkeiten der einzelnen Erhebungseinheiten in der Anpassungsgewichtung berücksichtigt werden und dass er bei unbesetzten Zellen weniger problematisch anzuwenden ist als der IPF-Algorithmus sowie der Poststratifizierung. Insgesamt eignen sich aber zumindest bei der Punktschätzung alle drei

Kalibrierungsverfahren zur Gewichtung aufgrund von selektivem Nonresponse.

4 Zwischenfazit

Durch die Auseinandersetzung mit der Literatur wird ersichtlich, dass Telefonstichproben für die Erhebung von Umfragedaten für die bundesdeutsche Bevölkerung noch immer eine wichtige Rolle spielen. Im historischen Kontext zeigt sich jedoch, dass eine reine Listenauswahl bspw. anhand eines Telefonverzeichnisses nicht mehr ausreicht, um verlässliche Aussagen über die Gesamtbevölkerung zu treffen. So stellten sowohl die Wiedervereinigung Deutschlands als auch die Aufhebung der Eintragspflicht die Stichprobenziehung für Festnetzstichproben vor große Herausforderungen. Auch wenn beide Ereignisse die Anwendung von Telefonstichproben (unter der Verwendung einer Listenauswahl) deutlich einschränkten, lagen als Ursachen unterschiedliche Faktoren zugrunde. Die Wiedervereinigung stellte die Umfrageforschung über Festnetzbefragungen vor das Problem, dass sehr schnell ein großer Anteil der Bevölkerung hinzukam, der nicht über einen Festnetzanschluss erreichbar war. So verfügten 1990 lediglich ca. 20 % der Haushalte aus den neu hinzugekommenen Bundesländern über einen Festnetzanschluss, während ca. 92 % aller Haushalte der westdeutschen Bundesländer über einen Festnetzanschluss erreichbar waren. Durch diese Diskrepanz wurde die Verwendung von telefonischen Befragungen in den frühen 1990er Jahren gravierend beeinträchtigt. Ein erstes Angleichen des Anteils der Haushalte, die über Festnetzanschlüsse erreichbar waren, erfolgte erst in den frühen 2000er-Jahren.

Deutlich langsamer machte sich der Einfluss der Aufhebung der Eintragspflicht für die Verwendung des Telefonverzeichnisses in der Auswahl von Erhebungseinheiten bemerkbar. Durch diese stand einer wachsenden Anzahl von Festnetzanschlüssen ein stetig geringer werdender Anteil an Haushalten gegenüber, die in einem Telefonverzeichnis geführt wurden. Erste Bedenken durch diese Beeinträchtigung wurden bspw. durch S. Häder (1996) geäußert. Einschränkungen in der Schätzung bevölkerungsrelevanter Parameter wurden genau dann befürchtet, wenn sich eingetragene und nicht-eingetragene Haushalte (sog. Non-Pubs) bezüglich ihrer soziodemografischen Charakteristika unterscheiden. Weiterhin bestand die Annahme, dass sich der zu Beginn vergleichsweise geringe Anteil an Haushalten, die zwar einen Festnetzanschluss besaßen, jedoch nicht in einem Telefonverzeichnis geführt war, durch die "Kosten" des Austragens begründen ließ. So wurde vermutet, dass die Entscheidung den eigenen Telefonanschluss aus einem Telefonverzeichnis austragen zu lassen, seltener vollzogen wurde. Bei einem Anschlusswechsel bspw. aufgrund eines Umzuges oder Anbieterwechsels wurde jedoch der neue Anschluss nicht mehr in ein Telefonverzeichnis eingetragen. Gerade anhand dieses Sachverhaltes kann bereits vermutet werden, dass sich Non-Pubs bspw. aufgrund ihrer Mobilität in ihren soziodemografischen Charakteristika von eingetragenen Haushalten unterscheiden. Diese Vermutung wurde bspw. durch Untersuchungen von Heckel (2002) bestätigt. So finden sich unter anderem deutlich

mehr Haushalte in ländlichen Regionen in einem Telefonverzeichnis als in urbanen Gebieten, die durch eine höhere Mobilität der Haushalte geprägt sind. In jüngeren Untersuchungen zeigt sich weiterhin, dass sich nicht nur der Anteil der Haushalte, die in einem Telefonverzeichnis geführt sind, deutlich verringert hat, sondern auch, dass sich das Eintragsverhalten zwischen den alten und neuen Bundesländern angeglichen hat. So waren 2015 bspw. durchschnittlich ca. 9,9 % aller Rufnummern, die sich in einem Block von 100 aufeinanderfolgenden Rufnummern befanden, in westdeutschen Bundesländern eingetragen (Eintragsdichte). In den neuen Bundesländern betrug dieser Wert 9,8 %.

Die Möglichkeit der Stichprobenziehung wird in diesem Kontext daher wesentlich durch dessen Auswahlgrundlage determiniert. Dass das Telefonverzeichnis als alleinige Auswahlgrundlage bereits zur Mitte der 1990er-Jahre unzulänglich war, diskutierten bereits Gabler und S. Häder (1997). Aufgrund der Gestaltung des deutschen Rufnummernraumes war jedoch das in der amerikanischen Literatur oftmals vorgeschlagene RDD-Verfahren nur schwer anwendbar. So diskutiert bspw. Schnell et al. (2013), dass die Gestaltung des amerikanischen Rufnummernraumes ein solches Vorgehen begünstigt, da dieser ausschließlich Rufnummern beinhaltet, die eine Länge von zehn Stellen aufweisen. Der deutsche Rufnummernraum besteht indes aus einer Kombination von Vorwahlbereichen, die zwei bis fünf Stellen aufweist,⁵³ sowie einer Rufnummer, die zwischen fünf und neun Stellen haben kann. Daher hebt S. Häder (2000) hervor, dass ein reiner RDD-Ansatz in Deutschland ineffizient sei, da eine Auswahlgrundlage über alle möglichen Kombinationen der Rufnummern und deren Länge gebildet werden müsste. Ein solches Verfahren hat deshalb eine Hitrate, die zu gering sei, um damit effizient telefonische Befragungen durchzuführen.

Aufgrund dieser Probleme wurde zur Mitte der 1990er-Jahre eine Methode durch Gabler und Häder entwickelt, die die Ziehung von Telefonstichproben mit gleichen Auswahlwahrscheinlichkeiten der Rufnummern und einem handhabbaren Aufwand ermöglicht. Basierend auf den eingetragenen Rufnummern eines Telefonverzeichnisses ermöglicht diese Methode das Ziehen von Rufnummern unabhängig davon, ob diese eingetragen oder nicht eingetragen sind. Mittlerweile ist dieses Verfahren für Festnetzstichproben gängige Praxis und wird in der in Kapitel 2.2 dargestellten Art und Weise von GESIS sowie in einer ähnlichen Form durch den ADM angewendet.

Mit der Durchsetzung des Mobiltelefons als Massentelekommunikationsmedium wurde die Umfrageforschung mittels Telefonbefragung gegen Ende der 1990er Jahre vor weitere Probleme gestellt. S. Häder (2000) berichtete bereits im Zuge seiner Delphi-Befragung, dass die Einschätzung der Experten dieses Bereiches bis 2005 eine zunehmende Verwendung dieses Telekommunikationsmediums prognostizierte. Damit verbunden ist auch die Einschätzung, dass der Anteil derjenigen, die ausschließlich mobile Telekommunikation verwenden, bis zu diesem

53 Die Angaben beziehen sich auf die Vorwahlbereiche ohne die deutschlandinterne "0".

Jahr ansteigen wird. Auch Fuchs berichtete bereits im Jahr 2002, dass sich das Mobilfunkgerät vor allem für diejenigen Einheiten der deutschen Population eigne, die durch eine hohe Mobilität geprägt sind.

Problematisch an der Durchsetzung des Mobilfunkgerätes war indes nicht die zunehmende Marktpenetration per se, die bereits 2005 darin resultierte, dass ca. 75 % der bundesdeutschen Population über Mobilfunk erreichbar waren, sondern, dass der Anteil der Haushalte zunahm, die mittlerweile ausschließlich über Mobilfunk erreichbar sind. Nach jüngeren Untersuchungen der Europäischen Kommission im Rahmen des Eurobarometers 2015 lässt sich der Anteil der Haushalte in Deutschland, die über mindestens einen Mobilfunkanschluss verfügen, auf ca. 91 % schätzen, während der Anteil der Mobile-Only-Population mit 15 % angegeben wird. Demnach ist (und war bereits Mitte der 2000er-Jahre) ein substanzieller Teil der deutschen Bevölkerung nicht mehr über eine gewöhnliche Festnetzstichprobe zu erreichen. Das Auslassen gerade dieser Population wäre im Rahmen der Umfrageforschung nur dann vertretbar, wenn sich Haushalte, die ausschließlich über Mobilfunk erreichbar sind, nicht systematisch von denjenigen, die über einen Festnetzanschluss verfügen, unterscheiden. Sowohl amerikanische als auch deutsche Untersuchungen kommen jedoch zu der Erkenntnis, dass Ergebnisse für Haushalte, die ausschließlich über Mobilfunk erreichbar sind, in ihren soziodemografischen Charakteristika deutlich von Haushalten mit Festnetzanschluss divergieren. So wurde bereits im Rahmen der CELLA 1- Erhebung durch Graeske und Kunz (2009) hervorgehoben, dass Mobile-Onlys deutlich häufiger alleinstehend, männlich mit niedrigem Einkommen und mittleren Bildungsabschluss sind. Daher ist zu vermuten, dass der Ausschluss gerade dieser Subpopulation in einer Erhebung zu einer verzerrten Schätzung von interessierenden Merkmalen führen würde. Auf der anderen Seite existieren weiterhin Haushalte, die ausschließlich über Festnetz erreichbar sind (Landline-Onlys). Der Ausschluss dieser 10 % der Haushalte lässt ähnliche negative Effekte auf die Schätzung anhand von Erhebungsdaten vermuten. Aufgrund dessen wird mittlerweile für bundesdeutsche CATI-Erhebungen empfohlen, Dual-Frame-Ansätze durchzuführen, die sich aus einer oder mehreren unabhängigen Festnetz- und Mobilfunkstichproben zusammensetzen.

Zur Durchführung einer solchen Dual-Frame-Erhebung werden in der Literatur zwei unterschiedliche Vorgehensweisen diskutiert. So beschreiben S. Häder et al. (2009a) in Anlehnung an die von der AAPOR hervorgebrachten *Considerations and Guidelines*, dass es die Möglichkeit gebe, einem Screening- oder einem Overlap-Ansatz zu folgen. Bei einem Screening-Ansatz wird im Rahmen einer der beiden Stichproben bewusst ausschließlich nach der Only-Population dieses Telekommunikationsmediums gescreent, während die restliche Population (Dual-User und Only-Einheiten des anderen Telekommunikationsmediums) über die andere Stichprobe befragt werden. Indes befragt ein Overlap-Ansatz Zielpersonen direkt über das jeweilige Telekommunikationsmedium, über das diese erreicht wurden. Die Entscheidung darüber, welcher der beiden Ansätze verwendet wird, ist Kennedy (2007) zufolge von der Varianz des aus dieser Erhebung

entstehenden Schätzwertes, der Verzerrung der geschätzten Parameter (bspw. aufgrund von Nonresponse) sowie von den Kosten des Ansatzes abhängig. Gerade die Kosten der Durchführung eines Screening-Ansatzes sind davon abhängig, wie groß der Anteil der Only-Haushalte des Telekommunikationsmediums ist, nach dem gescreent wird. So sind die mit einem solchen Ansatz verbundenen Kosten bei einem Screening nach Mobile-Onlys bspw. deutlich geringer, wenn der Anteil dieser Subpopulation in der Zielgesamtheit vergleichsweise hoch ist, da deutlich weniger kontaktierte Haushalte in der Mobilfunkstichprobe ausgescreent werden. Deshalb führen S. Häder et al. (2009a) an, dass von einem Screening-Ansatz aufgrund des vergleichsweise geringen Anteils von Mobile-Only-Haushalten abzuraten ist, da dies mit einem zu großen Aufwand verbunden wäre und die Hitrate eines solchen Ansatzes zu gering sei. Von Nachteil ist des Weiteren, dass bei diesem Ansatz Haushalte zu einem Zeitpunkt, zu dem diese bereits eine hohe Teilnahmebereitschaft zeigen, ausgescreent werden. Daher wird für die Durchführung bundesweiter Dual-Frame-Erhebungen i. d. R. ein Overlap-Ansatz verwendet. Auch im Rahmen dieser Arbeit wird bei der Beschreibung verschiedener Gewichtungsansätze angenommen, dass die Daten aus einer Erhebung stammen, die einen solchen Ansatz verfolgt.

Der Einsatz von Mobiltelefonen als weiteres Telekommunikationsmedium bei telefonischen Erhebungen bringt jedoch auch zusätzliche Herausforderungen für die Umfrageforschung mit sich. Neben technischen Problemen, die bspw. durch den Einsatz von Mobilfunkstichproben in der Verbindung mit Predictive Dialern und der Fehlinterpretation diverser Signale entstehen können, liegt eine große Herausforderung in dem Umgang mit und bestimmten Nutzungsgewohnheiten von Mobiltelefonen, sowie deren generellen Eigenschaften. So wird im Allgemeinen davon ausgegangen, dass ein Mobiltelefon ein personalisierter Gegenstand ist, der direkt einer Erhebungseinheit zugeordnet werden kann. Im Gegensatz dazu wird ein Festnetzanschluss i. d. R. mit einem Haushalt verbunden. Die Ziehung über ein Mobiltelefon ist daher als einstufiger Prozess zu verstehen, während eine Festnetzstichprobe ein zweistufiges Verfahren ist, bei der zuerst der Haushalt ausgewählt und dann innerhalb des Haushaltes eine Erhebungseinheit ermittelt wird. Damit verbunden besteht bei Mobiltelefonen die Möglichkeit, über eine Mobilfunkstichprobe bei einem positiven Kontakt dennoch keine Erhebungseinheit eines bestimmten Mindestalters zu erreichen. Bei Festnetzstichproben tritt dieses Problem seltener auf, da sich für gewöhnlich mindestens eine Person des Haushaltes ausfindig machen lässt, die nicht minderjährig ist.

Weiterhin können technische Unterschiede zwischen Festnetz- und Mobilfunkanschlüssen die Befragung über das Mobilfunknetz erschweren. So ist bspw. das Vorhandensein einer bestehenden Funkverbindung des Mobilfunkgerätes ("Signal") Voraussetzung für eine erfolgreiche Befragung. In Abhängigkeit der Situation, innerhalb derer die Befragung stattfindet, können somit unterschiedliche Faktoren eintreten, die die Qualität der Befragung beeinflussen oder diese sogar abrupt enden lassen. Ebenso kann sich die Befragungssituation durch den mobilen Charakter des Telekommunikationsmediums stark von derjenigen einer Befragung über

das Festnetz unterscheiden, was dazu führen mag, dass eine Erhebungseinheit nicht bereit ist, bestimmte Angaben zu machen. Die Situation innerhalb derer sich eine Erhebungseinheit gerade befindet, kann weiterhin zu einer Gefährdung derselben führen. So kann eine Befragung über einen Mobilfunkanschluss bspw. an öffentlichen Plätzen oder während des Lenkens eines Fahrzeuges stattfinden. Beides kann die Auskunftsbereitschaft einer Erhebungseinheit und die Interviewqualität negativ beeinflussen und erscheint bei Befragungen über das Festnetz unwahrscheinlich. Jedoch verweisen Kühne und S. Häder et al. (2009a) darauf, dass auch ein Großteil der Mobilfunkinterviews zu Hause durchgeführt wird.

Aufgrund der technischen Gegebenheiten von Mobilfunkanschlüssen lassen sich weiterhin niedrigere Responseraten im Rahmen von Mobilfunkstichproben feststellen. Dies liegt oftmals darin begründet, dass ein größerer Anteil von Signalen schwer zuzuordnen ist und es deshalb schwieriger zu determinieren ist, ob ein Anschluss geschaltet ist oder nicht. In solchen Fällen zeigen jedoch Kunz und Fuchs (2011) sowie Struminskaya et al. (2011), dass die Verwendung von sog. HLR-Lookup-Verfahren sowohl zur Klärung des Status einer Mobilfunknummern als auch zur Kostenersparnis im Rahmen der Erhebung führen kann. Bestärkt werden diese Befunde durch die Ergebnisse der Untersuchung Vermln, die darlegen, dass dieses Verfahren, in Abhängigkeit des gewählten Service-Anbieters, durchaus dazu beitragen kann, den wahren Status einer Mobilfunkrufnummer noch vor dem eigentlichen Erhebungsbeginn zu ermitteln und somit nicht-vergebene Anschlüsse bereits im Vorfeld zu identifizieren.

Frame-Imperfection, die im Sinne von Under- und Overcoverage bei telefonischen Befragungen oftmals zu beobachten sind, werden häufig versucht durch Maßnahmen der Stichprobenziehung zu reduzieren. Eine solche Maßnahme ist bspw. die Durchführung von Dual-Frame-Erhebungen im eigentlichen Sinne. Um unverzerrte Schätzergebnisse zu erhalten, sollten solche Maßnahmen jedoch auch im Zuge der Designgewichtung berücksichtigt werden.

Die Designgewichtung solcher Dual-Frame-Erhebungen kann auf mehrere Arten erfolgen. Grundlage jeder der hier diskutierten Verfahren ist jedoch der Horvitz-Thompson-Schätzer, dessen Designgewichte sich als Inverse der Inklusionswahrscheinlichkeit bestimmen lassen. Vielmehr von Relevanz ist daher die Frage, ob eine gemeinsame Inklusionswahrscheinlichkeit der Dual-User über beide Ziehungsrahmen berechnet werden soll oder ob eine Kombination nach einzelnen Stichproben angestrebt wird. Ausgangspunkt ist hierbei immer die Verwendung eines Overlap-Ansatzes, da die Inklusionswahrscheinlichkeit einer Einheit bei einem Screening-Ansatz sowie die daraus resultierenden Designgewichte vergleichsweise einfach zu bestimmen sind. Begründet ist dies dadurch, dass die drei Domains (Landline-Only, Dual-User und Mobile-Only) als disjunkte Schichten betrachtet werden können und daher einzeln gewichtet werden können. Bei einem Overlap-Ansatz trifft dies nicht mehr zu, da sowohl Einheiten der Festnetz- als auch der Mobilfunkstichprobe in der Domain der Dual-User enthalten sind.

Gerade in der englischsprachigen Literatur wird oftmals ein sog. Composite-Ansatz bei der

Gewichtung von Dual-Frame-Erhebungen verwendet, der mittels Konvexkombination einen designgewichteten Schätzer erzeugt. Bei diesem Ansatz werden in einem ersten Schritt die Designgewichte separat nach Auswahlgrundlage für die jeweilige Erhebungseinheit berechnet. Für Einheiten der beiden Only-Domains entspricht das Designgewicht der jeweiligen Inversen der Inklusionswahrscheinlichkeit dieser Einheit für die entsprechende Auswahlgrundlage. Die Designgewichte der Dual-User werden auf ähnliche Art und Weise bestimmt, jedoch wird dieses Designgewicht nochmals mit einem Konvexfaktor (ϕ bzw. $1 - \phi$) multipliziert. Das Designgewicht wird daher nur anhand der Inklusionswahrscheinlichkeit der Einheit für die jeweilige Auswahlgrundlage, über die diese in die Erhebung gelangt ist, berechnet. Der Konvexfaktor sollte für Dual-User zwischen 0 und 1 liegen, wodurch die Summe dieser beiden Faktoren, die an die Designgewichte der Dual-User heran multipliziert wird, 1 entspricht.

Zur Wahl von ϕ werden in der Literatur mehrere Ansätze diskutiert. Das einfachste Vorgehen zur Ermittlung des Composite-Faktors ist dabei das von Hartley 1962 vorgeschlagene Fixed-Weight-Adjustment. Dabei wird bereits im Vorfeld der Wert dieses Faktors festgesetzt. So könnte, wie bspw. Brick et al. (2006) vorschlägt, für ϕ der Wert 0,5 festgelegt werden, unter der Annahme, dass der Bereich der Dual-User doppelt erfasst wird und dieses Vorgehen das Designgewicht (für Einheiten des Überlappungsbereiches) um die Hälfte reduziert. Andere vergleichsweise einfache Vorgehensweisen bestimmen ϕ anhand des Verhältnisses einer der beiden Stichproben zum gemeinsamen Umfang beider Stichproben oder berechnen dieses Verhältnis anhand der effektiven Stichprobengrößen. All diesen Verfahren ist gemein, dass dadurch eine interne Konsistenz der Gewichte erreicht wird, da für alle betrachteten Variablen der Erhebung die gleichen Designgewichte bestehen. Im Gegensatz zu diesen Ansätzen steht die Verwendung eines optimalen Schätzers, der ϕ so festlegt, dass der daraus resultierende Schätzer varianzminimierend für eine betrachtete Variable ist. Problematisch an diesem Vorgehen ist einmal, dass dafür die Varianzen und die Kovarianz für eine entsprechende Variable aus den beiden Populationen (Festnetz- und Mobilfunkpopulation) bekannt sein oder aus der Stichprobe geschätzt werden müssen. Weiterhin unerwünscht ist, dass durch die Anwendung dieses Verfahrens unterschiedliche Designgewichte für jede betrachtete Variable berechnet werden müssen. Weitere, komplexere Verfahren stellen bspw. der von Burmeister (1972) vorgestellte Ansatz oder die Verwendung des Pseudo-Maximum-Likelihood Schätzers dar. Beiden Ansätzen ist gemein, dass diese die Populationsgröße der beiden Auswahlgrundlagen verwenden. Da im Rahmen von Telefonstichproben ein solcher Wert aufgrund der technischen Gegebenheiten und der großen Anzahl nicht vergebener Anschlüsse diese Angabe oftmals fehlt, müsste die Populationsgröße, die durch eine bestimmte Auswahlgrundlage abgebildet wird, geschätzt werden.

Eine andere Vorgehensweise ist die Verwendung des gerade in der deutschen Anwendung von Dual-Frame-Erhebungen populären Single-Frame-Ansatzes. Bei dieser Art der Designgewichtung wird der Überlappungsbereich, der aus beiden Auswahlgrundlagen (und Stichproben)

entsteht, als eine Art "Super Stratum" verstanden. Der Umgang mit diesem wird daher durch die Berechnung der gemeinsamen Inklusionswahrscheinlichkeit einer Erhebungseinheit über beide Auswahlgrundlagen erreicht. Demnach müssen im Zuge der Befragung alle relevanten Parameter, die zur Bestimmung der Inklusionswahrscheinlichkeit einer Einheit erforderlich sind, für die Stichprobenziehung aus beiden Stichproben erfasst werden. Das von Gabler und Häder im Zuge der CELLA-Erhebungen entwickelte Vorgehen, dass anhand von Gleichung 3.6 beschrieben wurde, ist in Deutschland mittlerweile gängige Praxis. Hierbei wird die gemeinsame Inklusionswahrscheinlichkeit einer Erhebungseinheit über beide Auswahlgrundlagen beschrieben, unter der Nebenannahme, dass die Wahrscheinlichkeit einer Einheit über beide Auswahlgrundlagen gleichzeitig in die Erhebung zu erlangen, vernachlässigbar ist. Die dadurch berechnete gemeinsame Inklusionswahrscheinlichkeit wird dann als Inverse zur Bestimmung des Designgewichtes des Horvitz-Thompson-Schätzers verwendet. Zu beachten bei diesem Ansatz ist jedoch, dass die Anforderungen an die Befragung einer Zielperson höher sind, da die relevanten Parameter zur Berechnung der Inklusionswahrscheinlichkeit über beide Stichproben erhoben werden müssen. Eine Konvexkombination benötigt indes lediglich die relevanten Parameter, die zur Bestimmung der Inklusionswahrscheinlichkeit einer Einheit benötigt werden, über die diese tatsächlich in die Erhebung gelangt ist.

Beide Vorgehensweisen wurden im Zuge dieser Arbeit anhand eines synthetischen Datensatzes überprüft. Dafür wurde mittels des Statistikprogrammes *R* in einem ersten Schritt eine künstliche Population geschaffen, die über alle relevanten Parameter, die zur Bestimmung der Inklusionswahrscheinlichkeiten benötigt werden, verfügt und bei der für jede Einheit der Population alle Parameter bekannt sind. Die künstliche Population beinhaltet sowohl Only- als auch Dual-User-Einheiten. Anhand dieser Population wurde die Schätzung einer Substichprobe nach den unterschiedlichen Gewichtungsansätzen unter dem Aspekt der Erwartungstreue des Schätzers und des modellbasierten Designeffektes untersucht. Das Ergebnis dieser Untersuchung zeigt, dass beide Vorgehensweisen erwartungstreue Schätzergebnisse liefern. Im Rahmen dieser Untersuchung war der modellbasiert-geschätzte Designeffekt des Schätzers, der anhand des Composite-Ansatzes, der ϕ mittels effektiver Stichprobengröße berechnet, am geringsten unter den beiden, die Konvexkombinationen verwenden.

Der Schätzer, der durch die Designgewichtung ermittelt wird, ist genau dann unverzerrt, wenn alle Einheiten der Zielpopulation eine positive Auswahlwahrscheinlichkeit aufweisen und keine weiteren systematischen Fehlerquellen eintreten. Jedoch ist in einem Großteil der Erhebungen unter anderem ein Nonresponse bestimmter Erhebungseinheiten zu beobachten. Ist dieser systematisch, so kann er die Schätzergebnisse weitergehend verzerren. In solchen Fällen kann eine nachträgliche Kalibrierung der Daten helfen, die Lücke zwischen den aus den Erhebungsdaten geschätzten Werten und deren tatsächlicher Verteilung in der Zielpopulation zu schließen oder zu verringern. In den meisten Ansätzen, die sich der Anpassungsgewichtung widmen, wird von

einem Zusammenhang bestimmter (Hilfs-) Variablen und der Verweigerung der Beantwortung der Untersuchungsvariable ausgegangen. Ein solcher Missing At Random (MAR) –Zusammenhang unterstellt somit eine systematische Relation zwischen soziodemografischen Charakteristika einer Erhebungseinheit und deren Bereitschaft, an einer Erhebung teilzunehmen. Grundsätzlich kann Nonresponse zu einer Reduktion der Genauigkeit und Präzision der Schätzergebnisse beitragen, da die effektive Stichprobengröße hierdurch sinkt. Weiterhin kann ein solcher Nonresponse eine höhere Variation der Gewichte mit sich führen. Der bedeutendste Einfluss wird jedoch den systematischen Unterschieden zwischen Respondenten und Nicht-Respondenten in der Begutachtung von Fehlerquellen zugesprochen.

Jedoch darf die Kalibrierung nicht als Allheilmittel betrachtet werden, die eine mangelhafte Stichprobenpraxis ausgleicht. Vielmehr besteht deren Aufgabe darin, die Verzerrung aufgrund einer systematischen Teilnahmeverweigerung bestimmter Erhebungseinheiten zu reduzieren. Weiterhin kann eine Kalibrierung auch fernab des Nonresponses zur nachträglichen Schichtung verwendet werden, wenn Informationen, die für eine Schichtung erforderlich sind, im Vorfeld der Stichprobenziehung nicht verfügbar sind. Im Rahmen der vorliegenden Arbeit wurden die drei gängigsten Kalibrierungsverfahren Poststratifizierung, Raking und die Verwendung des GREG-Schätzers vorgestellt. Anhand der Erhebung Influenza 2014 konnte gezeigt werden, dass alle drei Verfahren vergleichbare Ergebnisse (in der Punktschätzung) liefern, dass jedoch auch nach einer Kalibrierung eine Abweichung zwischen den gewichteten und kalibrierten Variablen einer Erhebung und der tatsächlichen Ausprägung einer Variable (auf die nicht angepasst wurde) in der Zielpopulation herrscht. Die Wahl des Kalibrierungsverfahrens wird i. d. R. durch die Art und Weise, wie die Daten für die Zielpopulation und in der Erhebung verfügbar sind sowie durch die Anzahl der zur Kalibrierung verwendeten Hilfsvariablen bestimmt. Dabei gilt zu beachten, dass eine Kalibrierung für jegliche Variablen, nach denen Untersuchungen der Subpopulationen vorgenommen werden sollen, erfolgt.

Bei der Verwendung einer Poststratifizierung ist es erstrebenswert, dass innerhalb der Erhebungsdaten keine oder nur wenige nicht besetzte Zellen in der Verteilungsmatrix der Kreuzkombination der einzelnen Anpassungsvariablen vorhanden sind. Dies gilt genau dann, wenn diese Zellen in der Zielpopulation besetzt sind. Die Gefahr, dass ein solcher Sachverhalt vorzufinden ist, steigt mit der Anzahl der Anpassungsvariablen. Weiterhin wird für eine solche Anpassung eine vollständige Datenmatrix als Kreuztabelle aller Anpassungsvariablen der Zielpopulation benötigt.

Sind diese Vorgaben nicht erfüllt, kann durch ein Rakingverfahren erreicht werden, dass zumindest die Randverteilungen der Anpassungsvariablen in der Erhebung und der Zielpopulation übereinstimmen. Das bekannteste dieser Verfahren stellt das iterative Randsummenverfahren dar. Bei diesem werden die Randsummen der Erhebung und der Zielpopulation iterativ durch eine Soll-durch-Ist-Gewichtung angepasst, bis die dadurch berechneten Gewichte sich nicht mehr (stark) verändern. Der Vorzug dieses Vorgehens besteht darin, dass nicht mehr eine vollständige

Kreuztabelle der Hilfsvariablen der Erhebung und der Zielpopulation benötigt wird. Somit kann vermieden werden, dass einzelne Zellen der Anpassungsvariablen kaum oder nicht besetzt sind. Jedoch wird hierdurch lediglich eine Anpassung auf die Randverteilung und nicht auf die einzelnen Kreuzkombinationen der Anpassungsvariablen erwirkt.

Der verallgemeinerte Regressionsschätzer (GREG) ist eine Kombination aus Design- und Anpassungsgewichtung, der anhand eines Vektors von Hilfsvariablen eine Kalibrierung der Erhebungsdaten erreicht. Ein Vorteil dieses Verfahrens besteht darin, dass dieser direkt die Inklusionswahrscheinlichkeit der Erhebungseinheiten in der Kalibrierung berücksichtigt. Unabhängig von dem Vorhandensein von Nonresponse kann dieser auch zur Reduktion der Varianz des gewichteten Schätzers beitragen. Zusätzlich hat dieses Verfahren den Vorzug, dass lediglich die Totalwerte der Ausprägungen der Hilfsvariablen in der Population benötigt werden und, dass dieser weniger problematisch bei nicht besetzten Zellen anzuwenden ist. Die Kalibrierung anhand eines GREG-Schätzers kann sich jedoch rechen- und arbeitsintensiver gestalten. Grundsätzlich eignen sich jedoch alle drei Verfahren zur Berücksichtigung von selektivem Nonresponse.

Dass eine Kalibrierung von Erhebungsdaten durchaus sinnvoll sein kann, zeigt die Untersuchung der unterschiedlichen Gewichtungsverfahren anhand der Erhebung Influenza 2014. Jedoch lässt sich auch bei dieser Betrachtung feststellen, dass noch immer eine Abweichung zwischen den kalibrierten Schätzergebnissen einer Erhebung und der tatsächlichen Verteilungen in der Zielpopulation existiert, wenn auf dieses Merkmal nicht angepasst wurde. Begründet werden kann diese Abweichung anhand einer Mehrzahl von Ursachen. So besteht die Vermutung, dass die MAR-Annahme nicht vollständig zutreffend ist. Auch ist anzunehmen, dass selbst wenn die MAR-Annahme zutreffend ist, die Hilfsvariablen, die zur Kalibrierung verwendet werden, nicht vollständig den Einfluss auf das Teilnahmeverhalten begründen.

Dass das Telekommunikationsmedium, über das die Stichprobe realisiert wurde, ebenfalls Erhebungseinheiten mit bestimmten soziodemografischen Charakteristika besser erreicht, wurde spätestens mit der Diskussion der Befunde der CELLA-Erhebungen verdeutlicht. So stellen sowohl Graeske und Kunz (2009) als auch Schneiderat und Schlinzig (2012) fest, dass sich besonders Mobile-Only-Haushalte in ihren soziodemografischen Eigenschaften von denjenigen der restlichen Erhebungen unterscheiden.

Fraglich ist jedoch, ob das gewählte Telekommunikationsmedium, über das die Stichprobe realisiert wurde, Einfluss auf das Antwortverhalten einer Erhebungseinheit hat. Auch hier zeigen die Ergebnisse der CELLA-Erhebung, dass Erhebungseinheiten, die bei einer Dual-Frame-Erhebung über die Mobilfunkstichprobe gezogen wurden, schwerer zu erreichen sind, bei einem erfolgreichen Kontakt jedoch eine höhere Teilnahmebereitschaft zeigen (vgl. S. Häder et al. (2009b), S. 73). Weitgehend unbeachtet bleibt jedoch die Frage, ob dies ebenfalls für die Domain der Dual-User gilt. Hierbei stellt sich die Frage, ob sich für bestimmte Erhebungseinheiten ein Telekommunikationsmedium finden lässt, über das diese Einheit besser zu erreichen und

daher eher in der Erhebung zu finden ist. Weiterhin gilt zu eruieren, ob eine Präferenz für ein bestimmtes Telekommunikationsmedium auch die Teilnahmebereitschaft erhöht, wenn eine Erhebungseinheit gerade über dieses Medium kontaktiert wird.

Fuchs (2002) bemerkt in diesem Kontext, dass eine Mobilfunkstichprobe gerade für Erhebungseinheiten, die eine höhere Mobilität an den Tag legen, eher geeignet ist, um diese für eine Erhebung zu erreichen. Auch Hoops und Michael (2015) sowie Sand (2015a) zeigen, dass es in der Domain der Dual-User Unterschiede bezüglich der Erreichbarkeit und den Nutzungsgewohnheiten zu geben scheint. Diese Unterschiede können einen Einfluss auf die Möglichkeit, eine bestimmte Zielperson zu erheben, haben sowie deren Teilnahmebereitschaft beeinflussen. Um die Erreichbarkeit von Dual-Usern über ein bestimmtes Telekommunikationsmedium besser beurteilen zu können, sollte dieses Kontinuum entsprechend der Nutzungsgewohnheiten und der Erreichbarkeit über die beiden Medien, in Subdomains unterteilt werden. Sand (2014) empfiehlt diesbezüglich zwischen Mobile-Mainlys, wahren Dual-Usern und Landline-Mainlys zu unterscheiden. Mobile-Mainlys sind diejenigen Erhebungseinheiten, die aufgrund ihrer Nutzungsgewohnheiten und ihrer grundsätzlichen Erreichbarkeit eher über eine Mobilfunkstichprobe gezogen werden können. Bei wahren Dual-Usern besteht keine (eindeutige) Präferenz für ein bestimmtes Telekommunikationsmedium, während Landline-Mainlys eher über die Festnetzstichprobe in die Erhebung gelangen (können).

Eine solche Unterteilung bietet weiterhin die Möglichkeit, einen stärkeren Fokus auf die Erreichbarkeit nach Medium im Rahmen der Designgewichtung zu setzen, um so genauere Schätzergebnisse zu erhalten. Gerade diesem Vorhaben widmet sich der folgende Teil dieser Arbeit. Dazu wird im nächsten Kapitel erstmals das Vorgehen zur Unterteilung nach Erreichbarkeit und Nutzungsgewohnheiten vorgestellt und begründet. Der Kern dieses Kapitels ist jedoch die Einführung eines Gewichtungsansatzes, der gerade diese unterschiedlichen Kategorien im Zuge der Designgewichtung berücksichtigt. Dieses Verfahren wird dann anhand geeigneter Erhebungen sowie Simulationen in Kapitel 6 ausführlich betrachtet. Kapitel 7 zeigt eine weitere Alternative zu dem hier vorgestellten Ansatz auf und untersucht diese genauer. Zuletzt werden die entsprechenden Modelle hinsichtlich ihrer Präzision in Kapitel 8 eingehend untersucht.

5 Einführung eines Composite-Modells zur Berücksichtigung von Nutzungsgewohnheiten der Dual-User

Nach der Beschreibung unterschiedlicher Gewichtungsverfahren für Dual-Frame-Erhebungen wird innerhalb dieses Kapitels ein neuartiger Ansatz zur Gewichtung von dieser Erhebungen vorgestellt, der neben der Auswahlwahrscheinlichkeit unterschiedlicher Stichprobenelemente auch deren Responseverhalten und Erreichbarkeit in Abhängigkeit des zum Kontakt verwendeten Auswahlrahmens berücksichtigt. Im Rahmen dieser Arbeit wird daher erstmals dargestellt und überprüft, wie sich ein solcher Gewichtungsansatz bei einer Dual-Frame-Erhebung in Deutschland gestalten könnte und welche Kenntnisse zu dessen Umsetzung erforderlich sind. Dazu wird zunächst anhand der Erhebung CELLA 2 untersucht, ob sich das Antwortverhalten und die Erreichbarkeit in Abhängigkeit des Auswahlrahmens unterscheidet. Dies dient der weiterführenden Erläuterung, ob das hier entwickelte Modell zur Einteilung der Dual-User anhand von Nutzungsgewohnheiten bei einer gleichzeitigen Berücksichtigung innerhalb eines Gewichtungsansatzes, erfolgversprechend sein könnte.

Um die daraus erlangten Erkenntnisse direkt anzuwenden, wird im Anschluss daran ein konkretes Modell zur Einteilung der Dual-User erarbeitet. Das im Zuge dieser Arbeit entwickelte Modell sieht vor, diese Dual-User entsprechend ihrer Nutzungsgewohnheiten in geeignete Subkategorien (Subdomains) einzuteilen. Ein solches Modell kann dabei sowohl für spätere Analyse zwecke der Zielpopulation herangezogen werden als auch nähere Kenntnisse der Zielpersonen in der Stichprobe vermitteln. Auch wenn Nutzungsgewohnheiten und die Erreichbarkeit von Dual-Usern womöglich eher als Kontinuum zu verstehen sind, wird im Rahmen des hier vorgestellten Modells eine Einteilung anhand von drei Subdomains vorgenommen. Die genaue Ausgestaltung dieser Einteilung wird in Kapitel 5.2 erläutert.

Zuletzt wird im Zuge dieses Kapitels ein Gewichtungsansatz erarbeitet, der die gerade gewonnenen Erkenntnisse im Zuge der Gewichtung von Dual-Frame-Telefonstichproben näher berücksichtigt. Der Gewichtungsansatz zielt demnach darauf ab, durch die Inklusion der Empfangsbereitschaft und Nutzungsgewohnheiten der Dual-User nach den jeweiligen Telekommunikationsmedien, Schätzwerte hervorzubringen, die genauer und/oder präziser sind als diejenigen, die sich unter der Verwendung der in Kapitel 3.3 vorgestellten Ansätze ergeben würden. Weiterhin bildet das hier vorgestellte Konzept zur Gewichtung die Grundlage zur späteren Analyse der Ansätze sowie für Erweiterungen des Gewichtungsverfahrens der folgenden Kapitel.

5.1 Darstellung und Beurteilung des Nutzungsverhaltens von Dual-Usern nach Kommunikationsmedium

Die Notwendigkeit zur Integration der Mobile-Only-Population im Rahmen von bundesweiten CATI-Erhebungen in Deutschland und die damit verbundene Verwendung von Dual-Frame-Ansätzen wurde bereits in Kapitel 2 ausführlich dargelegt. So führt das Nicht-Beachten dieser Population aufgrund von systematischen Unterschieden im Bereich der soziodemografischen Charakteristika häufig zu Verzerrungen in der Punkt- und/oder Varianzschätzung. Ebenfalls deutlich wurde die Notwendigkeit der Verwendung von designgewichteten Schätzwerten aufgrund der unterschiedlichen Inklusionswahrscheinlichkeiten einzelner Einheiten einer solchen Erhebung in Kapitel 3 hervorgehoben.

Eine weitere Erkenntnis der vorherigen Kapitel ist, dass eine Designgewichtung alleine, bspw. wegen Nonresponse, oftmals nicht ausreichend ist, um unverzerrte Schätzergebnisse zu erhalten. Eine Kalibrierung ist daher oftmals vonnöten. Dennoch können Schätzwerte nach der Durchführung einer Kalibrierung noch immer stark von den Populationswerten abweichen. So führen etwa Gabler et al. (2011) Ergebnisse der CELLA 2-Erhebung an, bei denen auch nach der Kalibrierung eine erhebliche Diskrepanz zwischen den kalibrierten Schätzwerten der Erhebung und den Angaben der amtlichen Statistik besteht. Exemplarisch wird dies in Tabelle 14 verdeutlicht.

Tabelle 14 CELLA 2: Familienstand vor und nach Designgewichtung und Kalibrierung (in %)

Familienstand	Ungewichtet	Horvitz-Thompson	GREG	MZ 2009
Verheiratet, zusammenlebend	43,6	46,2	47,3	52,8
Eingetragene Partnerschaft	0,1	0,1	0,1	0,0
Verheiratet, getrenntlebend	2,6	2,2	2,5	2,0
Ledig	38,4	36,9	29,5	29,7
Geschieden	9,9	9,1	10,9	7,5
Verwitwet	4,9	5,0	9,3	8,0
k.A.	0,5	0,5	0,4	0,0
Gesamt	100	100	100	100

Quelle: Gabler et al. (2011), S. 164.

Der Vergleich der Schätzwerte des Familienstandes mit den Daten der amtlichen Statistik zeigt, dass durch die Kalibrierung zwar eine Annäherung an die Angaben der amtlichen Statistik erfolgt, diese jedoch abhängig von der jeweiligen Kategorie unterschiedlich stark ist und noch immer

Abweichungen bestehen.

Die Daten der CELLA 2 - Erhebung wurden an die Hilfsvariablen *Alter*, *Bildungsabschluss* und *Geschlecht* angepasst. Die dafür notwendigen Daten stammen aus dem Mikrozensus 2009. Eine starke Annäherung nach der Kalibrierung erfolgt dabei vor allem bei denjenigen Haushalten, die der Kategorie *Ledig* oder *Verheiratet, zusammenlebend* angehören, während es bei den restlichen Ausprägungen dieser Variable noch immer (starke) Abweichungen zu geben scheint (vgl. Gabler et al. (2011), S. 163ff.).

Demnach ist anzunehmen, dass Nonresponse nicht ausschließlich anhand von soziodemographischen Charakteristika bspw. durch die Korrelation des Untersuchungsmerkmals mit bestimmten Hilfsvariablen erklärt werden kann, sondern auch weitere Quellen Aufschluss zu dessen Entstehen geben können. So ist denkbar, dass diese, wie bereits in Kapitel 3.4 geschildert, neben **soziodemografischen Ursachen** auch durch **Topic Salience** entstehen kann. Weiterhin ist es denkbar, dass bestimmte Eigenschaften des Telekommunikationsmediums das Teilnahmeverhalten einer Zielperson beeinflussen. Diese **Device-Specific Nonresponse** kann durch mehrere Ursachen entstehen. So kann das Antwort- und Teilnahmeverhalten einer Einheit einmal durch bestimmte Eigenschaften beeinflusst werden, die mit einem Telekommunikationsgerät verbunden sind. Weiterhin kann dieses Verhalten durch bestimmte Nutzungsgewohnheiten und die Vertrautheit im Umgang mit einem bestimmten Medium in Verbindung stehen. Die Folge ist eine unscharfe Trennung von Device-Specific Nonresponse und soziodemografischer Ursachen der Antwortverweigerung. Ein technischer Aspekt, der in diesem Kontext das Antwortverhalten beeinflussen kann, ist bspw. die Verfügbarkeit eines Signals. Weiterhin gilt bei Dual-Frame-Telefonstichproben zu beachten, dass Mobiltelefone nicht stationär an einen bestimmten Ort gebunden sind. Der Aufenthaltsort einer Erhebungseinheit der Mobilfunkstichprobe kann ebenfalls Einfluss auf das Antwortverhalten nehmen.

Wie bereits in Kapitel 2.3 geschildert, wurde schon 2002 vermutet, dass vor allem die mobilere Teilpopulation in Telefonerhebungen besser über einen Mobilfunkanschluss erreichbar ist, auch wenn diese grundsätzlich über einen Festnetzanschluss verfügt. Der Zusammenhang zwischen dem Telekommunikationsmedium und soziodemografischen Charakteristika kann ebenfalls anhand der CELLA 2 - Erhebung beobachtet werden. Aus Kapitel 2.3.2 ist bekannt, dass es sich bei Mobile-Onlys vermehrt um jüngere Einpersonenhaushalte handelt, die grundsätzlich aktiver sind als ihr Gegenstück in der Festnetzstichprobe. Trifft dies nun ebenfalls verstärkt für Dual-User zu, die durch die Mobilfunkstichprobe kontaktiert wurden, so lässt sich auch ein anderes Antwortverhalten gerade dieser Gruppe vermuten.

Gerade die Unterschiede im Antwortverhalten, die sich in Abhängigkeit des gewählten Telekommunikationsmediums einstellen, sind näher zu untersuchen, da sie in vielen derzeit angewendeten GewichtungsROUTINEN nicht oder nicht ausreichend berücksichtigt werden. Anhand der Kontaktergebnisse der CELLA 2 - Erhebung, die in den Tabellen 15 und 16 getrennt nach

Stichproben dargestellt sind,⁵⁴ lassen sich deutliche Unterschiede feststellen.

Tabelle 15 **CELLA 2: Kontaktergebnisse der Erhebung (Festnetzstichprobe)**

Kontaktergebnis	Festnetz	Anteil Brutto-SP	Anteil angepasst
Bruttostichprobe	30.360	100 %	••
Ausfälle (neutral)	16.359	53,9 %	••
• Nicht-vergebene Rufnummern	13.004	42,8 %	••
• Andere	3.355	11,1 %	••
Stichprobe (bereinigt)	14.001	46,1 %	100 %
• Vollst. abgeschlossen	1.500	4,9 %	10,1 %
• Teilw. abgeschlossen	41	0,1 %	0,3 %
• Verweigerung/Abbruch	5.221	17,2 %	37,3 %
• Anrufbeantworter/Mailbox (NC)	215	0,7 %	1,5 %
• Andere	1.196	3,9 %	8,5 %
• Unbekannt ob HH	5.828	19,2 %	41,6 %
Haushalt/Zielperson (Kontakt)	6.762	22,3 %	48,3 %
AAPOR Ref 1 / 2 / 3		36,7 / 51 / 63,3	
AAPOR Con 1 / 2 / 3		56,5 / 78,4 / 97,3	

⁵⁴ Die hier dargestellten Kontaktergebnisse wurden in Anlehnung an die Ergebnisse von Schneiderat und Schlinzig (2011) neu berechnet und zusammengefasst.

Tabelle 16 CELLA 2: Kontaktergebnisse der Erhebung (Mobilfunk)

Kontaktergebnis	Festnetz	Anteil Brutto-SP	Anteil angepasst
Bruttostichprobe	46.966	100 %	••
Ausfälle (neutral)	25.422	54,1 %	••
• Nicht-vergebene Rufnummern	21.053	44,8 %	••
• Andere	4.369	9,3 %	••
Stichprobe (bereinigt)	21.544	45,9 %	100 %
• Vollst. abgeschlossen	1.507	3,2 %	7,0 %
• Teilw. abgeschlossen	90	0,2 %	0,4 %
• Verweigerung/ Abbruch	4.735	10,1 %	22,0 %
• Anrufbeantworter/ Mailbox (NC)	1.155	2,5 %	5,5 %
• Andere	2.144	4,6 %	10,0 %
• Unbekannt ob HH	11.913	25,4 %	55,3 %
Haushalt/Zielperson (Kontakt)	6.332	13,5 %	21,4 %
AAPOR Ref 1 / 2 / 3		21,3 / 35,7 / 48,1	
AAPOR Con 1 / 2 / 3		38,7 / 65,2 / 87,8	

Ein Test auf Homogenität der beiden Stichproben liefert einen χ^2 -Wert von 1.582,255. Der entsprechende P-Wert ist $p < 2,2 \cdot 10^{-16}$. Daher kann festgehalten werden, dass die beiden Stichproben nicht der gleichen Verteilung folgen und bezüglich der Anteile nach den jeweiligen Ausprägungen stark divergieren. Einheiten der Festnetz- und Mobilfunkstichprobe unterscheiden sich in ihrem Antwortverhalten demnach grundsätzlich (in dieser Erhebung).

Betrachtet man bspw. das Verhältnis des Anteils der (eindeutigen) Antwortverweigerungen zu dem Anteil der insgesamt kontaktierten Einheiten,⁵⁵ so zeigt sich, dass weniger Kontaktversuche notwendig sind, um eine Einheit aus dem Festnetz zu erheben. Allerdings ist ebenfalls ersichtlich, dass nach einem erfolgreichen Kontaktversuch der Anteil der Einheiten, die die Teilnahme verweigern, in der Mobilfunkstichprobe deutlich geringer ist. Demnach kann vermutet werden, dass Einheiten einer Mobilfunkstichprobe zwar mehr Aufwand benötigen, um erhoben zu werden, bei einem erfolgreichen Kontakt allerdings eine höhere Teilnahmebereitschaft besteht. Zu ähnlichen Ergebnissen führt der Vergleich der Contact- und Refusal-Rate, die nach AAPOR-Standard in den Tabellen 15 und 16 berichten und in allen unterschiedlichen Berechnungsgrundlagen für die

⁵⁵ Dieser wurde (vereinfacht) durch die Summe aller abgeschlossenen, teilweise abgeschlossenen und verweigernten Interviews geteilt durch die Stichprobengröße berechnet.

Festnetzstichprobe höher liegen (vgl. American Association for Public Opinion Research (2011), S. 47 ff.; Barron et al. (2008), S. 4058 ff.).

Der gesteigerte Aufwand, eine Einheit über das Mobilfunknetz zu erheben, kann einerseits mit der geringeren Hitrate einer solchen Stichprobe begründet werden, andererseits lassen sich zusätzliche Faktoren vermuten, die eine Rolle spielen. Diese Erkenntnisse decken sich weitgehend mit Befunden aus früheren Erhebungen, die eine Mobilfunkstichprobe beinhalten. So wird bereits in Kapitel 2.4.3 darauf verwiesen, dass durchschnittlich mehr Kontaktversuche benötigt werden, um eine Erhebungseinheit über eine Mobilfunkstichprobe zu erreichen. Im gleichen Zug ist jedoch zu erwähnen, dass bei einem erfolgreichen Kontakt über das Mobilfunknetz weniger häufig verweigert wird.

Als mögliche Begründung der Unterschiede im Antwortverhalten zwischen Einheiten, die über Festnetz oder Mobilfunk kontaktiert wurden, kann daher eine Abhängigkeit vom jeweiligen Telekommunikationsmedium sowie dessen Erreichbarkeit angeführt werden. Weiterhin lässt sich annehmen, dass die Erreichbarkeit von Einheiten, gerade im Dual-User-Bereich, individuell unterschiedlich zwischen den jeweiligen Telekommunikationsmedien ist. Zumindest in einigen Fällen kann hierbei eine eindeutige Präferenz für ein bestimmtes Telekommunikationsmedium bestehen. Demnach ist Device-Specific Nonresponse in Dual-Frame-Erhebungen nicht auszuschließen und sollte bei solchen Erhebungen entsprechend Berücksichtigung finden. So befinden sich möglicherweise Einheiten in der Erhebung, deren Teilnahme abhängig vom Telekommunikationsmedium ist, über welches diese kontaktiert wurden. Um Verzerrungen aufgrund dieser Fehlerquelle zu vermeiden oder zu verringern, sollte ein Gewichtungsansatz auch diese Fehlerquellen stärker berücksichtigen. Ein solcher Ansatz wird in den folgenden beiden Abschnitten genauer erläutert. Hierbei führt Abschnitt 5.2 ein grundsätzliches Schema zur Einteilung nach präferierten Telekommunikationsmedien anhand der Nutzungsgewohnheiten und der Erreichbarkeit ein. Abschnitt 5.3 beschreibt einen Gewichtungsansatz, der gerade eine solche Device-Specific Nonresponse stärker berücksichtigt.

5.2 Einteilung der Dual-User nach Nutzungsgewohnheiten

Die zuvor geschilderten Ergebnisse lassen vermuten, dass im Rahmen von Dual-Frame-Erhebungen neben Only-Populationen und Dual-Usern noch weitere Unterscheidungen getroffen werden sollten. Demnach scheint die Population der Dual-User nicht homogen bezüglich der Präferenz eines bestimmten Telekommunikationsmediums zu sein. Danach gibt es scheinbar Dual-User, die zwar theoretisch über beide Auswahlrahmen in die Stichprobe gelangen können, aber ein bestimmtes Telekommunikationsmedium (und damit verbunden die dazugehörige Auswahlgrundlage) präferieren. Diese Einheiten sollten mit weniger Aufwand gerade über das präferierte Medium zu erheben sein. Eine genauere Unterteilung der Population der Dual-User

nach deren Nutzungsgewohnheiten erscheint daher sinnvoll. Eine solche Unterteilung sollte anschließend auch in einer entsprechenden Designgewichtung Berücksichtigung finden. Erste Untersuchungen zur genaueren Unterscheidung von Dual-Usern nach Nutzungsgewohnheiten einzelner Telekommunikationsmedien wurden bspw. durch Brick et al. (2011) durchgeführt.

Die in Kapitel 3 dargestellten Ansätze berücksichtigen gerade diese Unterschiede nicht. So unterstellt der *Single-Frame-Ansatz*, dass es für Einheiten aus der Population der Dual-User irrelevant ist, über welche Auswahlgrundlage diese in die Stichprobe gelangt sind, solange für beide Stichproben die Inklusionswahrscheinlichkeit bekannt ist. Implizit wird in diesem Ansatz daher angenommen, dass der Auswahlrahmen, über den eine Zielperson in die Erhebung gelangt ist, austauschbar ist und Zielpersonen über beide Stichproben entsprechend ihrer Inklusionswahrscheinlichkeit, gleich gut erreichbar sind (vgl. S. L. Lohr und J. N. K. Rao (2000), S. 272.; Clark et al. (2007), S. 3237.).

Folgend wird nicht mehr bloß zwischen Onlys und Dual-User unterschieden, sondern eine genauere Untergliederung der Population im Überlappungsbereich der beiden Auswahlgrundlagen vorgenommen. Wurde zuvor die Zielpopulation U in einer Dual-Frame-Telefon-erhebung durch die beiden Populationen U^F und U^C erreicht, ergaben sich drei Domains; die der Festnetz-Onlys ($U^F \setminus C$), die der Dual-User ($U^{F \cap C}$) sowie die der Mobile-Only ($U^C \setminus F$). Da diese Unterteilung jedoch nicht die Nutzungsgewohnheiten der Dual-User berücksichtigt, wird die Population U in l ($l \in \{1, 2, 3, 4, 5\}$) Teilpopulationen (u_l) untergliedert. Speziell lässt sich die Zielpopulation daher in die nachstehenden Teilpopulationen einteilen:

- **Festnetz-Only (u_1):** Diejenigen Einheiten, die ausschließlich über die Auswahlgrundlage Festnetz in die Erhebung gelangen können.
- **Festnetz-Mainly (u_2):** Diejenigen Einheiten, die zwar (theoretisch) über beide Auswahlrahmen in die Erhebung gelangen können, jedoch ihren Festnetzanschluss *dominant* verwenden.
- **Dual-User (u_3):** Einheiten, die über beide Auswahlrahmen in die Erhebung gelangen können und keines der Telekommunikationsmedien dominant verwenden.
- **Mobile-Mainly (u_4):** Diejenigen Einheiten, die zwar (theoretisch) über beide Auswahlrahmen in die Erhebung gelangen können, jedoch ihren Mobilfunkanschluss *dominant* verwenden.
- **Mobile-Only (u_5):** Diejenigen Einheiten, die ausschließlich über die Auswahlgrundlage Mobilfunk in die Erhebung gelangen können.

In dieser Untergliederung meint *dominant*, dass eine Einheit ein bestimmtes Telekommunikationsmedium aufgrund der Nutzungsgewohnheiten präferiert und daher regelmäßiger/öfter ver-

wendet. Zur Beurteilung dieser Nutzungsgewohnheiten gelten folglich die Erreichbarkeit einer Einheit über ein bestimmtes Telekommunikationsmedium sowie deren Empfangsbereitschaft. Weiterhin besteht die Möglichkeit, eine Präferenz für ein bestimmtes Medium direkt bei der Zielperson zu erfragen. Es wird angenommen, dass eine Erhebungseinheit, die über dasjenige Telekommunikationsmedium kontaktiert wird, das diese dominant verwendet, mit einem geringeren Aufwand erhoben werden kann (bspw. durch weniger Anwahlversuche) und/oder eine höhere Teilnahmebereitschaft zeigt. Die Einteilung in die jeweiligen Teilpopulationen kann hierbei nach der in Abbildung 11 dargestellten Art und Weise erfolgen.

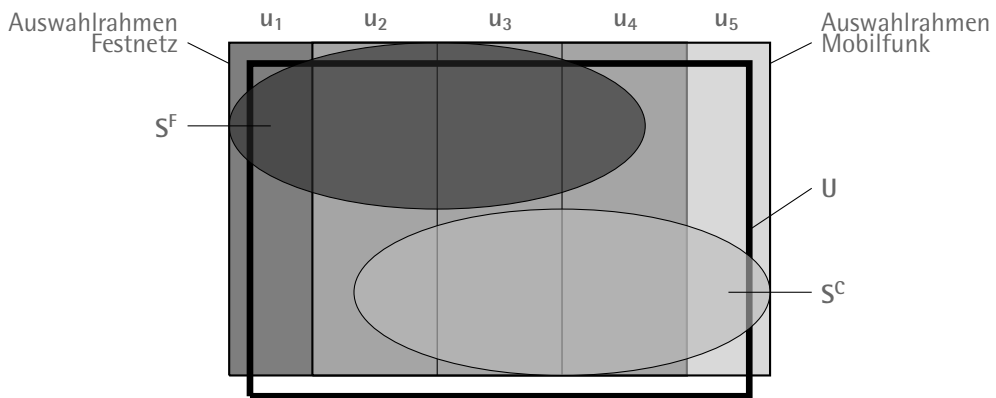


Abbildung 11 Neuzuordnung der Dual-Frame-Erhebung

Diese wurde der zuvor erläuterten Abbildung 1 aus Abschnitt 2.3.1 nachempfunden und entsprechend der soeben geschilderten Einteilung erweitert. U gibt somit die Zielpopulation einer CATI-Erhebung wieder, die durch einen Dual-Frame-Ansatz erhoben werden soll. Die beiden disjunkten Stichproben S^F und S^C enthalten nun die dem Telekommunikationsmedium entsprechende Only-Population sowie Einheiten beider Mainly-Populationen (u_2 und u_4) und der Population der wahren Dual-User (u_3). Ebenfalls ersichtlich wird, dass im Rahmen einer solchen Erhebung das Potenzial besteht, Einheiten zu erheben, die nicht Teil der Zielpopulation sind (*Overcoverage*) sowie dass Einheiten der Zielpopulation nicht durch eine der beiden Stichproben erfasst werden können (*Undercoverage*; diejenigen ohne Telekommunikationsmedium). Der Unterschied zu der Darstellung in Abbildung 1 besteht demnach darin, dass bei einer Dual-Frame-Erhebung nicht mehr lediglich nach Onlys und Dual-Usern unterschieden wird, sondern der Bereich der Dual-User deutlich genauer nach Nutzungsgewohnheiten unterteilt wird. Dennoch bleibt die Annahme, dass Einheiten beider Mainly-Population über beide Stichproben erhoben werden können.

5.3 Das Domain-Specific-Composite-Design-Weighting-Modell

Mit der exakteren Einteilung nach Nutzungsgewohnheiten wird zwar eine genauere Untersuchung des Antwortverhaltens der Einheiten und bestimmter Untersuchungsmerkmale, die von Interesse sind, ermöglicht, jedoch sollte ein solcher Ansatz ebenfalls dazu in der Lage sein, Verzerrungen, bspw. aufgrund von *Device-Specific Nonresponse*, zu verringern. Daher wird folgend ein Gewichtungsansatz vorgestellt, der, in Anlehnung an die Arbeit von Brick et al. (2011), die Nutzungsgewohnheiten von Dual-Usern stärker in den Fokus setzt. Dazu wird anstelle des *Single-Frame-Ansatzes*, der in bundesdeutschen Erhebungen oftmals verwendet wird, ein (erweiterter) Composite-Ansatz verwendet. Die Gewichtung erfolgt dabei in Abhängigkeit der jeweiligen Teilpopulation. Der Schätzer $\hat{Y}_{DSC}$ setzt sich aus der Summe der Schätzwerte der einzelnen Teilbereiche nach

$$\hat{Y}_{DSC} = \sum_{l \in L} \hat{Y}_l$$

zusammen. L entspricht der Menge der Subdomains. In dem vorliegenden Fall gilt $L = \{1, \dots, 5\}$. Im Gegensatz zu den vorherigen Ansätzen, berechnen sich jedoch die Schätzer in Abhängigkeit der Subdomain unterschiedlich. Die Idee ist, für beide Mainly-Kategorien eine Konvexkombination durchzuführen, bei der der Composite-Faktor die Device-Specific Nonresponse berücksichtigt. Die Domain-Totalwerte berechnen sich daher nach

$$\begin{aligned}\hat{Y}_1 &= \sum_{i \in S^F \cap u_1} \frac{1}{\pi^F} y_i, \\ \hat{Y}_2 &= \sum_{i \in S^F \cap u_2} \lambda \frac{1}{\pi^F} y_i + \sum_{i \in S^C \cap u_2} (1 - \lambda) \frac{1}{\pi^C} y_i, \\ \hat{Y}_3 &= \sum_{i \in S^F \cap u_3} \phi \frac{1}{\pi^F} y_i + \sum_{i \in S^C \cap u_3} (1 - \phi) \frac{1}{\pi^C} y_i, \\ \hat{Y}_4 &= \sum_{i \in S^F \cap u_4} \mu \frac{1}{\pi^F} y_i + \sum_{i \in S^C \cap u_4} (1 - \mu) \frac{1}{\pi^C} y_i, \\ \hat{Y}_5 &= \sum_{i \in S^C \cap u_5} \frac{1}{\pi^C} y_i.\end{aligned}$$

In der Schreibweise von Gleichung 3.9 ist

$$\begin{aligned}\hat{Y}_{DSC} &= \hat{Y}_{S^F \cap u_1} + \lambda \hat{Y}_{S^F \cap u_2} + \phi \hat{Y}_{S^F \cap u_3} + \mu \hat{Y}_{S^F \cap u_4} \\ &\quad + (1 - \lambda) \hat{Y}_{S^C \cap u_2} + (1 - \phi) \hat{Y}_{S^C \cap u_3} + (1 - \mu) \hat{Y}_{S^C \cap u_4} + \hat{Y}_{S^C \cap u_5}\end{aligned}\tag{5.1}$$

bzw.

$$\hat{Y}_{DSC} = \sum_{i \in S^F} c_{i,DSC}^F \frac{1}{\pi_F} y_i + \sum_{i \in S^C} c_{i,DSC}^C \frac{1}{\pi_C} y_i, \quad (5.2)$$

mit

$$c_{i,DSC}^F = \begin{cases} 1 & \forall i \in S^F \cap u_1 \\ \lambda & \forall i \in S^F \cap u_2 \\ \phi & \forall i \in S^F \cap u_3 \\ \mu & \forall i \in S^F \cap u_4 \end{cases} \quad (5.3)$$

und

$$c_{i,DSC}^C = \begin{cases} (1 - \lambda) & \forall i \in S^C \cap u_2 \\ (1 - \phi) & \forall i \in S^C \cap u_3 \\ (1 - \mu) & \forall i \in S^C \cap u_4 \\ 1 & \forall i \in S^C \cap u_5. \end{cases} \quad (5.4)$$

Bei diesem Ansatz ändert sich das Designgewicht für die jeweiligen Only-Populationen (u_1 und u_5) nicht und gleicht sowohl dem "einfachen" Composite- als auch dem Single-Frame-Ansatz. Bei den Only-Populationen entspricht die Inklusionswahrscheinlichkeit für die jeweilig andere Stichprobe ohnehin Null. Mit der Domain der wahren Dual-User wird ebenfalls in einer mit dem Composite-Ansatz vergleichbaren Weise umgegangen. Allerdings ist die Domain der wahren Dual-User in dem vorliegenden Ansatz kleiner als die Domain, die alle Einheiten beinhaltet, die sowohl über Festnetz als auch über Mobilfunk erreichbar sind. Ebenfalls möglich ist, den Totalwert dieser Domain unter Verwendung des Single-Frame-Ansatzes zu schätzen. Da die Annahme, dass wahre Dual-User mit dem gleichen Aufwand bzw. entsprechend ihrer Inklusionswahrscheinlichkeit der jeweiligen Stichproben, in die Erhebung gelangen, kann ϕ daher bspw. proportional zur Stichprobengröße gewählt werden. Dies gilt, da angenommen wird, dass wahre Dual-User keine Präferenz für ein bestimmtes Telekommunikationsmedium besitzen oder (durchgehend) leichter über eines der beiden Medien zu erreichen sind.

In dem vorliegenden Ansatz sollten λ und μ für den Bereich der Landline- und Mobile-Mainly das unterschiedliche Teilnahmeverhalten in Abhängigkeit des zum Kontakt verwendeten Telekommunikationsmedium widerspiegeln. ϕ hat die gleiche Bedeutung für den Bereich der wahren Dual-User. Weiterhin gilt, dass $\lambda, \mu, \phi \in [0,1]$ und $\lambda + (1 - \lambda) = 1$, $\mu + (1 - \mu) = 1$, $\phi + (1 - \phi) = 1$. Der einfachste Fall wäre dabei, die beiden Composite-Faktoren im Mainly-Bereich (und für die wahren Dual-User) so zu wählen, dass $\lambda = \mu = \phi$ gilt. Dabei wäre implizit die Annahme getroffen, dass es keine Unterschiede bezüglich der Erreichbarkeit

oder der Nutzungsgewohnheiten innerhalb der Gruppe der Dual-User gibt und $\hat{Y}_{DSC}$ würde dem "einfachen" Composite-Ansatz entsprechen. Ein vereinfachender Fall entstünde, wenn $\lambda = \mu = \phi = 0,5$ angenommen wird. Hierbei kann der daraus generierte Schätzwert mit demjenigen des Single-Frame-Ansatz vergleichbar sein. Demnach sollten die beiden Composite-Faktoren so gewählt werden, dass das unterschiedliche Teilnahmeverhalten in Abhängigkeit der Stichprobe, über die eine Mainly-Einheit kontaktiert wurde, entsprechend reflektiert wird. Daher sollte für λ und μ gelten, dass $\lambda > 0,5$ und $\mu < 0,5$ ist. Da unter den wahren Dual-Usern keine eindeutige Präferenz für ein Telekommunikationsmedium festzustellen sein sollte, ist anzunehmen, dass ϕ (ungefähr) dem Verhältnis der abgeschlossenen Festnetz- zu den abgeschlossenen Mobilfunkinterviews entspricht.⁵⁶

Bei einem "gewöhnlichen" Overlap-Ansatz, der ausschließlich Onlys und Mainlys unterscheidet (ohne wahre Dual-User zu berücksichtigen), empfehlen Brick et al. (2011) den (einzigen) Composite-Faktor (v) so zu wählen, dass dieser die Verzerrung aufgrund des Nonresponse minimiert. Die Verzerrung innerhalb des Dual-User-Bereiches⁵⁷ für den Populationsmittelwert $\bar{Y}_{F \cap C}$ in dem von den Autoren untersuchten Modell wird dabei durch

$$Bias(\bar{Y}_{F \cap C}) = \iota N_{F \cap C} (\bar{Y}_{u_2} - \bar{Y}_{u_4}) \cdot \left(v \frac{r_{SF \cap u_2}}{r_{SF}} + (1 - v) \frac{r_{SC \cap u_4}}{r_{SC}} - 1 \right)$$

berechnet. Hierbei wird der Bereich der wahren Dual-User nicht beachtet. Einheiten aus dem Überlappungsbereich sind demnach wahlweise Landline- oder Mobile-Mainly. ι entspricht hier dem Anteil der Dual-User an der Gesamtstichprobe, $N_{F \cap C}$ der Anzahl der Dual-User in der Population, während r_{SF} und r_{SC} die Responserate der jeweiligen Stichprobe wiedergibt. $r_{SF \cap u_2}$ und $r_{SC \cap u_4}$ entsprechen somit der Responserate der Stichprobe in der entsprechenden Mainly-Teilpopulation. Den Autoren zufolge ist die Verzerrung innerhalb des Überlappungsbereiches folglich Null, wenn $\bar{Y}_{u_2} = \bar{Y}_{u_4}$ oder $v r_{SF \cap u_2} / r_{SF} + (1 - v) r_{SC \cap u_4} / r_{SC} = 1$ gilt. Die erste der beiden Annahmen stammt aus dem Single-Frame-Ansatz, während letztere die Verzerrung in Abhängigkeit der relativen Responserate und des Composite-Faktors betrachtet (vgl. Brick et al. (2011), S. 5f.). Zur Bestimmung eines geeigneten Composite-Faktors empfehlen die Autoren, diesen durch die Berücksichtigung der jeweiligen Responseraten zu berechnen.

In Analogie zu der von Brick et al. (2011) vorgeschlagenen Vorgehensweise könnte die Bestimmung der beiden Composite-Faktoren in dem Domain-Specific-Composite-Weighting-Ansatz unter der Einbeziehung der relativen Responseraten vorgenommen werden. Diesbezüglich gilt jedoch zu beachten, dass das von den Autoren vorgeschlagene Modell lediglich einen Composite-Faktor über den gesamten Überlappungsbereich verwendet und keine Konvex-Kombination innerhalb einzelner Teilpopulationen der Dual-User vorsieht. Daher können die

⁵⁶ Eine schwächere Annahme wäre $\lambda \geq 0,5$ und $\mu \leq 0,5$.

⁵⁷ Die Autoren gehen davon aus, dass für die Only-Populationen unverzerrt geschätzt werden kann.

beiden Composite-Faktoren (theoretisch) nach

$$\lambda_0 = \frac{r_{SF}(r_{SC} - r_{SC \cap u_2})}{r_{SC} \cdot r_{SF \cap u_2} - r_{SF} \cdot r_{SC \cap u_2}},$$

$$\phi_0 = \frac{r_{SF}(r_{SC} - r_{SC \cap u_3})}{r_{SC} \cdot r_{SF \cap u_3} - r_{SF} \cdot r_{SC \cap u_3}}$$

und

$$\mu_0 = \frac{r_{SF}(r_{SC} - r_{SC \cap u_4})}{r_{SC} \cdot r_{SF \cap u_4} - r_{SF} \cdot r_{SC \cap u_4}}$$

bestimmt werden. $r_{SC \cap u_2}$, $r_{SF \cap u_2}$, $r_{SC \cap u_3}$, $r_{SF \cap u_3}$, $r_{SC \cap u_4}$ und $r_{SF \cap u_4}$ geben die Responseraten der jeweiligen Einheiten der Subpopulationen in Abhängigkeit der verwendeten Stichprobe wieder. Bei der praktischen Umsetzung scheitert dieses Vorgehen jedoch daran, dass sich die Responseraten nach Subpopulation nicht eindeutig bestimmen lassen. Das liegt darin begründet, dass nicht-erfolgreiche Kontaktversuche (sowie eine direkte Antwortverweigerung) bei Telefonstichproben keinen Hinweis auf die Zugehörigkeit einer Einheit zu einer bestimmten Teilpopulation erlauben. Eine Responserate in Dual-Frame-Erhebungen kann demzufolge nur für die beiden einzelnen Stichproben, jedoch nicht ohne Weiteres für Subpopulationen der Zielpopulation berechnet werden (vgl. Sand (2015a), S. 147f.).

Eine Möglichkeit wäre, die Kontaktergebnisse proportional zu der Zusammensetzung nach Verwendungskategorie zuzuordnen. Eine solche Vorgehensweise führt jedoch zu den selben Responseraten, die auch für die gesamte Einsatzstichprobe gelten. Auch eine Aufteilung der Non-Contacts, die sich proportional zu der Summe der Designgewichte nach Verwendungskategorie verhält, liefert wenig plausible Ergebnisse. Der Grund dafür liegt vermutlich darin, dass eine Einteilung der Non-Contacts in eine bestimmte Verwendungskategorie, die nur anhand der Anschlussart erfolgt, ohne dass Informationen über die Verteilung der Nutzungsgewohnheiten in der Population existieren, nicht möglich ist. Dies begründet sich in der Gestaltung des Erhebungsprozesses. So besteht die Stichprobe i. d. R. (initial) lediglich aus Festnetz- und Mobilfunknummern. Ohne eine Kenntnis über die Verteilung der jeweiligen Subpopulationen Landline-Mainly, wahre Dual-User und Mobile-Mainly in der Zielpopulation kann bei einem erfolglosen Kontaktversuch⁵⁸ keine Einordnung dieser Rufnummern in die entsprechende Kategorie erfolgen. Vielmehr ist dies nur dann möglich wenn ein Kontakt zu der Zielperson erfolgt und die Zugehörigkeit zu einer der Subdomains im Rahmen der Befragung ermittelt wird. Da zur Berechnung der Responserate in einer bestimmten Subdomain jedoch auch die Kenntnis über die Anzahl der Non-Contacts in dieser erforderlich ist, scheitert ein solches Vorgehen.

Ein weniger komplexes Verfahren zur Bestimmung der Composite-Faktoren ist die Berechnung anhand der Stichprobengröße der jeweiligen Domain. Im Idealfall sollten dafür die Brutto-

⁵⁸ Darunter zählt sowohl ein Non-Contact als auch eine Verweigerung der Teilnahme.

stichprobengrößen verwendet werden. Da sich diese jedoch aufgrund der zuvor geschilderten Probleme nicht einer bestimmten Teilpopulation zuordnen lassen, verwendet ein einfacher Ansatz die realisierten Nettofallzahlen nach Teilpopulation. Die Composite-Faktoren können dann anhand von

$$\lambda = \frac{n_{S^F \cap u_2}}{n_{S^F \cap u_2} + n_{S^C \cap u_2}}, \quad (5.5)$$

$$\phi = \frac{n_{S^F \cap u_3}}{n_{S^F \cap u_3} + n_{S^C \cap u_3}} \quad (5.6)$$

und

$$\mu = \frac{n_{S^F \cap u_4}}{n_{S^F \cap u_4} + n_{S^C \cap u_4}} \quad (5.7)$$

bestimmt werden. Hierbei entsprechen $n_{S^F \cap u_2}$, $n_{S^C \cap u_2}$, $n_{S^F \cap u_3}$, $n_{S^C \cap u_3}$, $n_{S^F \cap u_4}$ und $n_{S^C \cap u_4}$ der Anzahl der realisierten Interviews nach Stichprobe und Verwendungskategorie.

Um eine Beurteilung des vorgestellten Modells zu ermöglichen, wird es im folgenden Abschnitt auf die Daten der beiden Erhebungen CELLA 2 und *Influenza 2014* angewendet und entsprechende Punktschätzer ermittelt. Kapitel 7 erweitert es dann um alternative Anwendungen der Einteilung nach Nutzungskategorie, während Kapitel 8 die Varianzschätzung des Modells vorstellt.

6 Untersuchung des Composite-Modells anhand der Erhebungen CELLA 2 und Influenza 2014

Nach der Beschreibung des Composite-Modells zur Berücksichtigung von Nutzungsgewohnheiten innerhalb der Gewichtung von Dual-Frame-Erhebungen wird das Modell nun anhand von ausgewählten Studien im folgenden Kapitel überprüft. Hierfür werden die beiden Erhebungen CELLA 2 und Influenza 2014 verwendet, die bereits in Kapitel 2 erläutert wurden. Ziel des Kapitels ist es, Erkenntnisse darüber zu erlangen, ob und unter welchen Bedingungen ein solches Modell zu einer erhöhten Genauigkeit der Schätzergebnisse beitragen kann. Dazu erfolgt erst eine Einteilung der Ergebnisse der jeweiligen Studien in die entsprechenden Verwendungskategorien. Im Anschluss werden die Schätzergebnisse des Composite-Modells mit den ungewichteten Stichprobenergebnissen, den Schätzergebnissen unter Verwendung des Gewichtungmodells nach Gabler et al. (2011) sowie mit Daten der amtlichen Statistik verglichen. Darüber hinaus werden sowohl Paradata als auch ausgewählte Items der beiden Studien untersucht, um weiterführende Aufschlüsse über die Erreichbarkeit von Stichprobenelementen in Abhängigkeit ihrer Nutzungsgewohnheiten zu erlangen. Diese Auseinandersetzung erfolgt separat nach Erhebung. Zuletzt werden noch Items zur Einordnung in die jeweiligen Nutzungskategorien untersucht. Zu Beginn wird das Modell jedoch anhand des synthetischen Datensatzes aus Kapitel 3.5.1 überprüft.

6.1 Untersuchung anhand des synthetischen Datensatzes

Zur Beurteilung des in Kapitel 5.3 vorgestellten Ansatzes musste der synthetische Datensatz dahin gehend erweitert werden, dass eine Einteilung der Population anhand von Nutzungsgewohnheiten ermöglicht wird. Dazu wurde die bereits beschriebene künstlich erzeugte Population mit einem Umfang von insgesamt 1.000.000 Einheiten um die beiden Variablen *Nutzungskategorie* und *Präferenz* erweitert. Nutzungskategorie stellt eine Zuteilung der einzelnen Einheiten in eine der fünf zuvor beschriebenen, disjunkten Domains dar, während die Variable Präferenz die Wahrscheinlichkeit wiedergibt, dass ein bestimmtes Telekommunikationsmedium verwendet wird. Die Einteilung nach Nutzungskategorien ergibt folgende Verteilung der Population:

Landline-Only	Landline-Mainly	True Dual-User	Mobile-Mainly	Mobile-Only
135.456	114.956	496.741	152.956	99.891
(13,5 %)	(11,5 %)	(49,7 %)	(15,3 %)	(10 %)

Die Variable Präferenz gibt, in Abhängigkeit der Nutzungskategorie, die Wahrscheinlichkeit wieder, ob im Rahmen einer Stichprobe eine Einheit über das Medium Festnetz erreicht wird. Dafür wurde angenommen, dass dies bei Landline-Onlys mit einer Wahrscheinlichkeit von 1, bei Landline-Mainlys mit einer Wahrscheinlichkeit von 0,75, bei wahren Dual-Usern mit 0,5, bei Mobile-Mainlys mit 0,25 und für Mobile-Onlys mit einer Wahrscheinlichkeit von 0 geschieht. Entsprechend berechnet sich dies für die Mobilfunkstichprobe durch 1 minus die jeweilige Ausprägung der Präferenz. Anhand dieser Variable wurden die Ziehungswahrscheinlichkeiten neu berechnet. Dies erfolgte durch die Multiplikation der Variable Präferenz mit der (originären) Ziehungswahrscheinlichkeit im Festnetz bzw. 1 minus der Präferenz mit der (originären) Ziehungswahrscheinlichkeit im Mobilfunk. Anschließend wurde unter der Verwendung dieser (neuen) Wahrscheinlichkeiten eine Festnetzstichprobe im Umfang von 6.000 Einheiten sowie eine Mobilfunkstichprobe im Umfang von 4.000 Einheiten gezogen. Die 10.000 Stichprobeneinheiten lassen sich wie folgt in die jeweiligen Nutzungskategorien aufteilen:⁵⁹

Tabelle 17 Zusammensetzung der Stichprobe nach Nutzungskategorie und Medium

Medium/Kat.	Landline-Only	Landline-Mainly	True Dual-User	Mobile-Mainly	Mobile-Only
Festnetz	1.576 (100 %)	1.009 (81,2 %)	2.939 (58,0 %)	476 (33,4 %)	0 (0 %)
Mobilfunk	0 (0 %)	233 (18,8 %)	2.124 (42,0 %)	951 (66,6 %)	692 (100 %)
Gesamt	1.576	1.242	5.063	1.427	692

Die Anteile, die in Klammern angegeben sind, entsprechen den erhobenen Einheiten nach Medium innerhalb der jeweiligen Verwendungskategorie. Weiterhin kann der Anteil der Festnetzstichprobe als Konvexfaktor der späteren Gewichtung verstanden werden.

Anhand der zugrunde liegenden Stichprobe wurde dann nochmals die Untersuchungsvariable, die bereits in Kapitel 3.5.1 beschrieben wurde, unter Verwendung des Single-Frame-Ansatzes, des Composite-Ansatzes sowie des Domain-Specific-Composite-Ansatzes (DSC-Ansatz) geschätzt. Die Ergebnisse können in Tabelle 18 abgelesen werden.

⁵⁹ Es handelt sich hierbei um die ungewichtete, deskriptive Darstellung der Stichprobenzusammensetzung.

Tabelle 18 Vergleich der unterschiedlichen Schätzverfahren I (Anteilswerte)

Variable 1	Stichprobe	SF-Ansatz	CW-Ansatz	DSC-Ansatz	Population
1	0,256	0,260	0,254	0,259	0,25
2	0,148	0,150	0,150	0,150	0,15
3	0,497	0,492	0,499	0,493	0,5
4	0,099	0,098	0,096	0,098	0,1
Designeffekt	●●	1,43	1,43	1,52	●●
n_{eff}	●●	7.000	7.000	6.570	●●

Bei der Anwendung der unterschiedlichen Schätzverfahren auf die Variable 1 (siehe auch Kapitel 3.5.1) des synthetischen Datensatzes zeigt sich, dass in der vorliegenden Stichprobe der (reine) Composite-Ansatz den beiden anderen Schätzverfahren zu bevorzugen ist. Jedoch muss festgehalten werden, dass sowohl die Werte der deskriptiven Untersuchung, als auch die drei Schätzwerte nahe beieinander liegen. Begründet liegt dies darin, dass alle drei Verfahren erwartungstreue Schätzer sind und in der Simulation keinerlei Frame-Imperfections oder Non-response zugelassen werden. Die Erwartungstreue der jeweiligen Schätzfunktionen ergibt sich aus den bereits in den Kapiteln 3.3.1, 3.3.2 und 5.3 angeführten Gründen. So basieren alle drei der unterschiedlichen Ansätze grundsätzlich auf der Anwendung des Ansatzes von Horvitz und Thompson. Lediglich die Art und Weise der Bestimmung der Auswahlwahrscheinlichkeit sowie die Kombination der einzelnen Designgewichte unterscheidet sich in den drei verschiedenen Gewichtungsverfahren. So wird im Zuge des Single-Frame-Ansatzes die gemeinsame Inklusionswahrscheinlichkeit einer Zielperson über beide Auswahlrahmen verwendet, während der Composite-Ansatz die Designgewichte auf der Basis der Inklusionswahrscheinlichkeit einer Zielperson lediglich über den Erhebungsrahmen berechnet, über den diese erhoben wurde und im Anschluss eine Konvexkombination (siehe Kapitel 3.3.2) aller Dual-User vornimmt. Letzteres geschieht, um den Bereich der Dual-User nicht zu überschätzen. Da der DSC-Ansatz grundsätzlich eine Erweiterung des Composite-Ansatzes darstellt, bei dem der Bereich der Dual-User in drei disjunkte Schichten unterteilt wird, für die dann eigenständige Konvexfaktoren berechnet werden, ist die erwartungstreue dieses Schätzers ebenfalls grundsätzlich gegeben. Der Horvitz-Thompson-Schätzer, der hier angewendet wird, ist daher immer dann erwartungstreu, wenn im Zuge der Schätzung keinerlei weitere verzerrende Ursachen wie bspw. Nonresponse und Frame-Imperfections, eintreten. Da innerhalb dieser Simulation als vereinfachte Grundannahme kein Nonresponse auftritt und alle Elemente über mindestens einen der beiden Auswahlrahmen erreicht werden können, kann davon ausgegangen werden, dass die berechneten Designgewichte nach allen drei Vorgehensweisen unverzerrte Schätzwerte für die Gesamtpopulation liefern.

Hierbei gilt zu beachten, dass diese (idealen) Voraussetzungen i. d. R. bei realen Erhebungen nicht vorzufinden sind und zumindest mit Nonresponse gerechnet werden muss. Um eine Einschätzung der Anwendbarkeit und der Präzision der jeweiligen Verfahren zu erhalten, wurden diese Aspekte jedoch ausgelassen, da sonst die Entscheidung über bestimmte Eigenschaften der drei Schätzer eine abkömmliche Komplexität erhalten würde.

Bezüglich der einzelnen Punktschätzer liefert der Domain-Specific-Composite-Ansatz ebenfalls Ergebnisse, die näher an der Verteilung der Population sind, als der Single-Frame-Ansatz, jedoch mit einem höheren, modellbasierten Designeffekt, der auf eine größere Variation der Gewichte und eine höhere Varianz bzw. eine geringere Präzision des Schätzers schließen lässt. Weiterhin gilt zu bemerken, dass die Ausprägungen der Variablen 1 einer Einheit unabhängig von deren Zugehörigkeit zu einer der jeweiligen Subdomains sind.

Ein besonderes Interesse liegt bei der Untersuchung des Domain-Specific-Composite-Ansatzes jedoch bei dessen Genauigkeit bezogen auf Variablen, deren Ausprägungen abhängig von Nutzungsgewohnheiten bzw. der Verwendung eines bestimmten Telekommunikationsmediums sind. Wie bereits erläutert, wurde in mehreren Untersuchungen gezeigt, dass Mobile-Only-Haushalte sich bezüglich bestimmter soziodemografischer Charakteristika von Dual- Usern und Landline-Only-Haushalten unterscheiden (vgl. Meier (1999), S. 134f.; Graeske und Kunz (2009), S. 59f.; Link et al. (2007), S. 827f.). Auch Untersuchungen bezüglich derjenigen Haushalte, die vornehmlich ihren Mobilfunkanschluss nutzen, zeigen Unterschiede bezüglich diverser Charakteristika (vgl. Fuchs (2002), S. 122f.; Hoops und Michael (2015), S. 2f.). Um dies näher zu untersuchen, wurde die Population um eine Variable erweitert, deren Ausprägungen direkt mit der Nutzungskategorie zusammenhängen. Gewählt wurde hierfür eine Variable, deren Ausprägung entweder 25, 50 oder 75 ist. Eine solche Variable könnte bspw. die Ausprägung der Altersvariable widerspiegeln. Bezogen auf die gesamte Population ist der Anteil der einzelnen Ausprägungen in diesem Beispiel jeweils 1/3. Innerhalb der einzelnen Nutzungskategorien unterscheidet sich dieser Anteil jedoch. Die Zusammensetzung der Population in Abhängigkeit der Verwendungskategorie kann anhand von Tabelle 19 nachvollzogen werden.

Tabelle 19 Variable 2 nach Nutzungskategorie (Anteilswerte)

Verwendungskategorie/ Variable 2	25	50	75
Landline-Only	0	0,333	0,667
Landline-Mainly	0,111	0,334	0,555
True Dual-User	0,333	0,333	0,334
Mobile-Mainly	0,554	0,336	0,110
Mobile-Only	0,665	0,335	0

Auch für diese Variable wurden unter der Verwendung der unterschiedlichen Verfahren geeignete Schätzwerte anhand der Stichprobenergebnisse ermittelt. Diese können anhand von Tabelle 20 nachvollzogen werden.

Tabelle 20 Vergleich der unterschiedlichen Schätzverfahren II (Anteilswerte)

Variable 2	Stichprobe	SF-Ansatz	CW-Ansatz	DSC-Ansatz	Population
25	0,317	0,294	0,292	0,288	0,330
50	0,330	0,329	0,331	0,331	0,334
75	0,353	0,377	0,376	0,381	0,337
Designeffekt	●●	1,43	1,43	1,52	●●
n_{eff}	●●	7.000	7.000	6.570	●●

Verwunderlich an den in Tabelle 20 dargestellten Ergebnissen ist, dass scheinbar die Verwendung der reinen Stichprobenergebnisse die Verteilung der Variable 2 in der Population am ehesten widerspiegelt, während der Domain-Specific-Composite-Ansatz gerade diejenigen Ergebnisse liefert, die am wenigsten mit den Populationswerten übereinstimmen. Unter den drei Gewichtungungsverfahren ist der reine Composite-Ansatz derjenige, der noch am ehesten mit der Verteilung der Grundgesamtheit übereinstimmt. Jedoch liegen auch bei dieser Variablen die Werte nahe beieinander.

Es zeigt sich somit auch an dem zweiten Beispiel, dass der in Kapitel 5 vorgestellte Ansatz in der Anwendung auf einen synthetischen Datensatz nicht die präzisesten Ergebnisse liefert. Aufgrund dessen kann nicht ausgeschlossen werden, dass die Berücksichtigung der Präferenzen anhand des hier verwendeten Ansatzes unpassend ist. Bei beiden Variablen wäre an diesem Beispiel der (reine) Composite-Ansatz zu bevorzugen. Jedoch berücksichtigt die Untersuchung an diesem Datensatz keinerlei Frame-Imperfections oder Nonresponse. Weiterhin wird bei diesem Datensatz ein sehr einfacher Zusammenhang zwischen Nutzungsgewohnheiten und der Variable 2 zugrunde gelegt. Der Zusammenhang zwischen Nutzungsgewohnheiten und den Ausprägungen bestimmter Variablen kann in einem Datensatz aus einer realen Erhebung deutlich komplexer sein. Daher wird folgend der Domain-Specific-Composite-Ansatz auf die beiden Erhebungen CELLA 2 und Influenza angewendet.

6.2 Anwendung auf die Daten der CELLA 2-Erhebung

Die 3.007 abgeschlossenen Interviews der CELLA 2-Erhebung, die 2010 erhoben wurden, teilen sich in 1.500 Festnetz- und 1.507 Mobilfunkinterviews auf. Von den 3.007 befragten Personen

entstammen dabei 178 der Landline- und 170 der Mobile-Only-Population.⁶⁰ Demnach sind 2.659 Befragte Dual-User. Die Befragung erfolgte unter der Verwendung eines Overlap-Ansatzes. Somit lässt sich grundsätzlich das in Kapitel 5 dargestellte Gewichtungsverfahren auf den Datensatz anwenden. Daher wird folgend zuerst die Einteilung in die jeweiligen Nutzungskategorien vorgestellt und anschließend die Anwendung des Modells beschrieben sowie die entsprechenden Ergebnisse präsentiert.

6.2.1 CELLA 2: Einteilung in Verwendungskategorien

Da zum Zeitpunkt der CELLA 2-Erhebung noch keine Einteilung der Dual-User nach deren Nutzungsgewohnheiten vorgesehen war, wurden auch keine Variablen erhoben, die dies direkt ermöglichen. Um dennoch die Nutzungsgewohnheiten beurteilen zu können, wurden für die folgende Untersuchung geeignete Proxys gesucht, die die Erreichbarkeit und die Empfangsbereitschaft des Mobilfunkgerätes wiedergeben. Hierzu wurden folgende Variablen verwendet:⁶¹

- **v8a:** *Haben Sie Ihr Handy immer an, d. h. empfangsbereit?
Oder schalten Sie es manchmal aus?*

Immer an	1.661
Manchmal aus	1.168
tnz	178

- **v8b:** *Wie oft haben Sie Ihr Handy empfangsbereit?*

Täglich	824
Mehrmals in der Woche	197
Mehrmals im Monat	68
Seltener	79
tnz	1.839

- **v8c:** *Wir möchten es gern noch etwas genauer wissen.
Wie viele Stunden am Tag haben Sie Ihr Handy durchschnittlich empfangsbereit?*

⁶⁰ Die Angaben unterscheiden sich von denjenigen, die durch Schneiderat und Schlinzig (2011) gemacht wurden. Begründet liegt dies im Umgang mit denjenigen Personen, die angaben, über keinen Festnetzanschluss zu verfügen, aber dennoch eine *Homezone*-Rufnummer (o. ä.), also eine Festnetzrufnummer, die über das Mobiltelefon erreichbar ist, besitzen. Diese zählen in der hier dargestellten Betrachtung nicht zu den Mobile-Onlys.

⁶¹ Für die Variablen wurde nach dem in Abschnitt 3.5.1 bereits erwähnten Vorgehen eine Imputation durchgeführt.

1 – 4 Std.	135
5 – 8 Std.	152
9 – 12 Std.	222
13 – 16 Std.	181
17 – 20 Std.	95
21 – 24 Std.	236
tnz	1.986

- **v9a:** *Wenn Sie unterwegs sind, nehmen Sie das Handy dann für gewöhnlich mit?
Tun Sie das immer, meistens, gelegentlich, selten oder nie?*

Immer	1.949
Meistens	615
Gelegentlich	150
Selten	88
Nie	27
tnz	178

Die ersten drei Fragen zielen darauf ab, wie oft ein Mobiltelefon grundsätzlich empfangsbereit ist und ob demnach die Möglichkeit besteht, dass eine Zielperson (oder gar ein Haushalt) überhaupt über eine Mobilfunkstichprobe zu einem bestimmten Zeitpunkt kontaktiert werden kann. Dabei liegt die Vermutung nahe, dass, je seltener ein Mobilfunkanschluss tatsächlich empfangsbereit ist, es umso schwieriger ist, einen Dual-User über eine Mobilfunkstichprobe auszuwählen. Die Variable *v9a* kann wiederum Aufschlüsse über die grundsätzliche Erreichbarkeit einer Zielperson über einen Mobilfunkanschluss sowie evtl. auch über deren Vertrautheit in der Nutzung eines Mobiltelefons geben. Ähnliche Befunde beschreibt bereits Fuchs (2002) bei der Untersuchung einer Mobilfunkstichprobe. So schildert der Autor, dass durchschnittlich mehr Anwahlversuche benötigt wurden, um Zielperson zu kontaktieren, die während der Befragung angaben, ihr Mobilfunkgerät nicht immer mit sich zu führen und/oder nicht immer eingeschaltet zu haben (vgl. Fuchs (2002), S. 131f.).

Weiterhin wurden zur Einteilung der Verwendungskategorien die beiden Variablen *t7* und *t2* verwendet. *t7* gibt die Anzahl der Festnetzanschlüsse einer Einheit wieder. *t2* gibt an, ob eine Einheit über eine virtuelle Festnetzrufnummer erreichbar ist. Beide Items gemeinsam dienen zur Eruierung, ob eine Zielperson ausschließlich über eine virtuelle Festnetzrufnummer in die Festnetzstichprobe gelangen kann. Diese Personen werden dann, da eine solche Rufnummer nicht an einen physischen Anschluss gebunden ist, den Mobile-Mainly zugeordnet. Wie jedoch bereits in Kapitel 3.3.1 dargelegt wurde, ist die Verwendung dieser Angabe nicht unproblematisch. Wenn davon ausgegangen wird, dass Zielpersonen bereits bei der exakten Nennung der Anzahl der

Festnetzanschlüsse Probleme haben, so ist ebenfalls zu vermuten, dass diese Abfrage nicht genau denjenigen Anteil der Einheiten mit virtueller Festnetznummer erfasst. Denkbar ist daher, dass eine virtuelle Festnetznummer schlichtweg "vergessen" wird, da sie sich nicht in der Nutzung befindet oder, dass es zu Verwechslungen kommt. So ist es bspw. denkbar, dass ein Anschluss, der an ein schnurloses Festnetztelefon gebunden ist, bei dieser Frage genannt wird. Da jedoch in der Abfrage, ob eine virtuelle Festnetznummer vorhanden ist, explizit Beispiele für solche Angebote genannt werden, trifft der letztere Fall vermutlich weniger häufig zu. Demnach kann der Anteil derjenigen, die über eine virtuelle Festnetznummer verfügen, als Untergrenze angenommen und eher davon ausgegangen werden, dass der tatsächliche Anteil untererfasst wird.

Zur Einteilung nach diesen Variablen weiterhin von Interesse ist deren Relation zueinander. Aufgrund dessen wurde anschließend anhand von entsprechenden χ^2 -Statistiken auf die paarweise Unabhängigkeit der jeweiligen Variablen geprüft. Hierfür wurde lediglich die Teilstichprobe der Dual-User verwendet. Weiterhin geben die folgenden Statistiken lediglich deskriptive Ergebnisse wieder.⁶² Für eine generalisierbare Aussage über das Verhältnis der jeweiligen Variablen müssten die Statistiken auf der Basis von gewichteten Ergebnissen erfolgen. Da jedoch hier die genauere Zuteilung der Zielpersonen innerhalb der Stichprobe Untersuchungsgegenstand ist, ist die Verwendung der ungewichteten Fallzahlen angemessen. Tabelle 21 berichtet die zugehörigen p-Werte der einzelnen Teststatistiken. Die gesamte Statistik kann im Anhang A.1.1 nachvollzogen werden.

Tabelle 21 CELLA 2: Paarweise Unabhängigkeit der Items zur Einordnung von Nutzungsgewohnheiten

Variable/ p-Wert	v8a	v9a	t7	t2
v8a	•	$2,2 \cdot e^{-16}$	0,037	0,002
v9a	$2,2 \cdot e^{-16}$	•	0,142 [◇]	0,001 [◇]
t7	0,037	0,142 [◇]	•	0,001 [◇]
t2	0,002	0,001 [◇]	0,001 [◇]	•

Da es sich bei den Variablen v8b und v8c um Items, denen eine bestimmte Filterführung zugrunde liegt, handelt, wurden diese aus der Betrachtung ausgeschlossen, da sie zumindest für Variable v8a per Definition nicht unabhängig sein können. Die mit [◇] gekennzeichneten Werte wurden anhand einer Monte-Carlo-Simulation mit 2.000 Iterationen berechnet, da entweder aufgrund von Stichprobennullen oder aufgrund der Beschaffenheit der jeweiligen Kontingenztabellen eine direkte Berechnung auf der Basis der Stichprobenergebnisse vermeintlich zu fehlerhaften Schätzwerten geführt hätte. Dies begründet sich damit, dass die entsprechende Teststatistik der

⁶² Die in diesem Kapitel vorliegenden χ^2 -Tests beruhen auf der Häufigkeitsverteilung als absolute Werte.

Annahme unterliegt, dass die Verteilung der Fallzahlen der Variablen approximativ normalverteilt sind. Wenn ein Großteil der erwarteten Fallzahlen vergleichsweise gering ist, kann dies zu einer mangelhaften Approximation führen. Daher wird in diesen Fällen ein iteratives Resampling angewendet, um die Annahmen der Teststatistik nicht zu verletzen.

Die Auswertung der paarweisen χ^2 -Statistiken zeigt, dass die Nullhypothese der Unabhängigkeit der betrachteten Variablen in einem Großteil der Fälle auf einem höchst signifikanten Niveau abgelehnt werden kann. Einzig bei dem Test auf Unabhängigkeit der beiden Variablen t7 und v9a trifft diese Aussage nicht zu. Hier kann nicht davon ausgegangen werden, dass diese beiden Variablen nicht voneinander unabhängig sind. Betrachtet man, was diese beiden Variablen messen⁶³ scheint dies jedoch plausibel. So wäre es fraglich festzustellen, dass die Häufigkeit das ein Mobiltelefon mitgeführt wird, einen Einfluss auf die Anzahl der Festnetzzurufnummern hat. Insgesamt zeigt sich, dass die Unabhängigkeitstests, die die beiden Variablen beinhalten, die die Anzahl bestimmter Rufnummern messen, auch wenn diese signifikant zur Ablehnung der Nullhypothese führen, einen höheren p-Wert aufweisen, als dies bspw. beim Vergleich der beiden Variablen v8a und v9a zu beobachten ist. Dennoch zeigt sich, dass auch die Anzahl der Festnetznummern und vor allem auch die Zahl der virtuellen Festnetznummern nicht unabhängig von der täglichen Empfangsbereitschaft des Mobiltelefons einer Zielperson sind. Nicht wenig verwunderlich ist auch, dass basierend auf den vorliegenden Tests die Empfangsbereitschaft eines Mobilfunkgerätes auch auf die Erreichbarkeit (anhand von v9a gemessen) Einfluss nehmen kann. Grundsätzlich lässt sich daher nicht ausschließen, dass diese Variablen einen besseren Einblick in die Nutzungsgewohnheiten von Dual-Usern geben können. Daher wird auf der Basis dieser Variablen eine Einteilung nach Subdomains vorgenommen, wohl wissentlich, dass es sich hierbei lediglich um Proxi-Variablen handelt, die lediglich Hinweise auf die tatsächliche Präferenz für ein bestimmtes Telekommunikationsmedium, die im Zuge dieser Erhebung jedoch nicht gemessen wurde, geben können.

Da die Einteilung in Verwendungskategorien nach diesen Variablen bis zum heutigen Zeitpunkt wenig erforscht ist, wurden im Anschluss mehrere Modelle gebildet und dann zur Schätzung nach dem im vorherigen Kapitel vorgestellten Verfahren verwendet. Bei diesem Vorgehen wurde der Bereich der wahren Dual-User ($S^F \cap u_3 \cup S^C \cap u_3$) schrittweise verkleinert. Insgesamt konnten anhand der hier präsentierten Variablen sieben unterschiedliche Modelle gebildet werden.⁶⁴ Von diesen sieben führen vier Modelle zu (eindeutigen) Unterschieden in der Aufteilung nach Verwendungskategorie. Diese werden nachfolgend kurz erläutert.

63 Variable t7 misst die Anzahl der Festnetzzurufnummern. Variable v9a gibt an, wie häufig eine Zielperson ihr Mobiltelefon mitführt.

64 Kombinatorisch sind (mindestens) 1.920 unterschiedliche Modelle unter dem Auslassen fehlender Werte möglich, jedoch ist eine solche Einteilung weder sinnvoll, noch führt jede der Kombinationen zu voll besetzten Zellen. Deshalb wurde sich auf lediglich sieben Modelle beschränkt.

- Modell1:** *Landline-Mainlys:* Diejenigen, deren Mobiltelefon nicht täglich eingeschaltet ist und höchstens gelegentlich mitgeführt wird.
- Mobile-Mainlys:* Diejenigen, deren Mobiltelefon immer eingeschaltet ist und immer mitgeführt wird. Weiterhin sind diese Befragten lediglich über eine virtuelle Festnetznummer bei der Festnetzstichprobe kontaktierbar.
- Modell2:** *Landline-Mainlys:* Diejenigen, deren Mobiltelefon nicht täglich eingeschaltet ist.
- Mobile-Mainlys:* Diejenigen, deren Mobiltelefon immer eingeschaltet ist und immer mitgeführt wird. Des Weiteren verfügen die Einheiten über maximal zwei Festnetzrufnummern, von denen eine virtuell ist.
- Modell3:** *Landline-Mainlys:* Diejenigen, deren Mobiltelefon nicht täglich oder täglich maximal vier Stunden eingeschaltet ist.
- Mobile-Mainlys:* Diejenigen, deren Mobiltelefon immer eingeschaltet ist und immer mitgeführt wird. Darüber hinaus verfügen die Einheiten maximal über einen (echten/physischen) Festnetzanschluss, bzw. sind nur über eine Festnetzrufnummer erreichbar.
- Modell4:** *Landline-Mainlys:* Diejenigen, deren Mobiltelefon nicht täglich oder täglich maximal zwölf Stunden eingeschaltet ist.
- Mobile-Mainlys:* Diejenigen Dual-User, die nicht in die Kategorie der Landline-Mainlys fallen.

Durch die vier Modelle ergeben sich dann die in Tabelle 22 dargestellten Verteilungen nach Verwendungskategorie im Rahmen der CELLA 2-Erhebung.

Tabelle 22 CELLA 2: Dual-User nach Verwendungskategorie

Modell/Teilpopulation	u_2	u_3	u_4
Modell 1	114	2.447	98
Anteil Gesamt-SP	3,8 %	81,1 %	3,3 %
Modell 2	335	2.169	155
Anteil Gesamt-SP	11,1 %	72,1 %	5,2 %
Modell 3	386	1.504	769
Anteil Gesamt-SP	12,8 %	50,0 %	25,6 %
Modell 4	658	0	2.001
Anteil Gesamt-SP	21,9 %	0 %	66,5 %

Die hier dargestellte Modellbildung kann dahin gehend verstanden werden, dass Modell 1 den Anteil der wahren Dual-User unter den Vorgaben der Proxys maximiert, während Modell 4 gerade diese Teilpopulation negiert. Weiterhin wurde die Obergrenze der täglichen Empfangsbereitschaft des Mobiltelefons der Landline-Mainly frei gewählt, da hierfür keine Erfahrungswerte vorhanden sind. Anhand der Einteilung dieser Items ist lediglich die Abschätzung der Größe der jeweiligen Verwendungskategorie möglich. So ist es bspw. durchaus denkbar, dass auch ein Landline-Mainly über ein Mobiltelefon verfügt, das durchgehend empfangsbereit ist. Auf der anderen Seite scheint es wenig plausibel, dass ein Mobile-Mainly (im Alltag) über einen Mobilfunkanschluss verfügt, der nicht mindestens täglich empfangsbereit ist. Dennoch stellt die hier gewählte Einteilung im besten Fall eine Annäherung an die tatsächliche Verteilung in der Population dar.

Eine genauere Betrachtung der Zusammensetzung der Erhebung nach Teilpopulation und der Stichprobe, die zur Kontaktaufnahme der jeweiligen Einheiten verwendet wurde, wird daher anhand von Tabelle 23 gegeben.

Tabelle 23 CELLA 2: Zusammensetzung innerhalb der Teilpopulation nach Modell und Stichprobe

Teilpopulation	Stichprobe	Mod 1	Mod 2	Mod 3	Mod 4
u_1	S^F	178 (100 %)	178 (100 %)	178 (100 %)	178 (100 %)
	S^C	••	••	••	••
	$S^F \cup S^C$	178 (100 %)	178 (100 %)	178 (100 %)	178 (100 %)
u_2	S^F	89 (78,1 %)	255 (76,1 %)	287 (74,4 %)	412 (62,6 %)
	S^C	25 (21,9 %)	80 (23,9 %)	99 (25,6 %)	246 (37,4 %)
	$S^F \cup S^C$	114 (100 %)	335 (100 %)	386 (100 %)	658 (100 %)
u_3	S^F	1.198 (49,0 %)	1.005 (46,3 %)	686 (45,6 %)	••
	S^C	1.249 (51,0 %)	1.164 (53,7 %)	818 (54,4 %)	••
	$S^F \cup S^C$	2.447 (100 %)	2.169 (100 %)	1.504 (100 %)	••
u_4	S^F	35 (35,7 %)	62 (40,0 %)	349 (45,4 %)	910 (45,5 %)
	S^C	63 (64,3 %)	93 (60,0 %)	420 (54,6 %)	1.091 (54,5 %)
	$S^F \cup S^C$	98 (100 %)	155 (100 %)	769 (100 %)	2.001 (100 %)
u_5	S^F	••	••	••	••
	S^C	170 (100 %)	170 (100 %)	170 (100 %)	170 (100 %)
	$S^F \cup S^C$	170 (100 %)	170 (100 %)	170 (100 %)	170 (100 %)

Die Zusammensetzung nach Art der Stichprobe in Abhängigkeit des Modells erfüllt dabei die Annahme, dass innerhalb einer Mainly-Kategorie mehr Interviews über das von Einheiten dieser Teilpopulation dominant genutzte Telekommunikationsmedium abgeschlossen werden. Auch für die Teilpopulation u_3 wird die Vermutung, dass deren Einheiten aufgrund der fehlenden Präferenz eines Telekommunikationsmediums mit einem (annähernd) gleichen Aufwand über beide Stichproben befragt werden können, zumindest für die beiden ersten Modelle bestätigt. Der Anteil der hier geführten Interviews nach Stichprobe ist für u_3 in Modell 1 und 2 annähernd gleich und unterhalb derjenigen, die mit den dominanten Telekommunikationsmedien der entsprechenden Mainly-Kategorie geführt wurden. Somit kann in einem nächsten Schritt die Berechnung der Schätzwerte nach Gleichung 5.2 durchgeführt werden.

Ausgehend von dieser Modellbildung ist weiterhin von großem Interesse, wie sich die Einheiten der jeweiligen Subdomains in ihren soziodemografischen Charakteristika unterscheiden. Deren Verteilung wurde daher einzeln untersucht. Die Tabellen 24,25 und 26 präsentieren dafür die Verteilungen nach Geschlecht, Alter und Bildungsabschluss der Zielpersonen innerhalb dieser Subdomains, um potenzielle Unterschiede in deren Komposition nach Nutzungsgewohnheiten

sowie der entsprechenden Einteilung in den verschiedenen Modellen darzustellen. Die Ergebnisse der Untersuchung der Haushaltsgröße sowie des Familienstandes sind weiterhin im Anhang in Kapitel A.2.1 zu finden.

Innerhalb der gesamten Erhebung waren 51,8 % aller Zielpersonen männlich und 48,2 % weiblich. Über die vier Modelle hinweg zeigt sich, dass diese Verteilung am ehesten mit derjenigen der wahren Dual-User übereinstimmt. Von Interesse ist hierbei, dass der Anteil der männlichen Zielpersonen anzusteigen scheint, je dominanter das Mobilfunkgerät verwendet wird. So sind Personen aus der Domain der Landline-Only überwiegend weiblich, während Mobile-Only vorwiegend männlich sind.

Tabelle 24 CELLA 2: Geschlecht nach Nutzungsgewohnheit und Modell (in %)

	Geschlecht	Modell 1	Modell 2	Modell 3	Modell 4
u_1	Männlich	41,6	41,6	41,6	41,6
	Weiblich	58,4	58,4	58,4	58,4
u_2	Männlich	49,1	43,9	45,3	47,0
	Weiblich	50,9	56,1	54,7	53,0
u_3	Männlich	52,4	53,5	54,1	0,0
	Weiblich	47,6	46,5	45,9	0,0
u_4	Männlich	45,9	49,0	51,4	53,7
	Weiblich	54,1	51,0	48,6	46,3
u_5	Männlich	61,2	61,2	61,2	61,2
	Weiblich	38,8	38,8	38,8	38,8

Bezüglich der Stichprobenzusammensetzung der Landline-Mainly lässt sich feststellen, dass der Anteil männlicher Zielpersonen steigt. Weiterhin interessant scheint die tendenzielle Abnahme männlicher Teilnehmer bei den Modellen 2 und 3. Der erneute Anstieg im Zuge von Modell vier lässt sich voraussichtlich dadurch begründen, dass dieses Modell die Gruppe der wahren Dual-User negiert und somit wiederum mehr männliche Zielpersonen dieser Domain zugeordnet werden. Da es sich hierbei um ein dichotomes Merkmal handelt, ist es auch wenig verwunderlich, dass der Anteil männlicher Zielpersonen innerhalb der Mobile-Mainlys in den Modellen 2 und 3 ansteigt (und oberhalb desjenigen der Landline-Mainly liegt). Dies trifft insbesondere deshalb zu, da sich dieser Anteil für die wahren Dual-User über die Modelle hinweg nur geringfügig verändert. Innerhalb des ersten Modells liegt weiterhin der Anteil der männlichen Zielpersonen der Mobile-Mainly deutlich unter demjenigen der Landline-Mainly sowie dem gesamten Anteil männlicher Teilnehmer an der Erhebung. Weiterhin lässt sich zumindest für die ersten drei Modelle feststellen, dass der Anteil männlicher Teilnehmer (teilweise) deutlich

unterhalb desjenigen der gesamten Erhebung ist. Dabei liegt jedoch die Vermutung nahe, dass der Gesamtanteil männlicher Zielpersonen in der Erhebung durch den vergleichsweise hohen Anteil der männlichen Mobile-Only beeinflusst wird.

In der gesamten Erhebung lagen 16 % im Alter von 16 und 25 Jahren, 26,8 % zwischen 26 und 39 Jahren, 21,7 % zwischen 40 und 49 Jahren, 17,4 % zwischen 50 und 59 Jahren, 10,2 % zwischen 60 und 69 Jahren und 7,9 % waren mindestens 70 Jahre alt. Wenig verwunderlich bei der Betrachtung der Altersverteilung nach Subdomain ist die Beobachtung, dass Landline-Onlys durchschnittlich deutlich älter zu sein scheinen, als der Stichprobendurchschnitt und Mobile-Only entsprechend jünger sind. Somit fällt auch der Anteilswert innerhalb der jüngeren Alterskategorien erheblich niedriger für Landline-Onlys aus, als dies in der gesamten Stichprobe zu beobachten ist, während der Anteil der Personen in höheren Alterskategorien erheblich oberhalb liegt. Genau die umgekehrte Beobachtung lässt sich für die Mobile-Onlys feststellen.

Tabelle 25 CELLA 2: Alter nach Nutzungsgewohnheit und Modell (in %)

	Alter	Modell 1	Modell 2	Modell 3	Modell 4
u_1	16 - 25	1,7	1,7	1,7	1,7
	26 - 39	11,8	11,8	11,8	11,8
	40 - 49	18,0	18,0	18,0	18,0
	50 - 59	18,0	18,0	18,0	18,0
	60 - 69	19,1	19,1	19,1	19,1
	70 +	31,5	31,5	31,5	31,5
u_2	16 - 25	4,4	5,1	4,9	8,1
	26 - 39	9,6	14,3	13,5	20,4
	40 - 49	24,6	20,6	21,2	20,8
	50 - 59	19,3	20,9	21,2	20,4
	60 - 69	20,2	19,4	19,9	16,7
	70 +	21,9	19,7	19,2	13,7
u_3	16 - 25	16,3	17,3	16,3	0,0
	26 - 39	27,7	28,4	28,2	0,0
	40 - 49	22,5	23,1	24,6	0,0
	50 - 59	17,9	17,7	17,2	0,0
	60 - 69	9,6	8,8	8,7	0,0
	70 +	5,9	4,7	5,0	0,0
u_4	16 - 25	22,4	22,6	21,2	18,7
	26 - 39	39,8	41,9	32,9	29,7
	40 - 49	17,3	16,1	18,6	22,9
	50 - 59	7,1	9,0	16,5	16,7
	60 - 69	8,2	6,5	7,4	7,7
	70 +	5,1	3,9	3,4	4,2
u_5	16 - 25	30,0	30,0	30,0	30,0
	26 - 39	32,4	32,4	32,4	32,4
	40 - 49	15,3	15,3	15,3	15,3
	50 - 59	12,9	12,9	12,9	12,9
	60 - 69	4,7	4,7	4,7	4,7
	70 +	4,7	4,7	4,7	4,7

Zielpersonen der Subdomain der wahren Dual-User entsprechen noch am ehesten der Altersverteilung über die gesamte Stichprobe innerhalb der ersten drei Modelle. Jedoch lässt sich auch hier

ein höherer Anteil gerade jüngerer Zielpersonen feststellen, während der Anteil von Zielpersonen in höheren Alterskategorien geringer ausfällt. Insgesamt kann anhand dieser Altersverteilungen beobachtet werden, dass mit zunehmend dominanter Verwendung des Mobilfunkgerätes über die Subdomains, der Anteil gerade jüngerer Zielpersonen steigt. Gerade das Umgekehrte trifft auf die Nutzung des Festnetzgerätes zu. So liegt der Anteil der jüngeren Zielpersonen innerhalb der Mobile-Mainlys oberhalb des gesamten Anteilswertes der Erhebung und der Anteil älterer Zielpersonen ist, ähnlich der Landline-Onlys, für Landline-Mainlys höher. Dies lässt sich zumindest für die ersten drei Modelle eindeutig feststellen. Da die Subdomain der wahren Dual-User innerhalb von Modell 4 negiert wird, ist eine solche Aussage nur eingeschränkt möglich. Das durchschnittliche Alter der Landline-Onlys innerhalb der Stichprobe liegt dabei bei 58,8 Jahren. Das der Mobile-Only entspricht 37 Jahren. In allen vier Modellen lässt sich feststellen, dass das Durchschnittsalter von Landline-Mainly zu Mobile-Mainly sinkt. So liegt dieses für Landline-Mainlys zwischen 55,5 Jahren in Modell 1 und 49,4 Jahren in Modell 4. Für Mobile-Mainlys beträgt das durchschnittliche Alter in der Stichprobe zwischen 38,1 Jahren in Modell 1 und 40,7 in Modell 4. Ähnliches trifft ebenso für die wahren Dual-User zu, deren Durchschnittsalter bei 42,5 Jahren in Modell 1 und bei 41,9 in Modell 3 liegt.

Im Rahmen der CELLA 2-Erhebung konnten insgesamt 3,6 % keinen Schulabschluss nachweisen, 18,3 % hatten einen Hauptschulabschluss (oder Ähnliches), 26,9 % die mittlere Reife (oder Ähnliches), 6 % einen Abschluss der polytechnischen Oberschule, 43,7 % (Fach-) Hochschulreife und 1,7 % einen anderen Schulabschluss vorzuweisen. Wie bereits in Kapitel 2.3.2 hervorgehoben, lassen sich unter den Zielpersonen, die lediglich über einen Mobilfunkanschluss verfügen, häufiger einen niedrigen oder mittleren Bildungsabschluss aufweisen. Bei der Betrachtung der Tabelle 26 kann Ähnliches für Zielpersonen aus der Domain Landline-Only festgestellt werden. So zeigt sich auch hier ein deutlich höherer Anteil gerade von niedrig gebildeten Zielpersonen, als dies in der Gesamtstichprobe zu finden ist.

Tabelle 26 CELLA 2: Bildungsabschluss nach Nutzungsgewohnheit und Modell (in %)

	Schulabschluss	Modell 1	Modell 2	Modell 3	Modell 4
u_1	Ohne	5,1	5,1	5,1	5,1
	Hauptschulabschluss	34,8	34,8	34,8	34,8
	Mittlere Reife	19,7	19,7	19,7	19,7
	Polytechnischen Oberschule	6,2	6,2	6,2	6,2
	FH/HS - Reife	32,0	32,0	32,0	32,0
	Anderer Schulabschluss	2,2	2,2	2,2	2,2
u_2	Ohne	0,9	2,7	2,3	2,6
	Hauptschulabschluss	21,1	19,7	20,2	18,5
	Mittlere Reife	24,6	25,7	25,4	24,5
	Polytechnischen Oberschule	5,3	4,8	4,9	4,0
	FH/HS - Reife	46,5	45,4	45,1	48,9
	Anderer Schulabschluss	1,8	1,8	2,1	1,5
u_3	Ohne	3,1	3,0	3,2	0,0
	Hauptschulabschluss	16,7	16,6	16,0	0,0
	Mittlere Reife	27,1	26,9	25,1	0,0
	Polytechnischen Oberschule	5,7	5,8	4,9	0,0
	FH/HS - Reife	45,8	46,0	48,9	0,0
	Anderer Schulabschluss	1,6	1,7	1,9	0,0
u_4	Ohne	5,1	4,5	3,1	3,2
	Hauptschulabschluss	10,2	9,7	16,0	16,0
	Mittlere Reife	29,6	32,9	31,9	28,0
	Polytechnischen Oberschule	9,2	7,7	8,1	6,4
	FH/HS - Reife	42,9	43,2	39,8	44,7
	Anderer Schulabschluss	3,1	1,9	1,2	1,7
u_5	Ohne	8,2	8,2	8,2	8,2
	Hauptschulabschluss	27,6	27,6	27,6	27,6
	Mittlere Reife	31,2	31,2	31,2	31,2
	Polytechnischen Oberschule	8,2	8,2	8,2	8,2
	FH/HS - Reife	24,1	24,1	24,1	24,1
	Anderer Schulabschluss	0,6	0,6	0,6	0,6

Auch innerhalb der Subdomain Mobile-Mainly lässt sich über alle vier Modelle hinweg ein überdurchschnittlicher Anteil von Zielpersonen mit mittlerer Reife finden. Da der Anteil der

Zielpersonen mit (Fach-) Hochschulreife annähernd vergleichbar mit demjenigen der gesamten Stichprobe ist, lässt sich dies (zumindest für die ersten beiden Modelle) anhand des geringeren Anteils an Zielpersonen mit niedrigerem formalen Bildungsabschluss begründen. Weiterhin fällt auf, dass der Anteil von Zielpersonen ohne formalen Bildungsabschluss ebenfalls oberhalb des Stichprobendurchschnitts liegt. Im Gegensatz hierzu sind die Beobachtungen aus der Subdomain der Landline-Mainly abweichend. Hier ist sowohl der Anteil der Personen mit (Fach-) Hochschulreife als auch derjenige der Zielpersonen mit Hauptschulabschluss höher als in der Gesamtstichprobe. Dementsprechend fallen die Anteile derjenigen mit mittlerer Reife sowie ohne Schulabschluss geringer aus. Zielpersonen, die in diesen Modellen den wahren Dual-Usern zugeordnet wurden, zeigen hingegen zumindest bezüglich der niedrigeren formalen Bildungsabschlüsse eine annähernd vergleichbare Verteilung auf, wie diese in der gesamten Stichprobe zu finden ist. Jedoch liegt auch hier der Anteil formal höher Gebildeter oberhalb des Stichprobenanteilswertes.

Insgesamt lässt sich demnach keine eindeutige Aussage darüber treffen, welche Tendenz Zielpersonen aus den Subdomains des Überlappungsbereiches bezüglich ihrer formalen Bildungsabschlüsse aufweisen. Im Gegensatz zu den beiden Only-Kategorien, bei denen nachweislich weniger Personen mit formal höherem Bildungsabschluss und mehr Personen mit einem niedrigeren Bildungsabschluss zu finden waren als in der Gesamtstichprobe, ist eine Aussage über die Zielpersonen der Mainly-Subdomains erschwert. So finden sich unter den Landline-Mainlys zwar vermehrt Personen mit Hauptschulabschluss als in der Gesamtstichprobe, aber auch mehr Personen mit (Fach-) Hochschulreife. Ähnliches gilt auch für diejenigen der Mobile-Mainly-Kategorie. Hier sind zwar vermehrt Personen mit mittlerem Bildungsabschluss zu finden, was tendenziell mit den Anteilen der Mobile-Onlys vergleichbar ist, jedoch ist der Anteil der Zielpersonen mit höherem Bildungsabschluss mit demjenigen der Gesamtstichprobe vergleichbar. Somit lässt sich nicht eindeutig sagen, dass Mainly-Personen in ihren Bildungsabschlüssen tendenziell mit den respektiven Onlys vergleichbar sind, während die Verteilung der wahren Dual-User eher mit derjenigen der Gesamtstichprobe übereinstimmt.

Die Verteilung der Haushaltsgröße nach Subdomain, wie sie im Anhang innerhalb von Tabelle 50 dargestellt wird, zeigt nicht bloß, dass der Anteil der Einpersonenhaushalte unter den Mobile-Only-Befragten deutlich höher als in der Gesamterhebung ist, sondern auch, dass sich dies ebenso, wenn auch nicht in einem solch "gravierendem" Maß, für Landline-Only-Befragte festhalten lässt. Allerdings finden sich unter den Mobile-Onlys in jeglicher Art von Mehrpersonenhaushalten weniger Zielpersonen, als dies für die gesamte Erhebung zu beobachten ist, während zumindest der Anteil der Zweipersonenhaushalte unter den Landline-Onlys (leicht) oberhalb des Wertes der Gesamterhebung ist.

Unter den wahren Dual-Usern entspricht die Verteilung der Haushaltsgröße in den ersten drei Modellen noch am ehesten der Verteilung, die sich auch in der gesamten Erhebung zeigt, wobei

sich tendenziell ein leicht erhöhter Anteil von Mehrpersonenhaushalten einstellt. Indes ähnelt sowohl die Verteilung der Landline- als auch der Mobile-Mainly derjenigen der entsprechenden Only-Haushalte. So zeigt sich in beiden ein höherer Anteil von Einpersonenhaushalten, der jedoch nicht so ausgeprägt wie derjenige der Only-Haushalte ist. Im Fall der Mobile-Mainly geschieht dies ebenso wie unter den Mobile-Only zulasten aller weiteren Ausprägungen der Mehrpersonenhaushalte. Unter den Landline-Mainly kann indes ein erheblich höherer Anteil an Zweipersonenhaushalten gefunden werden.

Die Verteilung des Familienstandes in Abhängigkeit der Subdomain (siehe: Tabelle 51) zeigt ein deutlich dichotomeres Bild der beiden Only-Kategorien, als dies bei anderen soziodemografischen Merkmalen der Erhebung CELLA 2 zu beobachten ist. So ist ein deutlich höherer Anteil der Landline-Onlys verheiratet, während unter den Befragten der Mobile-Onlys ein wesentlich höherer Anteil angab, ledig zu sein. Tendenziell lassen sich ähnliche Befunde auch innerhalb der beiden Mainly-Kategorien beobachten. Interessant ist weiterhin, dass der Anteil an verheirateten Zielpersonen unter den Landline-Mainly nochmals oberhalb dessen der Landline-Only liegt. Bezüglich des Familienstandes der wahren Dual-User entspricht deren Verteilung annähernd derjenigen der gesamten Erhebung.

Wie bereits in Kapitel 2.3.2 beschrieben wurde, konnte auch an dieser Untersuchung gezeigt werden, dass sich gerade Only-Haushalte in ihren soziodemografischen Merkmalen von deren Verteilung in der gesamten Erhebung abheben. Während der Fokus in Kapitel 2.3.2 auf den Mobile-Onlys lag und gezeigt wurde, dass diese eher Männer jüngeren bis mittleren Alters sind, eher einen mittleren Bildungsabschluss vorweisen und überwiegend ledig in Einpersonenhaushalten leben, wurden hier auch die Merkmale der Landline-Only nochmals herausgearbeitet. Im Gegensatz zu Mobile-Onlys ist hier der Anteil weiblicher Befragter ebenso wie das durchschnittliche Alter höher. Weiterhin weisen Befragte dieser Subdomain überwiegend einen niedrigeren Bildungsabschluss vor, stammen aus Ein- oder Zweipersonenhaushalten und sind deutlich häufiger verheiratet.

Von besonderem Interesse ist jedoch gerade die Verteilung der soziodemografischen Merkmale derjenigen, die aus einer der Subdomains des Überlappungsbereiches stammen. Hier konnte gerade für die wahren Dual-User festgehalten werden, dass deren Verteilung oftmals derjenigen ähnelt, die sich über die gesamte Erhebung einstellt. Interessant wäre hierbei zu prüfen, ob die Wahrscheinlichkeit, dass eine Zielperson keine eindeutige Präferenz für ein bestimmtes Kommunikationsmedium aufweist, gleichverteilt über die untersuchten Merkmale ist, oder ob dies lediglich dadurch zustande kommt, dass sich die "gegenläufigen" Verteilungen der Only- (und Mainly-) Haushalte gegenseitig "ausgleichen", sodass die entsprechenden Mittelwerte der Erhebung koinzidenziell denjenigen ähneln, die eine "mittlere" Präferenz für beide Telekommunikationsmedien haben. Weiterhin kann dies evtl. durch einen Größeneffekt begründet werden, bedenkt man, dass die Anzahl der Befragten in dieser Subdomain für die ersten drei Modelle

diejenige der anderen Kategorien deutlich übersteigt. Dadurch kann auch mehr Variation in die Verteilung der soziodemografischen Charakteristika gelangen.⁶⁵

Die Verteilung der soziodemografischen Charakteristika der Mainly-Befragten hingegen scheint zumindest tendenziell derjenigen der Only-Haushalte des entsprechenden Telekommunikationsmediums zu folgen. Jedoch setzt sich hier der Modus bestimmter Ausprägungen innerhalb der Merkmale weniger deutlich von deren Verteilung der gesamten Erhebung ab. So sind Landline-Mainly der Erhebung CELLA 2 tendenziell in einem höheren Alter zu finden, leben vermehrt in Ein- oder Zweipersonenhaushalten, sind vorwiegend weiblich und eher verheiratet. Im Gegensatz hierzu sind Mobile-Mainly tendenziell jünger, ledig, männlich und leben in Einpersonenhaushalten.

Das einzige untersuchte Charakteristikum, dass hier nicht der Verteilung der Only-Haushalte ähnelt, ist der Bildungsabschluss. So kann zwar unter den Mobile-Mainly-Befragten auch ein höherer Anteil von Befragten mit mittlerem Bildungsabschluss gefunden werden, jedoch überwiegt hier der Anteil derjenigen mit einem formal hohen Bildungsabschluss. Ähnliches lässt sich auch für die Befragten aus der Subdomain der Landline-Mainly feststellen. So ist der Anteil der Befragten mit niedrigem Bildungsabschluss zwar ebenfalls oberhalb desjenigen der Gesamterhebung, jedoch liegt auch hier der Modus auf einem hohen formalen Bildungsabschluss. Insgesamt zeigt sich jedoch, dass sich auch Befragte der Mainly-Domains in der Verteilung soziodemografischer Merkmale von der gesamten Erhebung abheben.

6.2.2 CELLA 2: Berechnung der Composite-Faktoren und Schätzung

Die Composite-Faktoren λ , ϕ und μ , die zur Berechnung des Domain-Specific-Composite-Schätzers benötigt werden, wurden durch die Angabe des Anteils der abgeschlossenen Festnetzinterviews nach Nutzungskategorie bereits in Tabelle 23 für jedes der vier Modelle dargestellt. Eine weitere Möglichkeit zur Berechnung der Composite-Faktoren ist die Verwendung der effektiven Stichprobengröße. Diese Methode wurde bereits in Kapitel 3.3.2 eingehend für den "einfachen" Fall einer Composite-Gewichtung beschrieben, bei der lediglich der Overlap-Bereich Berücksichtigung findet. Eine solche Methode kann ebenso auf die einzelnen Domains im Domain-Specific-Composite-Ansatz angewandt werden. Daher wurden zusätzlich noch λ^* , ϕ^* und μ^* unter der Verwendung der effektiven Stichprobengröße berechnet. Hierfür wurde der modellbasierte Designeffekt sowie die effektive Stichprobengröße in Abhängigkeit der jeweiligen (Dual-User) -Domain berechnet. Diese Composite-Faktoren werden nachfolgend in Tabelle 27 gegenübergestellt.

⁶⁵ Um dies genauer zu untersuchen, fehlen jedoch wichtige Erkenntnisse, die im Vorfeld erarbeitet werden müssen. So wäre vorab zu klären, wie eine Einteilung der jeweiligen Subdomains am sinnvollsten durchzuführen wäre und welche Items dafür besonders geeignet sind. Diese beiden Fragen werden erst im Zuge dieser Arbeit genauer untersucht. Demnach müsste der Versuch der Beantwortung dieser Forschungsfrage nachgelagert erfolgen.

Tabelle 27 CELLA 2: Composite-Faktoren nach Modell und Berechnungsgrundlage

C.-Faktor	Modell1	Modell2	Modell3	Modell4
λ	0,781	0,761	0,744	0,626
ϕ	0,490	0,463	0,456	●●
μ	0,357	0,40	0,454	0,455
λ^*	0,740	0,719	0,702	0,563
ϕ^*	0,422	0,397	0,375	●●
μ^*	0,318	0,344	0,414	0,388

Unter Verwendung dieser Composite-Faktoren wurden anschließend verschiedene Variablen geschätzt und mit den Ergebnissen des Single-Frame-Ansatzes und der amtlichen Statistik verglichen.

Bezug nehmend auf die Schätzung der Bildungsvariable, die in Tabelle 11 bereits dargestellt wurde, erfolgte die Schätzung dieser Variable auch unter der Verwendung des Domain-Specific-Composite-Ansatzes. Die Resultate dieser sowie der vorherigen Schätzungen können anhand von Tabelle 28 und 29 nachvollzogen werden.

Tabelle 28 CELLA 2: Höchster allgemeiner Schulabschluss nach Modell und Methode 1 (in %)

Bildungsabschluss	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	MZ 2010
Ohne/noch kein	4,0	3,9	3,9	3,8	3,8	7,6
Haupt-/Volksschulabschl.	20,8	20,9	20,7	20,7	20,7	37,0
Polytechn. OS	6,3	6,0	6,0	6,0	6,0	7,1
Mittlere Reife	26,6	26,3	26,1	26,2	26,3	21,7
FH/HS - Reife	40,8	41,4	41,6	41,6	41,5	25,8
Andere	1,6	1,6	1,6	1,6	1,6	0,4
Designeffekt	1,38	1,52	1,53	1,54	1,53	●●

Tabelle 29 CELLA 2: Höchster allgemeiner Schulabschluss nach Modell und Methode 2 (in %)

Bildungsabschluss	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	MZ 2010
Ohne/noch kein	4,0	3,9	3,9	3,9	3,9	7,6
Haupt-/Volksschulabschl.	20,8	20,9	20,8	20,8	20,8	37,0
Polytechn. OS	6,3	6,0	6,0	6,1	6,1	7,1
Mittlere Reife	26,6	26,2	26,0	26,1	26,2	21,7
FH/HS – Reife	40,8	41,3	41,6	41,5	41,4	25,8
Andere	1,6	1,6	1,6	1,6	1,6	0,4
Designeffekt	1,38	1,47	1,49	1,49	1,47	●●

Der Vergleich der Bildungsvariable des Datensatzes der CELLA 2-Erhebung unter der Berücksichtigung der unterschiedlichen Modelle des Domain-Specific-Composite-Ansatzes mit der amtlichen Statistik sowie dem Single-Frame-Schätzer bestätigt die Ergebnisse der vorherigen Simulation. So zeigt sich, dass, auch wenn die geschätzten Werte grundsätzlich nahe beieinanderliegen, der Single-Frame-Schätzer Ergebnisse hervorbringt, die näher an der Verteilung der amtlichen Statistik liegen. Zusätzlich ist der modellbasierte Designeffekt dieses Schätzers geringer, was (ohne Klumpung der Einheiten) auf eine geringere Variation der Gewichte und eine höhere Präzision hindeutet.

Der Vergleich der beiden Methoden bevorzugt die Verwendung der effektiven Stichprobengröße zur Bestimmung der Composite-Faktoren gegenüber der Verwendung der reinen Sub-Stichprobengrößen. Dies lässt sich an den Ergebnissen der Schätzung zeigen, die (geringfügig) näher an den Angaben der amtlichen Statistik liegen. Weiterhin führt eine solche Verwendung zu einem geringeren Designeffekt.

Bezüglich der unterschiedlichen Schätzergebnisse in Abhängigkeit der Zuordnung in die jeweiligen Nutzungskategorien lässt sich keines der vier Modelle eindeutig favorisieren. Jedoch ist zu vermuten, dass aufgrund der Zuordnung das vierte Modell als wenig plausibel erscheint, da dieses die Möglichkeit, dass Einheiten der Dual-User-Population keine Präferenz für ein bestimmtes Telekommunikationsmedium haben, vollständig negiert. Anhand der Untersuchung weiterer Variablen, die im Anhang in Kapitel A.3 dargestellt werden, zeigt sich jedoch, dass Modell 2 bzw. Modell 3 am ehesten zu bevorzugen sind.

Grundsätzlich lässt sich jedoch nicht ausschließen, dass die Einteilung in der Nutzungsgewohnheiten aufgrund der verwendeten Proxis fehlgeleitet ist und die vermeintlich schlechteren Schätzergebnisse aufgrund einer fehlerhaften Einteilung in die jeweiligen Domains zustande kommen. Daher wird der Domain-Specific-Composite-Ansatz folgend anhand des Datensatzes

der Erhebung Influenza 2014 nochmals untersucht. Im Rahmen dieser Erhebung wurden zusätzlich Fragen inkorporiert, die der Einteilung nach Nutzungsgewohnheiten dienen.

6.3 Anwendung auf die Daten der Erhebung Influenza

Die 2014 durchgeführte Dual-Frame Erhebung Influenza setzt sich aus mehreren Teilstichproben zusammen, deren Rufnummern alle innerhalb des Zeitraumes von Mai bis Juni aus den Auswahlgrundlagen von GESIS gezogen wurden. Die jeweiligen Festnetz- und Mobilfunkstichproben beruhen hierbei auf der gleichen Auswahlgrundlage. Insgesamt erzielte die Erhebung 1.519 Interviews in den drei unabhängigen Stichproben Mobilfunk (367 abgeschlossene Interviews), Festnetz (647 abgeschlossene Interviews) sowie einem Oversample im Festnetz, bei dem ausschließlich nach Personen, die mindestens 60 Jahre alt sind, gescreeent wurde (505 abgeschlossene Interviews). Weiterhin wurden im Rahmen dieser Erhebung Fragen, die der Einschätzung der Nutzungsgewohnheiten einer Erhebungseinheit dienen sollen, inkludiert. Die sieben Items, die dafür verwendet werden können, wurden durch GESIS in Kooperation mit der TU Dresden entwickelt. Weiterhin waren diese nur Bestandteil der "regulären" Dual-Frame Erhebung und wurden nicht im Oversample abgefragt. Begründet wird diese Entscheidung damit, dass sich die Nutzungsgewohnheiten der Subpopulation, die mindestens 60 Jahre alt ist, vermeintlich von der Gesamtpopulation systematisch unterscheiden. Überdies beinhalten die Erhebungsdaten drei Interviews, die im Rahmen von Qualifizierungsanrufen zur Überprüfung der Ergebnisse des Predictive-Dialers abgeschlossen wurden. Da die Qualifizierungsanrufe außerhalb der eigentlichen Feldzeit liegen, wurden auch diese aus der hier vorliegenden Betrachtung ausgeschlossen. Aufgrund dessen reduziert sich die Anzahl der hier betrachteten Interviews auf 1.011.⁶⁶ Enthalten sind hierbei 102 Interviews mit Einheiten die zur Landline-Only-Population, 63 Interviews mit Einheiten die zur Mobile-Only-Population und 846 Interviews mit Einheiten die zur Dual-User-Population gehören. Folgend wird zuerst die Einteilung in die in Kapitel 5.2 dargestellten Nutzungskategorien anhand der eigens dafür entwickelten Items beschrieben.

6.3.1 Influenza 2014: Einteilung nach Verwendungskategorie

Im Gegensatz zu der zuvor geschilderten Erhebung CELLA 2 wurden im Rahmen dieser Erhebung Items erhoben, durch die eine Einteilung in Nutzungskategorien direkt ermöglicht werden soll. Überdies wurde noch einer der in der CELLA-Erhebung verwendeten Items, der als Proxy zur Einschätzung des Nutzungsverhaltens verwendet wurde, erhoben. Die Erhebung dieses Items dient dem späteren Vergleich der Einteilung anhand von Proxys und den spezifisch auf das

⁶⁶ 1.519 abzüglich der 505 Interviews des Oversamples und der drei Interviews, die im Rahmen der Qualifizierungsanrufe getätigt wurden.

Nutzungsverhalten abzielende Fragen. Der Vergleich wird bemüht, um einschätzen zu können, ob eine Einteilung anhand der Proxys zu ähnlichen Ergebnissen führt. Daher lässt sich durch diesen Vergleich abschätzen, ob die zuvor gewählten Einteilungen im Rahmen der CELLA 2-Erhebung adäquat gewesen sind.

Um die Einteilung aller Dual-User in eine entsprechende Nutzungskategorie zu ermöglichen, wurden des Weiteren fehlende Werte, die bei Dual-Usern auftraten anhand des in Kapitel 3.5.1 beschriebenen Verfahrens imputiert.⁶⁷ Die entsprechenden Items sowie deren Verteilung auf die jeweiligen Antwortkategorien stellen sich wie folgt dar:

- **Mainly1:** *Für wie wahrscheinlich halten Sie es, dass wir Sie in den letzten zwei Stunden auf Ihrem Handy erreicht hätten?*

Sehr wahrscheinlich	382
eher wahrscheinlich	148
teils/teils	111
eher unwahrscheinlich	103
sehr unwahrscheinlich	163
tnz	102
NA	2

- **Mainly2:** *Für wie wahrscheinlich halten Sie es, dass wir Sie in den letzten zwei Stunden zu Hause auf ihrem Festnetz erreicht hätten?*

Sehr wahrscheinlich	329
eher wahrscheinlich	129
teils/teils	162
eher unwahrscheinlich	134
sehr unwahrscheinlich	181
tnz	63
NA	13

- **Mainly3:** *Ganz allgemein, wenn man Sie erreichen möchte, wie geht das am besten?*

⁶⁷ Als weitere Hilfsvariablen wurden dabei die Stichprobe, aus der eine Einheit in die Erhebung gelangt ist, die Anzahl der Festnetz- und Mobilfunknummern sowie das Geschlecht, die Erwerbssituation und der Migrationshintergrund einer Einheit aus der Dual-User-Domain verwendet.

deutlich eher auf dem Festnetz	152
eher auf dem Festnetz	129
Handy und Festnetz sind gleich wahrscheinlich	172
eher auf dem Handy	232
deutlich eher auf dem Handy	161
tnz	165

- **Mainly4:** *Nun geht es um die Empfangsbereitschaft Ihres Handys, also dass Ihr Handy eingeschaltet ist.*

Wie oft ist Ihr Handy gewöhnlich empfangsbereit?

Immer	596
Täglich	169
Mehrmals in der Woche	58
Seltener	82
tnz	102
NA	4

- **Mainly4A:** *Wir möchten es gern noch etwas genauer wissen.*

Wie viele Stunden am Tag haben Sie Ihr Handy durchschnittlich empfangsbereit?

1 - 4 Std.	14
5 - 8 Std.	32
9 - 12 Std.	68
13 - 16 Std.	20
17 - 20 Std.	10
21 - 24 Std.	44
tnz	806
NA	17

- **Mainly5:** *Abgesehen von diesem Anruf, sind Sie in den letzten zwei Stunden auf Ihrem Handy angerufen worden?*

Ja	142
Nein	215
tnz	646
NA	8

■ **Mainly6:** *Abgesehen von diesem Anruf, sind Sie in den letzten zwei Stunden auf Ihrem Festnetz angerufen worden?*

Ja	250
Nein	373
tnz	366
NA	22

Anhand dieser Items wurde folgend eine Einteilung in die jeweiligen Domains der Dual-User, die das Nutzungsverhalten dieser widerspiegeln soll, durchgeführt.

Analog zum Vorgehen, das auch schon zu der Beurteilung der Proxys im Zuge der Anwendung der Cella 2-Daten verwendet wurde, kann auch an dieser Stelle anhand von Tabelle 30 die Teststatistik der paarweisen χ^2 -Tests auf Unabhängigkeit abgelesen werden.

Tabelle 30 **Influenza 2014: Paarweise Unabhängigkeit der Items zur Einordnung von Nutzungsgewohnheiten**

Variable/ p-Wert	Mainly1	Mainly2	Mainly3	Mainly4
Mainly1	•	$4,2 \cdot e^{-12}$	$2,2 \cdot e^{-16}$	$2,2 \cdot e^{-16}$
Mainly2	$4,2 \cdot e^{-12}$	•	$2,2 \cdot e^{-16}$	0,004
Mainly3	$2,2 \cdot e^{-16}$	$2,2 \cdot e^{-16}$	•	$2,2 \cdot e^{-16}$
Mainly4	$2,2 \cdot e^{-16}$	0,004	$2,2 \cdot e^{-16}$	•

Wie auch zuvor wurden die jeweiligen Statistiken ausschließlich anhand der Substichprobe der Dual-User berechnet. Da die Items *Mainly5* und *Mainly6* lediglich denjenigen gestellt wurde, die über das jeweilige Telekommunikationsmedium erhoben wurden, ist deren Betrachtung in der vorliegenden Tabelle ausgeschlossen. Jedoch sind deren sowie die weiteren vollständigen Teststatistiken im Anhang A.1.2 nachzulesen. Dabei wurde die Teststatistik für diese beiden Variablen ausschließlich für diejenigen der Substichprobe berechnet, die auch über das entsprechende Telekommunikationsmedium erhoben wurden.

Im Gegensatz zu der Betrachtung im Zuge der CELLA 2-Daten liegen Variablen vor, die die tatsächlichen Präferenzen für ein bestimmtes Telekommunikationsmedium messen sollen. Auch hier ist auffällig, dass für jeglichen paarweise durchgeführten Test, die Nullhypothese, dass beide Variablen unabhängig seien, hochsignifikant abgelehnt werden kann. Rein intuitiv sind diese Ergebnisse auch durchaus plausibel, bedenkt man, dass die eine grundsätzliche Erreichbarkeit über ein bestimmtes Telekommunikationsmedium (*Mainly3*) im Zusammenhang mit der besseren Erreichbarkeit der letzten beiden Stunden (*Mainly1* und *Mainly2*) steht.

Wie bereits beschrieben, wurden die Variablen *Mainly5* und *Mainly6* lediglich auf der Basis der Stichprobe untersucht, durch deren Telekommunikationsmedium auch der Kontakt erfolgt ist. Im Gegensatz zu den vorherigen Ergebnissen über den gesamten Bereich der Dual-User, zeigt sich hier, dass eine paarweise Unabhängigkeit mit anderen Variablen nicht durchgehend abgelehnt werden kann. So lässt sich anhand der in Anhang A.1.2 abgebildeten Untersuchungsergebnisse feststellen, dass für beide Variablen lediglich die Unabhängigkeit mit der grundsätzlichen Erreichbarkeit signifikant abgelehnt werden kann. Weiterhin kann gezeigt werden, dass zwischen weiteren eingehenden Anrufen über das Festnetz (*Mainly6*) innerhalb der zwei Stunden vor der Befragung und der besseren Erreichbarkeit über das Festnetz im selben Zeitraum, die Unabhängigkeit auf einem Signifikanzniveau von $\alpha = 0,1$ abgelehnt werden kann. Mit einem p-Wert von 0,1023 ist dieser nur geringfügig oberhalb der Schwelle für die entsprechende Testentscheidung, wenn man sich dem analogen Test für das Telekommunikationsmedium Mobilfunk widmet. Grundsätzlich fallen die entsprechenden p-Werte für die beiden Variablen *Mainly5* und *Mainly6* jedoch höher aus, was vermutlich auch der geringen Stichprobengröße zuzuschreiben ist. Jedoch sind auch diese Ergebnisse nicht kontraintuitiv. So wäre die Ablehnung der Unabhängigkeit zwischen einem eingehenden Anruf über das Mobilfunknetz bei einer Zielperson, die gewöhnlich deutlich eher über dieses Telekommunikationsmedium zu erreichen ist, zu erwarten. Dies kann auch durch Aspekte des sozialen Umfeldes einer Zielperson begründet werden, da anzunehmen ist, dass Personen aus diesem Umfeld eine Person eher über dieses Medium kontaktieren, wenn diese wissen, dass die Zielperson eher über dieses Medium zu erreichen ist. Das Gleiche ist auch für das Telekommunikationsmedium Festnetz anzunehmen. Auch wenn diese beiden Variablen durchaus interessante Einblicke in die Nutzungsgewohnheiten geben können, besteht das Problem, dass sie nur für einen bestimmten Teil der Dual-User abgefragt wurden. Daher wurde auf deren Einbeziehung in der Einteilung der Dual-User nach Nutzungsgewohnheiten verzichtet.

Da auch bei den zugrunde liegenden Items vorerst ungewiss ist, wie ein *Mainly*-Haushalt exakt zu bestimmen ist, erfolgt die Einteilung analog zum Vorgehen in der CELLA 2-Erhebung. Demnach wurde in einem ersten Modell zur Einteilung ein minimaler Anteil an Zielpersonen der *Mainly*-Domains angenommen, der dann in weiteren Modellen sukzessive ausgeweitet wurde. Die Grundlage der Einteilung der Dual-User nach Nutzungsgewohnheiten in die jeweiligen Domains anhand der entsprechenden Modelle war:

- Modell1:** *Landline-Mainlys:* Diejenigen, die angeben, dass sie (grundsätzlich) deutlich eher auf dem Festnetz erreichbar sind.
- Mobile-Mainlys:* Diejenigen, die angeben, dass sie (grundsätzlich) deutlich eher auf dem Handy erreichbar sind.
- Modell2:** *Landline-Mainlys:* Diejenigen, die angeben, dass sie (grundsätzlich) deutlich eher auf dem Festnetz erreichbar sind. Weiterhin sind diejenigen inkludiert, die eher auf dem Festnetz erreichbar sind und angeben, dass sie sehr wahrscheinlich innerhalb der letzten beiden Stunden über Festnetz kontaktierbar gewesen wären, während sie sehr unwahrscheinlich über Mobilfunk innerhalb der letzten beiden Stunden erreichbar waren.
- Mobile-Mainlys:* Diejenigen, die angeben, dass sie (grundsätzlich) deutlich eher auf dem Handy erreichbar sind. Weiterhin sind diejenigen inkludiert, die eher auf dem Handy erreichbar sind und angeben, dass sie sehr wahrscheinlich innerhalb der letzten beiden Stunden über Mobilfunk kontaktierbar gewesen wären, während sie sehr unwahrscheinlich über Festnetz innerhalb der letzten beiden Stunden erreichbar waren.
- Modell3:** *Landline-Mainlys:* Diejenigen, die angeben, dass sie (grundsätzlich) deutlich eher auf dem Festnetz erreichbar sind. Hinzu kommen diejenigen, die eher auf dem Festnetz erreichbar sind und angeben, dass sie sehr wahrscheinlich innerhalb der letzten beiden Stunden über Festnetz kontaktierbar gewesen wären.
- Mobile-Mainlys:* Diejenigen, die angeben, dass sie (grundsätzlich) deutlich eher auf dem Handy erreichbar sind. Hinzu kommen diejenigen, die eher auf dem Handy erreichbar sind und angeben, dass sie sehr wahrscheinlich innerhalb der letzten beiden Stunden über Mobilfunk kontaktierbar gewesen wären.
- Modell4:** *Landline-Mainlys:* Diejenigen, die angeben, dass sie (grundsätzlich) deutlich eher oder eher auf dem Festnetz erreichbar sind.
- Mobile-Mainlys:* Diejenigen, die angeben, dass sie (grundsätzlich) deutlich eher oder eher auf dem Handy erreichbar sind.

Unter der Verwendung einer solchen Aufteilung ergibt sich die Verteilung der Nutzungsgewohnheiten nach Tabelle 31.

Tabelle 31 Influenza 2014: Dual-User nach Verwendungskategorie

Modell/ Teilpopulation	u_2	u_3	u_4
Modell 1	152	533	161
Anteil Gesamt-SP	15,0 %	52,7 %	15,9 %
Modell 2	163	499	185
Anteil Gesamt-SP	16,1 %	49,3 %	18,3 %
Modell 3	206	368	272
Anteil Gesamt-SP	20,4 %	36,4 %	26,9 %
Modell 4	281	172	393
Anteil Gesamt-SP	27,8 %	17,0 %	38,9 %

Analog zur Vorgehensweise bei der Modellbildung der CELLA 2-Erhebung wurde auch hier der Anteil der wahren Dual-User schrittweise verkleinert. Begründet wird dieses Vorgehen damit, dass zwar anhand der entwickelten Items eine Einteilung in die entsprechenden Mainly-Domains ermöglicht werden soll, dass allerdings nicht bekannt ist, wie genau ein Mainly-Haushalt seine eigene Erreichbarkeit über ein entsprechendes Medium einschätzen würde. Im Gegensatz zur Modellbildung in der CELLA 2-Erhebung wurde jedoch darauf verzichtet, ein Modell zu bilden, das den Anteil der wahren Dual-User negiert, da dies wenig plausibel erscheint und sich durchaus Beispiele konstruieren lassen, bei denen keine klare Präferenz eines bestimmten Telekommunikationsmediums eines Dual-Users erkennen lässt. So ist es bspw. denkbar, dass eine Einheit eines Einpersonenhaushaltes über Tag (bspw. während seines Arbeitstages) ausschließlich über den Mobilfunkanschluss erreichbar ist und am Abend, da zu diesem Zeitpunkt zum Beispiel das Handy ausgeschaltet ist, nur über den Festnetzanschluss kontaktiert werden kann. Eine solche Einheit hätte zwar zu bestimmten Tageszeiten eine eindeutig bessere Erreichbarkeit über ein bestimmtes Telekommunikationsmedium, jedoch lässt sich über den Verlauf eines Tages nicht eindeutig bestimmen, ob eine Präferenz für ein Medium vorliegt.

Daher stellt auch diese Aufteilung lediglich eine Annäherung an die Aufteilung nach Nutzungsgewohnheiten in der Population dar. Jedoch lässt sich gerade durch die Inklusion der ersten beiden Fragen zur Erreichbarkeit in den letzten beiden Stunden eine Aussage darüber treffen, ob sich die grundsätzliche Erreichbarkeit und die "akute" Erreichbarkeit unterscheiden. Ein Fall, wie er im soeben dargestellten Beispiel beschrieben wurde, könnte daher unter Umständen nachvollzogen werden. Daher wird die gemeinsame Verteilung der grundsätzlichen Erreichbarkeit (M3) und der Erreichbarkeit der letzten beiden Stunden nach Mobilfunk (M1) und Festnetz (M2) in den Tabellen 32 und 33 gegenübergestellt.

Tabelle 32 Grundsätzliche Erreichbarkeit vs. Erreichbarkeit der letzten 2 Stunden (Mobilfunk)

M3/ M1	Sehr wahrsch.	eher wahrsch.	teils/ teils	eher unwahrsch.	sehr unwahrsch.
deutlich eher Festnetz	21 (13,8 %)	13 (8,6 %)	13 (8,6 %)	30 (19,7 %)	75 (49,3 %)
eher Festnetz	31 (24,0 %)	21 (16,3 %)	18 (14,0 %)	30 (23,3 %)	29 (22,5 %)
gleich wahrscheinlich	88 (51,2 %)	32 (18,6 %)	26 (15,1 %)	7 (4,1 %)	19 (11,0 %)
eher Handy	111 (47,8 %)	48 (20,7 %)	27 (11,6 %)	22 (9,5 %)	24 (10,3 %)
deutlich eher Handy	100 (62,1 %)	27 (16,8 %)	17 (10,6 %)	6 (3,7 %)	11 (6,8 %)

Tabelle 33 Grundsätzliche Erreichbarkeit vs. Erreichbarkeit der letzten 2 Stunden (Festnetz)

M3/ M2	Sehr wahrsch.	eher wahrsch.	teils/ teils	eher unwahrsch.	sehr unwahrsch.
deutlich eher Festnetz	81 (53,3 %)	23 (15,1 %)	18 (11,8 %)	13 (8,6 %)	17 (11,2 %)
eher Festnetz	54 (41,9 %)	32 (24,8 %)	17 (13,2 %)	11 (8,5 %)	15 (11,6 %)
gleich wahrscheinlich	69 (40,1 %)	28 (16,3 %)	31 (18,0 %)	20 (11,6 %)	24 (14,0 %)
eher Handy	53 (22,8 %)	26 (11,2 %)	50 (21,6 %)	42 (18,1 %)	61 (26,3 %)
deutlich eher Handy	32 (19,9 %)	9 (5,6 %)	29 (18,0 %)	39 (24,2 %)	52 (32,3 %)

Die dargestellten Verteilungen geben die bedingten Anteilswerte der Antworten zu M1 oder M2 wieder. Demnach sind diese als die Anteile der Einheiten, die eine bestimmte Erreichbarkeit in den letzten beiden Stunden angegeben haben, gegeben der Antwort auf die grundsätzliche Erreichbarkeit, zu verstehen. Zumindest für die hier vorliegende Erhebung zeigt sich ein direkter Zusammenhang zwischen der Erreichbarkeit der letzten beiden Stunden und der grundsätzlichen Erreichbarkeit über ein bestimmtes Medium. Dies zeigen auch die Ergebnisse der Chi-Quadrat-Tests auf Unabhängigkeit.⁶⁸ Für M1 und M3 ergibt sich dabei ein χ^2 -Wert von 219,0 (bzw. ein P-Wert von $p < 2.2^{-16}$) und für M2 bezogen auf M3 ein Wert von 216,5 (bzw. ein P-Wert von $p < 2.2^{-16}$). Eine Unabhängigkeit kann demnach ausgeschlossen werden.

Gerade bei denjenigen, die im Rahmen der Erhebung angegeben haben, dass sie deutlich eher über eines der beiden Telekommunikationsmedien erreichbar sind, zeigt sich auch innerhalb des Zweistundenfensters eine deutlich bessere Erreichbarkeit gerade für dieses Medium. Dies

⁶⁸ Die vorliegenden Tests beruhen auf der Häufigkeitsverteilung als absolute Werte.

stellt sich sowohl für diejenigen ein, die grundsätzlich besser über Festnetz erreichbar sind, sowie für diejenigen, die grundsätzlich besser über ihren Mobilfunkanschluss erreichbar sind. Wenig verwunderlich erscheint auch, dass die Erreichbarkeit der letzten beiden Stunden über das Telekommunikationsmedium, über das eine Zielperson nicht grundsätzlich deutlich besser erreichbar ist, einen wesentlich höheren Anteil an Einheiten aufweist, die es als unwahrscheinlich oder sehr unwahrscheinlich einschätzen, gerade über dieses Medium erreicht worden zu sein.

Auch für diejenigen, die während der Befragung angaben, keine Präferenz für ein bestimmtes Telekommunikationsmedium zu haben, treten die Ergebnisse wie erwartet ein. So lässt sich für die Mehrheit derjenigen, die diese Antwortkategorie in M3 gewählt haben, feststellen, dass die Erreichbarkeit über die letzten beiden Stunden bei beiden Telekommunikationsmedien als sehr wahrscheinlich eingeschätzt wird. Die Darstellung in den beiden vorherigen Tabellen erlaubt für diese Gruppe jedoch nicht, eine Aussage darüber zu treffen, wie hoch der Anteil dieser Zielpersonen war, die grundsätzlich keine Präferenz für ein bestimmtes Telekommunikationsmedium aufweisen und es sowohl für Festnetz als auch für Mobilfunk angaben, in den letzten beiden Stunden sehr wahrscheinlich erreichbar gewesen zu sein. Bei einer weitergehenden Untersuchung derjenigen, die keine eindeutige Präferenz für ein bestimmtes Telekommunikationsmedium aufweisen, zeigt sich, dass ca. 27 % dieser Befragten auch in den letzten beiden Stunden vor ihrer Befragung sowohl über Festnetz als auch über ihr Handy sehr wahrscheinlich zu erreichen gewesen sind. Demnach kann gerade für diese Gruppe vermutet werden, dass die Erreichbarkeit über ein bestimmtes Telekommunikationsmedium auch vom Zeitpunkt der Befragung abhängig ist. Weiterhin gab ein höherer Anteil der Befragten dieser Gruppe an, in den letzten beiden Stunden sehr wahrscheinlich über ihr Handy erreichbar gewesen zu sein.

Für diejenigen, die lediglich eine vage grundsätzliche Präferenz für ein bestimmtes Telekommunikationsmedium geäußert haben ("eher" anstelle von "deutlich eher" bei M3) stellen sich prinzipiell die gleichen Ergebnisse ein, die sich auch bei einer klaren Präferenz beobachten lassen. Dennoch zeigt sich, dass hier auch ein höherer Anteil angibt, über das andere, nicht präferierte Telekommunikationsmedium in den letzten beiden Stunden erreichbar gewesen zu sein. Gerade für Mobilfunkanschlüsse scheint dies in einem höheren Maß zuzutreffen. Daher scheint das graduelle Vorgehen, das in der Beschreibung der unterschiedlichen Modelle dargelegt wird, adäquat, da nicht direkt ersichtlich wird, wie groß der Anteil derjenigen in diesen beiden Gruppen ist, der zu einer der beiden Mainly-Populationen gehört und wie groß der Anteil der wahren Dual-User in dieser Gruppe ist. Treibend hierbei könnten weiterhin Probleme des Under- und Overreporting der Befragten sein. So kann nicht ausgeschlossen werden, dass eine befragte Zielperson entweder ihre Erreichbarkeit in den letzten beiden Stunden über ein spezifisches Telekommunikationsmedium über- oder unterschätzt oder seine grundsätzliche Präferenz für ein bestimmtes Telekommunikationsmedium die eigene Einschätzung zur grundsätzlichen Erreichbarkeit durch ein bestimmtes Telekommunikationsmedium beeinflusst.

Weiterhin von Interesse ist der Zusammenhang zwischen der Einschätzung der Erreichbarkeit über ein bestimmtes Telekommunikationsmedium und die tatsächliche Erreichbarkeit über gerade dieses. Dabei geht, wie bereits zuvor geschildert, die Überlegung voraus, dass Dual-User, die eine Präferenz für ein bestimmtes Medium äußern, auch eher über gerade dieses erreichbar sind. Als Konsequenz sollte sich hieraus ergeben, dass diejenigen, die durch die unterschiedlichen Modellbildungen als Mainlys identifiziert wurden, anteilig weniger Kontaktversuche bis zum erfolgreichen Interview benötigen. Daher wird folgend anhand der Abbildungen 12 und 13 die Anzahl der benötigten Kontaktversuche in Abhängigkeit des Mediums und der Verwendungskategorie nach Modell 1 näher untersucht. Aufgrund der Erkenntnisse aus Tabelle 32 und 33 wurde Modell 1 als Minimalmodell gewählt, da in diesem die Zuordnungsprobleme, die durch die Aussprache einer "weniger eindeutigen", besseren Erreichbarkeit ("eher besser erreichbar") entstehen können, wegfallen. Eine Darstellung der Anzahl der Kontakte getrennt nach Medium unter der Zuordnung nach dem vierten Modell findet sich im Anhang.

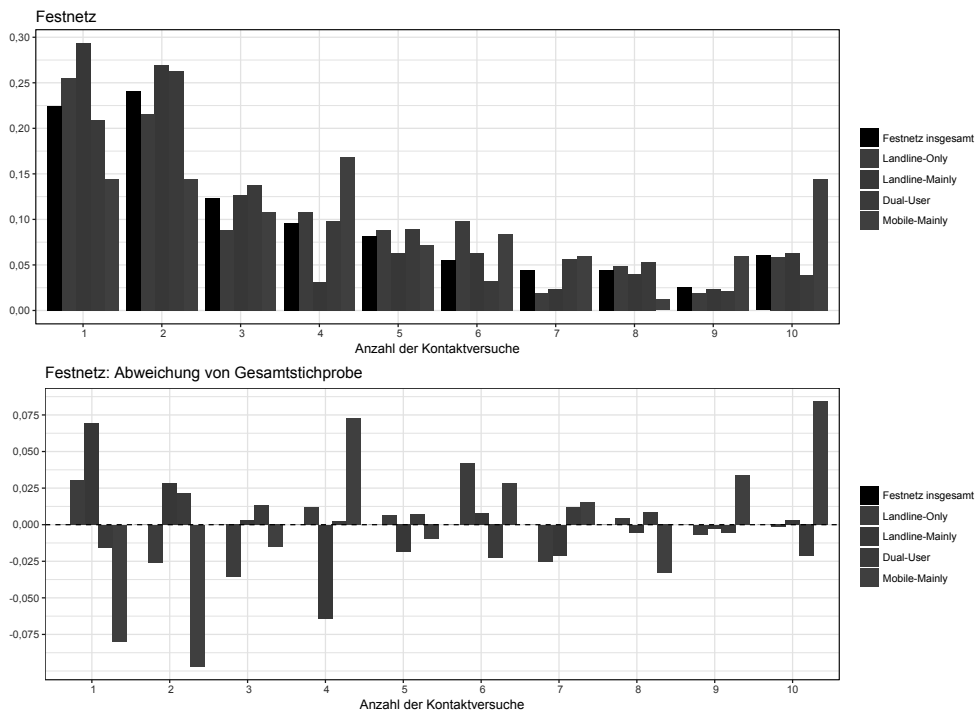


Abbildung 12 Festnetz: Anzahl der Kontakte nach Modell 1

Anhand der abgeschlossenen Festnetzinterviews nach der Anzahl der Kontakte sowie des ersten Modells wird ersichtlich, was bereits in Kapitel 5 vermutet wurde. So lässt sich eindeutig ein höherer Anteil an abgeschlossenen Interviews nach den ersten (beiden) Kontaktversuchen unter den Landline-Onlys und Landline-Mainlys finden, als dies über die gesamte Festnetzstichprobe

zu beobachten ist. Gerade für diejenigen Einheiten, die im Rahmen dieses Modells als Landline-Mainly eingeordnet werden, ist dies zu beobachten. Ebenfalls zeigt sich, dass der Anteil der Landline-Mainlys, die mehr als drei und weniger als neun Kontaktversuche bis zu einem erfolgreich abgeschlossenen Interview benötigen, deutlich abnimmt und unterhalb des Anteils der gesamten Festnetzstichprobe liegt. Lediglich bei zehn oder mehr Kontaktversuchen ist wieder ein Anstieg des Anteils der Landline-Mainlys zu beobachten, der oberhalb desjenigen der gesamten Stichprobe ist. Vermutet werden kann bei zehn oder mehr Kontaktversuchen jedoch, dass diese ohnehin nur Personen beinhaltet, die den Teil der Zielpopulation widerspiegeln, die schwer zu erreichen sind. Durchschnittlich wurden für ein erfolgreiches Interview ca. 3,7 Anrufe benötigt. Die durchschnittliche Anzahl der Kontaktversuche für Landline-Onlys unterscheidet sich hiervon nicht wesentlich, während die durchschnittliche Anzahl der Kontaktversuche der Festnetzstichprobe mit einem Wert von 3,4 nur leicht abweicht. Allerdings wird dieser Wert stark durch den Anteil einer hohen Anzahl von Kontaktversuchen beeinflusst. Daher eignet sich der Median deutlich eher zum Vergleich der Anzahl der Kontaktversuche. Dieser liegt sowohl für die gesamte Stichprobe als auch für Landline-Onlys bei einem Wert von drei. Der Median der Landline-Mainlys in diesem Modell liegt im Gegensatz dazu bei zwei. Daher entspricht diese Untersuchung der zuvor geäußerten Vermutung, dass Landline-Mainlys weniger Kontaktversuche für ein erfolgreiches Interview benötigen, wenn diese auch über das Festnetz kontaktiert werden.

Die Ergebnisse der wahren Dual-User entsprechen ebenso wie diejenigen der Mobile-Mainly den Erwartungen der zuvor geäußerten Vermutung. So entspricht der Median der wahren Dual-User demjenigen der gesamten Festnetzstichprobe und auch die durchschnittliche Anzahl der Kontaktversuche weicht mit einem Wert von 3,4 nur leicht vom Gesamtdurchschnitt ab.

Sowohl anhand der Abbildung 12 als auch am Median der Mobile-Mainlys mit einem Wert von 4 zeigt sich, dass diese eine erheblich höhere Anzahl von Kontakten benötigen, um ein erfolgreiches Interview abzuschließen, wenn diese über das Telekommunikationsmedium befragt werden, das sie nicht dominant verwenden.⁶⁹

69 Vergleichbare Ergebnisse stellen sich auch in der Betrachtung des vierten Modells ein.

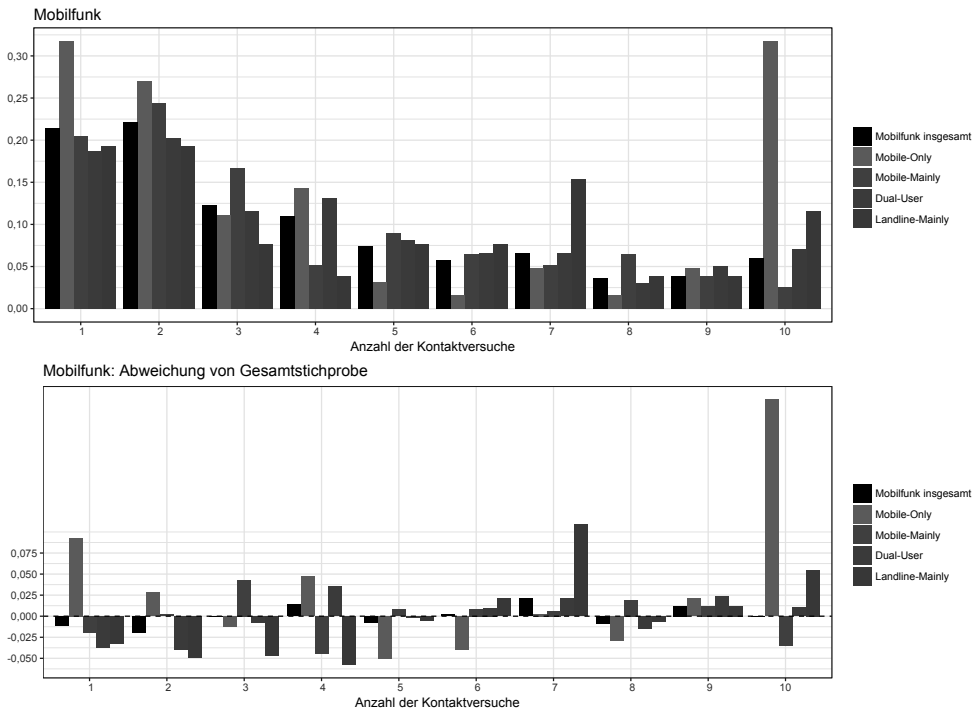


Abbildung 13 Mobilfunk: Anzahl der Kontakte nach Modell 1

Die gleiche Untersuchung der Mobilfunkstichprobe bringt weniger deutliche Ergebnisse. Der Grund dafür kann durch den verhältnismäßig hohen Anteil der Mobile-Onlys, die entweder direkt beim ersten Kontaktversuch erreicht werden und/oder denjenigen, die zehn Kontaktversuche benötigten, begründet werden. Für beide Ausprägungen lässt sich ein deutlich höherer Anteil der Mobile-Onlys beobachten, der wiederum die relative Verteilung der Kontaktversuche der gesamten Mobilfunkstichprobe in den beiden Ausprägungen beeinflusst. Grundsätzlich zeigt sich jedoch auch für diese Stichprobe unter der gegebenen Modellbildung, dass Einheiten, die zu den Mobile-Mainly gezählt werden, durchschnittlich weniger Kontaktversuche benötigen, während Landline-Mainlys eine höhere Anzahl von Kontakten brauchen, bis ein Interview abgeschlossen ist. Jedoch sind die Ergebnisse weniger deutlich als in der Festnetzstichprobe. Dies kann evtl. mit der in Kapitel 5 geäußerten Annahme, dass Einheiten in Mobilfunkstichproben schwieriger erreichbar sind, begründet werden. Der Vergleich der mittleren Anzahl von Kontaktversuchen zwischen der gesamten Stichprobe und den Mobile-Mainlys liefert wenig Unterschiede und liegt bei beiden bei einem Wert von drei, während durchschnittlich nur geringfügig weniger Kontaktversuche für einen erfolgreichen Kontakt zu einer Einheit aus der Mobile-Mainly-Population benötigt werden (3,7 vs. 3,9). Aber der Median der Anzahl der Kontaktversuche der Landline-Mainly-Einheiten liegt mit einem Wert von 4,5 deutlich höher. So zeigt sich auch in dieser

Stichprobe, dass Einheiten über das Telekommunikationsmedium, das diese nicht dominant verwenden, deutlich aufwendiger zu erreichen sind.

Grundsätzlich scheint die Verwendung der dargestellten Items zur Bildung unterschiedlicher Modelle demnach zu Ergebnissen zu führen, die mit den in Kapitel 5 geäußerten Annahmen übereinstimmt. Jedoch zeigt sich gerade bei der Äußerung einer "schwachen" besseren Erreichbarkeit über ein bestimmtes Medium die Schwierigkeit, zwischen Mainly-Einheit und wahren Dual-User zu differenzieren. Daraus ersichtlich wird die Notwendigkeit einer graduellen Anpassung der Größe der jeweiligen Bereiche. Daher wird anhand von Tabelle 34 eine genauere Darstellung der Erhebung nach Teilpopulation und Modell gegeben.

Tabelle 34 Influenza 2014: Zusammensetzung innerhalb der Teilpopulation nach Modell und Stichprobe

Teilpopulation	Stichprobe	Mod 1	Mod 2	Mod 3	Mod 4
u_1	S^F	102 (100 %)	102 (100 %)	102 (100 %)	102 (100 %)
	S^C	••	••	••	••
	$S^F \cup S^C$	102 (100 %)	102 (100 %)	102 (100 %)	102 (100 %)
u_2	S^F	126 (82,9 %)	134 (82,7 %)	168 (81,6 %)	219 (77,9 %)
	S^C	26 (17,1 %)	28 (17,3 %)	38 (18,4 %)	62 (22,1 %)
	$S^F \cup S^C$	152	162 (100 %)	206 (100 %)	281 (100 %)
u_3	S^F	335 (62,9 %)	316 (63,3 %)	228 (62,0 %)	107 (62,2 %)
	S^C	198 (37,1 %)	183 (36,7 %)	140 (38,0 %)	65 (37,8 %)
	$S^F \cup S^C$	533 (100 %)	499 (100 %)	368 (100 %)	172 (100 %)
u_4	S^F	83 (51,6 %)	94 (50,8 %)	148 (54,4 %)	218 (55,5 %)
	S^C	78 (48,4 %)	91 (49,2 %)	124 (45,6 %)	175 (44,5 %)
	$S^F \cup S^C$	161 (100 %)	185 (100 %)	272 (100 %)	393 (100 %)
u_5	S^F	••	••	••	••
	S^C	63 (100 %)	63 (100 %)	63 (100 %)	63 (100 %)
	$S^F \cup S^C$	63 (100 %)	63 (100 %)	63 (100 %)	63 (100 %)

Im Gegensatz zu der Erhebung CELLA 2 wurde der Anteil der abgeschlossenen Interviews nicht so gewählt, dass ungefähr die Hälfte aller Befragungen über eines der beiden Medien erfolgte. Vielmehr wurde hier eine Aufteilung von ca. 60 % Festnetz- und 40 % Mobilfunkinterviews verwendet.⁷⁰ Daher ist es interessant zu beobachten, dass die Aufteilung der wahren Dual-User

⁷⁰ In den beiden betrachteten Stichproben der Erhebung wurden ca. 63,9 % aller Interviews via Festnetz und ca. 36,1

ebenfalls diesen Angaben über alle vier Modelle entspricht. Diese Ergebnisse sind vergleichbar mit den Beobachtungen, die bereits in der CELLA 2-Erhebung getroffen wurden. Das gleiche lässt sich auch für die Verteilung innerhalb der Landline- und Mobile-Mainly-Domains über alle vier verwendeten Modelle beobachten. Hier zeigt sich ebenso wie in der Erhebung CELLA 2, dass der Anteil der abgeschlossenen Interviews über das Telekommunikationsmedium, das dominant von der jeweiligen Teilpopulation verwendet wird, deutlich höher ist, als dieser in der Gesamtstichprobe zu vermuten ist. Demnach sind auch bei einer disproportionalen Aufteilung der abgeschlossenen Interviews nach Telekommunikationsmedium konform mit den Beobachtungen aus der CELLA 2 - Erhebung sowie den in Kapitel 5 vorgestellten Annahmen. Um weitere Erkenntnisse bezüglich des Domain-Specific-Composite-Ansatzes zu erlangen, werden auch für die Erhebung Influenza 2014 entsprechende Composite-Faktoren ermittelt und Schätzergebnisse präsentiert.

Auch bei der hier vorliegenden Modellbildung ist es von gesteigertem Interesse, ob sich Zielpersonen innerhalb des Überlappungsbereiches bezüglich ihrer soziodemografischen Charakteristika unterscheiden. Analog zum Vorgehen bei der Erhebung CELLA 2 wurden daher auch für die Zielpersonen der Erhebung Influenza 2014 die entsprechenden Verteilungen nach Subdomain untersucht. Die Tabellen 35, 36 und 37 geben in dieser Betrachtung die Verteilung nach Geschlecht, Alter und Bildungsabschluss wieder. Die Verteilung der Haushaltsgröße kann im Anhang unter Kapitel A.2.2 nachvollzogen werden.

Im Rahmen dieser Erhebung wurden insgesamt 46 % männliche und 54 % weibliche Zielpersonen befragt. Ähnlich wie im Zuge der Erhebung CELLA 2 bereits beobachtet, ist auch hier der Anteil der Frauen mit 66,6 % innerhalb der Landline-Onlys deutlich höher als in der Gesamtstichprobe. Auch die Geschlechtsverteilung der Mobile-Only zeigt ein ähnliches Ergebnis wie dasjenige der CELLA 2-Erhebung. So ist auch hier ein deutlich höherer Anteil männlicher Zielpersonen anzufinden, als in der gesamten Erhebung.

% über Mobilfunkanschlüsse geführt.

Tabelle 35 Influenza 2014: Geschlecht nach Nutzungsgewohnheit und Modell (in %)

	Geschlecht	Modell 1	Modell 2	Modell 3	Modell 4
u_1	Männlich	33,3	33,3	33,3	33,3
	Weiblich	66,7	66,7	66,7	66,7
u_2	Männlich	35,5	35,8	36,8	38,4
	Weiblich	64,5	64,2	63,2	61,6
u_3	Männlich	45,8	45,5	44,8	36,6
	Weiblich	54,2	54,5	55,2	63,4
u_4	Männlich	58,4	57,8	58,9	56,2
	Weiblich	41,6	42,2	41,1	43,8
u_5	Männlich	61,9	61,9	61,9	61,9
	Weiblich	38,1	38,1	38,1	38,1

Mit der Ausnahme des vierten Modells ist die Geschlechterverteilung der wahren Dual-User nahe an derjenigen, die sich über die gesamte Erhebung berechnen lässt und unterscheidet sich nur unwesentlich. Die Geschlechtsverteilung der wahren Dual-User im vierten Modell hingegen ist eher mit derjenigen der Landline-Only zu vergleichen. Bezüglich der Verteilung innerhalb der beiden Mainly-Kategorien lässt sich über alle vier Modelle eine ähnliche Verteilung wie diejenige der respektiven Only-Kategorien feststellen. So sind auch unter den Landline-Mainlys mehr weibliche Zielpersonen zu finden, als in der gesamten Erhebung. Ähnliches lässt sich über Mobile-Mainly und den Anteil männlicher Zielpersonen feststellen. Allerdings ist in beiden Mainly-Kategorien festzustellen, dass die Disparität zwischen den Anteilen der beiden Ausprägungen nicht so stark ist, wie dies in den Only-Kategorien vorzufinden ist. Diese Befunde bleiben über alle vier Modelle bestehen.

Der Anteil der Zielpersonen in einem Alter zwischen 18 und 29 Jahren innerhalb der Erhebung Influenza 2014 betrug 14%. 15% aller erhobenen Zielpersonen waren zwischen 30 und 39 Jahren, 18,9% zwischen 40 und 49 Jahren, 21,2% zwischen 50 und 59 Jahren, 16,3% zwischen 60 und 69 Jahren, 12,2% zwischen 70 und 79 Jahren und 2,4% waren mindestens 80 Jahre alt. Das durchschnittliche Alter innerhalb dieser Erhebung betrug ca. 50,1 Jahre. Auch hier zeigt sich, ähnlich wie im Zuge der CELLA 2-Erhebung, dass Zielpersonen, die den Landline-Onlys zuzuordnen sind, durchschnittlich älter und Mobile-Onlys durchschnittlich jünger sind als in der gesamten Erhebung. Somit ist auch der Anteil der Zielpersonen innerhalb der höheren Alterskategorien deutlich höher bei Landline-Onlys. Gleiches lässt sich auch für den Anteil der "jüngeren" Alterskategorien innerhalb der Mobile-Onlys festhalten. Das Durchschnittsalter der Mobile-Only betrug hierbei 37,4 Jahre. Das der Landline-Only lag bei 64,5 Jahren.

Tabelle 36 Influenza 2014: Alter nach Nutzungsgewohnheit und Modell (in %)

	Alter	Modell 1	Modell 2	Modell 3	Modell 4
u_1	18 - 29	3,9	3,9	3,9	3,9
	30 - 39	3,9	3,9	3,9	3,9
	40 - 49	6,9	6,9	6,9	6,9
	50 - 59	21,6	21,6	21,6	21,6
	60 - 69	18,6	18,6	18,6	18,6
	70 - 79	32,4	32,4	32,4	32,4
	80+	12,7	12,7	12,7	12,7
u_2	18 - 29	2,0	2,5	1,8	2,8
	30 - 39	4,6	4,9	4,3	5,7
	40 - 49	19,1	18,5	19,0	17,8
	50 - 59	20,4	20,4	21,5	22,1
	60 - 69	27,6	27,8	27,0	26,0
	70 - 79	22,4	22,2	22,7	22,4
	80+	3,9	3,7	3,7	3,2
u_3	18 - 29	12,6	12,6	12,2	5,2
	30 - 39	15,2	15,0	15,5	15,7
	40 - 49	21,4	21,8	21,5	25,6
	50 - 59	24,6	23,6	23,9	22,1
	60 - 69	16,1	16,6	16,7	20,9
	70 - 79	9,2	9,2	9,2	9,3
	80+	0,9	1,0	1,0	1,2
u_4	18 - 29	26,7	24,9	26,5	24,4
	30 - 39	27,3	26,5	25,9	22,6
	40 - 49	19,9	19,5	20,0	20,6
	50 - 59	15,5	19,5	17,8	22,1
	60 - 69	8,1	7,0	7,6	8,1
	70 - 79	2,5	2,7	2,2	2,0
	80+	0,0	0,0	0,0	0,0
u_5	18 - 29	39,7	39,7	39,7	39,7
	30 - 39	25,4	25,4	25,4	25,4
	40 - 49	14,3	14,3	14,3	14,3
	50 - 59	7,9	7,9	7,9	7,9
	60 - 69	7,9	7,9	7,9	7,9
	70 - 79	4,8	4,8	4,8	4,8
	80+	0,0	0,0	0,0	0,0

Auch für die Zielpersonen der beiden Mainly-Subdomains lassen sich vergleichbare Befunde festhalten. Ähnlich der Altersverteilung der Mobile-Onlys ist unter den Zielpersonen der Mobile-Mainlys ein höherer Anteil gerade jüngerer Zielpersonen zu finden, als über die gesamte Erhebung. Gleiches lässt sich über die Zielpersonen, die der Subdomain Landline-Mainly zugehörig sind und einem höheren Anteil in den oberen Alterskategorien festhalten. Diese Beobachtungen bleiben über alle vier Modelle bestehen. Für beide Mainly-Kategorien ist jedoch die Tendenz jüngerer, bzw. älterer Zielpersonen nicht so eindeutig, wie es für die entsprechenden Only-Kategorien vorzufinden ist. Interessant ist weiterhin, dass die Altersverteilung der wahren Dual-User hier nicht ähnlich derjenigen der gesamten Erhebung ist, sondern vielmehr einen höheren Anteil an Zielpersonen mittleren Alters nachweisen kann. Auch dies zeigt sich über alle vier der betrachteten Modelle. Diese Ergebnisse spiegeln sich auch im Mittelwert des Alters über die drei Subdomains im Überlappungsbereich wieder. So liegt dieses innerhalb der vier Modelle für Landline-Mainly zwischen 59,4 und 59,8 Jahren. Für Mobile-Mainlys ist dieses deutlich geringer, mit einem Wert zwischen 39,9 und 41,5, während das durchschnittliche Alter der wahren Dual-User mit Werten zwischen 49,2 und 49,4 nahe am Mittelwert über die gesamte Erhebung liegt.

Die Abfrage des höchsten Bildungsabschlusses der Zielpersonen innerhalb der Erhebung Influenza 2014 ergab, dass insgesamt 0,7% aller Befragten maximal 7 Jahre Schulbesuch vorzuweisen hatten, 15,6% aller Befragten hatten einen Hauptschulabschluss, 33,1% konnten die mittlere Reife vorweisen, 9,3% hatten eine Fachhochschul- und 41,2% allgemeine Hochschulreife. Ähnlich der Ergebnisse aus der Erhebung CELLA 2 kann festgestellt werden, dass der Anteil derjenigen Zielpersonen mit niedrigerem Bildungsabschluss unter den Befragten, die ausschließlich über das Festnetz erreicht werden können, deutlich höher als in der gesamten Stichprobe ist.

Tabelle 37 Influenza 2014: Bildungsabschluss nach Nutzungsgewohnheit und Modell (in %)

	Schulabschluss	Modell 1	Modell 2	Modell 3	Modell 4
u_1	Höchstens 7 Jahre Schulbesuch	2,9	2,9	2,9	2,9
	Haupt-/Volksschulabschluss	32,4	32,4	32,4	32,4
	Mittlere Reife	26,5	26,5	26,5	26,5
	Fachhochschulreife	7,8	7,8	7,8	7,8
	Abitur	30,4	30,4	30,4	30,4
u_2	Höchstens 7 Jahre Schulbesuch	0,7	0,6	0,6	0,4
	Haupt-/Volksschulabschluss	20,4	19,8	20,2	18,5
	Mittlere Reife	33,6	33,3	33,7	33,8
	Fachhochschulreife	5,9	6,8	6,1	6,8
	Abitur	39,5	39,5	39,3	40,6
u_3	Höchstens 7 Jahre Schulbesuch	0,4	0,4	0,4	0,0
	Haupt-/Volksschulabschluss	12,8	13,0	12,9	18,0
	Mittlere Reife	38,5	38,5	38,8	36,6
	Fachhochschulreife	9,2	9,2	9,6	14,0
	Abitur	39,2	38,9	38,4	31,4
u_4	Höchstens 7 Jahre Schulbesuch	0,0	0,0	0,0	0,5
	Haupt-/Volksschulabschluss	8,7	8,6	8,6	7,6
	Mittlere Reife	21,7	24,3	23,2	33,8
	Fachhochschulreife	13,0	11,9	11,4	9,2
	Abitur	56,5	55,1	56,8	48,9
u_5	Höchstens 7 Jahre Schulbesuch	1,6	1,6	1,6	1,6
	Haupt-/Volksschulabschluss	19,0	19,0	19,0	19,0
	Mittlere Reife	27,0	27,0	27,0	27,0
	Fachhochschulreife	11,1	11,1	11,1	11,1
	Abitur	41,3	41,3	41,3	41,3

Im Gegensatz zu den vorherigen Ergebnissen aus CELLA 2 ist dies jedoch weniger deutlich für Befragte der Mobile-Only zu beobachten. So ist hier zwar ein leicht erhöhter Anteil derjenigen, die über einen Hauptschulabschluss verfügen, festzustellen, der eindeutig höhere Anteil gerade von Personen mit niedrigem bis mittlerem Bildungsabschluss zeigt sich hier jedoch nicht. Vielmehr liegt hier der Anteil derjenigen mit gerade höherem Bildungsabschluss (leicht) oberhalb des Gesamtstichprobenanteilswertes.

Dieser höhere Anteil von Befragten mit mittlerem formalen Bildungsabschluss lässt sich indes innerhalb der Subdomain der wahren Dual-User über alle vier Modelle finden. Gleichzeitig ist innerhalb dieser Kategorie sowohl der Anteil derjenigen mit einem niedrigeren als auch einem höheren Bildungsabschluss geringer, als in der gesamten Erhebung. Ein ähnliches Bild lässt sich für die Befragten zeichnen, die im Zuge dieser vier Modelle in die Subdomain Landline-Mainly eingeordnet wurden. Auch hier liegt der Anteil der Befragten mit mittlerem Bildungsabschluss oberhalb desjenigen, der sich über alle Befragte der Erhebung Influenza 2014 einstellt. Allerdings ist der Unterschied zwischen Landline-Mainly und Gesamterhebung bei einem mittleren Bildungsabschluss weniger artikuliert. Gleichzeitig ist hier der Anteil derjenigen, die einen Hauptschulabschluss haben, oberhalb des Wertes, der sich für die Gesamterhebung berechnen lässt. Der Anteil derjenigen, die einen formal höheren Bildungsabschluss aufweisen, ist jedoch ähnlich gering wie unter den wahren Dual-Usern.

Konträr zu diesen Beobachtungen ist die Verteilung der Bildungsvariablen unter den Mobile-Mainly. So ist hier der Anteil derjenigen mit einem formal höheren Bildungsabschluss über alle vier Modelle erheblich höher als unter denjenigen, die ausschließlich über das Mobilfunknetz erreichbar sind und somit auch erheblich höher als der Anteil, der sich über die gesamte Befragung ergibt. Dementsprechend gering fallen auch die Anteile von Befragten mit niedrigem und mittlerem formalen Bildungsabschluss aus.

Interessant ist auch, dass der sich die Verteilung der Bildungsabschlüsse für diejenigen Befragten, die aus einer Subdomain der Mainly-Kategorien stammen, in Modell 4 tendenziell der Verteilung über alle Befragten annähert. Erklärt werden kann dieser Befund damit, dass die Modelle bei der Modellbildung zunehmend den Anteil der Mainlys erhöht. Somit beinhaltet Modell 4 entsprechend mehr Befragte des Überlappungsbereiches, wodurch sich die Variation des Bildungsabschlusses immer mehr an diejenige der gesamten Erhebung annähern kann.

Grundsätzlich zeigt sich hier, dass innerhalb der Erhebung Influenza 2014 diejenigen Befragten, die verstärkt oder ausschließlich ihr Mobiltelefon nutzen, einen höheren formalen Bildungsabschluss aufweisen. Dieser Unterschied zu der Gesamterhebung ist jedoch deutlich artikulierter für diejenigen, die dennoch über einen Festnetzanschluss verfügen.

Ähnlich der Verteilung der Haushaltsgröße der CELLA 2-Erhebung, stammen Befragte aus der Domain der Mobile-Only überwiegend aus Einpersonenhaushalten, während Landline-Only-Befragte häufiger aus Ein- oder Zweipersonenhaushalten stammen. Interessant hingegen ist die Verteilung dieses Charakteristikums unter denjenigen, die den wahren Dual-Usern zugeordnet wurden. So entspricht deren Verteilung gerade nicht der der Gesamterhebung. Vielmehr zeigt sich hier über alle vier Modelle, dass unter den Befragten vermehrt Personen zu finden sind, die aus Mehrpersonenhaushalten stammen. Ebenfalls verwunderlich ist, dass der Anteil der Mobile-Mainly, die aus Ein- oder Zweipersonenhaushalten stammen unterhalb desjenigen der Gesamterhebung ist, während der Anteil für Haushalte mit mindestens drei Personen denjenigen

der Gesamterhebung übersteigt. Eindeutig hingegen scheint die Feststellung, dass Befragte der Erhebung Influenza 2014, die in einem der vier Modelle den Landline-Mainly zugeordnet wurden, überdurchschnittlich häufig aus Zweipersonenhaushalten stammen.

Noch deutlicher als bei der Erhebung CELLA 2 ermöglicht der Datensatz der Erhebung Influenza 2014 Aussagen über die soziodemografischen Merkmale der Only-Haushalte zu treffen.⁷¹ So lässt sich für Befragte der Mobile-Only feststellen, dass diese überwiegend männlich sind, aus Einpersonenhaushalten stammen und jüngeren Alters sind. Abweichend von den Ergebnissen aus CELLA 2 ähnelt deren Verteilung der Bildungsvariablen jedoch derjenigen der Gesamterhebung. Landline-Onlys hingegen sind überwiegend weiblich, höheren Alters und stammen aus Ein- oder Zweipersonenhaushalten. Die Verteilung der Bildungsvariable ist hier abweichend von der der Gesamterhebung und es zeigt sich, dass der Anteil derjenigen mit einem niedrigen Bildungsabschluss deutlich höher ist.

Zumindest bezüglich der Alters- und Geschlechtsverteilung ähneln Befragte aus den beiden Mainly-Subdomains denen der Only-Befragten. So finden sich unter Mobile-Mainlys vermehrt jüngere Männer, während unter den Landline-Mainly deutlich häufiger Frauen höheren Alters zu beobachten sind. Ebenso wie innerhalb der Erhebung CELLA 2 ist deren Verteilung jedoch weniger stark artikuliert. Für Landline-Mainly kann innerhalb dieser Erhebung überdies gezeigt werden, dass diese im Vergleich zur Gesamterhebung häufiger aus Ein- oder Zweipersonenhaushalten stammen. Jedoch liegt hier der Modus (im Gegensatz zu den Landline-Onlys) eindeutig auf Zweipersonenhaushalten. Im Gegensatz zu Mobile-Onlys stammen jedoch auch Mobile-Mainlys vermehrt aus Zweipersonenhaushalten. Wahre Dual-User hingegen scheinen innerhalb dieser Erhebung überdurchschnittlich häufig aus Haushalten mit mindestens drei Personen zu stammen.

In der Verteilung der Bildungsvariablen ähneln Mobile-Mainly wiederum deutlich mehr den Mobile-Onlys. Allerdings ist auch hier, ebenso wie in der CELLA 2-Erhebung festzustellen, dass der Anteil der höher gebildeten Zielpersonen erheblich höher ist. Indes liegt der Anteil der Befragten mit niedrigerem formalen Bildungsabschluss innerhalb der Landline-Mainly oberhalb desjenigen der Gesamterhebung. Unter den wahren Dual-Usern lässt sich dies stattdessen für Befragte mit mittlerem Bildungsabschluss beobachten. Insgesamt zeigt sich an der Bildungsvariablen daher, dass die soziodemografischen Merkmale der Befragten in den Subdomains des Überlappungsbereiches weniger eindeutig derjenigen Verteilungen der Only-Haushalte folgt.

6.3.2 Influenza 2014: Berechnung der Composite-Faktoren und Schätzung

Analog zu dem Vorgehen, das bei der Erhebung CELLA 2 angewendet wurde, werden zu Beginn der Betrachtung der Schätzergebnisse der Erhebung Influenza 2014 unter der Verwendung des Domain-Specific-Composite-Ansatzes die jeweiligen Composite-Faktoren unter dem regulären

⁷¹ Auch wenn die Ergebnisse durchaus vergleichbar sind.

und dem modifizierten Ansatz⁷² in Tabelle 38 dargestellt. Entsprechend der zuvor dargelegten Erkenntnisse lassen sich die Composite-Faktoren λ , ϕ und μ anhand der in Tabelle 34 berechneten Anteile der abgeschlossenen Festnetzinterviews der jeweiligen Subpopulation ablesen.

Tabelle 38 Influenza 2014: Composite-Faktoren nach Modell und Berechnungsgrundlage

C.-Faktor	Modell1	Modell2	Modell3	Modell4
λ	0,829	0,827	0,816	0,779
ϕ	0,629	0,633	0,620	0,622
μ	0,516	0,508	0,544	0,555
λ^*	0,768	0,765	0,746	0,702
ϕ^*	0,555	0,559	0,551	0,548
μ^*	0,457	0,452	0,483	0,497

Tendenziell zeigt sich hier ein ähnliches Bild, wie es bereits in der Berechnung der Composite-Faktoren der CELLA 2-Erhebung berichtet wurde. So sind diese unter Verwendung des regulären Ansatzes innerhalb der Domain der wahren Dual-User nahe dem eigentlichen Anteil der abgeschlossenen Festnetzinterviews. Weiterhin ist in beiden Ansätzen der Composite-Faktor innerhalb der Domain der Landline-Mainly höher als derjenige der wahren Dual-User und niedriger als dieser innerhalb der Domain der Mobile-Mainly. Überdies sind die berechneten Werte unter der Verwendung der effektiven Stichprobengröße, ebenso wie im Rahmen von CELLA 2 beobachtet, geringer als im regulären Ansatz. Dies lässt sich für alle der vier verwendeten Modelle berichten.

Gleichbleibend zum Vorgehen, das im Rahmen dieser Gewichtungsansätze im Zuge der CELLA 2 Erhebung angewendet wurde, erfolgte die entsprechende Schätzung unter der Verwendung dieser Composite-Faktoren für mehrere soziodemografische Variablen. Im Gegensatz zu CELLA 2 wurde hierbei jedoch zum Vergleich der Zensus 2011 als Referenz der amtlichen Statistik herangezogen, da einige der untersuchten Variablen innerhalb der Erhebung nicht in einer geeigneten Form in den Daten des Mikrozensus zu finden gewesen sind. Ebenso wie im Rahmen der Erhebung CELLA 2 werden die Ergebnisse unter der Verwendung des Domain-Specific-Composite-Ansatzes exemplarisch anhand der Bildungsvariable in den Tabellen 39 und 40 dargestellt. Die Schätzergebnisse der Alters- und Geschlechtsverteilung sowie dem Migrationshintergrund und der Haushaltsgröße können im Anhang nachvollzogen werden.

⁷² Der modifizierte Ansatz berücksichtigt die effektiven Stichprobengrößen anstelle der realisierten.

Tabelle 39 Influenza 2014: Höchster allgemeiner Schulabschluss
nach Modell und Methode 1 (in %)

Bildungsabschluss	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	Zensus 2011
Höchstens 7 Jahre Schulbesuch	0,7	0,5	0,5	0,6	0,6	5,0
Haupt-/Volksschulabschluss	15,7	15,1	15,2	15,3	15,2	36,6
Mittlere Reife	33,1	33,1	33,1	33,3	33,4	29,0
Fachhochschulreife	9,6	9,9	9,8	9,8	9,7	8,3
Abitur	40,9	41,4	41,4	41,1	41,1	21,1
Designeffekt	1,43	1,45	1,45	1,45	1,46	●●

Tabelle 40 Influenza 2014: Höchster allgemeiner Schulabschluss
nach Modell und Methode 2 (in %)

Bildungsabschluss	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	Zensus 2011
Höchstens 7 Jahre Schulbesuch	0,7	0,6	0,6	0,6	0,6	5,0
Haupt-/Volksschulabschluss	15,7	15,3	15,4	14,9	15,2	36,6
Mittlere Reife	33,1	33,1	33,1	31,9	32,4	29,0
Fachhochschulreife	9,6	9,8	9,7	10,4	10,0	8,3
Abitur	40,9	41,2	41,2	42,1	41,8	21,1
Designeffekt	1,43	1,43	1,43	1,71	1,57	●●

Ebenso wie bei der Untersuchung der Daten der CELLA 2-Erhebung liegen auch bei der Untersuchung der Ergebnisse der Erhebung Influenza 2014 die geschätzten Werte unter der Verwendung des Single-Frame-Ansatzes und des Domain-Specific-Composite-Ansatzes nahe beieinander. Auch hier zeigt sich jedoch, dass der Single-Frame-Ansatz tendenziell zu bevorzugen ist, da dessen Schätzwerte insgesamt (geringfügig) näher an den Ergebnissen der amtlichen Statistik liegen. Im Rahmen der Domain-Specific-Composite-Schätzung kann i. d. R. beobachtet werden, dass der Ansatz, der zur Berechnung der Composite-Faktoren den effektiven Stichprobenumfang verwendet (Methode 2), sowohl für die Bildungsvariable als auch für die weiteren Variablen, die im Anhang in Kapitel A.4 dargestellt werden, geringfügig bessere Ergebnisse liefert, als unter Verwendung der regulären Stichprobengröße.

Weiterhin liegen die modellbasiert-geschätzten Designeffekte ebenso wie bei der Untersuchung der CELLA 2-Daten mit Ausnahme der letzten beiden Modelle unter der Verwendung der alternativen Methode nahe beieinander. Allerdings ist dieser unter der Verwendung des Single-

Frame-Ansatz geringer. Im Gegensatz zu den vorherigen Untersuchungen lassen sich jedoch im Rahmen der Ergebnisse der Erhebung Influenza 2014 die ersten beiden Modelle zur Einteilung der Dual-User in entsprechende Nutzungskategorien favorisieren. Somit ist davon auszugehen, dass eine konservativere Einteilung nach Nutzungsgewohnheiten, der eine eindeutig ausgesprochene Präferenz für ein bestimmtes Telekommunikationsmedium zugrunde liegt, bei der späteren Anwendung eines solchen Ansatzes zu Ergebnissen führt, die näher an den Angaben der amtlichen Statistik liegen.

Des Weiteren stechen in der Auseinandersetzung die Schätzergebnisse der Altersvariable hervor, die im Anhang abgebildet sind. Hier liefert die alternative Berechnungsmethode des Domain-Specific-Composite-Ansatzes unter der Verwendung der ersten beiden Modelle Ergebnisse, die näher an den Angaben der amtlichen Statistik liegen, als diejenigen, des Single-Frame-Ansatzes. Dennoch ist unter den vorliegenden Ergebnissen auch im Rahmen der Erhebung Influenza 2014 der Single-Frame-Ansatz insgesamt zu bevorzugen.

6.4 Vergleich der Ergebnisse

Auch wenn die Ergebnisse der Schätzungen unter der Verwendung der beiden Datensätze grundsätzlich ähnliche Ergebnisse bezüglich der Abweichung des DSF-Schätzers von den Angaben der amtlichen Statistik hervorbringen und ebenfalls bei beiden Datensätzen der gleiche Gewichtungsansatz bevorzugt wird, ist es eingehend relevant, ob sich die Erhebungen strukturell unterscheiden. Dies liegt darin begründet, dass sowohl die Einteilung der Modelle, als auch der Anteil der Befragten innerhalb der jeweiligen Subdomains unter Umständen stark voneinander abweichen kann. Dies kann unter Umständen auch durch die in Kapitel 3.4 dargestellten Ursachen von Nonresponse geschehen. Es ist daher bspw. nicht auszuschließen, dass beide Erhebungen in einem unterschiedlichen Ausmaß durch Topic Salience betroffen sind und sich deshalb soziodemografische Merkmale in ihrer Verteilung systematisch unterscheiden.

Zuerst fällt in diesem Kontext die Zusammensetzung der beiden Erhebungen nach Anteil der Befragten über eine der beiden Stichproben auf. So wurde im Zuge der Erhebung CELLA 2 eine Aufteilung gewählt, die nahezu gleich viele Befragte über beide Telekommunikationsmedien erhoben hat. Bei Influenza 2014 stammten lediglich ca. 36 % aller Befragten aus der Mobilfunkstichprobe. Weiterhin unterscheidet sich die Größe der Erhebung deutlich. Innerhalb der Erhebung CELLA 2 wurde annähernd die dreifache Anzahl von Personen befragt. Da weiterhin die Möglichkeit besteht, dass sich beide Erhebungen in der Zusammensetzung der einzelnen Befragten unterscheiden, werden in Tabelle 41 nochmals die Anteile der Merkmale Alter, Geschlecht und Bildungsabschluss gegenübergestellt. Diesbezüglich ist anzumerken, dass sich beide Erhebungen in der Definition der Zielpopulation unterscheiden. So wurden im Rahmen von CELLA 2 Personen befragt, die mindestens 16 Jahre alt sind, während innerhalb von Influenza 2014 die Zielpopu-

lation bei auf Personen, die mindestens 18 Jahre sind, festgelegt wurde. Weiterhin wurden die Bildungs- und Altersvariable unterschiedlich abgefragt. Um eine adäquate Gegenüberstellung zu gewährleisten, wurden die Ausprägungen der jeweiligen Variablen entsprechend angepasst. Aufgrund dessen ist jedoch alleine dadurch schon mit Unterschieden zu rechnen.⁷³

Tabelle 41 Vergleich soziodemografischer Merkmale der Erhebungen CELLA 2 und Influenza 2014 (in %)

Merkmal		Influenza 2014	CELLA 2
Geschlecht	Männlich	46	51,8
	Weiblich	54	48,2
Alter	16/18* - 39	29	42,8
	40 - 49	18,9	21,7
	50 - 59	21,2	17,4
	60 - 69	16,3	10,2
	70 +	14,6	7,9
Bildungsabschluss	Kein Abschluss/max. 7 Jahre	0,7	3,6
	Haupt-/Volksschulabschluss	15,6	18,3
	Mittlere Reife	33,1	32,9
	FH/HS-Reife	50,5	43,7
AAPOR RR3		17	13

Anhand der hier angeführten Tabelle wird deutlich, dass sich die Zusammensetzung der Befragten innerhalb der beiden Erhebungen durchaus unterscheidet. Wurden innerhalb von CELLA 2 bspw. mehr Männer befragt, so zeigt sich, dass im Rahmen der Erhebung Influenza 2014 mehr Frauen an der Erhebung teilgenommen haben. Auch bezüglich des Alters sind deutliche Unterschiede zu bemerken, die nicht alleine auf die abweichende Definition der Zielpopulation zurückzuführen sind. So wurde innerhalb der Erhebung CELLA 2 ein deutlich geringeres Durchschnittsalter erreicht. Grundsätzlich zeigt sich hier, dass innerhalb der Erhebung Influenza 2014 ein deutlich höherer Anteil an Personen in mittlerem bis höheren Alter befragt wurde. Auch innerhalb des Bildungsabschlusses lassen sich Unterschiede feststellen. Der Anteil der Befragten, die angaben, keinen formalen Bildungsabschluss vorzuweisen, ist innerhalb der CELLA 2-Erhebung erheblich höher als der (sehr geringe) Anteil derjenigen, die eine vergleichbare Angabe innerhalb von Influenza 2014 gemacht haben. Ein möglicher Grund hierfür kann in der zugrunde liegenden Zielpopulation liegen. So ist es gerade unter den Befragten, die zwischen

73 Da in CELLA 2 noch die Antwortmöglichkeit "Anderer Schulabschluss" (1,7 %) gegeben war, der sich keiner der Antwortkategorien aus Influenza 2014 zuordnen lässt, wurde dieser hier ausgelassen. Daher summieren sich die angegebenen Anteile bei der Betrachtung von CELLA 2 nicht auf 100 %.

16 und 18 Jahren alt sind, durchaus plausibel, dass ein nicht zu vernachlässigender Teil dieser Subpopulation noch keinen Abschluss besitzt. Aber auch unter der Auslassung der zugrunde liegenden Zielpopulationsdifferenzen zeigt sich, dass bei CELLA 2 ein höherer Anteil von Personen mit niedrigem formalen Bildungsabschluss befragt wurde, während ein höherer Anteil mit hohem formalen Bildungsabschluss innerhalb der Erhebung Influenza 2014 zu finden ist. Jedoch gilt für beide Erhebungen festzuhalten, dass, gemessen an der Verteilung der amtlichen Statistik, ein überdurchschnittlich hoher Anteil an Personen mit hohem Bildungsabschluss teilgenommen hat. Zuletzt lässt sich innerhalb der Responserate 3, die jeweils nach AAPOR-Standard berechnet wurde, eine Differenz von vier Prozentpunkten beobachten. Diese Differenz lässt sich jedoch hauptsächlich durch eine (deutlich) höhere Responserate im Festnetz innerhalb der Erhebung Influenza 2014 begründen. Die Responserate in beiden Mobilfunkstichproben ist annähernd gleich gering. Überdies sollte auch in diesem Kontext nicht vernachlässigt werden, dass eine Responserate per se nur einer von vielen Qualitätsindikatoren einer Erhebung ist.⁷⁴

Tendenziell stimmt bei beiden Erhebungen zumindest die "Richtung" der Verteilung der beiden Merkmale Alter und Geschlecht innerhalb der Only-Domains überein. Für beide Erhebungen kann demnach festgehalten werden, dass Landline-Onlys häufiger weiblich und in einem höheren Alter sind, während Mobile-Only eher Männer jüngeren Alters sind. Weiterhin zeigt sich bei beiden Erhebungen, dass Mobile-Onlys vermehrt in Einpersonenhaushalten leben, während Landline-Only häufiger in Ein- oder Zweipersonenhaushalten zu finden sind. Bezüglich der Verteilung der Bildungsvariablen hingegen kann lediglich für den Bereich der Landline-Only eine vergleichbare Tendenz beschrieben werden. Für diese sind in beiden Erhebungen vermehrt Personen mit niedrigem Bildungsabschluss vorzufinden. Innerhalb der Mobile-Only hingegen weichen die Ergebnisse voneinander ab. So befinden sich in der Erhebung CELLA 2 vermehrt Personen mit niedrigem und mittlerem Bildungsabschluss in dieser Kategorie, während innerhalb von Influenza 2014 zwar ein Anteil von Personen mit niedrigem Bildungsabschluss gefunden werden konnte, der oberhalb des Anteils der Gesamterhebung lag, jedoch auch ein höherer Anteil von Personen mit hohem formalen Bildungsabschluss festgestellt werden konnte.

In beiden Erhebungen konnte oftmals beobachtet werden, dass die Verteilung der soziodemografischen Charakteristika der Befragten aus den Mainly-Domains mit der entsprechenden Only-Domain vergleichbar, jedoch weniger stark artikuliert ist. Dies betrifft unter anderem die beiden Merkmale Geschlecht und Alter. So zeigt sich in beiden Erhebungen, dass Landline-Mainly tendenziell häufiger Frauen höheren Alters und Mobile-Mainly eher Männer jüngeren Alters sind. Als mögliche Ursache dieser Ergebnisse kann unter Umständen die Begründung, die bereits in Kapitel 2.3 für jüngere, mobilere Zielpersonen angeführt werden. So ist es nicht auszuschließen, dass auch unter den Mobile-Mainlys tendenziell eine höhere Mobilität besteht,

74 Weitere Qualitätsindikatoren können bspw. der Designeffekt, die Präzision der Schätzer sowie der damit verbundene Mean Squared Error (MSE) sein.

weswegen diese gerade nur mühsam über einen Festnetzanschluss kontaktiert werden. Gestützt wird diese Vermutung auch durch Abbildung 12, in der gezeigt wurde, dass Mobile-Mainlys durchschnittlich eine höhere Anzahl an Kontaktversuchen über das Festnetz benötigt haben.

Auch bei der Haushaltsgröße der Mainly-Befragten lassen sich in beiden Erhebungen ähnliche Ergebnisse festhalten. So scheinen Mobile-Mainly tendenziell eher aus Einpersonenhaushalten zu stammen, während Landline-Mainly häufiger Zweipersonenhaushalten zugehörig sind. Bezugnehmend auf die Verteilung der Bildungsvariablen konnte in beiden Erhebungen keine direkte Übereinstimmung gefunden werden. In der Erhebung CELLA 2 gehörten bspw. überdurchschnittlich viele Befragte der Landline-Mainly entweder zu den formal höher oder den formal niedrig gebildeten. Während der Anteil der formal niedrig gebildeten unter den Mobile-Mainly unterdurchschnittlich ist, der der Befragten mit mittlerem Bildungsabschluss oberhalb des Wertes der Gesamterhebung war und der Anteil der höher gebildeten mit demjenigen, der sich über alle Befragte ergibt, vergleichbar war. Landline-Mainlys der Erhebung Influenza 2014 hingegen sind formal eher niedrig gebildet. Mobile-Mainly dieser Erhebung weisen hingegen einen deutlich höheren Anteil an Befragten mit höherem Bildungsabschluss auf. In beiden Erhebungen ist die Verteilung beider Mainly-Kategorien dieser Variablen nur eingeschränkt mit derjenigen der Only-Domains vergleichbar.

Für wahre Dual-User konnte in beiden Erhebungen oftmals beobachtet werden, dass deren Verteilung mit der der gesamten Erhebung vergleichbar ist. Als mögliche Ursache dafür kann einerseits die Größe der jeweiligen Kategorien in allen Modellen, die wahre Dual-User enthalten, gesehen werden. Andererseits kann die Vermutung, dass die fehlende Präferenz für ein bestimmtes Medium gleich wahrscheinlich über einige soziodemografische Merkmale verteilt ist, nicht ausgeschlossen werden. Für ersteres Argument spricht die deutlich höhere Anzahl von Befragten innerhalb der Modelle in dieser Subdomain, die dazu führen kann, dass Ausreißer einen geringeren Einfluss auf die Verteilung haben.

Merkmale, bei denen dies nicht zutrifft, sind Bildungsabschluss (in beiden Erhebungen) sowie Alter und die Haushaltsgröße in Influenza 2014. Die Haushaltsgröße der wahren Dual-User weist bspw. einen deutlich höheren Anteil Personen in Mehrpersonenhaushalten auf, die häufiger mittleren Alters sind. Bezüglich des Bildungsabschlusses unterscheidet sich die Verteilung in beiden Erhebungen voneinander und von derjenigen der Gesamterhebung. Innerhalb von Influenza 2014 war hier ein überdurchschnittlich hoher Anteil von Personen mit mittlerem Bildungsabschluss zu finden. Bei CELLA 2 hingegen war der Anteil der formal höher gebildeten Personen geringfügig höher. Dies geschah zulasten der formal niedrigeren Bildungsabschlüssen.

Auch wenn die Ergebnisse der Betrachtung von Befragten innerhalb des Überlappungsreiches nicht vollständig für beide Erhebungen übereinstimmen, kann immerhin eine Ausrichtung der Verteilung der Mainly-Befragten festgehalten werden, die oftmals der der Onlys des jeweiligen Telekommunikationsmediums folgt. Häufig ist diese jedoch weniger stark artikuliert.

Die fehlende Übereinstimmung innerhalb der beiden Erhebungen kann auf mehrere Ursachen zurückgeführt werden. Einmal kann dies, wie zuvor bereits gezeigt, an den grundsätzlichen Unterschieden in den Verteilungen der soziodemografischen Charakteristika liegen. Weiterhin ist hier anzuführen, dass die Einteilung in CELLA 2 lediglich anhand von Proxys erfolgte, deren Eignung noch folgend genauer zu untersuchen ist.

Ein Vergleich der Ergebnisse, die aus der Anwendung des Domain-Specific-Composite-Ansatzes auf die Erhebungsdaten von CELLA 2 und Influenza 2014 gewonnen wurden, zeigt, dass diese bezüglich ihrer Präzision sowie ihrer Erwartungstreue grundsätzlich zu ähnlichen Resultaten führen. Besonders hinsichtlich des Vergleiches mit dem Single-Frame-Schätzer führen beide Untersuchungen zu dem Resultat, dass dieser, aufgrund eines geringeren modellbasierten Designeffektes und einer geringeren Abweichung der Punktschätzer von den Angaben der amtlichen Statistik, zu bevorzugen ist. So liefert der SF-Schätzer unter der Verwendung der Daten aus CELLA 2 grundsätzlich Ergebnisse, die näher an der amtlichen Statistik liegen, als diejenigen, die sich unter der Verwendung von einem der vier Modelle der Einteilung nach Nutzungsgewohnheiten unter der Verwendung des DSC-Ansatzes ergeben haben. Auch für die Untersuchung, die unter der Verwendung der Erhebung Influenza 2014 vorgenommen wurde, lässt sich dies für einen Großteil der Schätzergebnisse bestätigen. Die einzige Ausnahme bildet hier die Altersvariable unter der Verwendung der effektiven Stichprobengröße. Bei dieser scheinen die Ergebnisse unter der Verwendung der ersten beiden Vorgehen zur Einteilung nach Nutzungsgewohnheiten näher an der amtlichen Statistik zu liegen, als dies unter der Punktschätzung nach SF-Ansatz zu beobachten ist. Eine mögliche Ursache dieser Ergebnisse könnte darin liegen, dass die Altersstruktur in der Stichprobe ungewichtet bereits auffällig scheint. So wurde in Tabelle 41 bereits dargestellt, dass im Zuge dieser Erhebung gerade Personen höheren Alters deutlich vermehrt befragt wurden. Im Zuge der Schätzung mittels Designgewichtung scheint dieser Anteil zu sinken. Hier fällt jedoch auf, dass die entsprechenden Anteile nun unterhalb der Verteilung der amtlichen Statistik sind. Mögliche Ursachen hierfür sind die Anzahl der Rufnummern, unter denen ein bestimmter Haushalt erreicht werden kann sowie die Haushaltsgröße. Da gerade die effektive Haushaltsgröße vermeintlich geringer bei Personen niedrigen Alters ausfällt, erhalten diese unter Umständen ein geringeres Gewicht aufgrund einer höheren Auswahlwahrscheinlichkeit im Festnetz. Da dies aufgrund der Gestaltung des Single-Frame-Ansatzes in alle Inklusionswahrscheinlichkeiten eingeht, während im DSC-Ansatz die Haushaltsgröße lediglich für Befragte der Festnetzstichprobe relevant ist, kann sich dies gerade in der Schätzung des Alters bemerkbar machen.

Weiterhin liefert der SF-Ansatz sowohl in der Untersuchung unter der Verwendung von synthetischen Daten als auch in der Anwendung im Rahmen der beiden Erhebungen einen merklich geringeren modellbasierten Designeffekt. Dies lässt sich unter der Verwendung eines Großteils der gebildeten Modelle zur Einteilung nach Nutzungsgewohnheiten sowie unter der

Verwendung der regulären und der effektiven Stichprobengröße zur Berechnung der Composite-Faktoren beobachten. Die einzige Ausnahme bildet dabei die Verwendung der ersten beiden Modelle zur Einteilung nach Nutzungsgewohnheiten der Erhebung Influenza 2014 für beide Vorgehensweisen zur Bestimmung der Composite-Faktoren. Unter der Verwendung der regulären Stichprobengröße sind die Designeffekte dieser beiden Modelle nur unwesentlich höher als unter der Schätzung durch den SF-Ansatz. Bei der Verwendung der effektiven Stichprobengröße sind diese sogar annähernd identisch.

Gerade diese beiden Vorgehensweisen zur Einteilung der Nutzungsgewohnheiten stechen auch deshalb hervor, da diese im Rahmen der untersuchten Variablen der Erhebung Influenza 2014 eindeutig zu favorisieren sind. Im Gegensatz hierzu konnte unter der Verwendung unterschiedlicher Vorgehensweisen zur Einteilung von Mainly-Haushalten in der Anwendung auf die Daten der CELLA 2-Erhebung kein Modell gefunden werden, das eindeutig zu favorisieren ist.

Da die Einteilung nach den unterschiedlichen Verwendungskategorien der Dual-User zum Zeitpunkt der CELLA 2-Erhebung noch nicht geplant war, wurde in der vorliegenden Untersuchung versucht, eine solche Einteilung anhand von Proxys durchzuführen, die als geeignet erscheinen, um Aussagen über die Nutzungsgewohnheiten von Telekommunikationsmedien der Dual-User zu treffen. Um zu überprüfen, ob sich eine Einteilung nach solchen Proxys tatsächlich eignet, wurde das Item *M4*, das die Empfangsbereitschaft des Mobiltelefons eines Dual-Users abfragt, in der Erhebung Influenza 2014 mit aufgenommen. Zu überprüfen gilt dabei, ob die Einteilung nach der tatsächlichen Erreichbarkeit über ein bestimmtes Telekommunikationsmedium mit der Einteilung nach einem solchen Proxy übereinstimmt. Da für diese Überprüfung jedoch die Information fehlt, wie oft das Mobiltelefon auch mitgeführt wird, kann ein direkter Vergleich nach den beiden Vorgehensweisen an dieser Stelle nicht erfolgen. Dennoch ist die Überprüfung der Empfangsbereitschaft eines Mobiltelefons und der Verwendungskategorie von großem Interesse. Überdies gestattet ein solcher Vergleich zumindest einen Einblick, ob die in der Auseinandersetzung mit den Daten der Erhebung CELLA 2 gewählte Einteilung geeignet ist. Daher werden in Tabelle 42 die Einteilung nach Modell 1 innerhalb der Erhebung Influenza 2014 sowie die Empfangsbereitschaft gegenübergestellt.⁷⁵

⁷⁵ Aufgrund von fehlenden Werten bei der Beantwortung von *M4* innerhalb der Mobile-Only Domain ist die Gesamtzahl der Beobachtungen um vier Elemente reduziert.

Tabelle 42 Influenza 2014: Vergleich der Einteilung nach Modell 1 und der Empfangsbereitschaft des Mobiltelefons (absolut)

M4/Modell 1	Landline-Only	Landline-Mainly	Dual User	Mobile-Mainly	Mobile-Only
immer	0	64	359	135	38
täglich	0	19	106	25	19
mehrmals in der Woche	0	16	40	1	1
seltener	0	53	28	0	1
tnz	102	0	0	0	0

Betrachtet man sowohl die Domain der Mobile-Only als auch diejenige der Mobile-Mainly, so zeigt sich, dass ein Großteil der Befragten ihr Mobiltelefon immer, oder zumindest täglich eingeschaltet hat. Dies entspricht auch der intuitiven Annahme, die sich ebenfalls im Vorgehen zur Einteilung im Rahmen der CELLA 2 Daten wiederfindet. Diese bestand darin, dass eine Person, die dominant einen Mobilfunkanschluss verwendet, auch dauerhaft oder zumindest die meiste Zeit über dieses erreichbar ist. Jedoch findet sich in der Betrachtung der hier vorliegenden Gegenüberstellung auch die bereits geäußerte Vermutung wieder, dass es durchaus Landline-Mainlys gibt, deren Mobiltelefon durchgehend empfangsbereit ist. Verwunderlich ist dabei jedoch, dass die Anzahl der Landline-Mainly-Haushalte, bei denen dies zutrifft, nahezu vergleichbar ist mit derjenigen, deren Mobiltelefon nicht täglich empfangsbereit ist. Mögliche Ursache dessen könnte in dem vermehrten Aufkommen von Smartphones liegen, da sich durch diese die Funktion des Mobiltelefons vom reinen Telekommunikationsmedium hin zum Multimedia-Gerät geändert haben kann. So ist nicht auszuschließen, dass Personen, die besser über ihren Festnetzanschluss erreicht werden können, dennoch andere Funktionen ihres Mobilfunkgerätes, wie bspw. mobiles Internet oder die Funktion als Wecker, dauerhaft nutzen und deshalb ein Mobilfunkgerät dauerhaft empfangsbereit ist. Eine weitere mögliche Erklärung ist die Zunahme der Vertrautheit im Umgang eines Mobiltelefons. So besteht die Möglichkeit, dass zwischen den beiden Erhebungen CELLA 2 und Influenza 2014 durch die zunehmende Verbreitung (und Nutzung) von Mobiltelefonen auch eine höhere Vertrautheit geschaffen wurde und gerade Probleme, wie die Akkulaufzeiten dadurch weniger häufig anfallen. Um einen besseren Eindruck über die Verteilung der Variable M4 nach Domain zu erhalten, stellt Tabelle 43 die bedingten Anteile der Empfangsbereitschaft nach Domain in Modell 1 dar.

Tabelle 43 Influenza 2014: Empfangsbereitschaft nach Domain in Modell 1 (in %)

M4/Modell 1	Landline-Only	Landline-Mainly	Dual User	Mobile-Mainly	Mobile-Only
immer	0	42,1	67,4	83,9	64,4
täglich	0	12,5	19,9	15,5	32,2
mehrmals i. d. Woche	0	10,5	7,5	0,6	1,7
seltener	0	34,9	5,3	0	1,7
tnz	100	0	0	0	0

Grundsätzlich lässt sich auch hier die in Kapitel 6.2.1 geäußerte Annahme, dass diejenigen, die vorwiegend oder ausschließlich über einen Mobilfunkanschluss erreichbar sind, auch eine höhere Empfangsbereitschaft des Mobilfunkanschlusses aufweisen, feststellen. So weisen ca. 96,6 % aller Mobile-Onlys sowie ca. 99,4 % aller Mobile-Mainlys einen Mobilfunkanschluss auf, der zumindest auf täglicher Basis empfangsbereit ist. Dieser Anteil ist überdies deutlich höher, als die ca. 64,6 % der Befragten aus der Landline-Mainly-Domain, deren Mobilfunkanschluss zumindest täglich empfangsbereit ist. Dennoch scheint auch dieser Anteil vergleichsweise hoch, gerade unter der Berücksichtigung, dass dadurch ca. 8,24 % (83 Personen) der Landline-Mainlys, die durch die Einteilung anhand der Vorgehensweise, die für die Influenza 2014-Daten gewählt wurde, unter der vorherigen Einteilung, die durch Proxys im Rahmen der CELLA 2-Erhebung angewendet wurde, für die ersten drei Modelle einer anderen Domain zugewiesen worden wären. Eine Einteilung, die alleine auf der Basis dieser Proxy-Variable beruht, scheint demnach problematisch. Jedoch gilt es festzuhalten, dass nicht auszuschließen ist, dass sich das Nutzungsverhalten von Mobiltelefonen in dem Zeitraum zwischen der Erhebung CELLA 2 und der Erhebung Influenza 2014 geändert haben kann. So kann vermutet werden, dass gerade durch die schnelle Verbreitung von Smartphones Mobilfunkgeräte grundsätzlich häufiger eingeschaltet sind. Eine solche Vermutung legt nahe, dass Mobiltelefone aufgrund von weiterführenden Anwendungen wie bspw. der Nutzung von mobilem Internet sowie mobilen Applikationen häufiger verwendet werden, ohne dass diese zur Kommunikation genutzt werden. Gerade diese neueren Verwendungsarten sind deshalb in der aktuelleren Erhebung Influenza 2014 eher zu beobachten. Jedoch stellt dies lediglich eine Mutmaßung dar. Wie sich eine solche Änderung in der Nutzung von Mobiltelefonen auf die Erreichbarkeit einer Zielperson und die grundsätzliche Bereitschaft, über ein solches Telekommunikationsmedium an einer Befragung teilzunehmen, auswirkt, bleibt ungeklärt. Es ist jedoch anzunehmen, dass eine Einteilung anhand der in CELLA 2 verwendeten Proxys zum heutigen Zeitpunkt aufgrund der geänderten Nutzungsgewohnheiten und dem Aufkommen von Smartphones, die den Status des Mobiltelefons als reines Telekommunikationsmedium verändert haben und weitere Einsatzmöglichkeiten bieten, nicht

mehr zeitgemäß ist. Dies zeigt sich alleine in den Beobachtungen aus Tabelle 42.

Ein χ^2 -Test auf Unabhängigkeit der beiden Variablen über den gesamten Datensatz und alle fünf Domains liefert einen Wert von $\chi^2 = 1024,1$ (bzw. $p\text{-Wert} \leq 2,2 \cdot 10^{-16}$).⁷⁶ Jedoch wird dieser Wert stark durch die Ausprägungen der Only-Einheiten beeinflusst.⁷⁷ Die Verwendung dieser Testergebnisse würde daher evtl. zu fehlgeleiteten Schlüssen führen. Aufgrund dessen wurde nochmals auf Unabhängigkeit der beiden Variablen innerhalb des Überlappungsbereiches der Ziehungsrahmen getestet. Dabei ergibt sich ein Wert von $\chi^2 = 152,6$ bei ähnlichem $p\text{-Wert}$. Somit lässt sich durchaus eine Abhängigkeit der beiden Variablen feststellen, was sich auch in den Darstellungen der beiden Tabellen 42 und 43 beobachten lässt. So ist trotz der beschriebenen Ergebnisse der Anteil von Personen mit mindestens täglich eingeschaltetem Mobilfunkgerät noch immer höher für diejenigen in den Domains, die ausschließlich oder bevorzugt über einen Mobilfunkanschluss erreichbar sind.

Auch wenn sich grundsätzlich in der Auseinandersetzung der Schätzergebnisse des Domain-Specific-Composite-Ansatzes zeigt, dass dessen Verwendung i. d. R. Ergebnisse hervorbringt, die dem Single-Frame-Ansatz dahin gehend unterlegen sind, dass sie (wenn auch nur geringfügig) weniger genau sind und einen höheren Designeffekt aufweisen, sollte das Konzept, den Überlappungsbereich in unterschiedliche Domains nach Nutzungsgewohnheiten einzuteilen und dies auch in der Gewichtung zu berücksichtigen, nicht vorschnell verworfen werden. So zeigt sich u. a. in der Betrachtung der Kontaktergebnisse, dass es durchaus Unterschiede nach verschiedenen Domains gibt. So kann beobachtet werden, dass diejenigen, die ein Telekommunikationsmedium dominant verwenden, auch durchschnittlich weniger Kontaktversuche bis zu einem erfolgreichen Interviewabschluss benötigen. Um daher den Ansatz, die Nutzungsgewohnheiten und die Erreichbarkeit von Einheiten innerhalb des Überlappungsbereiches im Rahmen eines Gewichtungskonzepts nicht vorschnell auszuschließen, stellt das folgende Kapitel alternative Vorgehensweisen zur Berücksichtigung der Nutzungsgewohnheiten vor.

⁷⁶ Die vorliegenden Tests beruhen auf der Häufigkeitsverteilung als absolute Werte.

⁷⁷ Da Landline-Onlys kein Mobiltelefon besitzen, kann dieses auch nicht empfangsbereit sein.

7 Der Domain-Specific-Single-Frame-Ansatz

Untersuchungsgegenstand des Kapitels ist ein alternatives Vorgehen zur Verwendung der unterschiedlichen Nutzungskategorien der Zielpersonen im Überlappungsbereich einer Dual-Frame-Erhebung im Zuge der Gewichtung. Die Einteilung der Zielpersonen erfolgt analog zu dem Vorgehen, das in Kapitel 5.2 genauer erläutert wurde. Ziel des Kapitels ist es zu untersuchen, ob die Informationen über Nutzungsgewohnheiten in einem zu Kapitel 5.3 abweichenden Gewichtungsansatz zu einer Verbesserung der Schätzung beitragen können. Die Untersuchung des weiterführenden Ansatzes erfolgt an den bereits beschriebenen Datengrundlagen der beiden Erhebungen CELLA 2 und Influenza 2014 sowie unter der Zuhilfenahme des in Kapitel 6.1 beschriebenen, synthetischen Datensatzes.

7.1 Alternative Gewichtungsverfahren zur Berücksichtigung der Nutzungsgewohnheiten

Die Ergebnisse der Auseinandersetzung mit der Einteilung nach Nutzungsgewohnheiten sowie der Anwendung des Domain-Specific-Composite-Ansatzes haben zu Ergebnissen geführt, die durchaus unterschiedlich geartet sind. So zeigt bspw. die Betrachtung der Anzahl der Kontaktversuche, wie sie in der Untersuchung der Erhebungsdaten von Influenza 2014 erfolgt ist, dass sich nach der Einteilung in die Subdomains der Landline-Mainly (u_2), Dual-User (u_3) sowie der Mobile-Mainly (u_4) durchaus Unterschiede feststellen lassen. Diese stellen sich ähnlich der in Kapitel 5.2 geäußerten Vermutungen ein. So werden durchschnittlich weniger Kontaktversuche benötigt, um Mobile-Mainlys via Mobilfunk zu kontaktieren und ebenso werden weniger Anwahlversuche benötigt, um ein Interview mit einem Landline-Mainly über das Festnetz abzuschließen. Als Referenz dieser Betrachtung galt jeweils die durchschnittliche sowie die mittlere Anzahl von Kontaktversuchen der gesamten Stichprobe über ein bestimmtes Telekommunikationsmedium. Dahin gehend erscheint eine Einteilung nach Subdomains durchaus sinnvoll und angebracht.

Auf der anderen Seite zeigt die Betrachtung des Domain-Specific-Composite-Ansatzes, dass, unabhängig vom Vorgehen zur Einteilung der Einheiten im Überlappungsbereich der beiden Stichproben in die jeweiligen Subdomains, die Schätzergebnisse unter der Verwendung des Single-Frame-Ansatzes i. d. R. näher an den Angaben der amtlichen Statistik liegen. Weiterhin weist der SF-Ansatz überwiegend einen geringen Designeffekt⁷⁸ auf. Die einzigen Ausnahmen bilden hierbei die ersten beiden Modelle zur Einteilung in die jeweiligen Subdomains im Zuge der Schätzung auf der Basis der Influenza 2014-Daten. Ausgehend von diesen Ergebnissen scheint die Berücksichtigung der Nutzungsgewohnheiten, die durch die jeweiligen Subdomains wieder-

⁷⁸ Für die modellbasierte Schätzung des Designeffektes siehe Kapitel 6.

gegeben werden, im Rahmen der Gewichtung durch den DSC-Ansatz weniger empfehlenswert zu sein.

Somit stellen sich in der Betrachtung unterschiedliche Resultate ein. Bezug nehmend auf die Kontaktergebnisse scheint es durchaus sinnvoll zu sein, eine solche Einteilung vorzunehmen. Entgegengesetzt dazu scheint die Anwendung des DSC-Ansatzes i. d. R. nicht dazu zu führen, dass die Schätzergebnisse näher an der tatsächlichen Verteilung der Zielpopulation liegen. Intuitiv liegt die Vermutung nahe, dass ein Schätzer, der mehr Informationen über die zugrunde liegenden Stichprobenelemente miteinbezieht, auch Ergebnisse hervorbringt, die präziser und/oder weniger verzerrt sind. Im Fall des DSC-Ansatzes scheint dies jedoch nicht zuzutreffen. Daher ist anzunehmen, dass evtl. weitere Möglichkeiten bestehen, durch die die Information der Nutzungsgewohnheiten in einem Gewichtungsansatz verwendet werden kann, der dann tatsächlich weniger verzerrte Punktschätzer und/oder eine höhere Präzision verspricht. Daher wird folgend ein weiteres Gewichtungskonzept, der *Domain-Specific-Single-Frame-Ansatz (DSF)*, diskutiert. Dieses stellt eine Kombination der Einteilung nach Nutzungsgewohnheiten sowie dem Single-Frame-Ansatz dar.

In Anbetracht der Untersuchung der Kontaktergebnisse der Erhebungsdaten aus Influenza 2014 scheint es durchaus sinnvoll zu sein, die jeweiligen Erhebungseinheiten nach Nutzungsgewohnheiten in verschiedenen Subdomains zu gliedern. Wie bereits in Kapitel 5.2 erläutert, kann angenommen werden, dass die Wahrscheinlichkeit, dass eine Einheit, die über ein bestimmtes Telekommunikationsmedium erfolgreich kontaktiert wird, in einem direkten Zusammenhang damit steht, ob diese Einheit das entsprechende Telekommunikationsmedium bevorzugt verwendet. Diese Präferenz, die sich in der Wahrscheinlichkeit widerspiegelt, dass eine Zielperson i , die über beide Telekommunikationsmedien verfügt, eher das bevorzugte Medium beantwortet, wird künftig mit ξ_i bezeichnet. Für ξ_i wird angenommen, dass $0 \leq \xi_i \leq 1$ zutrifft. Die Wahrscheinlichkeit, dass eine Einheit in diesem Kontext, bei einer gegebenen Präferenz, gerade das nicht präferierte Telekommunikationsmedium beantwortet, wenn sie über dieses kontaktiert wird, lässt sich durch $(1 - \xi_i)$ ausdrücken. Aufgrund des konvexen Charakters dieser Annahme und um die künftigen Darstellungen zu vereinfachen, beschreibt ξ_i^F die Wahrscheinlichkeit einer Einheit i , eine Befragung zu beantworten, wenn diese über das Festnetz erfolgt. Demnach sollte eine Einheit aus der Subdomain Landline-Mainly für ξ_i^F einen Wert aufweisen, der oberhalb des Erwartungswertes der gesamten Zielpopulation liegt und eine Mobile-Mainly-Einheit einen Wert unterhalb dieses Wertes.

Ähnlich der bereits in Kapitel 5.2 geäußerten Annahmen, wird bei dieser Vorgehensweise unterstellt, dass die Wahrscheinlichkeit dafür, dass eine Zielperson in der Stichprobe enthalten ist, nicht nur von den Parametern abhängt, wie sie in den Gleichungen 3.4 und 3.5 angegeben wurden, sondern auch von der Zugehörigkeit zu einer der Subdomains, bzw. von der individuellen Präferenz der Person i für die Telekommunikationsmedien Festnetz und Mobilfunk. Demnach

kann die Wahrscheinlichkeit, dass eine Einheit i in die Stichprobe gelangt ist, die auch die Nutzungsgewohnheiten berücksichtigt, für das Festnetz durch

$$\pi_i^{F,DSF} = \xi_i^F \cdot \frac{m^F}{M^F} \cdot \frac{k_i^F}{z_i} = \xi_i^F \cdot \pi_i^F \quad (7.1)$$

sowie für den Mobilfunk durch

$$\pi_i^{C,DSF} = (1 - \xi_i^F) \cdot \frac{m^C}{M^C} \cdot k_i^C = (1 - \xi_i^F) \cdot \pi_i^C \quad (7.2)$$

angegeben werden.

Unter den gleichen Annahmen, die auch im Zuge von Gleichung 3.6 bereits bemüht wurden, entspricht die gemeinsame Inklusionswahrscheinlichkeit einer Zielperson i annähernd

$$\pi_i^{DSF} \approx \xi_i^F \cdot \frac{m^F}{M^F} \cdot \frac{k_i^F}{z_i} + (1 - \xi_i^F) \cdot \frac{m^C}{M^C} \cdot k_i^C = \xi_i^F \cdot \pi_i^F + (1 - \xi_i^F) \cdot \pi_i^C = \pi_i^{F,DSF} + \pi_i^{C,DSF}. \quad (7.3)$$

Auch hier wird unterstellt, dass der Term $\pi_i^{*F \cap C} = \xi_i^F \cdot \pi_i^F \cdot ((1 - \xi_i^F) \cdot \pi_i^C)$ vernachlässigbar ist.

Weiterhin wird vereinfachend angenommen, dass ξ_i^F , also die von der Präferenz des Mediums Festnetz ausgehende Wahrscheinlichkeit, dieses zu beantworten, nicht als Kontinuum interpersonell variiert, sondern lediglich durch die Zugehörigkeit einer Einheit zu einer der fünf Subdomains determiniert wird. Diese Annahme unterstellt, dass alle Einheiten einer bestimmten Subdomain die gleiche präferenzabhängige Wahrscheinlichkeit aufweisen, dass eine Kontaktaufnahme über das Festnetz auch beantwortet wird. Die Wahrscheinlichkeit ein Festnetzgerät bei einem Kontaktversuch zu beantworten (ξ_i^F) wird somit für jede Zielperson i innerhalb einer Subdomain l als konstant angenommen. Mit Ausnahme der Only-Befragten stellt dies somit eine starke Vereinfachung dar. Jedoch wäre die individuelle Messung und Bestimmung dieser Wahrscheinlichkeit im Zuge von Befragungen vermutlich zu aufwendig, um im Anschluss eine praktikable Gewichtungstrategie zu entwickeln. Die Inklusionswahrscheinlichkeit einer Einheit i entspricht daher

$$\pi_{i,l}^{DSF} \approx \xi_l^F \cdot \pi_i^F + (1 - \xi_l^F) \cdot \pi_i^C.$$

Folgt man diesem Vorgehen, so lässt sich ein entsprechendes Designgewicht der Einheit i unter der Verwendung des Horvitz-Thompson-Schätzers der Schätzer des Totatwertes Y nach

$$\hat{Y}_{DSF} = \sum_{l \in L} \sum_{i \in S^F \cup S^C} d_{i,l}^{DSF} y_i = \sum_{l \in L} \sum_{i \in S^F \cup S^C} \frac{y_i}{\pi_{i,l}^{DSF}} \quad (7.4)$$

bestimmen. L bezeichnet hier die Menge aller möglichen Subdomains.

Problematisch in der Anwendung dieses Vorgehens ist die Messung von ξ_l^F . So ist davon auszugehen, dass sich dieses nicht direkt im Zuge einer Befragung bestimmen lässt. Berücksichtigt man die Erkenntnisse, die aus den beiden CELLA-Erhebungen entstammen und vergegenwärtigt sich, dass ein nicht zu vernachlässigender Teil der damaligen Befragten sich bereits mit Schwierigkeiten konfrontiert sah, wenn es darum ging, die korrekte Anzahl der Rufnummern, unter denen diese erreichbar sind, anzugeben, so stellt sich weiterhin die Frage, wie zuverlässig die Einschätzung einer Wahrscheinlichkeit, einen Festnetzkontakt zu beantworten, ist.

Eine weitere Möglichkeit, diese Wahrscheinlichkeit abzuschätzen, stellt unter der Berücksichtigung der in Kapitel 5.3 bereits diskutierten Annahmen, die Verwendung der in diesem Zusammenhang vorgestellten Konvexfaktoren $c_{i,DS}^F$ und $c_{i,DS}^C$ (bzw. c_l^F und c_l^C), dar. Aufgrund der Eigenschaften dieser gilt hierbei $c_{i,DS}^C = 1 - c_{i,DS}^F$. Durch deren Ergänzung ergibt sich aus Gleichung 7.1 als Schätzwert der Inklusionswahrscheinlichkeit der Einheit i in Subdomain l

$$\hat{\pi}_{i,l}^{DSF} = c_l^F \cdot \pi_i^F + (1 - c_l^F) \pi_i^C. \quad (7.5)$$

Dabei wird angenommen, dass c_l^F sich ebenso wie in Gleichung 5.3 darstellt. Demnach kann das Designgewicht dieses Ansatzes nach

$$\hat{d}_{i,l}^{DSF} = \frac{1}{\hat{\pi}_{i,l}^{DSF}} \quad (7.6)$$

geschätzt werden, wodurch sich der Schätzer des Totalwertes durch Einsetzen in 7.4 anhand von

$$\hat{Y}_{DSF} = \sum_{l \in L} \sum_{i \in SF \cup SC} \hat{d}_{i,l}^{DSF} y_i \quad (7.7)$$

ermitteln lässt.

Dieses Vorgehen wird folgend, ebenso wie die Eignung dieses Vorgehens, anhand des synthetischen Datensatzes sowie den beiden Erhebungen diskutiert. Dabei wird vorerst zur Bestimmung der Composite-Faktoren einmal das Verhältnis der abgeschlossenen Festnetzbefragungen zu der Anzahl der gesamten Befragungen innerhalb einer Subdomain verwendet. Weiterhin findet auch die alternative Methode, die die effektive Stichprobengröße verwendet und bereits in den vorherigen Kapiteln diskutiert wurde, Anwendung.

7.2 Anwendung des Domain-Specific-Single-Frame-Ansatz

Ähnlich zum Vorgehen bei der Betrachtung des Domain-Specific-Composite-Ansatzes wird der Domain-Specific-Single-Frame-Ansatz beginnend anhand des Simulationsdatensatzes überprüft. Von Vorteil bei diesem Datensatz ist neben der vollständigen Kenntnis über die Zielpopulation

auch, dass bei diesem die tatsächliche, präferenzbedingte Wahrscheinlichkeit einer Einheit im Überlappungsbereich bekannt ist. Dies ermöglicht einmal die Überprüfung von Gleichung 7.4, als Schätzer eines Totalwertes, und weiterhin, ob c_i^F als Annäherung geeignet ist.

Wie bereits in Kapitel 6.1 beschrieben, wurde mit der Einführung der Nutzungsgewohnheiten der Einheiten des synthetischen Datensatzes auch eine Wahrscheinlichkeit festgelegt, mit der diese bei einem Kontakt über ein bestimmtes Medium in Abhängigkeit der zugehörigen Subdomain antworten. Diese Wahrscheinlichkeit verhält sich wie im vorherigen Kapitel durch ξ_i^F geschildert.

Weiterhin wurde die Simulation dahin gehend ausgeweitet, dass die Schätzung der entsprechenden Parameter über 100 Iterationen erfolgt ist. Die Ergebnisse unter der Anwendung des DSF-Ansatzes im Vergleich zum SF-Ansatz, der Composite-Gewichtung sowie der tatsächlichen Verteilung in der Zielpopulation können in Tabelle 44 nachvollzogen werden. Weiterhin stellt diese auch die Ergebnisse unter der Verwendung des subdomain-spezifischen Composite-Faktors (c_i^F) dar. Dies erfolgt sowohl unter der Verwendung der regulären Stichprobengröße ($\widehat{DSF}_1$) und der effektiven Stichprobengröße ($\widehat{DSF}_2$).

Tabelle 44 Vergleich der unterschiedlichen Schätzverfahren inklusive DSF-Ansatz für Variable 1 (in %)

Variable 1	Stichprobe	SF	CW	DSC	DSF	$\widehat{DSF}_1$	$\widehat{DSF}_2$	Population
1	24,9	25,1	25,1	25,1	25,1	25,1	25,1	25,0
2	15,2	15,0	15,0	15,0	15,0	15,0	15,0	15,0
3	49,5	49,9	49,9	49,9	49,9	49,9	49,9	50,0
4	10,4	10,0	10,0	10,0	10,0	10,0	10,0	10,0
Designeffekt	●●	1,41	1,42	1,50	1,17	1,18	1,17	●●
n_{eff}	●●	7.108	7.065	6.661	8.561	8.470	8.541	●●

Der Vergleich der unterschiedlichen Gewichtungsansätze anhand derjenigen Variable der synthetischen Population, die unabhängig von der Zugehörigkeit einer Einheit zu einer bestimmten Subdomain ist (Variable 1), zeigt, bezogen auf die Punktschätzer, wenig bis keine Unterschiede in der Abweichung von der tatsächlichen Verteilung in der Population. Interessant ist jedoch die Betrachtung des modellbasierten Designeffektes. Dieser fällt unter der Verwendung des DSF-Ansatzes sowie dessen Annäherungen zur Bestimmung der präferenzbedingten Wahrscheinlichkeit anhand der abgeschlossenen Festnetzbefragungen in einer Subdomain sowie durch die Verwendung der effektiven Stichprobengröße der Festnetzinterviews im Verhältnis zu deren Gesamtumfang in der Subdomain, deutlich geringer aus. Die drei neu eingeführten Verfahren

unterscheiden sich indes nur geringfügig in der Höhe dieses Wertes.

Von größerem Interesse ist jedoch, wie sich die Schätzverfahren in ihrer Punktschätzung unterscheiden, wenn die Zugehörigkeit zu einer bestimmten Subdomain Einfluss auf die Ausprägung einer Variable hat. Dies wurde in dem synthetischen Datensatz anhand der zweiten Variable simuliert. Die Ergebnisse des Vergleichs der unterschiedlichen Schätzverfahren werden in Tabelle 45 dargestellt.

Tabelle 45 Vergleich der unterschiedlichen Schätzverfahren inklusive DSF-Ansatz Variable 2 (in %)

Variable 2	Stichprobe	SF	CW	DSC	DSF	$\widehat{DSF}_1$	$\widehat{DSF}_2$	Population
25	30,2	28,8	28,5	28,3	30,5	30,5	30,5	33,0
50	33,8	33,3	33,4	33,4	33,3	33,3	33,3	33,4
75	36,0	37,9	38,1	38,4	36,2	36,2	36,2	33,7
Designeffekt	●●	1,41	1,42	1,50	1,17	1,18	1,17	●●
n_{eff}	●●	7.108	7.065	6.661	8.561	8.470	8.541	●●

Auch hier zeigt sich, dass sich die drei Varianten des DSF-Ansatzes nur wenig unterscheiden. Jedoch liefern diese deutlich genauere Punktschätzer als der reguläre SF-Ansatz, die Composite-Gewichtung und der in Kapitel 5.3 beschriebene Ansatz. In Anbetracht der exakteren Punktschätzung sowie des geringeren modellbasierten Designeffektes, der eine höhere Präzision der Ansätze vermuten lässt, sind diese deshalb zumindest in der vorliegenden Untersuchung eindeutig vorzuziehen.

Da sich die in diesem Kapitel vorgestellten Gewichtungsansätze weiterhin wenig voneinander unterscheiden, kann zumindest in der Anwendung auf den synthetischen Datensatz vermutet werden, dass die beiden Approximationen für die präferenzbedingte Wahrscheinlichkeit, als adäquat erachtet werden können.

Die Überprüfung des vorgestellten Gewichtungskonzeptes anhand des synthetischen Datensatzes hat allerdings seine Limitationen, wenn es um die Anwendung auf tatsächliche Erhebungsdaten geht. So wurde bereits festgehalten, dass es in dieser Stichprobe zu keinerlei Antwortausfällen aufgrund von Nonresponse kommt. Ein Zustand, der in der Praxis als sehr unwahrscheinlich gilt. Weiterhin steht die Art und Weise, wie die Auswahl einer bestimmten Zielperson in einer Subdomain erfolgt, in direkter Verbindung zu der hier verwendeten, präferenzbedingten Wahrscheinlichkeit. Diese wurde bei der Stichprobenziehung an die Auswahlwahrscheinlichkeit einer Einheit, wie sie in Kapitel 3.3.1 beschrieben wurde, multipliziert. Somit ist zu überprüfen, ob der Zusammenhang zwischen der Auswahlwahrscheinlichkeit einer Einheit

und deren präferenzbedingten Wahrscheinlichkeit sich tatsächlich ähnlich gestaltet, wie in der vereinfachten Modellwelt des synthetischen Datensatzes. Aufgrund dessen wird dieser Ansatz anhand der beiden in dieser Arbeit verwendeten Erhebungen nochmals getestet.

Bei beiden Erhebungen ist zu beachten, dass die tatsächliche, präferenzbedingte Wahrscheinlichkeit innerhalb der jeweiligen Subdomains nicht bekannt ist. Weiterhin ist zu vermuten, dass diese sich eher in einem bestimmten Kontinuum bewegt. Dies steht im Gegensatz zu der Untersuchung des synthetischen Datensatzes, bei dem diese Wahrscheinlichkeit für alle Einheiten einer bestimmten Subdomain gleich gewesen ist. Da die "echte", präferenzbedingte Wahrscheinlichkeit darüber hinaus unbekannt ist, werden an dieser Stelle lediglich die beiden Ansätze zur Approximation der Wahrscheinlichkeit verwendet.

Aufgrund der bereits in Kapitel 6.2.2 beschriebenen Probleme der Erhebung CELLA 2 bei der Zuordnung der einzelnen Erhebungseinheiten in unterschiedliche Subdomains, kann an dieser Stelle lediglich vermutet werden, welche Einteilung für die Anwendung des vorgestellten Gewichtungsansatzes am geeignetsten erscheint. Dies begründet sich dadurch, dass nur Proxys zur Einteilung zur Verfügung stehen und eine Indifferenz bezüglich der Auswahl unter den ersten drei Modellen zur Zuordnung besteht. Da unter der Verwendung der jeweiligen Composite-Faktoren, die auch hier zum Tragen kommen, der modellbasierte Designeffekt am geringsten gewesen ist und dieses allgemein als das am konservativsten zu werten ist, soll Modell 1 zur Anwendung herangezogen werden. Die Betrachtung der anderen Modelle zwei und drei erfolgt indes im Anhang dieser Arbeit. Da das vierte Modell ohnehin als unwahrscheinlich zu werten ist, wird dieses aus der näheren Betrachtung ausgeschlossen. Tabelle 46 stellt die Bildungsvariable der Erhebung sowie deren Schätzergebnisse unter den beiden Approximationen $\widehat{DSF}_1$ und $\widehat{DSF}_2$ dar. Die Ergebnisse werden abermals mit denjenigen der Schätzung mittels SF-Ansatz, des DSC-Ansatzes sowie mit den Angaben des Mikrozensus 2010 verglichen. Der Vergleich weiterer Variablen befindet sich im Anhang dieser Arbeit.

Tabelle 46 **CELLA 2: Höchster allgemeiner Schulabschluss
mit DSF-Schätzer nach Modell 1 (in %)**

Bildungsabschluss	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	MZ 2010
Ohne/noch kein	4,0	3,9	3,6	3,6	7,6
Haupt-/Volksschulabschl.	20,9	20,9	19,5	19,4	37,0
Polytechn. OS	6,2	6,0	6,1	6,1	7,1
Mittlere Reife	26,4	26,2	26,7	26,7	21,7
FH/HS - Reife	40,9	41,3	42,5	42,6	25,8
Andere	1,5	1,6	1,5	1,6	0,4
Designeffekt	1,38	1,47	1,15	1,14	●●

Im Gegensatz zu der Untersuchung der unterschiedlichen Gewichtungsansätze anhand des synthetischen Datensatzes wird bei den ermittelten Schätzwerten aus den Erhebungsdaten von CELLA 2 deutlich, dass der DSF-Schätzer in der gewählten Modellspezifikation zur Einteilung nach Subdomain nicht durchgehend Punktschätzer hervorbringt, die näher an der Verteilung der amtlichen Statistik liegen. Zwar unterscheiden sich der DSF- und der SF-Schätzer hinsichtlich der hier berechneten Schätzwerte nicht stark, jedoch zeigt sich für die Bildungsvariable, dass der SF-Schätzer Werte hervorbringt, die näher an der Verteilung der amtlichen Statistik liegen. Allerdings gilt weiterhin anzumerken, dass die ermittelten Werte unter der Verwendung der beiden DSF-Ansätze ebenfalls einen deutlich geringeren, modellbasierten Designeffekt aufweisen. Demnach ist zu vermuten, dass zumindest die Präzision dieses Ansatzes höher ist als unter der Verwendung des SF- sowie des DSC-Ansatzes. Zwischen der Verwendung der "regulären" und der effektiven Stichprobengröße zur Bestimmung der Konvexfaktoren lässt sich indes in der Anwendung des DSF-Ansatzes kein nennenswerter Unterschied feststellen. Dies gilt sowohl für die berechneten Punktschätzer der Bildungsvariablen als auch für den Designeffekt.

Ein vergleichbares Bild zeichnet sich auch unter der Anwendung der Modelle 2 und 3 ab, die im Anhang dieser Arbeit nachvollzogen werden können. Keines dieser beiden Modelle scheint hierbei dem ersten Modell in seinen ermittelten Punktschätzern eindeutig überlegen zu sein. Des Weiteren liegt der Designeffekt, wenn auch nur marginal, oberhalb desjenigen des hier betrachteten Modell 1.

Die Anwendung auf weitere soziodemografische Merkmale innerhalb der Erhebungsdaten von CELLA 2 zeigt indes ein anderes Bild, als dies noch unter der Verwendung des DSC-Ansatzes zu beobachten gewesen ist. So lassen sich hier durchaus Variablen beobachten, deren geschätzte Verteilung unter der Verwendung des DSF-Ansatzes näher an derjenigen der amtlichen Statistik liegt, als unter der Verwendung des SF-Ansatzes. So scheinen gerade die Punktschätzungen

unter der Verwendung des DSF-Ansatzes für die beiden Variablen Migrationshintergrund und Haushaltsgröße eindeutig, wenn auch geringfügig, derjenigen des SF-Ansatzes und insbesondere des DSC-Ansatzes überlegen zu sein. Für die beiden Variablen Familienstand und Nettoeinkommen kann dies zumindest für einige der jeweiligen Ausprägungen gezeigt werden, obschon dies nicht für die gesamte Verteilung der beiden Variablen zutrifft. Für die Verteilung der Variable Geschlecht hingegen ist wieder der SF-Ansatz zu bevorzugen.

Im Gegensatz zu der vorherigen Untersuchung des DSC-Ansatzes, kann innerhalb der Anwendung des DSF-Ansatzes auf die Erhebungsdaten von CELLA 2 gezeigt werden, dass eine Berücksichtigung der Nutzungsgewohnheiten innerhalb dieses Schätzers (zumindest für einige Variablen), durchaus sinnvoll sein kann. Dies macht sich einmal anhand einer geringeren Differenz zwischen den durch den DSF-Schätzer ermittelten Verteilungen und derjenigen der amtlichen Statistik bemerkbar, weiterhin jedoch anhand des deutlich geringeren geschätzten Designeffektes. Gerade dieser lässt für die Schätzwerte des DSF-Ansatzes eine höhere Präzision vermuten, als dies unter der Anwendung der anderen beiden hier dargestellten Ansätze zu beobachten ist. Damit einhergehend ist somit auch eine höhere effektive Nettostichprobengröße.

Die Anwendung des DSF-Ansatzes auf die Erhebungsdaten von Influenza 2014 kann anhand von Tabelle 47 sowie den Ergebnissen des Kapitels A.6 des Anhangs nachvollzogen werden. Analog zum Vorgehen im Zuge der Anwendung auf die Daten aus CELLA 2, wurde auch hier eine Schätzung des DSF-Ansatzes unter der Verwendung der regulären Stichprobengröße ($\widehat{DSF}_1$) sowie der effektiven Stichprobengröße ($\widehat{DSF}_2$) zur Berechnung der Konvexfaktoren verwendet. Diese wurden sowohl mit dem SF-Ansatz als auch mit dem DSC-Ansatz bezüglich ihres Designeffektes sowie der Übereinstimmung der geschätzten Verteilung soziodemografischer Merkmale mit der amtlichen Statistik verglichen. Als Quelle der Verteilung der amtlichen Statistik diene, wie auch in den vorherigen Kapiteln, der Zensus 2011.

Da bereits in Kapitel 6.3.2 festgestellt wurde, dass im Zuge der Einteilung nach Nutzungsgewohnheiten die ersten beiden Modelle diejenigen Schätzwerte hervorbringen, die am ehesten mit den Verteilungen der amtlichen Statistik übereinstimmen, wurden diese beiden Modelle auch für die Auseinandersetzung mit dem in diesem Kapitel angewendeten Schätzverfahren verwendet. Tabelle 47 stellt dazu die Ergebnisse bei der Schätzung der Verteilung des Bildungsabschlusses unter der Anwendung des ersten Modells zur Einteilung den weiteren Ansätzen gegenüber. Die Anwendung von Modell 2 findet sich ebenso wie die Ergebnisse der Schätzung der Variablen Migrationshintergrund, Bildungsabschluss und Alter, für beide Modelle, im Anhang.

Tabelle 47 Influenza 2014: Bildungsabschluss mit DSF-Schätzer nach Modell 1 (in %)

Bildungsabschluss	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	Zensus 2011
Höchstens 7 Jahre Schulbesuch	0,7	0,6	0,6	0,6	5,0
Haupt-/Volksschulabschluss	15,7	15,3	13,9	14,0	36,6
Mittlere Reife	33,1	33,1	34,3	34,2	29,0
Fachhochschulreife	9,6	9,8	9,4	9,4	8,3
Abitur	40,9	41,2	41,9	41,9	21,1
Designeffekt	1,43	1,45	1,29	1,27	●●

Ähnlich der Auseinandersetzung der Daten aus CELLA 2, sind die Ergebnisse der geschätzten Verteilung der Bildungsvariablen unter der Verwendung des DSF-Ansatzes weniger nahe an derjenigen der amtlichen Statistik, als diejenige des SF-Ansatzes. Die Unterschiede sind ebenfalls nur geringfügig. Jedoch kann alleinig anhand des Vergleichs mit der amtlichen Statistik, der DSF-Ansatz nicht gegenüber dem SF- sowie dem DSC-Ansatz präferiert werden. Auch hier zeigt sich allerdings ein deutlich geringerer Schätzwert des Designeffektes, als dies unter der Verwendung der anderen beiden Ansätze zu beobachten ist. Demnach kann auch im Fall der Erhebung Influenza 2014 vermutet werden, dass die Schätzwerte des DSF-Ansatzes über eine höhere Präzision verfügen.

Im Gegensatz zu den Ergebnissen aus der Untersuchung im Rahmen der CELLA 2-Erhebung lassen sich jedoch kaum Unterschiede in der Anwendung auf die Daten aus Influenza 2014 zwischen den beiden Verfahren zur Berechnung der Composite-Faktoren feststellen. Auch wenn diese nur geringfügig erscheinen, ist jedoch der Ansatz, der die effektive Stichprobengröße zur Bestimmung der Composite-Faktoren verwendet unter diesen beiden zu bevorzugen. Dies liegt sowohl an einer (marginal) höheren Übereinstimmung der geschätzten Verteilung mit den Ergebnissen der amtlichen Statistik als auch an einem geringeren Schätzwert des Designeffektes. Ebenso trifft dies für die im Anhang angeführten Untersuchungen mit Ausnahme der Variable Migrationshintergrund zu.

Beide der an dieser Stelle untersuchten Modelle zur Einteilung nach Nutzungsgewohnheiten unterscheiden sich hinsichtlich der geschätzten Verteilungen der soziodemografischen Merkmale nur unwesentlich. Jedoch zeigt sich, wenn auch nur geringfügig, dass in dem vorliegenden Fall die Einordnung nach Modell 1 zu bevorzugen ist.⁷⁹

Im Gegensatz zu den Ergebnissen aus CELLA 2 sowie der Anwendung auf den synthetischen Datensatz fällt weiterhin auf, dass zumindest in der Punktschätzung der DSF-Ansatz in beiden

⁷⁹ Hierfür sprechen sowohl die Ergebnisse der Schätzungen als auch der geschätzte Designeffekt. Allerdings zeigen sich diese Unterschiede lediglich ab der dritten Nachkommastelle, weshalb sie in der gewählten Darstellung der Ergebnisse nicht direkt nachvollziehbar sind.

Varianten und für beide Modelle der Einteilung, Schätzwerte hervorbringt, die, gemessen an den Angaben der amtlichen Statistik, weniger exakt sind, als die durch den SF-Ansatz geschätzten Werte. Jedoch scheint hier der DSC-Ansatz in der Punktschätzung näher an der Verteilung der amtlichen Statistik zu sein, als der DSF-Ansatz. Dies trifft für alle untersuchten Variablen mit Ausnahme des Migrationshintergrundes zu. Bei dieser Variable liegen die geschätzten Werte des DSF-Ansatzes näher an der Verteilung der amtlichen Statistik als der SF-Ansatz. Dennoch kann auch bei der Anwendung auf die Daten der Erhebung Influenza 2014 hervorgehoben werden, dass der geschätzte Designeffekt geringer als derjenige der anderen beiden Ansätze ist.

Insgesamt zeigt sich bei der Auseinandersetzung mit dem DSF-Ansatz unter der Anwendung auf beide Erhebungen, dass die hervorgebrachten Schätzwerte nicht eindeutig denjenigen des SF-Ansatzes in ihrer Punktschätzung überlegen sind. Während bei der Anwendung auf den synthetischen Datensatz der DSF-Ansatz evident zu bevorzugen war, da einerseits die geschätzte Verteilung beider Variablen (im Rahmen der Punktschätzung) näher an der tatsächlichen Verteilung der Zielpopulation lag und andererseits der geschätzte Designeffekt deutlich unterhalb desjenigen der anderen beiden Ansätze war, trifft dies für beide Erhebungen nicht unmittelbar zu. So lässt sich innerhalb von Influenza 2014 lediglich für die (dichotome) Variable Migrationshintergrund zeigen, dass der DSF-Ansatz Schätzergebnisse liefert, die näher an der Verteilung der amtlichen Statistik liegen. Für CELLA 2 hingegen lässt sich dies zumindest für die geschätzte Verteilung der beiden Variablen Migrationshintergrund und Haushaltsgröße vollständig und für die beiden Variablen Familienstand und Haushaltsgröße zu Teilen beobachten. Jedoch liefert der DSF-Ansatz in beiden Erhebungen Ergebnisse, die denjenigen des DSC-Ansatzes vorzuziehen sind.

Die Disparität zwischen der Anwendung auf Erhebungsdaten und derjenigen auf den synthetischen Datensatz kann unter Umständen darin begründet liegen, dass vergleichsweise starke Vereinfachungen innerhalb der künstlichen Population bestehen. Dazu zählen bspw. die vollständige Kenntnis der Präferenzen, die dadurch ermöglicht, eine klare Zuordnung der Stichprobenelemente in die jeweiligen Subdomains vorzunehmen sowie eine direkte, lineare Abhängigkeit der zweiten Variable und den Präferenzen der Stichprobenelemente. Diese Vereinfachungen sind oftmals auf eine tatsächliche Erhebung nicht vollständig übertragbar. Hinzu kommen bei einer "realen" Erhebung noch weitere Probleme wie bspw. Nonresponse und eine beschränkte Kenntnis über die wahren Größen der jeweiligen Sub-Populationen. Es ist demnach durchaus möglich, dass durch eine nicht fehlerfreie Einordnung der Zielpersonen der Stichprobe in die jeweiligen Subdomains, aus denen diese stammen, zu diesen Unterschieden geführt haben kann. Gerade im Zuge der Erhebung CELLA 2, bei der zur Einordnung lediglich Proxys verwendet werden konnten, ist dies durchaus denkbar. Weiterhin zeigen die bereits in Kapitel 2 dargestellten Erkenntnisse aus den CELLA-Erhebungen, dass Zielpersonen bestimmte Eigenschaften ihrer Telekommunikationsmedien falsch benennen. Es ist demnach nicht auszuschließen, dass auch die in Influenza 2014 abgefragten Nutzungsgewohnheiten teilweise falsch von Zielpersonen

eingeschätzt wurden, was wiederum die Schätzergebnisse des DSF-Ansatzes beeinflussen kann. Zusätzlich ist nicht auszuschließen, dass aufgrund der bereits in der Stichprobe der Erhebung Influenza 2014 deutlich überproportional gezogenen Personen mit höherem Bildungsabschluss und im höheren Alter, die Verteilung gerade dieser und damit korrelierter Variablen beeinflusst wird.

Dennoch muss festgehalten werden, dass der Unterschied zwischen den Punktschätzungen des SF- und des DSF-Ansatzes nur geringfügig ist. Berücksichtigt man weiterhin, dass der DSF-Ansatz in allen drei Anwendungsbeispielen einen deutlich geringeren geschätzten Designeffekt hervorbringt, ist die Entscheidung zwischen diesen beiden Ansätzen nicht mehr eindeutig. Dies liegt darin begründet, dass ein niedrigerer Designeffekt eine geringere Varianz bzw. eine höhere Präzision der Schätzwerte vermuten lässt. Demnach ist es durchaus möglich, dass Konfidenzintervalle für Schätzwerte basierend auf dem DSF-Ansatz kleiner ausfallen als diejenigen des SF-Ansatzes.

Bezüglich der Wahl des Vorgehens zur Bestimmung der Konvexfaktoren konnte gezeigt werden, dass dasjenige Verfahren, das die effektive Stichprobengröße verwendet, zu bevorzugen ist. Auch wenn die Unterschiede zwischen beiden Vorgehensweisen nur geringfügig scheinen, war zumindest in allen drei Untersuchungen der geschätzte Designeffekt des Verfahrens mit der effektiven Stichprobengröße geringer. Weiterhin konnte anhand des synthetischen Datensatzes gezeigt werden, dass dieser Designeffekt auch demjenigen entspricht, der sich unter der Kenntnis der wahren Präferenzen in der Anwendung des DSF-Ansatzes einstellt.

Um zu klären, welcher Ansatz endgültig zu bevorzugen ist, wird im folgenden Kapitel ein Vergleich des *Mean Squared Errors (MSE)* durchgeführt. Dieser Vergleich dient dazu, genauer zu determinieren, ob die geringere Übereinstimmung des DSF-Ansatzes mit den Angaben der amtlichen Statistik eventuell durch die vermeintlich höhere Präzision aufgewogen wird.

8 Beurteilung und Vergleich der verschiedenen Gewichtsmodelle

Im Rahmen des letzten, analytischen Kapitels wird als finaler Schritt in der Untersuchung der unterschiedlichen Verfahren zur Gewichtung von Dual-Frame-Erhebungen ein weiteres Maß zur Beurteilung der Güte eines Schätzwertes untersucht. Die hier verwendete *mittlere quadratische Abweichung* oder *Mean Squared Error (MSE)* setzt sich aus der Summe der quadrierten Verzerrung und der Varianz eines Schätzwertes zusammen. aufgrund der Ergebnisse des vorherigen Kapitels wird dieses Maß abschließend bemüht. So konnte zuvor gezeigt werden, dass bei der Anwendung auf Erhebungsdaten, der DSF-Ansatz in einigen Fällen Werte hervorbringt, die weniger nahe an der tatsächlichen Verteilung der Zielpopulation liegen, als der SF-Ansatz. Jedoch waren diese Unterschiede oftmals nur geringfügig, während gleichzeitig der geschätzte Designeffekt des DSF-Ansatzes deutlich niedriger war. Ein geringerer Design- effekt lässt indes eine höhere Präzision vermuten. Daher erscheint ein Maß, das sowohl die Verzerrung als auch die Varianz berücksichtigt, als adäquates Mittel zur Beurteilung der Ansätze. Um dies zu ermöglichen, wird folgend die Varianzschätzung bzw. -Approximation in einem solchen, komplexen Erhebungsverfahren erläutert, um anschließend die Berechnung des MSEs zu beschreiben und die Ergebnisse in dessen Anwendung auf die hier vorgestellten Gewichtungsansätze zu präsentieren und zu untersuchen. Die Anwendung erfolgt unter der Zuhilfenahme der bereits erläuterten Datengrundlagen. Abschließend werden dann anhand der vorliegenden Ergebnisse die jeweiligen Gewichtungsansätze beurteilt.

8.1 Varianzschätzung als Indikator der Genauigkeit

Wie in Kapitel 3 bereits hervorgehoben wurde, besteht das Ziel der Gewichtung von Erhebungsdaten (in einer komplexen Erhebung) vornehmlich darin, unterschiedliche Auswahlwahrscheinlichkeiten in der Schätzung von Populationsstatistiken miteinzubeziehen sowie Besonderheiten, die im Zuge des Erhebungsprozesses entstanden sind, wie bspw. Nonresponse, zu berücksichtigen. Weiterhin kann es wünschenswert sein, die Verteilung der Erhebungsdaten an diejenige, die in der Population vorherrscht, anzupassen (vgl. Gabler et al. (2011), S. 147.). Ziel eines jeden Gewichtungsprozesses ist es demnach, eine potenzielle Verzerrung, die im Rahmen der Erhebung und der Schätzung auf der Basis von Erhebungsdaten entstehen kann, zu verringern sowie die Schätzung eines möglichst genauen Abbilds der Zielpopulation zu gewährleisten. So ist die Designgewichtung bei komplexen Erhebungen genau dann notwendig, wenn unterschiedliche Auswahlwahrscheinlichkeiten für einzelne Zielpersonen vorliegen und Aussagen über eine zugrunde liegende Zielpopulation getroffen werden sollen. Eine ungewichtete Schätzung relevanter Statistiken

der Zielpopulation läuft bei unterschiedlichen Auswahlwahrscheinlichkeiten Gefahr, verzerrte Schätzwerte für die Zielpopulation hervorzubringen. Bei einer Dual-Frame-Telefonstichprobe, bei der kein Screening-Ansatz verwendet wird, ist demnach davon auszugehen, dass aufgrund mehrerer Faktoren, wie bspw. einer Mehrzahl von Anschlüssen, Mehrpersonenhaushalten und Dual-Usern, die über mehrere Telekommunikationsmedien in die Stichprobe gelangen können, unterschiedliche Auswahlwahrscheinlichkeiten vorliegen. Deshalb ist in einer solchen Erhebung i. d. R. eine Designgewichtung unumgänglich, wenn unverzerrte Aussagen über die zugrunde liegende Population getroffen werden sollen.

Neben der Eigenschaft, dass ein Schätzwert möglichst unverzerrt ist, ist jedoch auch die Präzision ein relevantes Maß zur Bemessung der "Qualität" der Ergebnisse. S. L. Lohr (2009) unterscheidet diesbezüglich die Genauigkeit eines Schätzers und dessen Präzision. Die Genauigkeit beschreibt dabei, wie nahe die geschätzten Werte an deren wahren Ausprägungen der Zielpopulation liegen. Die Präzision hingegen beschreibt, wie nahe die Schätzwerte aus unterschiedlichen Stichproben/Erhebungen (bei gleichem Schätzer) zueinander sind. Die Präzision wird i. d. R. anhand der Varianz der zu schätzenden Statistik ausgedrückt (vgl. S. L. Lohr (2009), S. 32). Zur Beurteilung der Schätzer, die sich aufgrund der unterschiedlichen Art und Weise zur Bestimmung der Inklusionswahrscheinlichkeit ergeben, muss daher auch deren Präzision berücksichtigt werden.

Unter der Verwendung des Horvitz-Thompson-Schätzers zur Designgewichtung berechnet sich die Varianz eines geschätzten Totalwertes ($\hat{Y}$) der Variablen Y bei einem festen Stichprobenumfang anhand von⁸⁰

$$Var(\hat{Y}_{HT}) = -\frac{1}{2} \sum_{i \in U} \sum_{\substack{j \in U \\ i \neq j}} (\pi_{ij} - \pi_i \cdot \pi_j) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2. \quad (8.1)$$

Hierbei entspricht π_{ij} der Inklusionswahrscheinlichkeit zweiter Ordnung (für die beiden Elemente i und j). Diese gibt die Wahrscheinlichkeit wieder, dass sowohl ein Element i als auch ein Element j gemeinsam Teil einer Stichprobe sind. Da dieser Ansatz die vollständige Kenntnis aller Inklusionswahrscheinlichkeiten erster und zweiter Ordnung der Population voraussetzt, wird oftmals zur Schätzung der Varianz des geschätzten Totalwertes

$$\widehat{Var}(\hat{Y}_{HT}) = -\frac{1}{2} \sum_{i \in S} \sum_{\substack{j \in S \\ i \neq j}} \frac{(\pi_{ij} - \pi_i \cdot \pi_j)}{\pi_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j} \right)^2 \quad (8.2)$$

nach Sen, Yates und Grundy verwendet. Dieser erfordert lediglich die Kenntnis dieser Parameter innerhalb der zugrundeliegenden Stichprobe. Gilt $\pi_{ij} > 0$ und $\pi_{ij} \leq \pi_i \cdot \pi_j \forall i, j$, so erlaubt

⁸⁰ An dieser Stelle wird die Varianz direkt in Form von Sen (1953) sowie Yates und Grundy (1953) verwendet, die eine feste Stichprobengröße vorsieht.

dieser Ansatz eine erwartungstreue, nicht-negative Schätzung der Varianz des geschätzten Totalwertes (vgl. Matei und Tillé (2005), S. 548; S. 551.; vgl. Bruch et al. (2011), S. 6f.).

Ein solcher Ansatz kann bspw. immer dann zum Tragen kommen, wenn die Daten einer Erhebung aus einer Stichprobe stammen. Weiterhin wäre dieses Verfahren in einer Dual-Frame-Erhebung genau dann anwendbar, wenn die einzelnen Stichproben bezüglich ihrer Zielpopulation disjunkt wären, wie es etwa bei einem Screening-Ansatz der Fall wäre. Die Varianz des Schätzers eines Totalwertes entspräche dann der Summe der einzelnen geschätzten Varianzen der jeweiligen Subdomains. Würde bspw. die Mobilfunkstichprobe ausschließlich nach Mobile-Onlys screenen, so ließe sich die Varianz anhand von

$$\widehat{Var}(\hat{Y}_{HT}) = \widehat{Var}(\hat{Y}_{HT}^{F \setminus C \cup F \cap C}) + \widehat{Var}(\hat{Y}_{HT}^{C \setminus F})$$

schätzen. Ähnliches gilt auch für die Schätzung der Varianz eines Overlap-Design, sofern der Single-Frame-Ansatz verwendet wird. Da im Zuge dieses Gewichtungsansatzes angenommen wird, dass die beiden Stichproben eine Art "Super-Stratum" bilden und für jede Zielperson die gemeinsame Auswahlwahrscheinlichkeit aus beiden Stichproben berechnet wird, lässt sich die Varianz innerhalb dieses Ansatzes für jede Subdomain anhand von Gleichung 8.2 schätzen.

Im Gegensatz dazu stehen die unterschiedlichen Composite-Ansätze, die die Schätzwerte separat nach Stichprobe berechnen und bei denen sich der gemeinsame Schätzwert beider Stichproben als (gewichtete) Summe der beiden Einzelnen bildet. Brick et al. (2011) bemerkt diesbezüglich jedoch, dass die Varianzschätzung bei einem Composite-Ansatz vergleichsweise einfach ist, solange der Konvexfaktor konstant ist. Bei konstantem Konvexfaktor lässt sich die Varianz durch

$$\widehat{Var}_{CW}(\hat{Y}) = \widehat{Var}(\hat{Y}_{S^F \cap U^F \setminus C} + \phi \cdot \hat{Y}_{S^F \cap U^F \cap C}) + \widehat{Var}((1 - \phi) \cdot \hat{Y}_{S^C \cap U^F \cap C} + \hat{Y}_{S^C \cap U^C \setminus F}) \quad (8.3)$$

schätzen (vgl. Brick et al. (2011), S. 2.).

Die im Zuge dieser Arbeit vorgestellten Gewichtungsansätze, die ebenfalls auf einer Konvexkombination beruhen, weisen jedoch mehr als einen konstanten Konvexfaktor auf. Allerdings sind die Konvexfaktoren der hier verwendeten Ansätze, die der Berücksichtigung der Nutzungsgewohnheiten sowie der Erreichbarkeit der jeweiligen Dual-User dienen, innerhalb der entsprechenden Subdomain, konstant. Da die Subdomains im Zuge dieser Arbeit als disjunkte Teilmengen der gesamten Zielpopulation beschrieben wurden, lässt sich ein Schätzer der Varianz

innerhalb des DSC-Ansatzes durch

$$\begin{aligned}
 \widehat{Var}_{DSC}(\hat{Y}) &= \widehat{Var}(\hat{Y}_{S^F \cap u_1}) + \widehat{Var}(\lambda \hat{Y}_{S^F \cap u_2}) + \widehat{Var}(\phi \hat{Y}_{S^F \cap u_3}) + \widehat{Var}(\mu \hat{Y}_{S^F \cap u_4}) \\
 &+ \widehat{Var}((1 - \lambda) \hat{Y}_{S^C \cap u_2}) + \widehat{Var}((1 - \phi) \hat{Y}_{S^C \cap u_3}) + \widehat{Var}((1 - \mu) \hat{Y}_{S^C \cap u_4}) + \widehat{Var}(\hat{Y}_{S^C \cap u_5}) \\
 &= \widehat{Var}(c_{l,DSC}^F \hat{Y}_{S^F \cap u_l}) + \widehat{Var}(c_{l,DSC}^C \hat{Y}_{S^C \cap u_l})
 \end{aligned} \tag{8.4}$$

bestimmen. $c_{l,DSC}^F$ und $c_{l,DSC}^C$ haben dabei die in 3.7 und 3.8 dargestellten Ausprägungen in Abhängigkeit der zugrundeliegenden Subdomain l .

Auch für den DSC-Ansatz lässt sich aufgrund der disjunkten Teilmengen der Zielpopulation die Varianz verhältnismäßig einfach herleiten. Da hier in Abhängigkeit der Subdomain l jeweils nur eine einzelne, gemeinsame Inklusionswahrscheinlichkeit bestimmt wird und die Berücksichtigung der Präferenzen einzelner Dual-User im Zuge der Berechnung dieser anhand von 7.3 erfolgt, lässt sich die Varianz eines Schätzwertes für den Totalwert anhand von 8.2 schätzen.⁸¹

Allen Schätzern der Varianz, die hier vorgestellt wurden, ist demnach gemein, dass diese (zumindest in der hier dargelegten Vorgehensweise) alle auf der Anwendung des Horvitz-Thompson-Schätzers und dessen Varianz (-Schätzer) beruhen. Sowohl 8.1 als auch 8.2 sind in der Anwendung auf eine komplexe Erhebung nach Bruch et al. (2011) mit mehreren Problemen behaftet. So bemerken die Autoren, dass die in beiden Formeln erscheinende Doppelsumme zu einer schnell ansteigenden Anzahl von Termen führt, was wiederum bewirkt, dass diese Varianzschätzer rekursiv berechnet werden müssen. Weiterhin relevant bei diesen Schätzern ist das Problem, dass für die meisten in der Praxis angewandten Stichprobenverfahren die Inklusionswahrscheinlichkeiten zweiter Ordnung (π_{ij}) nicht unproblematisch zu bestimmen sind. In diesem Kontext wird von den Autoren festgehalten, dass lediglich in seltenen Fällen, bei einfachen Stichprobenverfahren, etwa einer uneingeschränkten Zufallsauswahl ohne Zurücklegen, π_{ij} leicht zu bestimmen sei. So könnte am Beispiel der uneingeschränkten Zufallsauswahl ohne Zurücklegen die Inklusionswahrscheinlichkeit zweiter Ordnung anhand von $\pi_{ij} = n(n - 1)/N(N - 1)$ hergeleitet werden (vgl. Bruch et al. (2011), S. 7)

Nimmt man nun den Fall einer Dual-Frame-Telefonerhebung als Grundlage, so wird schnell ersichtlich, dass es sich hierbei nicht um eine vergleichbar einfache Stichprobe handelt. Zwar werden bspw. bei GESIS, dem Leibniz-Institut für Sozialwissenschaften, Rufnummern für Festnetz- und Mobilfunkstichproben i. d. R. uneingeschränkt zufällig gezogen, aber diese Stichproben sind nicht die einzigen Faktoren, die zuletzt die jeweiligen Auswahlwahrscheinlichkeiten erster Ordnung determinieren. So handelt es sich, wie in Kapitel 2.4.2 bereits genauer erläutert, bei Festnetzstichproben gewöhnlich um ein zweistufiges Verfahren, bei dem nach erfolgreichem Kontakt noch eine Person innerhalb des Haushaltes ermittelt wird. Hierbei besteht das Problem,

⁸¹ Jedoch wird π_{ij} anhand von 7.3 berechnet.

dass die Haushaltsgröße erst bei Kontakt festgestellt werden kann. Diese ist somit lediglich für die Erhebungsteilnehmer bekannt. Weiterhin hängt die Inklusionswahrscheinlichkeit einer Zielperson sowohl in der Festnetz- als auch in der Mobilfunkstichprobe von der Anzahl der Rufnummern, unter denen diese erreichbar sind, ab. Auch dies lässt sich erst im Zuge der Erhebung feststellen. Zusätzlich besteht das Problem unterschiedlicher Inklusionswahrscheinlichkeiten zwischen Personen aus Only-Haushalten und Dual-Usern. So verfügen Dual-User gewöhnlich alleine dadurch bereits eine höhere Inklusionswahrscheinlichkeit als eine Zielperson aus einem Only-Haushalt, da diese über mehrere Stichproben in die Erhebung gelangen können. Wenn zuletzt auch noch die Präferenz für eines der beiden Telekommunikationsmedien innerhalb der Berechnung der Inklusionswahrscheinlichkeit berücksichtigt wird, wie es bspw. der DSF-Ansatz vorsieht, erlangt der Erhebungsprozess bzw. die Berechnung der Auswahlwahrscheinlichkeit, eine zusätzliche Komplexität durch einen weiteren Faktor, der sich erst im Zuge der eigentlichen Befragung der Zielperson ermitteln lässt und der für die Zielpopulation unbekannt ist.

Die hier geschilderten Probleme verdeutlichen, dass es sich bei einer Dual-Frame-Telefon-Erhebung nicht um einen einfachen Auswahlprozess innerhalb der Stichprobe handelt und dies ohne die bereits getroffenen Annahmen, die zur Bestimmung der Inklusionswahrscheinlichkeit in Kapitel 3.3.1 bereits erläutert wurden. Berücksichtigt man indes auch, dass im Zuge einer Stichprobe mehrere Personen eines Festnetzhaushaltes oder die gleiche Person über mehrere Rufnummern kontaktiert würde, zwei Ereignisse, die zwar unwahrscheinlich, aber nicht ausgeschlossen sind, so wird offensichtlich, dass aus einer Vielzahl von Ursachen, die Bestimmung der Inklusionswahrscheinlichkeit zweiter Ordnung in einer Dual-Frame-Erhebung eine Komplexität erreicht, die die exakte Bestimmung dieser sehr rechenintensiv macht oder, aufgrund der erforderlichen Informationen, stark einschränkt.⁸²

Aufgrund der beschriebenen Probleme scheint eine direkte Schätzung der Varianz in Dual-Frame-Erhebungen zu komplex, um 8.2 anzuwenden. Um dennoch eine Schätzung zu ermöglichen, werden allerdings durch Matei und Tillé (2005) eine Vielzahl von Approximationen diskutiert, die gerade der Bestimmung der Inklusionswahrscheinlichkeit zweiter Ordnung Abhilfe schaffen können. Eine der in diesem Kontext am häufigsten verwendeten Approximationen der designbasierten Varianz des Horvitz-Thompson-Schätzers stellt dabei diejenige unter der Verwendung der Brewer-Familie dar. Diese kann so konstruiert werden, dass sie unter einer uneingeschränkten Zufallsstichprobe ohne Zurücklegen exakt design-unverzerrt ist. Dabei lässt

82 Gleiches gilt auch für die Bestimmung von Korrelationen einzelner, (design-) gewichteter Variablen. Da hierfür die Kovarianz zweier Variablen benötigt wird und diese grundsätzlich auf der gleichen Statistik beruht, lassen sich diese für komplexe Erhebungen, wenn überhaupt, nur sehr aufwendig bestimmen. Ein oftmals beobachteter Fehler wäre es indes, die Korrelation zweier Variablen in der Zielpopulation, die aus einer komplexen Erhebung stammen, entweder nur ungewichtet, entsprechend der Form einer deskriptiven Statistik der Stichprobe, zu berechnen oder das gleiche Verfahren für die gewichteten Ausprägungen anzuwenden. Um die Korrelation der Grundgesamtheit in diesem Fall zu bestimmen, ist es demnach erforderlich, die Kovarianz sowie die einzelnen Varianzen unter der Berücksichtigung der Inklusionswahrscheinlichkeiten zweiter Ordnung zu bestimmen.

sich die Varianz anhand von

$$\widehat{Var}_{BR}(\hat{Y}) = \sum_{i \in S} \left(\frac{1}{e_i} - \pi_i \right) \left(\frac{y_i}{\pi_i} - \sum_{j \in S} \frac{y_j}{n \cdot \pi_j} \right)^2 \quad (8.5)$$

schätzen. Dabei kann e_i anhand von vier unterschiedlichen Möglichkeiten bestimmt werden:

Brewer Schätzer 1:

$$e_i = \frac{n-1}{n - \frac{\sum_{j \in U} \pi_j^2}{n}} \quad (8.6)$$

Brewer Schätzer 2:

$$e_i = \frac{n-1}{n - \pi_i} \quad (8.7)$$

Brewer Schätzer 3:

$$e_i = \frac{\frac{(n-1)}{n}}{1 - \frac{2\pi_i}{n} + \frac{\sum_{j \in U} \pi_j^2}{n^2}} \quad (8.8)$$

Brewer Schätzer 4:

$$e_i = \frac{\frac{(n-1)}{n}}{1 - \frac{(2n-1)\pi_i}{n(n-1)} + \frac{\sum_{j \in U} \pi_j^2}{n(n-1)}} \quad (8.9)$$

Aus den Darstellungen der einzelnen Schätzer wird ersichtlich, dass mit der Ausnahme von 8.7 alle Varianzapproximationen nach Brewer die Kenntnis über die Inklusionswahrscheinlichkeiten erster Ordnung für alle Elemente der Zielpopulation erfordern (vgl. Matei und Tillé (2005), S. 555f., vgl. Brewer und Donadio (2003), S. 194.). Wie zuvor jedoch bereits erörtert, lassen sich einige der relevanten Parameter zur Berechnung der Inklusionswahrscheinlichkeit erst im Zuge der Erhebung ermitteln. Daher scheiden diejenigen Approximationen, die die Kenntnis der Inklusionswahrscheinlichkeit des gesamten Vektorraumes der Zielpopulation voraussetzen, in der Anwendung auf Telefonstichproben i. d. R. aus. Im Zuge der folgenden Berechnungen wird daher 8.7 in der Anwendung von 8.5 verwendet.⁸³

8.2 Vergleich des Mean Squared Errors für verschiedene Gewichtungsansätze

Innerhalb der Untersuchung eines Schätzers wird oftmals zwischen der Genauigkeit und der Präzision des Ansatzes unterschieden. Lohr (2009) beschreibt die Genauigkeit eines Schätzers über die Nähe des ermittelten Wertes zu dem in der Population vorzufindenden, wahren Wert.

⁸³ Im Zuge des synthetischen Datensatzes wäre auch eine andere Approximation vorstellbar, da für die Inklusionswahrscheinlichkeiten sämtlicher Elemente der Population bekannt sind. Jedoch soll sich folgend nur auf Ansätze beschränkt werden, die in einer realen Telefonerhebung anwendbar sind.

Die Präzision hingegen wird anhand der Varianz ausgedrückt (vgl. S. L. Lohr (2009), S. 32.). Read (1981) bemerkt diesbezüglich, dass diese beiden Gütekriterien anhand der quadrierten Verlustfunktion ausgedrückt werden können. Wird eine Variable Y (im Rahmen einer Erhebung) anhand von $\hat{Y}$ geschätzt, so lässt sich der quadrierte Verlust ($Bias(\hat{Y})^2$) anhand von

$$Bias(\hat{Y})^2 = (E(\hat{Y}) - Y)^2$$

ausdrücken. Der MSE ist nun gerade der Erwartungswert $E((\hat{Y} - Y)^2)$ des $Bias(\hat{Y})^2$. In diesem Zuge ist es gängig, eine Schätzfunktion zu bestimmen, die den geringsten MSE aufweist. Ist ein bestehender Schätzer unverzerrt, sodass $E(\hat{Y}) = Y$ gilt, dann entspricht

$$MSE(\hat{Y}) = Var(\hat{Y}).$$

Grundsätzlich kann dieser jedoch anhand von

$$MSE(\hat{Y}) = Var(\hat{Y}) + (E(\hat{Y}) - Y)^2 \quad (8.10)$$

ausgedrückt werden (vgl. Read (1981), S. 375f.).

Um eine bessere Beurteilung der unterschiedlichen Vorgehensweisen zur Schätzung innerhalb von Dual-Frame-Erhebungen zu ermöglichen, wurden die in dieser Arbeit verwendeten Schätzer vorab anhand des synthetischen Datensatzes für beide Variablen des Datensatzes untersucht. Da die Varianz der unterschiedlichen Schätzer jedoch unbekannt ist und auf einer Approximation beruht, wird lediglich ein Schätzwert des MSEs ($\widehat{MSE}$) verwendet. Tabelle 48 stellt dabei sowohl die approximierte Varianz (des geschätzten Totalwertes), als auch den $Bias(\hat{Y})^2$ (aus der Schätzung des Totalwertes) und den $\widehat{MSE}$ für den Single-Frame-Ansatz (SF), den Composite-Ansatz (CW), den Domain-Specific-Composite-Ansatz (DSC) sowie den Domain-Specific-Single-Frame-Ansatz ($\widehat{DSF}$) gegenüber. Für die letzten beiden Ansätze erfolgte dies zusätzlich unter der Verwendung der effektiven Stichprobengröße zur Bestimmung der Konvexfaktoren ($DSC^*, \widehat{DSF}^*$). Zusätzlich wurde im Rahmen des DSF-Ansatzes noch der Schätzer, der die wahren Präferenzen der Dual-User berücksichtigt, verwendet (DSF).

Tabelle 48 MSE der Variablen im synthetischen Datensatz (in 100 Mio.)

	Variable 1			Variable 2		
Schätzer	$\widehat{Var}_{BR}(\hat{Y})$	$Bias(\hat{Y})^2$	$\widehat{MSE}(\hat{Y})$	$\widehat{Var}_{BR}(\hat{Y})$	$Bias(\hat{Y})^2$	$\widehat{MSE}(\hat{Y})$
SF	7,66	4,51	12,17	4.540,64	36.515,34	41.055,98
CW	9,38	4,59	13,97	4.948,07	45.643,80	50.591,86
<i>DSC</i>	10,25	4,37	14,62	5.440,99	46.958,11	52.399,10
<i>DSC*</i>	9,82	4,43	14,25	5.327,99	43.800,28	49.128,27
DSF	11,07	1,14	12,21	5.075,78	8.910,42	13.986,20
$\widehat{DSF}$	11,88	1,10	12,98	5.429,03	8.729,06	14.158,09
$\widehat{DSF}^*$	11,24	1,13	12,37	5.150,34	8.857,74	14.008,08

Bei dieser Tabelle gilt es zu berücksichtigen, dass es sich um die Schätzung des Totalwertes handelt, für den alle numerischen Ausprägungen in der Population summiert wurden. Daher fallen auch die Schätzwerte des MSE bzw. der Varianz und des $Bias(\hat{Y})^2$ so hoch aus. Weiterhin gilt für den hier berichteten Bias anzumerken, dass es sich um eine Simulation handelt, bei der alle Ausprägungen in der Population bekannt sind. Daher ist Y in diesem Fall bekannt.

Bezüglich der ersten Variable, deren geschätzter MSE im Rahmen dieser Untersuchung für die unterschiedlichen Ansätze verglichen wurde, zeigt sich, dass noch immer der SF-Ansatz zu bevorzugen ist. Dies begründet sich insgesamt durch den niedrigsten MSE in dieser Untersuchung. Jedoch wird unter einer reinen Beschränkung auf den Schätzwert des MSEs ebenfalls ersichtlich, dass alle drei DSF-Ansätze Ergebnisse produzieren, die lediglich geringfügig von demjenigen des SF-Ansatzes abweichen. Dabei ist gerade derjenige DSF-Ansatz zu bevorzugen, der die tatsächlichen Präferenzen verwendet. Die Verwendung der effektiven Stichprobengrößen scheint in dieser Untersuchung allerdings vergleichbare Ergebnisse hervorzubringen. Betrachtet man die beiden Komponenten, aus denen sich der MSE zusammensetzt, so zeigt der direkte Vergleich des SF- und des DSF-Ansatzes, dass die Schätzergebnisse des DSF-Ansatzes zwar deutlich näher an den Populationswerten liegen, dass der Schätzer aber auch eine wesentlich höhere geschätzte Varianz aufweist als der SF-Ansatz. Demnach scheint für diese Variable der DSF-Ansatz zwar wesentlich genauer, jedoch weniger Präzise zu sein, als der SF-Ansatz.

Innerhalb der Untersuchung der zweiten Variablen, deren Ausprägung von den Präferenzen der Dual-User abhängig ist, lässt sich hingegen eindeutig der DSF-Ansatz gegenüber den restlichen Schätzverfahren bevorzugen. Wie bereits in Kapitel 7.2 erwähnt, kann dies mitunter in der Konstruktion des synthetischen Datensatzes und der Integration der Präferenzen der Dual-User in den Ziehungsprozess erklärt werden. Auch hier ist die geschätzte Varianz des DSF-Ansatzes nicht die geringste innerhalb der untersuchten Ansätze, während der $Bias(\hat{Y})^2$ abermals am

niedrigsten ist.

Interessant an diesen Ergebnissen ist daher, dass die nach der Brewer-Approximation geschätzte Varianz für beide Variablen höher als bei anderen Ansätzen ist. Vergleicht man dies mit den Ergebnissen aus dem vorherigen Kapitel und der Auseinandersetzung mit dem modellbasiert geschätzten Designeffekt, so widersprechen sich die beiden Befunde. Im Zuge der Betrachtung der Designeffekte ließ sich festhalten, dass der geringere Designeffekt des DSF-Ansatzes eine geringere Varianz als diejenige der anderen Ansätze vermuten lässt. Hier zeigt sich jedoch, dass diese zumindest im direkten Vergleich mit dem SF-Ansatz höher ist. Dafür können zwei mögliche Gründe angeführt werden. Einmal besteht die Möglichkeit, dass die gewählte Brewer-Approximation die Varianz überschätzt. Ein solches Ergebnis würde dann zu konservativeren Testentscheidungen bspw. im Zuge von inferenziellen Aussagen führen. Eine weitere Möglichkeit besteht darin, dass die modellbasierte Schätzung des Designeffektes fehlerhaft ist. In dem vorliegenden Beispiel würde der Designeffekt daher für den DSF-Ansatz unterschätzt und/oder derjenige der anderen Ansätze falsch eingeschätzt werden. Dies würde dann im vorliegenden Fall zu einem zu hohen Wert der geschätzten effektiven Nettostichprobe führen. Ganninger (2009) bemerkt diesbezüglich, dass die Genauigkeit und Präzision des modellbasierten Designeffektes durchaus sensitiv auf starke Unterschiede in den einzelnen Inklusionswahrscheinlichkeiten reagieren kann (vgl. Ganninger (2009), S. 138.). Eine fehlerhafte Schätzung des zuvor berichteten Designeffektes ist daher nicht auszuschließen.

Da unter den beiden Erhebungen nur vergleichsweise wenige metrische Variablen vorhanden waren, die eine sinnvolle Auseinandersetzung der MSE-Schätzungen ermöglichen, beschränkt sich die folgende Untersuchung lediglich auf die Erhebung Influenza 2014 und deren Ergebnisse bezüglich der Schätzung des Totalwertes der beiden Variablen *Haushaltsgröße* und *Alter*. Für diese Untersuchung wurden sowohl der SF-Ansatz als auch der DSC- und der DSF-Ansatz verwendet. Bezüglich des DSC- und des DSF-Ansatzes wurden die Schätzungen sowohl unter der Verwendung des Einteilungsmodells 1 für Dual-User (DSC_{mod1} ; DSF_{mod1}) als auch des zweiten Modells zur Einteilung verwendet (DSC_{mod2} ; DSF_{mod2}). Weiterhin wurde im Zuge des DSF-Ansatzes noch die alternative Vorgehensweise zur Bestimmung der Konvexfaktoren unter der Verwendung der effektiven Stichprobengröße (DSF_{mod1}^* ; DSF_{mod2}^*) herangezogen. Die Ergebnisse der Untersuchung können in Tabelle 49 abgelesen werden. Zur Bestimmung des Stichprobenschätzwertes Totalwerts für die Variable *Alter* wurde der Mittelwert der in Tabelle 95 angeführten Kategorien separat bestimmt und mit den entsprechenden designgewichteten absoluten Häufigkeiten multipliziert. Die sich daraus ergebende Summe wurde als Schätzwert verwendet.

Tabelle 49 Influenza 2014: MSE für Haushaltsgröße und Alter (in 10 Mrd.)

Schätzer	Haushaltsgröße			Alter		
	$\widehat{Var}_{BR}(\hat{Y})$	$Bias(\hat{Y})^2$	$\widehat{MSE}(\hat{Y})$	$\widehat{Var}_{BR}(\hat{Y})$	$Bias(\hat{Y})^2$	$\widehat{MSE}(\hat{Y})$
SF	87,28	191,01	278,29	10.656,74	9.736.433,74	9.747.090,48
DSC_{mod1}	118,90	80,77	199,67	14.930,93	9.678.127,93	9.693.058,87
DSF_{mod1}	207,58	2.032,20	2.239,79	19.981,46	11.691.797,13	11.711.778,59
DSF^*_{mod1}	199,90	1.058,16	1.258,06	18.967,13	10.941.543,71	10.960.510,84
DSC_{mod2}	119,20	46,65	165,85	14.937,52	9.761.304,87	9.776.242,39
DSF_{mod2}	207,20	2.036,76	2.243,97	20.057,14	11.750.275,81	11.770.332,95
DSF^*_{mod2}	199,59	1.068,93	1.268,52	19.053,31	11.006.537,58	11.025.590,89

Bereits in der Auseinandersetzung des vorherigen Kapitels mit dem DSF-Schätzer in der Anwendung auf die Erhebungsdaten aus Influenza 2014 wurde beobachtet, dass dieser, im Vergleich zur Anwendung auf die synthetischen Daten, dessen Einsatz weniger eindeutig bevorzugt. Die in dieser Untersuchung verwendeten Variablen brachten auch in der vorherigen Untersuchung bereits günstigere Ergebnisse für andere Gewichtungsansätze hervor. Gerade bei der Schätzung des Totalwertes kommt diesbezüglich noch hinzu, dass es in den (numerisch) höheren Kategorien der Variablen zu einer Unterschätzung der entsprechenden Anteilswerte kommt. Daher verwundert es wenig, dass der $Bias(\hat{Y})^2$ für die DSF-Schätzer deutlich oberhalb der anderen Ansätze liegt. Insgesamt kann anhand der hier vorliegenden Untersuchung nicht davon ausgegangen werden, dass einer der DSF-Ansätze gegenüber den anderen Verfahren zu bevorzugen ist.

Bezogen auf die Schätzung der Varianz zeigt sich ein Ergebnis, das mit demjenigen des synthetischen Datensatzes der ersten Variable vergleichbar ist. So ist diese bei allen DSF-Schätzern für beide Modelle oberhalb der anderen Ansätze.

Interessant indes ist, dass innerhalb dieser Untersuchung die Anwendung des DSC-Ansatzes für beide Variablen klar zu bevorzugen ist. So ist dessen $Bias(\hat{Y})^2$ und die geschätzte Varianz innerhalb der Haushaltsgröße geringer als diejenigen der anderen untersuchten Ansätze. Bezüglich der Altersvariable ist zumindest der $Bias(\hat{Y})^2$ sowie der geschätzte MSE geringer. Weiterhin zeigt sich, dass Einteilungsmodell 1 genauere Ergebnisse zu liefern scheint. Der Schluss, den DSC-Ansatz zu bevorzugen, kann innerhalb dieser Untersuchung allerdings auch durch die Wahl der Variablen beeinflusst worden sein. Die beiden Variablen wurden bereits in der vorherigen Untersuchung durch den DSC-Ansatz besser geschätzt.

Auch hier zeigt sich ein Widerspruch zwischen den modellbasiert geschätzten Designeffekten und dem approximierten Schätzer der Varianz. Demnach kann nicht ausgeschlossen werden, dass der modellbasierte Ansatz zur Bestimmung der Designeffekte diesen fehlerhaft wiedergibt.

8.3 Beurteilung der verschiedenen Gewichtungsverfahren

Anhand der vorliegenden Ergebnisse kann gezeigt werden, dass die Verwendung des SF-Ansatzes in einer Dual-Frame-Erhebung als probate Lösung zur Designgewichtung von Erhebungsdaten scheint. So zeigt sich sowohl innerhalb der Untersuchung des Schätzwertes des MSEs als auch in der im vorherigen Kapitel angeführten Auseinandersetzung, dass dieser bezüglich seiner Genauigkeit oftmals Schätzergebnisse liefert, die näher an der Verteilung der Populationswerte liegt, als dies unter anderen Ansätzen zu beobachten ist. Dennoch kann sowohl in der Auseinandersetzung mit dem geschätzten MSE als auch anhand der Beobachtungen der Kapitel 6 und 7, gezeigt werden, dass die Berücksichtigung der Nutzungsgewohnheiten innerhalb der Designgewichtung in einigen Fällen durchaus sinnvoll sein kann. Im Rahmen der Untersuchung der beiden Erhebungen unter der Verwendung einer der beiden Ansätze (DSC; DSF) ließen sich durchaus Beispiele finden, bei denen einer der beiden Ansätze Ergebnisse liefert, die gerade bezüglich ihrer Genauigkeit, dem SF-Ansatz zu bevorzugen sind. Betrachtet man die Kapitel 6 und 7 unabhängig von der Untersuchung des Schätzwertes des MSEs, so ist gerade der DSF-Ansatz aufgrund der höheren Genauigkeit in diesem Kontext zu bevorzugen. Unter der Berücksichtigung des MSEs hingegen liegt der Vorteil des DSC-Ansatzes in dessen geringerem Wert. Wie jedoch bereits diskutiert, kann hier die Variablenauswahl unter Umständen zu fehlgeleiteten Schlüssen führen. Dies liegt darin begründet, dass die Untersuchung des geschätzten MSEs bei Erhebungsdaten sich lediglich auf zwei Variablen stützt. Gerade für diese beiden Variablen sind allerdings auch die vorherigen Schätzergebnisse unter der Verwendung des DSC-Ansatzes günstiger. So ist nicht auszuschließen, dass die Verwendung weiterer Variable zu einer anderen Empfehlung führen könnte. Aufgrund des Mangels an metrischen Variablen innerhalb des Datensatzes ist dies an dieser Stelle jedoch nicht unmittelbar einzusehen.

Die Untersuchung der verschiedenen Gewichtungsansätze anhand des synthetischen Datensatzes deutet hingegen sowohl im Rahmen der Auseinandersetzung mit dem Schätzwert für den MSE als auch innerhalb der verschiedenen Schätzungen darauf hin, dass der DSF-Ansatz eindeutig zu bevorzugen ist, während der DSC-Ansatz erheblich ungenauere Schätzergebnisse hervorbringt. Hierbei gilt jedoch zu beachten, dass die Wahrscheinlichkeit in eine der beiden Stichproben zu gelangen, direkt abhängig von der Präferenz ist, was die Schätzergebnisse durchaus zugunsten des DSF-Ansatzes beeinflussen kann. Dennoch zeigen sowohl die Ergebnisse des vorherigen Kapitels als auch die Untersuchung des geschätzten MSEs, dass der DSF-Ansatz gerade dann, wenn die Ausprägung einer Variable direkt abhängig von der Zugehörigkeit zu einer (bestimmten) Subdomain sowie der Präferenz für ein bestimmtes Medium ist, genauere Ergebnisse hervorbringt, als der SF-Schätzer. Etwas breiter gefasst kann dies auch für die Anwendungsbeispiele innerhalb der Erhebungsdaten aufgefasst werden. So scheint die hier aufgeführte und konstruierte Population, die versucht, ein (stark vereinfachtes) Abbild einer

tatsächlichen Erhebung darzustellen, durchaus dazu in der Lage zu sein, einige Befunde, die in der Anwendung der unterschiedlichen Ansätze auf tatsächliche Erhebungsdaten aufgetaucht sind, begründen zu können. Mit diesen Befunden vereinbar scheint die Erkenntnis, dass es durchaus sinnvoll sein kann, Nutzungsgewohnheiten bei der Gewichtung mitzuberechnen, insbesondere, wenn eine Abhängigkeit dieser mit der Ausprägung einer bestimmten Variable besteht.

Allerdings muss an dieser Stelle ebenfalls festgehalten werden, dass die Unterschiede zwischen den geschätzten Verteilungen einzelner Merkmale unter der Zuhilfenahme der verschiedenen Ansätze oftmals nur gering ausfielen. Dies trifft zumindest in der Punktschätzung zu.

Unter den hier untersuchten Gewichtungsansätzen wurde weiterhin ersichtlich, dass die Verwendung des SF-Ansatzes für alle der hier angeführten Beispiele die geringste geschätzte Varianz (unter der Verwendung der Brewer-Approximation) hervorbringt. Demnach ist dieser bezüglich seiner Präzision zu bevorzugen. Wie bereits in Kapitel 8.2 angemerkt, steht dies im Widerspruch zu den Ergebnissen der Untersuchung des modellbasiert geschätzten Designeffektes. Dieser ist gerade in der Anwendung des DSF-Ansatzes am geringsten. Ein solches Ergebnis suggeriert i. d. R., dass der zugrunde liegende Schätzer auch eine geringere Varianz aufweist als diejenigen Ansätze mit einem höheren Designeffekt. Die hier approximierten Schätzungen der Varianz zeigen jedoch ein deutlich anderes Bild. So weist der DSF-Ansatz, der zuvor den geringsten Designeffekt aufweist, bei allen Beispielen eine (deutlich) höhere Varianz auf als andere untersuchte Ansätze. Demnach ist der niedrigere MSE in denjenigen Beispielen, bei denen ein DSF-Ansatz bevorzugt wird, eindeutig durch die höhere Genauigkeit der ermittelten Schätzwerte erwirkt. Damit kann an dieser Stelle vermutet werden, dass gerade dann, wenn das Untersuchungsmerkmal und die Auswahlwahrscheinlichkeit einer Erhebungseinheit mit deren Präferenz in Abhängigkeit steht, der DSF-Ansatz zwar genauere Schätzwerte (für die Punktschätzer) liefert, diese aber zulasten einer geringeren Präzision zustande kommen.

Insgesamt zeigen die hier präsentierten Ergebnisse, dass die Verwendung des SF-Ansatzes in den meisten Fällen günstige Schätzwerte liefern, da dieser sowohl bezüglich seiner Präzision als auch seiner Genauigkeit durchgehend unter denjenigen Ansätzen zu finden ist, die zu bevorzugen sind. Dennoch konnte, ebenso wie in den vorherigen Kapiteln, gezeigt werden, dass, zumindest bezogen auf die Punktschätzung, die Berücksichtigung der Nutzungsgewohnheiten von Dual-Usern durchaus vorteilhaft sein kann. Dies scheint gerade dann besonders zutreffend zu sein, wenn das zu untersuchende Merkmal und/oder die Auswahlwahrscheinlichkeit einer Zielperson abhängig von diesen Präferenzen ist. Jedoch hat die hier vorliegende Untersuchung ebenfalls gezeigt, dass die gesteigerte Genauigkeit und die damit verbundene, gesteigerte Komplexität des Gewichtungsansatzes oftmals mit einer geringeren Präzision bzw. einer höheren Varianz der Schätzwerte einhergeht. Somit gilt fallweise zu prüfen, ob eine solche Berücksichtigung tatsächlich zu präferierbaren Ergebnissen führt. Hier unberücksichtigt, aber dennoch

von Relevanz ist überdies, dass eine solche Einteilung nach Nutzungsgewohnheiten ebenfalls die Abfrage bestimmter Items bedarf. Daraus resultierend ist daher auch eine gesteigerte Komplexität der Befragung sowie ein höherer zeitlicher Aufwand. Daher sollte noch vor einer beginnenden Dual-Frame-Erhebung entschieden werden, ob die durch den SF-Ansatz bereits erreichten Ergebnisse genügen oder ob ein spezieller Bedarf nach einer Berücksichtigung von Nutzungsgewohnheiten besteht. Ein solcher Bedarf könnte sich bspw. genau dann ergeben, wenn im Vorfeld bereits bekannt ist, dass der Untersuchungsgegenstand stark durch die Präferenz für ein bestimmtes Telekommunikationsmedium beeinflusst wird. Falls dies nicht zutrifft, kann durchaus die Verwendung des bereits etablierten SF-Ansatzes angeraten werden. Bestärkt wird diese Empfehlung insbesondere dadurch, dass die Differenzen in der Genauigkeit zwischen den einzelnen Ansätzen oftmals nur geringfügig sind.

9 Fazit/Ausblick

Die Auseinandersetzung mit der Entwicklung der (sozialwissenschaftlichen) Umfrageforschung über den Modus der telefonischen Befragung bestätigt, dass diese Art der Erhebung auch derzeitig noch immer über eine hohe Relevanz verfügt. Trotz des Aufkommens neuer Arten der Erhebung, wie bspw. Online-Surveys oder der vermehrten Verwendung non-probabilistischer Verfahren, kann unter anderem anhand des Jahresberichtes des ADMs nachgewiesen werden, dass ein nicht zu vernachlässigender Anteil der durchgeführten Befragungen telefonisch durchgeführt wird.

Zur adäquaten Anwendung dieses Modus bedarf es jedoch auch einer angemessenen Auswahlgrundlage, die die Stichprobenziehung aller Personen der Zielpopulation ermöglicht. Die in Kapitel 2 angeführte Auseinandersetzung mit der Entwicklung der Telefonstichprobe in Deutschland legt dar, dass deren Anwendung eng mit der Möglichkeit in Verbindung steht, Festnetzstichproben mit einem akzeptablen Aufwand bereitzustellen. Überdies ist für diese Entwicklung bedeutend, dass die Auswahlgrundlage der Festnetzstichprobe (annähernd) die gesamte Zielpopulation abdeckt. Daher wurden diese anfangs anhand einer Listenauswahl basierend auf Telefonregistern gezogen. Sowohl durch politische als auch aufgrund gesetzlicher und soziodemografischer Änderungen wurde jedoch eine Abkehr von dieser Listenauswahl erzwungen, die die Notwendigkeit weitreichender Überlegungen zur Sicherstellung der weiteren Tauglichkeit dieses Befragungsmodus mit sich brachte. Gerade die Aufhebung der Eintragspflicht in Verbindung mit einem zunehmenden Anteil der Bevölkerung, der nicht in einem Telefonverzeichnis geführt wurde, und die Wiedervereinigung Deutschlands, führten somit zur Untauglichkeit des Telefonverzeichnisses als alleinige Grundlage der Stichprobenziehung. Die hieraus resultierende Notwendigkeit, eine Auswahlgrundlage zu schaffen, die mit einem vergleichsweise geringen Aufwand die Ziehung von sowohl eingetragenen als auch nicht eingetragenen Rufnummern ermöglicht, führte zu dem durch Gabler und Häder Mitte der 1990er Jahre entwickelten Verfahren. Dieses Verfahren erfüllt gleichzeitig die Nebenbedingung, dass alle Rufnummern mit gleicher Auswahlwahrscheinlichkeit gezogen werden. Aufgrund der Eigenschaften des Modells wird es noch immer für Festnetzstichproben in Deutschland in dieser oder ähnlicher Form angewendet.

Veranlasst durch die weite Verbreitung privat genutzter Mobilfunkanschlüsse zeigt sich jedoch gerade in den vergangenen Jahren, dass auch diese Auswahlgrundlage nicht mehr hinreichend ist, um die bundesdeutsche Bevölkerung als Zielpopulation abzudecken. Verursacht wird dies durch den Anteil der Population, der ausschließlich über das Mobilfunknetz erreichbar ist. Bereits in Kapitel 2.3.2 dieser Arbeit wurde darüber berichtet, dass sich gerade diese Haushalte in ihren soziodemografischen Merkmalen von denjenigen, die über das Festnetz ausgewählt werden können,

unterscheiden. Daher kann, nach der Auseinandersetzung der bereits bestehenden, einschlägigen Forschung, grundsätzlich für Untersuchungen, die die gesamte (bundesdeutsche) Bevölkerung als Zielpopulation haben, die Durchführung von Dual-Frame-Erhebungen angeraten werden. Solche Erhebungen setzen sich aus einer oder mehreren unabhängigen Festnetz- und Mobilfunkstichproben zusammen. Durch deren Anwendung kann zumindest gewährleistet werden, dass alle Zielpersonen, die telefonisch kontaktiert werden können, eine Auswahlwahrscheinlichkeit größer Null haben. Die Verwendung von Dual-Frame-Erhebungen dient somit der Reduktion von Verzerrungen der Schätzung aufgrund von Undercoverage.

Nicht einzig aufgrund der Verwendung von Dual-Frame-Erhebungen besteht die Notwendigkeit, bei der Schätzung von Populationsstatistiken auf der Basis von Erhebungsdaten aus telefonischen Befragungen eine Designgewichtung durchzuführen. So lassen sich auch für Erhebungen aus einer der beiden Auswahlgrundlagen i. d. R. Unterschiede in der Auswahlwahrscheinlichkeit einzelner Zielpersonen festhalten. Bei Festnetzstichproben liegt dies gewöhnlich in dem zweistufigen Vorgehen der Personenauswahl und der unterschiedlichen Anzahl von Rufnummern, unter denen ein Haushalt oder eine Zielperson erreichbar ist, begründet. Mobilfunkstichproben werden allgemein als direkte Personenstichproben verstanden, bei denen die Kontaktperson der Zielperson entspricht. Aber auch bei diesen besteht die Möglichkeit, dass eine Person über mehrere Anschlüsse in die Stichprobe gelangen kann.

Bei Dual-Frame-Erhebungen kommt, je nach Art und Weise der Durchführung der Erhebung, indes noch hinzu, dass Personen, die über beide Telekommunikationsmedien verfügen, eine höhere Auswahlwahrscheinlichkeit haben, als diejenigen, die nur über einen bestimmten Auswahlrahmen in die Stichprobe gelangen können. In Kapitel 3.3 konnte gezeigt werden, dass eine separate Gewichtung und Schätzung der Populationswerte auf der Basis der beiden einzelnen Stichproben bei einem Overlap-Ansatz dazu führen kann, dass gerade die Populationswerte im Überlappungsbereich überschätzt wird. In der deutsch- sowie auch in der englischsprachigen Literatur werden zum Umgang mit diesem Problem mehrere Lösungsansätze angeboten. Gerade im deutschsprachigen Raum ist, unter anderem aufgrund der Beiträge, die aus den Ergebnissen der CELLA-Erhebungen hervorgekommen sind, insbesondere der sog. Single-Frame-Ansatz prävalent. Bei diesem wird der Überlappungsbereich der beiden Auswahlrahmen als eine Art "Super-Stratum" betrachtet und für die jeweiligen Personen aus dieser Domain die gemeinsame Inklusionswahrscheinlichkeit über beide Auswahlgrundlagen berechnet. Problematisch an einem solchen Vorgehen kann sein, dass für alle Personen aus dieser Subpopulation alle Parameter, die zur Bestimmung der Inklusionswahrscheinlichkeit über beide Auswahlgrundlagen relevant sind, erhoben werden müssen. Das Erheben aller relevanten Parameter kann unterdessen die Komplexität der Befragung sowie deren Dauer erhöhen.

Ohne diese Angaben über beide Auswahlgrundlagen kommt hingegen der sog. Composite-Gewichtungsansatz aus. Bei diesem werden lediglich die Parameter benötigt, die für die Stich-

probe, aus der eine Zielperson stammt, relevant sind. Die designgewichteten Schätzwerte des Überlappungsbereiches werden mittels Konvexkombination nachfolgend summiert. Zur Wahl des in diesem Ansatz verwendeten Composite-Faktors, der zwischen 0 und 1 liegt, gibt es unterdessen mehrere Vorschläge. In Kapitel 3.3 wurden einige dieser eingehend erläutert und deren Vor- und Nachteile beschrieben. Problematisch an einigen der vorgeschlagenen Ansätzen ist die Verletzung der internen Konsistenz der Gewichtung. Eine solche Verletzung kann dazu führen, dass unterschiedliche Gewichtungsfaktoren für verschiedene Untersuchungsvariablen entstehen. In einer Erhebung, die mehrere Untersuchungsgegenstände beinhaltet, kann dies zu Komplikationen in der Anwendung führen. Ein vergleichsweise einfaches Vorgehen ist daher, den Composite-Faktor so zu bestimmen, dass er für alle Variablen gleich ist. Als mögliche Berechnungsgrundlage wurde daher im weiteren Verlauf dieser Arbeit das Verhältnis der abgeschlossenen Interviews einer Stichprobe an denjenigen der gesamten Erhebung verwendet. Eine zweite Möglichkeit, die in dieser Arbeit näher betrachtet wurde, ist die Verwendung der effektiven Stichprobengröße bei der Berechnung dieses Verhältnisses.

Ein in Kapitel 3.5.1 durchgeführter Vergleich der beiden Gewichtungsverfahren zeigt lediglich geringfügige Unterschiede zwischen den beiden Ansätzen. Begründet werden kann dieses Ergebnis damit, dass es sich bei beiden um näherungsweise erwartungstreue Schätzer handelt. Dennoch kann unter anderem anhand des modellbasiert geschätzten Designeffektes gezeigt werden, dass der Single-Frame-Ansatz unter diesen beiden zu bevorzugen ist.

Beiden Gewichtungsansätzen ist jedoch gemein, dass dem Telekommunikationsmedium, das mit der Auswahlgrundlage in Verbindung steht, selbst kein Einfluss auf das Teilnahmeverhalten der Zielpersonen eingeräumt wird. Implizit besteht somit die Annahme, dass die Auswahlgrundlage, die gewählt wurde, um eine Zielperson des Überlappungsbereiches zu kontaktieren, unter der Berücksichtigung der (einzelnen) Auswahlwahrscheinlichkeiten, austauschbar ist. Die innerhalb von Kapitel 5.1 angeführten Beispiele zeigen jedoch, dass es durchaus Unterschiede bezüglich des Teilnahmeverhaltens in Abhängigkeit des gewählten Telekommunikationsmediums gibt und, dass dieses über Festnetz- und Mobilfunkstichprobe nicht gleich verteilt ist. Die in der vorliegenden Arbeit erörterte Forschungsfrage beschäftigt sich daher gerade mit der Berücksichtigung des Teilnahmeverhaltens, das vom Telekommunikationsmedium abhängig ist. Dazu wurde die These angeführt, dass sich die Präferenzen und Nutzungsgewohnheiten von Dual-Usern bezüglich der unterschiedlichen Telekommunikationsmedien unterscheiden können. Um dieses Kontinuum besser messen und analysieren zu können, wurde vereinfachend angenommen, dass sich die Teilpopulation der Dual-User in drei Subdomains einteilen lässt. So unterscheidet das in 5.2 erläuterte Modell zwischen Landline-Mainly, wahren Dual-Usern und Mobile-Mainly. Bei der ersten Subdomain handelt es sich um diejenigen, die zwar grundsätzlich über beide Telekommunikationsmedien erreichbar sind, ihren Festnetzanschluss jedoch präferiert verwenden. Umgekehrtes wird für die Gruppe der Mobile-Mainly angenommen. Für wahre Dual-User

hingegen lässt sich keine eindeutige Präferenz für ein bestimmtes Telekommunikationsmittel feststellen.

Die in der vorliegenden Arbeit eingeführte Aufteilung der Zielpersonen innerhalb des Überlappungsbereiches der beiden Auswahlgrundlagen wurde anhand der beiden Erhebungen CELLA 2 und Influenza 2014 eingehend geprüft. Von Interesse war diesbezüglich einmal, anhand welcher Variablen eine solche Einteilung erfolgen kann und wie sich Personen innerhalb der jeweiligen Subdomain anhand ihrer soziodemografischen Merkmale unterscheiden.

Zum Zeitpunkt, als die Erhebung CELLA 2 durchgeführt wurde, ist eine solche Einteilung der Dual-User anhand ihrer Nutzungsgewohnheiten nicht als Untersuchungsgegenstand implementiert worden. Daher musste diese anhand von Proxys durchgeführt werden. Die Einteilung erfolgte deshalb unter der Verwendung der Fragen nach der täglichen Dauer, die das Mobiltelefon eines Dual-Users eingeschaltet ist, der Häufigkeit, dass dieses mitgeführt wird sowie der Information über die Anzahl der virtuellen und physischen Festnetzanschlüsse, über die ein Dual-User erreichbar ist. Anhand dieser Informationen wurden mehrere Modelle zur Einteilung mit sinkendem Anteil der wahren Dual-User gebildet. Das Vorgehen lässt sich durch die Unkenntnis der Tauglichkeit dieser Variablen sowie der Verteilung nach diesen drei Subdomains innerhalb der Population begründen. Allerdings konnte mit Ausnahme des letzten Modells, das keine wahren Dual-User beinhaltet, auch bezogen auf spätere Anwendungen, keines der Modelle eindeutig bevorzugt werden.

Konträr dazu wurde im Zuge von Influenza 2014 eine Einteilung anhand von dafür erhobenen Variablen vorgenommen. Um dies zu ermöglichen, wurden Dual-User nach ihrer Einschätzung darüber befragt, über welches Telekommunikationsmedium diese grundsätzlich besser erreichbar sind. Weiterhin wurde erfragt, wie diese ihrer Erreichbarkeit innerhalb der letzten beiden Stunden vor der Befragung getrennt nach Festnetz und Mobilfunk einschätzen. Da das Problem der unbekannten Verteilung der einzelnen Subdomains innerhalb der Population noch immer besteht, wurde auch bei Influenza 2014 eine Vorgehensweise zur Einteilung gewählt, die mit derjenigen der Erhebung CELLA 2 vergleichbar ist. Somit wurden vier unterschiedliche Einteilungsmodelle präsentiert, bei denen jeweils der Anteil der wahren Dual-User zugunsten der beiden Mainly-Kategorien verkleinert wurde. Im Gegensatz zu der Einteilung innerhalb der Erhebung CELLA 2 wurde hier deutlich, dass diejenigen Modelle, die eher einen geringeren Anteil von Mainlys beinhalten, favorisiert werden können. Die in dieser Arbeit vorliegenden Ergebnisse scheinen dafür zu sprechen, nur diejenigen den jeweiligen Mainly-Domains zuzuordnen, die angaben, deutlich besser über ein bestimmtes Telekommunikationsmedium erreichbar zu sein. Eine mögliche Ergänzung, die vergleichbare Resultate hervorbringt, wäre diesen Subdomains noch diejenigen Zielpersonen hinzuzufügen, die angaben, grundsätzlich (etwas) besser über ein bestimmtes Telekommunikationsmedium erreichbar zu sein und es gleichzeitig als sehr unwahrscheinlich einschätzten innerhalb der letzten beiden Stunden vor der Befragung über

das andere Medium erreicht worden zu sein.

Zur Überprüfung der Einteilung, die für die Zielpersonen aus CELLA 2 vorgenommen wurde, beinhaltete der Fragebogen von Influenza 2014 auch eine der Variablen, die zuvor verwendet wurde. Da lediglich eine der Variablen nochmals erhoben wurde, waren die Vergleichsmöglichkeiten der Einteilung jedoch nur eingeschränkt möglich. Dennoch konnte anhand des Vergleichs bereits gezeigt werden, dass es unter beiden Vorgehensweisen durchaus Unterschiede gibt. So konnte beobachtet werden, dass Personen, die innerhalb von Influenza 2014 einer bestimmten Subdomain zugeordnet wurden, anhand der Einteilung die für CELLA 2 durchgeführt wurde, nicht zu der gleichen Subdomain gehörten. Neben der Eignung dieser Proxys kann indes auch diskutiert werden, ob sich das Nutzungsverhalten von Mobiltelefonen unter Umständen innerhalb des Zeitraumes zwischen den beiden Erhebungen geändert hat. Für dieses Argument kann angeführt werden, dass durch die zunehmende Verbreitung von Smartphones eine Loslösung des Mobilfunkgerätes als reines Telekommunikationsmedium hin zum Multimediagerät stattgefunden haben kann. Daher ist durchaus bestreitbar, dass eine Person bspw. zur Nutzung mobiler Applikationen ihr Mobilfunkgerät dauerhaft empfangsbereit hat, auch wenn diese den Kontakt über ihren Festnetzanschluss präferiert.

Von gesteigertem Interesse innerhalb der hier vorliegenden Untersuchung war ebenfalls, in welchem Ausmaß Personen innerhalb der jeweiligen Subdomains hinsichtlich ihrer soziodemografischen Charakteristika differenzierbar sind. Für beide Erhebungen erfolgte daher nach der Einteilung eine Untersuchung der Verteilung der soziodemografischen Merkmale getrennt nach Subdomain sowie der Vergleich untereinander, zur Gesamterhebung und zu den Angaben der amtlichen Statistik. Die Verteilung der soziodemografischen Merkmale wurde sowohl anhand der Daten der Erhebung CELLA 2 als auch an denjenigen von Influenza 2014 geprüft.

In beiden Erhebungen konnte beobachtet werden, dass die Verteilung soziodemografischer Merkmale innerhalb der Mainly-Domains in ihrer Tendenz mit denjenigen der entsprechenden Only-Kategorien vergleichbar sind. Gewöhnlich ist diese Tendenz jedoch weniger stark artikuliert. Aus vorherigen Ergebnissen war bereits bekannt, dass Mobile-Onlys tendenziell eher jüngere, männliche Personen aus Einpersonenhaushalten mit niedrigem bis mittlerem formalem Bildungsabschluss sind. Landline-Only sind hingegen eher ältere, weibliche Personen aus Ein- oder Zweipersonenhaushalten. Für diejenigen, die in einer der beiden Erhebungen den Landline-Mainly zugeordnet wurden, gilt, dass diese eher Frauen höheren Alters sind, die in Zweipersonenhaushalten leben. Für Mobile-Mainly hingegen konnte festgestellt werden, dass diese eher jünger und männlich sind. Weiterhin lebt ein höherer Anteil dieser Personen in Einpersonenhaushalten.

Einzig die Untersuchung der Bildungsvariablen war in beiden Erhebungen problematisch. So konnte lediglich für Landline-Only festgestellt werden, dass für diese eher ein niedriger formaler Bildungsabschluss prävalent ist. Die Ergebnisse der Mobile-Only aus CELLA 2 konnten, bezogen auf die Bildungsvariable, hingegen nicht direkt in Influenza 2014 wiedergegeben werden. So

waren in dieser Kategorie zwar vermehrt Personen mit niedrigem Bildungsabschluss zu finden, jedoch war der Anteil der formal höher gebildeten ebenfalls deutlich oberhalb desjenigen der Gesamterhebung. Ähnliche Schwierigkeiten ließen sich für die beiden Mainly-Domains beobachten. In beiden Erhebungen war die Verteilung der Bildungsvariablen der Mainly-Domains nur eingeschränkt mit derjenigen der jeweiligen Onlys sowie über die Erhebungen vergleichbar.

Der Vergleich konnte auch zeigen, dass die Verteilung der soziodemografischen Charakteristika der wahren Dual-User oftmals ähnlich derjenigen der Gesamterhebung war. Als mögliche Begründung dieses Ergebnisses wurden in dieser Arbeit zwei Ursachen angeführt. Einmal ist es nicht auszuschließen, dass die fehlende Präferenz für ein bestimmtes Telekommunikationsmedium gleichverteilt über alle Ausprägungen der einzelnen soziodemografischen Charakteristika ist. Des Weiteren kann die Größe der Subdomain der wahren Dual-User einen Effekt auf diese Ergebnisse haben. Dadurch, dass die Subdomain der wahren Dual-User i. d. R. deutlich größer als die beiden Mainly-Kategorien war, ist nicht auszuschließen, dass einzelne Ausreißer in der Verteilung einen geringeren Effekt auf deren Gesamtverteilung hatten. Beide Ursachen sind jedoch einzeln oder in Kombination denkbar.

Um die Annahme zu prüfen, dass Mainlys mit einem geringeren Aufwand über das von ihnen präferierte Telekommunikationsmedium zu erheben sind, wurden die Kontaktergebnisse aus Influenza 2014 bis zum Abschluss eines erfolgreichen Interviews getrennt nach Domain untersucht. Grundsätzlich konnte diese Untersuchung die innerhalb von Kapitel 5.2 bereits geäußerten Vermutungen bestätigen. Für Landline-Mainly konnte gezeigt werden, dass diese tendenziell weniger Kontaktversuche benötigen als über die gesamte Festnetzstichprobe, wenn sie auch über das Festnetz kontaktiert wurden. Umgekehrt ergab sich bei der Untersuchung der Mobilfunkstichprobe, dass Landline-Mainly häufiger über ihren Mobilfunkanschluss kontaktiert werden mussten, wenn sie in der Stichprobe enthalten waren. Auch für Mobile-Mainly wurde ein vergleichbares Ergebnis innerhalb der Festnetzstichprobe beobachtet. Die Untersuchung der Mobilfunkstichprobe zeigte jedoch gerade für die Mobile-Mainly ein weniger deutliches Bild. Begründet wurde diese Abweichung anhand des vergleichsweise hohen Anteils der Mobile-Only die entweder direkt beim ersten Kontaktversuch ein erfolgreiches Interview abschlossen und/oder durch diejenigen, die dafür zehn Kontaktversuche benötigten. Dadurch wurde auch die relative Verteilung der Gesamtstichprobe entsprechend beeinflusst. Dennoch kann auch hier festgehalten werden, dass Mobile-Mainly im direkten Vergleich bspw. zu den Landline-Mainly weniger Kontaktversuche benötigen, wenn sie über ihren Mobilfunkanschluss kontaktiert wurden.

Von gesteigertem Interesse innerhalb dieser Arbeit war ebenfalls die Frage, wie die Kenntnis dieser Subdomains im Rahmen der Schätzung bzw. innerhalb des dazugehörigen Gewichtungsansatzes implementiert werden kann, um die Genauigkeit und/oder die Präzision der geschätzten Werte zu erhöhen. Um dies zu ermöglichen, wurden im Rahmen dieser Arbeit

zwei unterschiedliche Vorgehen vorgestellt. Der *Domain-Specific-Composite-Ansatz* stellt dabei eine Möglichkeit dar, den zuvor bereits beschriebenen Composite-Ansätze auf die jeweiligen Subdomains anzuwenden. Dazu wurden nach der Einteilung der Dual-User in die jeweiligen Teilpopulationen separat sowohl für Landline- und Mobile-Mainly als auch für wahre Dual-User Konvexfaktoren bestimmt. Als Berechnungsgrundlage wurde dafür der Anteil der abgeschlossenen Interviews im Festnetz an der Gesamtmenge der Befragungen innerhalb der Domain verwendet. Ein weiterer Konvexfaktor wurde anhand der geschätzten effektiven Stichprobengröße berechnet. Die originären Designgewichte der Festnetzstichprobe wurden für die Befragten, die aus dieser Stichprobe stammen, dann mit dem entsprechenden Faktor multipliziert. Für die Befragten der Mobilfunkstichprobe wurde das gleiche Verfahren mit dem Designgewicht der Mobilfunkstichprobe und dem Wert von 1 minus dem Konvexfaktor, angewendet. Aus der Summe beider Schätzwerte der einzelnen Stichproben wurde dann ein gemeinsamer Wert bestimmt.

Der Vergleich dieses Ansatzes mit dem "gewöhnlichen" Composite-Ansatz sowie dem Single-Frame-Ansatz zeigte jedoch sowohl in der Anwendung innerhalb der Simulation als auch auf die Daten der beiden Erhebungen, dass dieses Vorgehen Schätzwerte hervorbringt, die, wenn auch nur geringfügig, weniger genau als diejenigen des Single-Frame-Ansatzes sind. Das gleiche Bild zeichnet sich ebenfalls für den modellbasiert geschätzten Designeffekt ab. Dessen Wert war bei beiden Erhebungen geringer bei der Verwendung des Single-Frame-Ansatzes. Somit wurde vermutet, dass die Verwendung des Single-Frame-Ansatzes bezogen auf dessen Genauigkeit und Präzision zu bevorzugen ist. Dennoch konnte festgehalten werden, dass zumindest die Unterschiede innerhalb der Genauigkeit der Schätzung lediglich geringfügig sind.

Dem *Domain-Specific-Single-Frame-Ansatz* hingegen geht die Vermutung voraus, dass die Berücksichtigung der Präferenz eines Dual-Users für ein bestimmtes Telekommunikationsmedium im Zuge der Berechnung der Auswahlwahrscheinlichkeit gerade diese genauer abbilden kann. Die Idee dabei ist, dass neben den Faktoren, die die tatsächliche Auswahl einer Zielperson bestimmen, ebenfalls die Präferenz eines Dual-Users für ein bestimmtes Telekommunikationsmedium dessen Wahrscheinlichkeit beeinflusst, in der Stichprobe enthalten zu sein. Die Inklusionswahrscheinlichkeit einer Zielperson wird in diesem Ansatz in ähnlicher Form bestimmt wie die des gewöhnlichen Single-Frame-Ansatzes. Sie ergibt sich aus der Summe der beiden einzelnen Wahrscheinlichkeiten der jeweiligen Auswahlgrundlagen. Jedoch wird die Auswahlwahrscheinlichkeit nach Telekommunikationsmedium zuvor noch mit der entsprechenden Präferenz oder deren Schätzwert multipliziert.

Da diese Präferenz numerisch nur schwer in einer Erhebung zu messen ist, wurde anhand des synthetischen Datensatzes getestet, ob anstelle der tatsächlichen Präferenz auch die zuvor beschriebenen Konvexfaktoren für Festnetz und Mobilfunk als Schätzwert verwendet werden können. Der synthetische Datensatz eignete sich gerade deshalb für diese Überprüfung, da für

jedes Element der Auswahlgesamtheit die wahre Präferenz bekannt ist. Die Schätzungen unter der Verwendung der wahren Präferenz und unter der Verwendung der Composite-Faktoren weichen im Ergebnis nur marginal voneinander ab. Weiterhin ist die modellbasierte Schätzung des Designeffektes nahezu identisch.

Bezüglich der Punktschätzer konnte sowohl in der Anwendung auf die synthetischen Daten als auch bei derjenigen auf die beiden Erhebungen gezeigt werden, dass deren Genauigkeit in den meisten Fällen höher ist, als unter der Verwendung eines anderen Gewichtungsverfahrens. Dies scheint insbesondere dann zuzutreffen, wenn die zu schätzende Variable abhängig von der Zugehörigkeit einer Zielperson zu einer bestimmten Domain ist. Zusätzlich konnte gezeigt werden, dass der geschätzte Designeffekt unter der Verwendung des Domain-Specific-Single-Frame-Ansatzes geringer ausfällt. Unter der Betrachtung dieser Ergebnisse scheint die Anwendung dieses Verfahrens daher zu bevorzugen zu sein.

Da neben der Genauigkeit eines Schätzers ebenfalls dessen Präzision relevant ist, wurde abschließend eine Untersuchung der hier vorgestellten Ansätze anhand des geschätzten Mean Squared Errors vorgenommen. Dieser berücksichtigt unterdessen gerade die Varianz eines Schätzwertes. Im Zuge von komplexen Erhebungsverfahren und unter der Verwendung des Horvitz-Thompson-Schätzers ist diese jedoch nur schwer zu bestimmen. Deshalb wurde für die Varianz eine Approximation nach Brewer durchgeführt. Die Untersuchung der Schätzverfahren erfolgte dann unter der Verwendung der Daten des synthetischen Datensatzes sowie für Variablen der Erhebung Influenza 2014. Nur wenige metrische Variablen sind in dieser Erhebung vorhanden. Infolgedessen wurde die Untersuchung auf die beiden Variablen Haushaltsgröße und Alter beschränkt.

Als Ergebnis der Überprüfung bestätigt sich, dass die Anwendung des gewöhnlichen Single-Frame-Ansatzes in der Anwendung auf beide Datengrundlagen konstant gute Ergebnisse im Vergleich zu den restlichen Schätzern hervorbringt. So konnten sowohl die Komponente der quadrierten Abweichung als auch für die Varianz-Komponente des MSEs unter der Verwendung des Single-Frame-Ansatzes relativ geringe Werte beobachtet werden. Gerade die approximierte Varianz dieses Schätzers war innerhalb aller Anwendungen am geringsten unter allen untersuchten Ansätzen. Dies ist zumindest mit den Ergebnissen der Untersuchungen aus Kapitel 3 und 6 konform, da auch innerhalb dieser Untersuchungen für den Single-Frame-Ansatz ein geringerer Designeffekt geschätzt wurde. Widersprüchlich indes sind dazu die Ergebnisse aus der Untersuchung des Domain-Specific-Single-Frame-Ansatzes. Für diesen wurde, basierend auf den Ergebnissen aus Kapitel 7, eine höhere Genauigkeit und Präzision angenommen.

Anhand der vorliegenden Untersuchung des MSEs konnte lediglich in der Anwendung des synthetischen Datensatzes ein geringerer MSE als derjenige des Single-Frame-Ansatzes festgestellt werden. Weiterhin weist gerade dieser Ansatz erheblich höhere Werte der approximated Varianzschätzung auf. Demnach scheinen die in Kapitel 7 berichteten Schätzwerte des Design-

effektes in diesem speziellen Fall zu fehlgeleiteten Schlüssen zu führen.

Darüber hinaus ist der MSE des Domain-Specific-Single-Frame-Ansatzes in der Anwendung auf Erhebungsdaten der Erhebung Influenza 2014 insgesamt deutlich höher als derjenige der anderen Gewichtungsverfahren. Bei den beiden hier untersuchten Variablen scheint, gegenläufig zu den vorherigen Untersuchungsergebnissen, gerade die Anwendung des Domain-Specific-Composite-Ansatzes zu bevorzugen zu sein.

Die abweichenden Ergebnisse zwischen der alleinigen Betrachtung der Punktschätzer des DSF-Ansatzes und derjenigen unter der Verwendung des MSEs lassen sich zumindest bei der Erhebung Influenza 2014 auch anhand der Variablenauswahl begründen. Da innerhalb der beiden betrachteten Erhebungen nur wenige metrische Variablen vorhanden sind, beschränkt sich der Vergleich des MSEs ausschließlich auf die beiden Variablen Alter und Haushaltsgröße. Gerade für diese beiden Variablen konnte eine höhere Abweichung des DSF-Ansatzes zu den Angaben der amtlichen Statistik in den oberen Kategorien beobachtet werden. Da der MSE diese Abweichung in quadrierter Form berücksichtigt, macht sich diese gerade für den DSF-Ansatz stärker bemerkbar.

Als mögliche Begründung der Unterschiede zwischen der geschätzten Varianz und dem geschätzten Designeffekt lassen sich zwei Ursachen anführen. Einmal basieren beide Werte auf vereinfachten Annahmen, die sich für die jeweilige Schätzung ergeben. Somit ist es durchaus auch denkbar, dass der approximierte Varianzschätzer die tatsächliche Varianz überschätzt. Ein solches Resultat würde dann i. d. R. zu einer konservativeren Testentscheidung führen. Andererseits ist nicht auszuschließen, dass die Schätzwerte des Designeffektes ein fehlerhaftes Bild abgeben. So berichtet bspw. bereits Ganninger (2009), dass die modellbasierte Schätzung des Designeffektes durchaus sensitiv auf starke Unterschiede in den einzelnen Auswahlwahrscheinlichkeiten reagieren. Da gerade dies durch die zusätzliche Einführung der geschätzten Präferenzen geschieht, kann nicht ausgeschlossen werden, dass die geschätzten Designeffekte in ihrer Genauigkeit beeinflusst werden.

Abschließend lässt sich anhand der vorliegenden Untersuchungen demnach festhalten, dass die Anwendung des Single-Frame-Ansatzes grundsätzlich anzuraten ist, da dieser sowohl bezüglich seiner Genauigkeit als auch aufgrund seiner Präzision zuverlässig gute Schätzergebnisse hervorbringt. Jedoch konnte auch gezeigt werden, dass gerade unter der Betrachtung des MSEs für jegliche Untersuchungsgegenstände einer der Ansätze, der die Dual-User anhand von Präferenzen und Nutzungsgewohnheiten einteilt und in der Schätzung entsprechend berücksichtigt, favorisierbare Ergebnisse hervorbringt. Die Wahl des Ansatzes fällt indes in Abhängigkeit der Datengrundlage unterschiedlich aus. Somit müsste, wenn Präferenzen und Nutzungsgewohnheiten in der Schätzung berücksichtigt werden sollten, fallweise geprüft werden, welcher Ansatz verwendet wird.

Da die Präzision eines Schätzers überdies eine hohe Relevanz hat und gezeigt wurde, dass

diejenige des Domain-Specific-Single-Frame-Ansatz für die untersuchten Variablen am geringsten war, sollte dessen Wahl nicht leichtfertig getroffen werden. Weiterhin besteht die Gefahr, dass eine fallweise Entscheidung für ein bestimmtes Verfahren, in Abhängigkeit der betrachteten Variable, die interne Konsistenz der Schätzung beeinträchtigt. Daher scheint die Anwendung des regulären Single-Frame-Ansatzes unter den Gesichtspunkten der Schätzung ein zuverlässiger "Mittelweg" zu sein, um vergleichsweise genaue und präzise Schätzergebnisse zu erhalten.

Ein weiterer Vorzug besteht darin, dass keine zusätzlichen Variablen zur Einordnung von Dual-Usern einzig zum Zweck der Gewichtung erhoben werden müssen. Gerade eine solche Abfrage könnte, je nach Ausrichtung der Befragung, Zielpersonen irritieren und/oder die reine Befragungszeit erhöhen und dadurch eine negative Auswirkung auf die gesamten Ergebnisse der Befragung haben.

Andererseits konnte anhand der vorliegenden Ergebnisse durchaus gezeigt werden, dass es Unterschiede innerhalb der Dual-User bezüglich ihrer Präferenz für ein bestimmtes Telekommunikationsmedium und dessen Nutzungsgewohnheiten zu geben scheint. Weiterhin ließ sich beobachten, dass diese Information gleichwohl dienlich bei der Untersuchung der Kontaktsergebnisse zu sein scheint und, dass Dual-User i. d. R. einfacher über das Telekommunikationsmedium zu kontaktieren sind, das diese präferieren. Überdies wurde herausgearbeitet, dass sich Personen innerhalb der jeweiligen Subdomains sehr wohl in ihren soziodemografischen Merkmalen unterscheiden. Demnach scheint die Einteilung nach Nutzungsgewohnheiten und Präferenzen für bestimmte Telekommunikationsmedien durchaus sinnvoll für verschiedene Analysezwecke zu sein.

Gerade unter dem Aspekt des *Tailored Designs* ist die Erkenntnis, dass diejenigen, die eine Präferenz für ein bestimmtes Telekommunikationsmedium, oder weiter gefasst, für ein bestimmtes Erhebungsinstrument, haben, in Verbindung mit der Erkenntnis, dass sich Mainlys in ihren soziodemografischen Charakteristika unterscheiden, nicht uninteressant. So könnte, losgelöst von telefonischen Befragungen, eine weitere Untersuchung bezüglich der Kombination von Web-Surveys und Face-to-Face oder postalischen Befragungen vorgenommen werden. Wäre in diesem Zusammenhang anhand von soziodemografischen Charakteristika bspw. beobachtbar, welche der Befragten eher eine Tendenz zur Online-Befragung haben und welche eher postalisch oder persönlich-mündlich zu befragen sind, könnte dies vorteilhaft gerade für die Kosten der Befragung und u. U. für das Teilnahmeverhalten sein. So ist denkbar, dass auf der Basis einer Registerstichprobe für diejenigen, bei denen es aufgrund ihrer soziodemografischen Merkmale (die im Register enthalten sind) wahrscheinlicher ist, dass diese die Online-Befragung ohnehin bevorzugen, auch lediglich diese Option (nach der Rekrutierung) angeboten werden. Die genaue Implementierung eines solchen Vorgehens sowie deren Anwendung ist jedoch Gegenstand künftiger Untersuchungen.

Literatur

- AAPOR Cell Phone Task Force (2010). *New Considerations for Survey Researchers when Planning and Conducting RDD Telephone Surveys in the U.S. with Respondents Reached via Cell Phone*. Techn. Ber. AAPOR, American Association for Public Opinion Research.
- ADM, Arbeitskreis Deutscher Markt und Sozialforschungsinstitute (2015). *Jahresbericht 2014*.
- ADM, Arbeitskreis Deutscher Markt und Sozialforschungsinstitute (2016). *Richtlinie für die Befragung von Minderjährigen*.
- ADM, Arbeitskreis Deutscher Markt und Sozialforschungsinstitute (2017). *Jahresbericht 2016*.
- ADM, Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute (2012). *Forschungsprojekt Dual-Frame-Ansätze - Forschungsbericht*. Techn. Ber. ADM Forschungsprojekt Dual-Frame-Ansätze.
- American Association for Public Opinion Research, AAPOR (2011). *Standard Definitions*. Techn. Ber. AAPOR.
- Anders, M. (1990). "Praxis der Telefonbefragung". In: *Telefon und Gesellschaft, Band 2*. Forschungsgruppe Telekommunikation, S. 426–436.
- Barron, Martin, Meene Khare und Zhen Zhau (2008). *Cell Telephone Response Rates*. Techn. Ber. AAPOR – Section on Survey Research Methods.
- Beck, François, Stéphane Legleye und Patrick Peretti-Watel (Apr. 2005). "Aux abonnés absents : Liste rouge et téléphone portable dans les enquêtes en population générale sur les drogues". In: *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 86(1), S. 5–29. DOI: 10.1177/075910630508600103.
- Beckmann, Katharina (8. Sep. 2015). "Telefonausstattung Deutschland gemessen im F2F-Bus TNS Infratest". In: "Präsentation im Rahmen des 6. Treffens der Arbeitsgruppe Mobilsample".
- Bertschi, Stefan (Dez. 2007). "Günter Burkart: Handymania. Wie das Mobiltelefon unser Leben verändert hat". In: *Publizistik* 52(4), S. 547–548. DOI: 10.1007/s11616-007-0251-3.
- Blumberg, Stephen J. und Julian V. Luke (2012). *Wireless Substitution: Early Release of Estimates From the National Health Interview Survey*. Techn. Ber. National Center for Health Statistics.
- Blumberg, Stephen J., Julian V. Luke, Marcie L. Cynamon und Martin R. Frankel (2008). "Recent Trends in Household Telephone Coverage in the United States". In: *Advances in Telephone Survey Methodology*. Hrsg. von J. Lepkowski, C. Tucker, J. Brick und Andere. John Wiley & Sons, Inc., S. 56–86. DOI: 10.1002/9780470173404.ch3.
- Boyle, John M., Faith Lewis und Brian Tefft (Dez. 2009). "Cell Phone Mainly Households: Coverage and Reach for Telephone Surveys Using RDD Landline Samples". In: *Survey Practice* 2(9), S. 1–10. DOI: 10.29115/sp-2009-0037.

- Brewer, K. R. W. und Martin E. Donadio (2003). "The High Entropy Variance of the Horvitz-Thompson Estimator". In: *Statistics Canada - Survey Methodology* 29(2), S. 189–196.
- Brick, J. Michael (Jan. 2013). "Unit Nonresponse and Weighting Adjustments: A Critical Review". In: *Journal of Official Statistics* 29(3), S. 329–353. DOI: 10.2478/jos-2013-0026.
- Brick, J. Michael, Ismael Flores Cervantes, Sunghee Lee und Greg Norman (2011). "Nonsampling errors in dual frame telephone surveys". In: *Statistics Canada - Survey Methodology* 37, S. 1–12.
- Brick, J. Michael, Sarah Dipko, Stanley Presser, Clyde Tucker und Yangyang Yuan (2005). "Estimation Issues in Dual Frame Sample of Cell and Landline Numbers". In: *Proceedings of the Survey Research Methods Section of the American Statistical Association*, S. 2791–2798.
- Brick, J. Michael, Sarah Dipko, Stanley Presser, Clyde Tucker und Yangyang Yuan (2006). "Nonresponse Bias in a Dual Frame Sample of Cell and Landline Numbers". In: *Public Opinion Quarterly* 70(5), S. 780–793. DOI: doi:10.1093/poq/nfl031.
- Bruch, Christian, Ralf Münnich und Stefan Zins (2011). *Deliverable 3.1: Variance Estimation for Complex Surveys*. Techn. Ber. European Commission – Advanced Methodology for European Laeken Indicators AMELI.
- Bundesamt für Sicherheit in der Informationstechnik (2008). *Öffentliche Mobilfunknetze und ihre Sicherheitsaspekte*. Bundesamt für Sicherheit in der Informationstechnik.
- Bundesnetzagentur (2007). *Tätigkeitsbericht 2006/2007 Telekommunikation*.
- Bundesnetzagentur (2008). *Jahresbericht 2007*.
- Bundesnetzagentur, Hrsg. (2011). *Tätigkeitsbericht 2010/2011 Telekommunikation*.
- Bundesnetzagentur (2013). *Tätigkeitsbericht 2012/2013 – Telekommunikation*.
- Bundesnetzagentur (2015). *Tätigkeitsbericht 2014/2015 Telekommunikation*.
- Bundesnetzagentur (15. Nov. 2017a). *Aktuelle Roaming-Regelungen in der Europäischen Union*. URL: <https://www.bundesnetzagentur.de/DE/Sachgebiete/Telekommunikation/Verbraucher/WeitereThemen/InternRoaming/EUROaming/EUROaming-node.html> (besucht am 06. 12. 2017).
- Bundesnetzagentur (23. Dez. 2017b). *Mobile Dienste, zugeteilte RNB*. URL: https://www.bundesnetzagentur.de/DE/Sachgebiete/Telekommunikation/Unternehmen_Institutionen/Nummerierung/Rufnummern/MobileDienste/zugeteilte%20RNB/MobileDiensteBelegteRNB-Basepage.html?nn=268384#download=1.
- Burmeister, Leon F. (1972). "Estimators Selected from Multiple Overlapping Frames". Diss. Iowa State University.

- Callegaro, Mario, Oztas Ayhan, Siegfried Gabler, Sabine Häder und Ana Villar (2011). "Combining Landline and Mobile Phone Sample - A Dual Frame Approach". In: *GESIS Working Paper* 13, S. 5–30.
- Christian, Leah, Scott Keeter, Kristen Purcell und Aaron Smith (2010). "Assessing the Cell Phone Challenge to Survey Research in 2010". In: *Artikel im Rahmen der 65. Jahreskonferenz der American Association of Public Opinion Research*. Chicago.
- Clark, Jay, Marianne Winglee und Benmei Liu (2007). "Handling Imperfect Overlap Determination in a Dual-Frame Survey". In: *American Statistical Association - Proceedings of the Survey Research Methods Section*, S. 3233–3238.
- Conrad, Frederick G., Jessica S. Broome, José R. Benkő, Frauke Kreuter, Robert M. Groves, David Vannette und Colleen McClain (Dez. 2012). "Interviewer speech and the success of survey invitations". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176(1), S. 191–210. DOI: 10.1111/j.1467-985x.2012.01064.x.
- Destatis (Nov. 2016). *Bevölkerung nach Bildungsabschluss in Deutschland*. URL: <https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/BildungForschungKultur/Bildungsstand/Tabellen/Bildungsabschluss.html>.
- Engel, Tobias (2013). *Locating Mobile Phones using Signalling System No. 7*.
- Engel, Uwe und Björn Oliver Schmidt (2014). "Unit- und Item-Nonresponse". In: *Handbuch Methoden der empirischen Sozialforschung*. Springer Fachmedien Wiesbaden, S. 331–348. DOI: 10.1007/978-3-531-18939-0_23.
- Ericsson GmbH (2001). *Die Infrastruktur von Mobilfunk - Funktion und Elemente*.
- European Commission (2016). *Special Eurobarometer 438 - E-Communications and the Digital Single Market*.
- Feierabend, Sabine, Theresa Plankenhorn und Thomas Rathgeb (2015). *KIM-Studie 2014 - Kinder + Medien, Computer + Internet*. Medienpädagogischer Forschungsverbund Südwest. Stuttgart.
- Fuchs, Marek (2002). "Eine CATI-Umfrage unter Handy-Nutzern - Methodische Erfahrungen aus einem Vergleich unter Festnetz-Stichproben". In: *Telefonstichproben - Methodische Innovation und Anwendung in Deutschland*. Gabler, S.; Häder, S., S. 121–137.
- Fuchs, Marek (2007). "Mobile Web Survey: Möglichkeiten der Verknüpfung von Online-Befragungen und Handy-Befragungen". In: *Mobiltelefonie - Eine Herausforderung für die Umfrageforschung*. Gabler, S.; Häder, S., S. 105–126.
- Fuchs, Marek und Britta Busse (2013). "Prevalence of Cell Phone Sharing". In: *Survey Methods: Insights from the Field*. Retrieved from <http://surveyinsights.org/?p=1019>, S. 1–15. DOI: 10.13094/smif-2013-00004.
- Gabler, Siegfried und Öztas Ayhan (2007). "Gewichtung bei Erhebungen im Festnetz und über Mobilfunk: Ein Dual-Frame Ansatz". In: *Mobilfunktelefonie - Eine Herausforderung für die*

- Umfrageforschung*. Hrsg. von S. Gabler und Häder. ZUMA-Nachrichten Spezial Band 13: Mannheim, S. 39–45.
- Gabler, Siegfried und Sabine Häder (1997). "Überlegungen zu einem Stichprobendesign für Telefonumfragen in Deutschland". In: *ZUMA Nachrichten* 21, S. 7–18.
- Gabler, Siegfried und Sabine Häder (2009a). "Die Kombination von Mobilfunk- und Festnetzstichproben in Deutschland". In: *Umfrageforschung – Herausforderungen und Grenzen*. Hrsg. von M. Weichbold, J. Bacher und C. Wolf. Österreichische Zeitschrift für Soziologie Sonderheft 9/2009. VS Verlag für Sozialwissenschaften: Wiesbaden, S. 239–252. DOI: 10.1007/978-3-531-91852-5_12.
- Gabler, Siegfried und Sabine Häder (2009b). "Gewichtung für die CELLA-Studie". In: *Telefonbefragungen über das Mobilfunknetz – Konzept, Design und Umsetzung einer Strategie zur Datenerhebung*. Hrsg. von M. Häder und S. Häder. VS Verlag für Sozialwissenschaften: Wiesbaden, S. 51–55. DOI: 10.1007/978-3-531-91490-9_4.
- Gabler, Siegfried, Sabine Häder und Jürgen H. P. Hoffmeyer-Zlotnik (1998). "Einleitung". In: *Telefonstichproben in Deutschland*. Hrsg. von Siegfried Gabler, Sabine Häder und Jürgen H. P. Hoffmeyer-Zlotnik. VS Verlag für Sozialwissenschaften: Opladen; Wiesbaden, S. 9–19. DOI: 10.1007/978-3-322-83334-1.
- Gabler, Siegfried, Sabine Häder, Iris Lehnhoff und Elisabeth Mardian (2011). "Weighting for Unequal Inclusion Probabilities and Nonresponse in Dual Frame Telephone Surveys". In: *Telephone Surveys in Europe*. Hrsg. von S. Häder, M. Häder und M. Kühne. Springer Berlin Heidelberg: Heidelberg, S. 147–167. DOI: 10.1007/978-3-642-25411-6_11.
- Gabler, Siegfried und Sabine Häder. (Sep. 2002). "Idiosyncrasies in Telephone Sampling–The Case of Germany". In: *International Journal of Public Opinion Research* 14(3), S. 339–345. DOI: 10.1093/ijpor/14.3.339.
- Gajek, Henning (2015). *Gajek-Liste, Vorwahl.faq*.
- Ganninger, Matthias (2009). "Design Effects: Model-based versus Design-based Approach". Diss. Trier: Universität Trier.
- Glasser, Gerald J. und Gale D. Metzger (Feb. 1972). "Random-Digit Dialing as a Method of Telephone Sampling". In: *Journal of Marketing Research* 9(1), S. 59. DOI: 10.2307/3149607.
- Glemser, Axel (2007). "Mobilfunknutzung in Deutschland. Eine Herausforderung für die Stichprobenbildung in der Markt- und Sozialforschung". In: *Mobilfunktelefonie – Eine Herausforderung für die Umfrageforschung*. Gabler, S.; Häder, S., S. 7–23.
- Glemser, Axel, Gerd Meier und Christiane Heckel (2014). "Dual-Frame: Stichprobendesign für CATI-Befragungen im mobilen Zeitalter". In: *Stichproben-Verfahren in der Umfrageforschung – Eine Darstellung für die Praxis*. Hrsg. von ADM e.V. 2. Springer: Wiesbaden, S. 137–166.

- Graeske, Jennifer und Tanja Kunz (2009). "Stichprobenqualität der CELLA-Studie unter besonderer Berücksichtigung der Mobileonlys". In: *Telefonbefragungen über das Mobilfunknetz*. VS Verlag für Sozialwissenschaften, S. 57–70. DOI: 10.1007/978-3-531-91490-9_5.
- Gramlich, Tobias (2015). "Influenza 2014 - Auswertung der Kontaktprotokolle". Unveröffentlichtes Arbeitspapier.
- Groves, Robert M. (1979). "Actors and Questions in Telephone and Personal Interview Surveys". In: *Public Opinion Quarterly* 43(2), S. 190–205. DOI: 10.1086/268511.
- Groves, Robert M. (Aug. 1990). "Theories and Methods of Telephone Surveys". In: *Annual Review of Sociology* 16(1), S. 221–240. DOI: 10.1146/annurev.soc.16.1.221.
- Groves, Robert M. (Dez. 2011). "Three Eras of Survey Research". In: *Public Opinion Quarterly* 75(5), S. 861–871. DOI: 10.1093/poq/nfr057.
- Grund, Matthias (2002). "Mobilfunkteilnehmer - Ein Repräsentativitätsproblem". In: *Telefonstichproben - Methodische Innovation und Anwendung in Deutschland*. Gabler, S.; Häder, S., S. 111–120.
- Gutt, Eike (2010). *LTE - Neue Dimension mobiler Breitbandnutzung - Eine technische Einführung*.
- Häder, Michael (2000). *Mobilfunk verdrängt Festnetz: Übersicht zu den Ergebnissen einer Delphi-Studie zur Zukunft des Mobilfunks*. Techn. Ber. 5. Mannheim: ZUMA.
- Häder, Michael (2009a). "Befragungen über das Mobilfunknetz: Eine neue Technik im sozialwissenschaftlichen Methodenarsenal". In: *Telefonbefragungen über das Mobilfunknetz*. VS Verlag für Sozialwissenschaften, S. 13–16. DOI: 10.1007/978-3-531-91490-9_1.
- Häder, Michael (2009b). "Modell einer Mixed-Mode-Studie mit Mobilfunk- und Festnetzbefragung. Genereller Problemaufriss". In: *Telefonbefragungen über das Mobilfunknetz*. Hrsg. von M. Häder und S. Häder. VS Verlag für Sozialwissenschaften: Wiesbaden, S. 17–20. DOI: 10.1007/978-3-531-91490-9_2.
- Häder, Michael (2011). "Data Quality in Telephone Surveys via Mobile and Landline Phone". In: *Telephone Surveys in Europe*. Springer Berlin Heidelberg, S. 247–262. DOI: 10.1007/978-3-642-25411-6_16.
- Häder, Michael und Mike Kühne (2009a). "Fragereihenfolgeeffekte". In: *Telefonbefragungen über das Mobilfunknetz*. Hrsg. von M. Häder und S. Häder. VS Verlag für Sozialwissenschaften: Wiesbaden, S. 207–216. DOI: 10.1007/978-3-531-91490-9_17.
- Häder, Michael und Mike Kühne (2009b). "Mode-Effekte bei Einstellungsfragen mit direktem inhaltlichen Bezug". In: *Telefonbefragungen über das Mobilfunknetz*. Hrsg. von M. Häder und S. Häder. VS Verlag für Sozialwissenschaften: Wiesbaden, S. 193–199. DOI: 10.1007/978-3-531-91490-9_15.

- Häder, Michael und Mike Kühne (2009c). "Theoretischer Rahmen und Untersuchungsdesign". In: *Telefonbefragungen über das Mobilfunknetz*. VS Verlag für Sozialwissenschaften, S. 165–174. DOI: 10.1007/978-3-531-91490-9_12.
- Häder, Michael und Mike Kühne (2009d). "Zusammenfassung zu Mode-Effekten". In: *Telefonbefragungen über das Mobilfunknetz*. VS Verlag für Sozialwissenschaften, S. 241–242. DOI: 10.1007/978-3-531-91490-9_21.
- Häder, Michael und Robert Neumann (2015). "Einleitung". In: *Auswahlprobleme und Antwortverhalten bei Telefonbefragungen*. Hrsg. von M. Häder und R. Neumann. MV-Verlag, S. 1–10.
- Häder, Sabine (1994). "Auswahlverfahren bei Telefonumfragen". In: *ZUMA-Arbeitsbericht 03*, S. 1–38.
- Häder, Sabine (1996). "Wer sind die Nonpubs?: zum Problem anonymer Anschlüsse bei Telefonumfragen". In: *ZUMA Nachrichten 20*, S. 45–68.
- Häder, Sabine (2000). "Telefonstichproben". In: *ZUMA How-to-Reihe 6*, S. 1–13.
- Häder, Sabine, Siegfried Gabler und Christiane Heckel (2009a). "Stichprobenziehung für die CELLA-Studie". In: *Telefonbefragungen über das Mobilfunknetz*. VS Verlag für Sozialwissenschaften, S. 21–49. DOI: 10.1007/978-3-531-91490-9_3.
- Häder, Sabine, Michael Häder, Jennifer Graeske, Tanja Kunz und Götz Schneiderat (2009b). "Realisierung der Stichprobe". In: *Telefonbefragungen über das Mobilfunknetz*. VS Verlag für Sozialwissenschaften, S. 71–82. DOI: 10.1007/978-3-531-91490-9_6.
- Haines, Dawn E. und Kenneth H. Pollock (1998). "Estimating the Number of Active and Successful Bald Eagle Nests: An Application of the Dual Frame Method". In: *Environmental and Ecological Statistics* 5(3), S. 245–256. DOI: 10.1023/a:1009673403664.
- Hansen, Sue Ellen (2008). "CATI Sample Management Systems". In: *Advances in Telephone Survey Methodology*. Hrsg. von J. Lepkowski, C. Tucker, J. Brick und Andere. John Wiley & Sons, Inc.: New Jersey, S. 340–358. DOI: 10.1002/9780470173404.ch16.
- Hartley, H. O. (1962). "Multiple Frame Surveys". In: *Proceedings of the Social Statistics Section*. Hrsg. von American Statistical Association, S. 203–206.
- Heckel, Christiane (2002). "Erstellung der ADM-Telefonauswahlgrundlage". In: *Telefonstichproben – Methodische Innovationen und Anwendungen in Deutschland*. Hrsg. von S. Gabler und S. Häder. Waxmann Verlag: Münster, S. 11–31.
- Heckel, Christiane (2007). "Weiterentwicklung der ADM-CATI-Auswahlgrundlagen". In: *Mobilfunktelefonie – eine Herausforderung für die Umfrageforschung*. Hrsg. von S. Gabler und S. Häder. Druck & Kopie Hanel, S. 25–37.
- Heckel, Christiane (2014). "Projekt Regionalisierung Mobil". In: *5. Treffen der Arbeitsgruppe Mobilsample*. Mannheim.

- Heckel, Christiane, Axel Glemser und Gerd Meier (2014). "Das ADM-Telefonstichproben-System". In: *Stichproben-Verfahren in der Umfrageforschung - Eine Darstellung für die Praxis*. Hrsg. von ADM e.V. 2. Springer: Wiesbaden, S. 137–166.
- Heckel, Christiane und Kathrin Wiese (2011). "Sampling Frames for Telephone Surveys in Europe". In: *Telephone Surveys in Europe*. Springer Berlin Heidelberg, S. 103–119. DOI: 10.1007/978-3-642-25411-6_9.
- Hochstim, Joseph R. (Sep. 1967). "A Critical Comparison of Three Strategies of Collecting Data from Households". In: *Journal of the American Statistical Association* 62(319), S. 976–989. DOI: 10.1080/01621459.1967.10500909.
- Hoops, Christian und Tobias Michael (Sep. 2015). "Mobile and Hard-to-Reach Voters". In: *Survey Practice* 8(3), S. 1–7. DOI: 10.29115/sp-2015-0020.
- Hox, J. J. (Feb. 1994). "Hierarchical Regression Models for Interviewer and Respondent Effects". In: *Sociological Methods & Research* 22(3) (2), S. 300–318. DOI: 10.1177/0049124194022003002.
- Hüfken, Volker (2000). "Kontaktierung bei Telefonumfragen. Auswirkungen auf das Kooperations- und Antwortverhalten". In: *Methoden in Telefonumfragen*. Hüfken, V.: Wiesbaden, S. 11–31.
- Hunsicker, Stefan und Yvonne Schroth (2014). "Dual-Frame-Ansatz in politischen Umfragen". In: *Arbeitspapiere der Forschungsgruppe Wahlen e.V. 2*.
- Jung, M. (1990). "Ausschöpfungsprobleme bei repräsentativen Telefonumfragen". In: *Telefon und Gesellschaft*. Bd. 2. Forschungsgruppe Telefonkommunikation, S. 386–399.
- Kalsbeek, William D. und Robert P. Agans (2008). "Sampling and Weighting in Household Telephone Surveys". In: *Advances in Telephone Survey Methodology*. Hrsg. von J. Lepkowski, C. Tucker, J. Brick und Andere. John Wiley & Sons, Inc.: New Jersey, S. 27–55. DOI: 10.1002/9780470173404.ch2.
- Kennedy, Courtney K. (Jan. 2007). "Evaluating the Effects of Screening for Telephone Service in Dual Frame RDD Surveys". In: *Public Opinion Quarterly* 71(5), S. 750–771. DOI: 10.1093/poq/nfm050.
- Kennedy, Courtney K. (2010). *Nonresponse and Measurement Error in Mobile Phone Surveys*.
- Kish, Leslie (März 1962). "Studies of Interviewer Variance for Attitudinal Variables". In: *Journal of the American Statistical Association* 57(297), S. 92–115. DOI: 10.1080/01621459.1962.10482153.
- Kunz, Tanja und Marek Fuchs (2011). "Pre-Call Validation of RDD Cell Phone Numbers - A Field Experiment". In: *Paper presented at the Annual Conference of the American Association for Public Opinion Research*.
- Kuusela, Vesa, Mario Callegaro und Vasja Vehovar (2008). "The Influence of Mobile Telephones on Telephone Surveys". In: *Advances in Telephone Survey Methodology*. Hrsg. von

- J. Lepkowski, C. Tucker, J. Brick und Andere. John Wiley & Sons, Inc., S. 87–112. DOI: 10 . 1002/9780470173404.ch4.
- Lambert, David, Gary Langer und Mike McMenemy (2010). "Cell-Phone Sampling: An Alternative Approach". In: *Proceeding of the Annual Conference of the American Association for Public Opinion Research*. Chicago.
- Levine, Burton und Rachel Harter (Okt. 2014). "Optimal Allocation of Cell-Phone and Landline Respondents in Dual-Frame Surveys". In: *Public Opinion Quarterly* 79(1), S. 91–104. DOI: 10.1093/poq/nfu044.
- Link, M. W., M. P. Battaglia, M. R. Frankel, L. Osborn und A. H. Mokdad (Jan. 2007). "Reaching the U.S. Cell Phone Generation: Comparison of Cell Phone Survey Results with an Ongoing Landline Telephone Survey". In: *Public Opinion Quarterly* 71(5), S. 814–839. DOI: 10 . 1093 / poq/nfm051.
- Little, Roderick J. A. und Donald B. Rubin (1987). *Statistical Analysis with Missing Data*. John Wiley und Sons.
- Lohr, Sharon L. (2007). "Recent developments in multiple frame surveys". In: *American Statistical Association - Proceedings of the Survey Research Methods Section*, S. 3257–3264.
- Lohr, Sharon L. (2009). *Sampling - Design and Analysis*. CENGAGE Learning.
- Lohr, Sharon L. (2011). "Alternative survey sample designs: Sampling with multiple overlapping frames". In: *Statistics Canada - Survey Methodology* 37, S. 197–213.
- Lohr, Sharon L. und J. N. K. Rao (März 2000). "Inference from Dual Frame Surveys". In: *Journal of the American Statistical Association* 95(449), S. 271–280. DOI: 10.1080/01621459.2000.10473920.
- Lohr, Sharon und J. N. K. Rao (Sep. 2006). "Estimation in Multiple-Frame Surveys". In: *Journal of the American Statistical Association* 101(475), S. 1019–1030. DOI: 10 . 1198 / 016214506000000195.
- Lumley, Thomas (Feb. 2010). *Complex Surveys*. John Wiley & Sons, Inc. DOI: 10 . 1002 / 9780470580066.
- Lundström, Sixten und Carl-Erik Särndal (2001). *Estimation in the Presence of Nonresponse and Frame Imperfections*. Hrsg. von SCB-tryck. Statistics Sweden: Örebro.
- Lynn, Peter und Olena Kaminska (2011). "Factors Affecting Measurement Error in Mobile Phone Interviews". In: *Telephone Surveys in Europe*. Springer Berlin Heidelberg, S. 211–228. DOI: 10.1007/978-3-642-25411-6_14.
- Matei, Alina und Yves Tillé (2005). "Evaluation of Variance Approximations and Estimators in Maximum Entropy Sampling with Unequal Probability and Fixed Sample Size". In: *Journal of Official Statistics* 21(4), S. 543–570.

- Mecatti, Fulvia (2007). "A single frame multiplicity estimator for multiple frame surveys". In: *Statistics Canada – Survey Methodology* 33, S. 151–157.
- Meier, Gerd (1999). "Random-Telefonstichproben". In: *Stichproben-Verfahren in der Umfrageforschung*. Hrsg. von ADM e.V. 1. Aufl. VS Verlag für Sozialwissenschaften: Opladen, S. 93–101. DOI: 10.1007/978-3-663-11511-3_7.
- Meier, Gerd (2007). "Validierung eines Fragebogens zur Erfassung der Anzahl von Telefonnummern". In: *Mobiltelefonie – Eine Herausforderung für die Umfrageforschung*. Gabler, S.; Häder, S., S. 91–104.
- Meier, Gerd (2011). "The Impact of Introductions in Telephone Surveys". In: *Telephone Surveys in Europe*. Springer Berlin Heidelberg, S. 265–274. DOI: 10.1007/978-3-642-25411-6_17.
- Meier, Gerd, Axel Glemser und Christiane Heckel (Nov. 2013). "Random-Telefonstichproben". In: *Stichproben-Verfahren in der Umfrageforschung*. 2. Aufl. Springer Fachmedien Wiesbaden: Wiesbaden, S. 117–135. DOI: 10.1007/978-3-531-18882-9_6.
- Montaquila, Jill M., J. Michael Brick, Mary C. Hagedorn, Courtney Kennedy und Scott Keeter (2008). "Aspects of Nonresponse Bias in RDD Telephone Surveys". In: *Advances in Telephone Survey Methodology*. Hrsg. von J. Lepkowski, C. Tucker, J. Brick und Andere. John Wiley & Sons, Inc.: New Jersey, S. 561–586. DOI: 10.1002/9780470173404.ch25.
- Noelle-Neumann, Elisabeth und Thomas Petersen (2000). "Das halbe Instrument, die halbe Reaktion. Zum Vergleich von Telefon- und face-to-face Umfragen". In: *Methoden der Telefonumfragen*. Hufken, V: Wiesbaden, S. 183–200.
- O'Muircheartaigh, Colm und Pamela Campanelli (Feb. 1998). "The relative impact of interviewer effects and sample design effects on survey precision". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 161(1), S. 63–77. DOI: 10.1111 / 1467 - 985x.00090.
- Peytchev, Andy, Lisa R. Carley-Baxter und Michele C. Black (Dez. 2009). "Coverage Bias in Variances, Associations, and Total Error From Exclusion of the Cell Phone-Only Population in the United States". In: *Social Science Computer Review* 28(3), S. 287–302. DOI: 10.1177 / 0894439309353027.
- Peytchev, Andy und Benjamin Neely (2013). "RDD Telephone Surveys". In: *Public Opinion Quarterly* 77(1), S. 283–304. DOI: 10.1093/poq/nft003.
- Rammstedt, Beatrice, Christoph J. Kemper, Mira Céline Klein, Constanze Beierlein und Anastassiya Kovaleva (2017). "A Short Scale for Assessing the Big Five Dimensions of Personality: 10 Item Big Five Inventory (BFI-10)". In: *methods, data, analyses, Vol 7, No 2 (2013) 7(2)*, S. 233–249. DOI: 10.12758/mda.2013.013.
- Read, Campbell B. (1981). "Mean Squared Error". In: *Encyclopedia of Statistical Sciences*. S. Kotz und Johnson, N.

- Rogers, Theresa F. (März 1976). "Interviews by Telephone and in Person: Quality of Responses and Field Performance". In: *Public Opinion Quarterly* 40(1), S. 51–65. DOI: 10.1086/268267.
- Routo Messaging (2009). *HLR LOOKUP User Guide*.
- Sand, Matthias (2014). "Dual-Frame-Telefonstichproben – Entwicklung, Handhabung und Gewichtung". In: *GESIS – Technical Reports 2014 2*. DOI: 10.1007/978-3-658-10459-7_4.
- Sand, Matthias (2015a). "Dual-Frame-Telefonstichproben: Gewichtung im Falle von Device-Specific Nonresponse". In: *Nonresponse Bias*. Hrsg. von J. Schupp und C. Wolf. Springer Fachmedien Wiesbaden: Wiesbaden, S. 133–160. DOI: 10.1007/978-3-658-10459-7_4.
- Sand, Matthias (2015b). "Überarbeitung des GESIS Auswahlrahmens für Telefonstichproben". In: *Auswahlprobleme und Antwortverhalten bei Telefonbefragungen*. Hrsg. von M. Häder und R. Neumann. MV-Verlag, S. 13–38.
- Sand, Matthias und Siegfried Gabler (2018). "Gewichtung von (Dual-Frame –) Telefonstichproben". In: *Arbeitsbuch Telefonstichproben*. Hrsg. von Michael Häder. Hrsg. von Sabine Häder. Hrsg. von Patrick Schmich.
- Särndal, Carl-Erik und Sixten Lundström (2005). *Estimation in Surveys with Nonresponse*.
- Schmich, Patrick, Matthias Sand, Regina Prigge, Marieke Varga und Tobias Gramlich (2015). "Einsatz einer Dual-Frame-Stichprobe im Rahmen einer Gesundheitsbefragung". Unveröffentlichtes Arbeitspapier.
- Schneiderat, Götz und Sabine Häder (2007). "Anlage und Vorstudien des DFG-Projektes "Telefonbefragungen in der Allgemeinbevölkerung über das Mobilfunknetz". In: *Mobiltelefonie – Eine Herausforderung für die Umfrageforschung*. Hrsg. von ZUMA Mannheim. Bd. 13. ZUMA Nachrichten. Gabler, S. und Häder, S., S. 59–80.
- Schneiderat, Götz und Tino Schlinzig (2009a). "Das Mobilfunktelefonverhalten in der Allgemeinbevölkerung". In: *Telefonbefragungen über das Mobilfunknetz*. Hrsg. von M. Häder und S. Häder. VS Verlag für Sozialwissenschaften, S. 99–119. DOI: 10.1007/978-3-531-91490-9_8.
- Schneiderat, Götz und Tino Schlinzig (2009b). "Die Anwesenheit Dritter als intervenierende Größe für die Response-Rate". In: *Telefonbefragungen über das Mobilfunknetz*. Hrsg. von M. Häder und S. Häder. VS Verlag für Sozialwissenschaften, S. 121–132. DOI: 10.1007/978-3-531-91490-9_9.
- Schneiderat, Götz und Tino Schlinzig (2009c). "Nonrespondenten: Auswertung der Ergebnisse der Kurzbefragung". In: *Telefonbefragungen über das Mobilfunknetz*. Hrsg. von M. Häder und S. Häder. VS Verlag für Sozialwissenschaften, S. 133–143. DOI: 10.1007/978-3-531-91490-9_10.
- Schneiderat, Götz und Tino Schlinzig (2009d). "Teilnahmebereitschaft und Teilnahmeverhalten bei Telefonumfragen der Allgemeinbevölkerung über das Mobilfunknetz". In:

- Telefonbefragungen über das Mobilfunknetz*. Hrsg. von M. Häder und S. Häder. VS Verlag für Sozialwissenschaften, S. 83–97. DOI: 10.1007/978-3-531-91490-9_7.
- Schneiderat, Götz und Tino Schlinzig (2011). "Mobile- and Landline-Onlys in Dual-Frame-Approaches: Effects on Sample Quality". In: *Telephone Surveys in Europe*. Hrsg. von S. Häder, M. Häder und M. Kühne. Springer Berlin Heidelberg: Heidelberg, New York, Dordrecht, London, S. 121–143. DOI: 10.1007/978-3-642-25411-6_10.
- Schnell, Rainer, Paul Hill und Elke Esser (1999). *Methoden der empirischen Sozialforschung*. Hrsg. von 8. Oldenbourg Wissenschaftsverlag GmbH: München.
- Schnell, Rainer, Paul Hill und Elke Esser (2013). *Methoden der empirischen Sozialforschung*. Hrsg. von 8. Oldenbourg Wissenschaftsverlag GmbH: München.
- Skinner, C. J. und J. N. K. Rao (März 1996). "Estimation in Dual Frame Surveys with Complex Designs". In: *Journal of the American Statistical Association* 91(433), S. 349–356. DOI: 10.1080/01621459.1996.10476695.
- Statistisches Bundesamt (Destatis) (Juli 2016). *Ausstattung privater Haushalte mit Informations- und Kommunikationstechnik - Deutschland*. https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/EinkommenKonsumLebensbedingungen/AusstattungGebrauchsgueter/Tabellen/Infotechnik_D.html. Abgerufen am 21.07.2016.
- Struminskaya, Bella, Lars Kaczmirek, Ines Schaurer, Wolfgang Bandilla, Siegfried Gabler und Sabine Haeder, Hrsg. (März 2011). *Improved cost-effectiveness in mobile surveys using HLR-Lookup. Paper presented at the GOR 11 Conference*. GESIS.
- Traugott, Michael W., Robert M. Groves und James M. Lepkowski (1987). "Using Dual Frame Designs to Reduce Nonresponse in Telephone Surveys". In: *Public Opinion Quarterly* 51(4), S. 522–539. DOI: 10.1086/269055.
- Valliant, Richard, Jill A. Dever und Frauke Kreuter (2013). *Practical Tools for Designing and Weighting Survey Samples*. Springer New York: Heidelberg. DOI: 10.1007/978-1-4614-6449-5.
- Von Der Lippe, Elena, Patrick Schmich und Cornelia Lange (2011). "Advance letters as a way of reducing non-response in a National Health Telephone Survey: Differences between listed and unlisted numbers". In: *Survey Research Methods, Vol 5, No 3 (2011)* 5(3), S. 103–116. DOI: 10.18148/srm/2011.v5i3.4657.
- Waksberg, Joseph (März 1978). "Sampling Methods for Random Digit Dialing". In: *Journal of the American Statistical Association* 73(361), S. 40–46. DOI: 10.1080/01621459.1978.10479995.
- West, Brady T. und Kristen Olson (Jan. 2010). "How Much of Interviewer Variance is Really Nonresponse Error Variance?" In: *Public Opinion Quarterly* 74(5), S. 1004–1026. DOI: 10.1093/poq/nfq061.

- Winglee, Marianne, Inho Park, Keith Rust, Benmei Liu und Gary Shapiro (2007). "A Case Study in Dual-Frame Estimation Methods". In: *American Statistical Association - Proceedings of the Section on Survey Research Methods*, S. 3195–3202.
- Wolter, Kirk M., Xian Tao, Robert Montgomery und Philip J. Smith (2015). "Optimum Allocation for a Dual-Frame Telephone Survey". In: *Survey Methodology* 41(2), S. 389–401.
- Xia, Kanru, Steven Pedlow und Michael Davern (2010). "Dual-Frame Weights (Landline and Cell) for the 2009 Minnesota Health Access Survey". In: *American Statistical Association - Proceedings of the Survey Research Methods Section*, S. 3912–3922.
- Zipp, John F. und Joann Toth (2002). "She Said, He Said, They Said - The Impact of Spousal Presence in Survey Research". In: *Public Opinion Quarterly* 66(2), S. 177–208. DOI: 10.1086/340028.

A Anhang

A.1 Überprüfung der Variablen zur Einordnung in Subdomains auf Unabhängigkeit

A.1.1 χ^2 -Tests auf Unabhängigkeit der Proxies zur Einteilung nach Subdomain im Rahmen von CELLA 2

Variablen v8a und v9a

Pearson's Chi-squared test

data: dl\$v8a and dl\$v9a

X-squared = 198.84, df = 4, p-value < 2.2e-16

Variablen t7 und v9a

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: dl\$t.7. and dl\$v9a

X-squared = 26.907, df = NA, p-value = 0.1424

Variablen v8a und v8a

Pearson's Chi-squared test

data: dl\$t.7. and dl\$v8a

X-squared = 12.301, df = 5, p-value = 0.03088

Variablen t2 und t7

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: dl\$t2 and dl\$t.7.

X-squared = 154.48, df = NA, p-value = 0.0004998

Variablen v8a und t2

Pearson's Chi-squared test with Yates' continuity correction

data: dI\$t2 and dI\$v8a

X-squared = 9.5661, df = 1, p-value = 0.001982

Variablen v9a und t2

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: dI\$t2 and dI\$v9a

X-squared = 17.787, df = NA, p-value = 0.0009995

A.1.2 χ^2 -Tests auf Unabhängigkeit der Variablen zur Einteilung nach Subdomain im Rahmen von Influenza 2014

Variablen Mainly1 und Mainly2

Pearson's Chi-squared test

data: dI\$MAINLY1 and dI\$MAINLY2

X-squared = 88.776, df = 16, p-value = 4.201e-12

Variablen Mainly1 und Mainly3

Pearson's Chi-squared test

data: dI\$MAINLY1 and dI\$MAINLY3

X-squared = 219.04, df = 16, p-value < 2.2e-16

Variablen Mainly1 und Mainly4

Pearson's Chi-squared test

data: dI\$MAINLY1 and dI\$MAINLY4

X-squared = 163.6, df = 12, p-value < 2.2e-16

Variablen Mainly2 und Mainly3

Pearson's Chi-squared test

data: dl\$MAINLY2 and dl\$MAINLY3

X-squared = 116.5, df = 16, p-value < 2.2e-16

Variablen Mainly2 und Mainly4

Pearson's Chi-squared test

data: dl\$MAINLY2 and dl\$MAINLY4

X-squared = 28.825, df = 12, p-value = 0.004182

Variablen Mainly3 und Mainly4

Pearson's Chi-squared test

data: dl\$MAINLY3 and dl\$MAINLY4

X-squared = 206.79, df = 12, p-value < 2.2e-16

Nur in der Mobilfunkstichprobe:

Variablen Mainly1 und Mainly5

Pearson's Chi-squared test

data: mob\$MAINLY1 and mob\$MAINLY5

X-squared = 7.7224, df = 4, p-value = 0.1023

Variablen Mainly2 und Mainly5

Pearson's Chi-squared test

data: mob\$MAINLY2 and mob\$MAINLY5

X-squared = 3.3957, df = 4, p-value = 0.4939

Variablen Mainly3 und Mainly5

Pearson's Chi-squared test

data: mob\$MAINLY3 and mob\$MAINLY5

X-squared = 24.527, df = 4, p-value = 6.262e-05

Variablen Mainly4 und Mainly5

Pearson's Chi-squared test

data: mob\$MAINLY4 and mob\$MAINLY5

X-squared = 1.6652, df = 3, p-value = 0.6447

Nur in der Festnetzstichprobe:

Variablen Mainly1 und Mainly6

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: lan\$MAINLY1 and lan\$MAINLY6

X-squared = 2.9512, df = NA, p-value = 0.9305

Variablen Mainly2 und Mainly6

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: lan\$MAINLY2 and lan\$MAINLY6

X-squared = 13.658, df = NA, p-value = 0.06197

Variablen Mainly3 und Mainly6

Pearson's Chi-squared test with simulated p-value (based on 2000 replicates)

data: lan\$MAINLY3 and lan\$MAINLY6

X-squared = 15.462, df = NA, p-value = 0.02749

A.2 Verteilung ausgewählter soziodemografischer Merkmale nach Verwendungskategorie

A.2.1 CELLA 2: Verteilung ausgewählter soziodemografischer Merkmale nach Verwendungskategorie

Tabelle 50 CELLA 2: Haushaltsgröße nach Nutzungsgewohnheit und Modell (in %)

	Haushaltsgröße	Modell 1	Modell 2	Modell 3	Modell 4
u_1	0 Personen	0,0	0,0	0,0	0,0
	1 Person	35,4	35,4	35,4	35,4
	2 Personen	36,0	36,0	36,0	36,0
	3 Personen	13,5	13,5	13,5	13,5
	4 Personen	9,0	9,0	9,0	9,0
	5+ Personen	6,2	6,2	6,2	6,2
u_2	0 Personen	0,0	0,0	0,0	0,0
	1 Person	27,2	27,2	25,9	23,4
	2 Personen	43,0	40,3	41,2	38,3
	3 Personen	10,5	12,2	13,2	16,6
	4 Personen	13,2	15,2	14,0	15,3
	5+ Personen	6,1	5,1	5,7	6,4
u_3	0 Personen	0,1	0,1	0,1	0,0
	1 Person	21,9	21,2	21,0	0,0
	2 Personen	33,1	32,5	31,8	0,0
	3 Personen	19,9	20,8	21,2	0,0
	4 Personen	15,8	15,8	15,8	0,0
	5+ Personen	9,1	9,6	10,1	0,0
u_4	0 Personen	0,0	0,0	0,1	0,1
	1 Person	32,7	31,6	23,8	22,2
	2 Personen	28,6	31,0	32,5	31,7
	3 Personen	18,4	16,1	19,2	20,4
	4 Personen	16,3	15,5	16,5	15,8
	5+ Personen	4,1	5,8	7,8	9,6
u_5	0 Personen	0,6	0,6	0,6	0,6
	1 Person	46,5	46,5	46,5	46,5
	2 Personen	28,8	28,8	28,8	28,8
	3 Personen	11,2	11,2	11,2	11,2
	4 Personen	7,1	7,1	7,1	7,1
	5+ Personen	5,9	5,9	5,9	5,9
Erhebung	0 Personen	0,1	0,1	0,1	0,1
	1 Person	24,6	24,6	24,6	24,6
	2 Personen	33,3	33,3	33,3	33,3
	3 Personen	18,7	18,7	18,7	18,7
	4 Personen	14,8	14,8	14,8	14,8
	5+ Personen	8,5	8,5	8,5	8,5

Tabelle 51 CELLA 2: Familienstand nach Nutzungsgewohnheit und Modell (in %)

	Familienstand	Modell 1	Modell 2	Modell 3	Modell 4
u_1	Verheiratet, zusammen lebend	52,8	52,8	52,8	52,8
	Eingetragene Lebensgemeinschaft	0,0	0,0	0,0	0,0
	Verheiratet, getrennt lebend	0,0	0,0	0,0	0,0
	Ledig	14,0	14,0	14,0	14,0
	Geschieden	13,5	13,5	13,5	13,5
	Verwitwet	19,7	19,7	19,7	19,7
u_2	Verheiratet, zusammen lebend	62,3	60,0	60,9	55,3
	Eingetragene Lebensgemeinschaft	0,0	0,0	0,0	0,2
	Verheiratet, getrennt lebend	0,9	0,6	0,5	1,4
	Ledig	14,9	19,4	18,9	27,5
	Geschieden	10,5	10,7	10,6	9,3
	Verwitwet	11,4	9,3	9,1	6,4
u_3	Verheiratet, zusammen lebend	44,8	43,8	45,3	0,0
	Eingetragene Lebensgemeinschaft	0,1	0,1	0,1	0,0
	Verheiratet, getrennt lebend	2,5	2,6	2,6	0,0
	Ledig	39,4	40,8	40,1	0,0
	Geschieden	9,6	9,6	8,7	0,0
	Verwitwet	3,5	3,0	3,2	0,0
u_4	Verheiratet, zusammen lebend	35,7	32,9	37,3	41,9
	Eingetragene Lebensgemeinschaft	0,0	0,0	0,0	0,0
	Verheiratet, getrennt lebend	8,2	7,7	3,8	3,0
	Ledig	44,9	48,4	45,5	42,2
	Geschieden	4,1	4,5	10,4	9,5
	Verwitwet	7,1	6,5	3,0	3,2
u_5	Verheiratet, zusammen lebend	14,7	14,7	14,7	14,7
	Eingetragene Lebensgemeinschaft	0,6	0,6	0,6	0,6
	Verheiratet, getrennt lebend	4,1	4,1	4,1	4,1
	Ledig	62,4	62,4	62,4	62,4
	Geschieden	14,1	14,1	14,1	14,1
	Verwitwet	4,1	4,1	4,1	4,1
Erhebung	Verheiratet, zusammen lebend	44,0	44,0	44,0	44,0
	Eingetragene Lebensgemeinschaft	0,1	0,1	0,1	0,1
	Verheiratet, getrennt lebend	2,6	2,6	2,6	2,6
	Ledig	38,5	38,5	38,5	38,5
	Geschieden	10,0	10,0	10,0	10,0
	Verwitwet	4,9	4,9	4,9	4,9

A.2.2 Influenza 2014: Verteilung ausgewählter soziodemografischer Merkmale nach Verwendungskategorie

Tabelle 52 Influenza 2014: Haushaltsgröße nach Nutzungsgewohnheit und Modell (in %)

	Haushaltsgröße	Modell 1	Modell 2	Modell 3	Modell 4
u_1	1 Person	41,2	41,2	41,2	41,2
	2 Personen	41,2	41,2	41,2	41,2
	3 Personen	7,8	7,8	7,8	7,8
	4 Personen	4,9	4,9	4,9	4,9
	5 Personen	3,9	3,9	3,9	3,9
	6+ Personen	1,0	1,0	1,0	1,0
u_2	1 Person	25,7	25,9	26,4	24,6
	2 Personen	47,4	46,3	45,4	45,9
	3 Personen	8,6	8,6	8,0	11,0
	4 Personen	12,5	13,0	12,9	12,5
	5 Personen	5,3	5,6	6,1	5,3
	6+ Personen	0,7	0,6	1,2	0,7
u_3	1 Person	22,7	22,8	22,5	21,5
	2 Personen	37,5	37,9	38,2	40,7
	3 Personen	17,8	17,6	18,5	16,9
	4 Personen	16,1	15,8	15,9	16,9
	5 Personen	4,5	4,6	4,0	4,1
	6+ Personen	1,3	1,2	1,0	0,0
u_4	1 Person	21,1	20,5	21,1	22,4
	2 Personen	31,7	31,9	31,9	31,6
	3 Personen	21,7	22,2	20,5	21,1
	4 Personen	17,4	17,8	17,8	17,6
	5 Personen	4,3	3,8	4,9	4,3
	6+ Personen	3,7	3,8	3,8	3,1
u_5	1 Person	44,4	44,4	44,4	44,4
	2 Personen	30,2	30,2	30,2	30,2
	3 Personen	14,3	14,3	14,3	14,3
	4 Personen	7,9	7,9	7,9	7,9
	5 Personen	0,0	0,0	0,0	0,0
	6+ Personen	3,2	3,2	3,2	3,2
Erhebung	1 Person	26,1	26,1	26,1	26,1
	2 Personen	38,0	38,0	38,0	38,0
	3 Personen	15,8	15,8	15,8	15,8
	4 Personen	14,1	14,1	14,1	14,1
	5 Personen	4,3	4,3	4,3	4,3
	6+ Personen	1,7	1,7	1,7	1,7

A.2.3 Influenza 2014: Kontaktversuche nach Sub-Domain

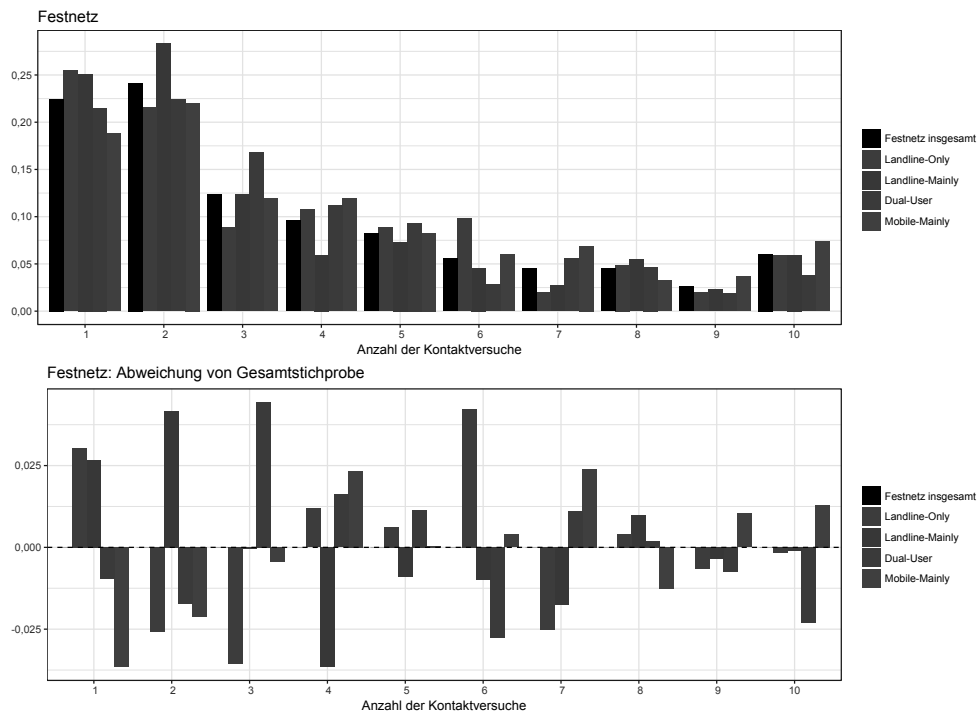


Abbildung 14 Festnetz: Anzahl der Kontakte nach Modell 5

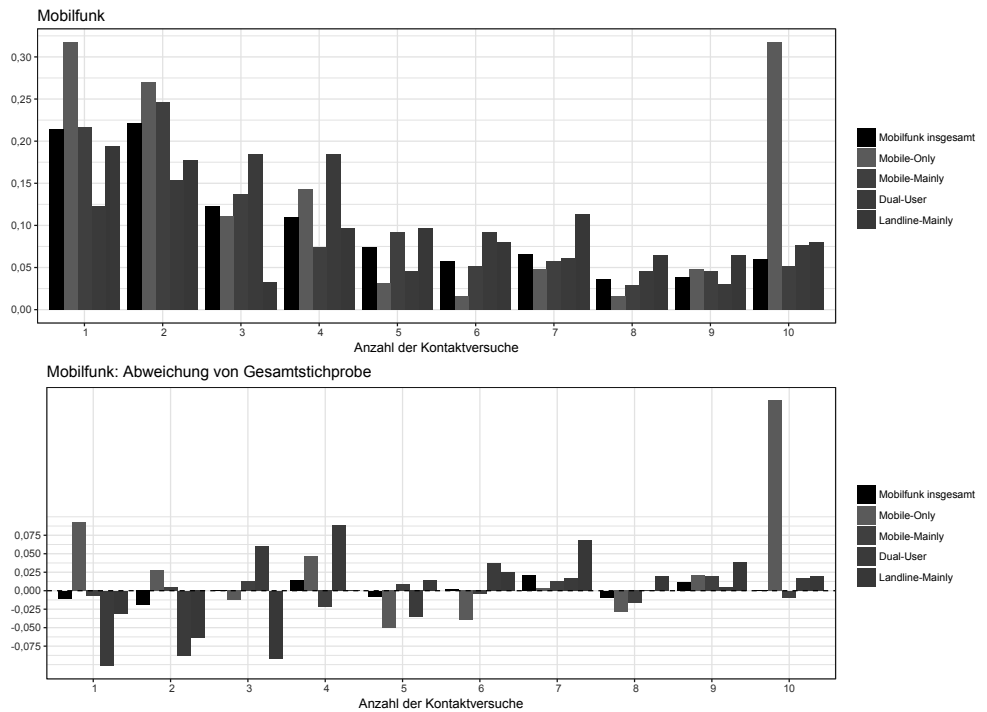


Abbildung 15 Mobilfunk: Anzahl der Kontakte nach Modell 5

A.3 Schätzergebnisse CELLA 2

Tabelle 53 CELLA 2: Geschlecht nach Modell und Methode 1 (in %)

Geschlecht	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	MZ 2010
Männlich	50,3	50,5	50,4	50,5	50,5	48,3
Weiblich	49,7	49,5	49,6	49,5	49,5	51,7
Designeffekt	1,38	1,52	1,53	1,54	1,53	●●

Tabelle 54 CELLA 2: Geschlecht nach Modell und Methode 2 (in %)

Geschlecht	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	MZ 2010
Männlich	50,3	50,7	50,7	50,8	50,7	48,3
Weiblich	49,7	49,3	49,3	49,2	49,3	51,7
Designeffekt	1,38	1,47	1,49	1,49	1,47	●●

Tabelle 55 CELLA 2: Migrationshintergrund nach Modell und Methode 1 (in %)

Migrationshintergrund	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	MZ 2010
Ja (dt. Staatsbürgerschaft)	94,1	94,1	94,3	94,3	94,2	93,3
Nein	5,9	5,9	5,7	5,7	5,8	6,7
Designeffekt	1,38	1,52	1,53	1,54	1,53	●●

Tabelle 56 CELLA 2: Migrationshintergrund nach Modell und Methode 2 (in %)

Migrationshintergrund	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	MZ 2010
Ja (dt. Staatsbürgerschaft)	94,1	94,1	94,2	94,3	94,1	93,3
Nein	5,9	5,9	5,8	5,7	5,9	6,7
Designeffekt	1,38	1,47	1,49	1,49	1,47	●●

Tabelle 57 CELLA 2: Haushaltsgröße nach Modell und Methode 1 (in %)

Haushaltsgröße	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	MZ 2010
0 Personen	0,1	0,1	0,1	0,1	0,1	●●
1 Person	20,9	19,5	19,5	19,4	19,4	19,2
2 Personen	34,6	34,3	34,6	34,6	34,5	34,0
3 Personen	19,2	19,4	19,3	19,4	19,3	18,3
4 Personen	15,6	16,5	16,4	16,4	16,6	18,6
5+ Personen	9,7	10,2	10,0	10,1	10,2	9,9
Designeffekt	1,38	1,52	1,53	1,54	1,53	●●

Tabelle 58 CELLA 2: Haushaltsgröße nach Modell und Methode 2 (in %)

Haushaltsgröße	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	MZ 2010
0 Personen	0,1	0,1	0,1	0,1	0,1	●●
1 Person	20,9	19,9	20,0	19,8	19,8	19,2
2 Personen	34,6	34,2	34,6	34,6	34,4	34,0
3 Personen	19,2	19,4	19,2	19,4	19,2	18,3
4 Personen	15,6	16,3	16,1	16,1	16,3	18,6
5+ Personen	9,7	10,2	10,0	10,1	10,2	9,9
Designeffekt	1,38	1,47	1,49	1,49	1,47	●●

Tabelle 59 CELLA 2: Familienstand nach Modell und Methode 1 (in %)

Familienstand	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	MZ 2010
Verheiratet	49,1	50,5	51,1	51,3	50,8	46,7
Eingetragene Lebenspartnerschaft	0,1	0,1	0,1	0,1	0,1	0,1
Geschieden	9,2	8,7	8,7	8,6	8,6	6,4
Verwitwet	5,1	4,8	5,0	4,9	4,9	7,7
Ledig	36,5	35,8	35,2	35,0	35,5	39,0
Designeffekt	1,38	1,52	1,53	1,54	1,53	●●

Tabelle 60 CELLA 2: Familienstand nach Modell und Methode 2 (in %)

Familienstand	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	MZ 2010
Verheiratet	49,1	50,0	50,7	50,8	50,4	46,7
Eingetragene Lebenspartnerschaft	0,1	0,1	0,1	0,1	0,1	0,1
Geschieden	9,2	8,8	8,8	8,8	8,7	6,4
Verwitwet	5,1	4,8	5,0	4,9	4,9	7,7
Ledig	36,5	36,2	35,4	35,4	35,9	39,0
Designeffekt	1,38	1,47	1,49	1,49	1,47	●●

Tabelle 61 CELLA 2: Privates Nettoeinkommen nach Modell und Methode 1 (in %)

Nettoeinkommen in Euro	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	MZ 2010
b.u. 2000	64,3	63,9	63,6	63,5	63,7	60,5
b.u. 4500	24,4	24,6	24,9	25,0	24,9	13,3
b.u. 10000	2,1	2,1	2,1	2,1	2,1	1,3
10000+	0,5	0,5	0,5	0,5	0,5	0,2
Kein eigenes Einkommen	8,7	8,8	8,8	8,8	8,8	18,7
Andere/Keine Angaben	●●	●●	●●	●●	●●	6,0
Designeffekt	1,38	1,52	1,53	1,54	1,53	●●

Tabelle 62 CELLA 2: Privates Nettoeinkommen nach Modell und Methode 2 (in %)

Nettoeinkommen in Euro	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	MZ 2010
b.u. 2000	64,3	63,9	63,6	63,4	63,7	60,5
b.u. 4500	24,4	24,6	25,0	25,2	24,9	13,3
b.u. 10000	2,1	2,1	2,2	2,2	2,1	1,3
10000+	0,5	0,5	0,5	0,5	0,5	0,2
Kein eigenes Einkommen	8,7	8,7	8,7	8,7	8,7	18,7
Andere/Keine Angaben	●●	●●	●●	●●	●●	6,0
Designeffekt	1,38	1,47	1,49	1,49	1,47	●●

A.4 Schätzergebnisse Influenza 2014

Tabelle 63 Influenza 2014: Geschlecht nach Modell und Methode 1 (in %)

Geschlecht	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	Zensus 2011
Männlich	46,4	46,3	46,3	46,2	46,2	48,3
Weiblich	53,6	53,7	53,7	53,8	53,8	51,7
Designeffekt	1,43	1,45	1,45	1,45	1,46	●●

Tabelle 64 Influenza 2014: Geschlecht nach Modell und Methode 2 (in %)

Geschlecht	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	Zensus 2011
Männlich	46,4	46,7	46,7	45,7	46,7	48,3
Weiblich	53,6	53,3	53,3	54,3	53,3	51,7
Designeffekt	1,43	1,43	1,43	1,71	1,57	●●

Tabelle 65 Influenza 2014: Migrationshintergrund nach Modell und Methode 1 (in %)

Migrationshintergrund	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	Zensus 2011
Ja (in Dtl. geboren)	87,7	88,1	88,1	88,1	88,0	82,3
Nein	12,3	11,9	11,9	11,9	12,0	17,7
Designeffekt	1,43	1,45	1,45	1,45	1,46	●●

Tabelle 66 Influenza 2014: Migrationshintergrund nach Modell und Methode 2 (in %)

Migrationshintergrund	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	Zensus 2011
Ja (in Dtl. geboren)	87,7	88,1	88,1	88,1	88,0	82,3
Nein	12,3	11,9	11,9	12,4	12,1	17,7
Designeffekt	1,43	1,43	1,43	1,71	1,57	●●

Tabelle 67 Influenza 2014: Haushaltsgröße nach Modell und Methode 1 (in %)

Haushaltsgröße	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	Zensus 2011
1 Person	21,8	20,7	20,8	20,7	20,6	20,4
2 Personen	34,3	34,4	34,5	34,5	34,2	35,9
3 Personen	19,3	18,8	18,6	18,5	18,8	19,5
4 Personen	16,1	17,9	18,0	17,8	18,0	15,3
5 Personen	5,9	5,6	5,5	5,7	5,7	5,5
6+ Personen	2,6	2,7	2,6	2,7	2,7	3,4
Designeffekt	1,43	1,45	1,45	1,45	1,46	●●

Tabelle 68 Influenza 2014: Haushaltsgröße nach Modell und Methode 2 (in %)

Haushaltsgröße	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	Zensus 2011
1 Person	21,8	21,3	21,4	20,0	20,9	20,4
2 Personen	34,3	34,5	34,6	32,9	33,1	35,9
3 Personen	19,3	18,6	18,4	19,4	19,2	19,5
4 Personen	16,1	17,5	17,6	19,4	18,3	15,3
5 Personen	5,9	5,5	5,5	5,8	5,9	5,5
6+ Personen	2,6	2,6	2,6	2,5	2,6	3,4
Designeffekt	1,43	1,43	1,43	1,71	1,57	●●

Tabelle 69 Influenza 2014: Alter nach Modell und Methode 1 (in %)

Altersgruppe	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	Zensus 2011
18-29	19,8	19,6	19,5	19,6	19,7	17,0
30-39	17,8	17,6	17,7	17,4	17,5	14,2
40-49	21,2	22,1	22,1	22,1	22,3	19,9
50-59	24,3	23,8	23,8	23,9	24,0	17,3
60-69	8,8	9,3	9,2	9,2	9,1	13,2
70-79	6,8	6,6	6,6	6,6	6,4	12,1
80+	1,2	1,1	1,1	1,1	1,1	6,3
Designeffekt	1,43	1,45	1,45	1,45	1,46	●●

Tabelle 70 Influenza 2014: Alter nach Modell und Methode 2 (in %)

Altersgruppe	SF-Schätzer	Modell1	Modell2	Modell3	Modell4	Zensus 2011
18-29	19,8	19,6	19,6	20,0	19,4	17,0
30-39	17,8	17,3	17,4	18,9	18,5	14,2
40-49	21,2	21,8	21,8	22,7	22,2	19,9
50-59	24,3	23,7	23,6	24,1	24,4	17,3
60-69	8,8	9,7	9,7	7,6	8,3	13,2
70-79	6,8	6,8	6,8	5,6	6,0	12,1
80+	1,2	1,1	1,1	1,2	1,2	6,3
Designeffekt	1,43	1,43	1,43	1,71	1,57	••

A.5 Schätzergebnisse DSF für CELLA 2

Tabelle 71 CELLA 2: Höchster allgemeiner Schulabschluss mit DSF-Schätzer nach Modell 2 (in %)

Bildungsabschluss	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	MZ 2010
Ohne/noch kein	4,0	3,9	3,7	3,6	7,6
Haupt-/Volksschulabschl.	20,9	20,8	19,5	19,5	37,0
Polytechn. OS	6,2	6,1	6,1	6,1	7,1
Mittlere Reife	26,4	26,0	26,7	26,7	21,7
FH/HS - Reife	40,9	41,6	42,5	42,6	25,8
Andere	1,5	1,6	1,5	1,5	0,4
Designeffekt	1,38	1,49	1,16	1,15	●●

Tabelle 72 CELLA 2: Höchster allgemeiner Schulabschluss mit DSF-Schätzer nach Modell 3 (in %)

Bildungsabschluss	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	MZ 2010
Ohne/noch kein	4,0	3,9	3,6	3,6	7,6
Haupt-/Volksschulabschl.	20,9	20,8	19,5	19,5	37,0
Polytechn. OS	6,2	6,2	6,1	6,1	7,1
Mittlere Reife	26,4	26,1	26,7	26,7	21,7
FH/HS - Reife	40,9	41,5	42,6	42,6	25,8
Andere	1,5	1,6	1,5	1,5	0,4
Designeffekt	1,38	1,49	1,16	1,15	●●

Tabelle 73 CELLA 2: Geschlecht mit DSF-Schätzer nach Modell 1 (in %)

Geschlecht	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	MZ 2010
Männlich	50,3	50,7	50,6	50,5	48,3
Weiblich	49,7	49,3	49,4	49,5	51,7
Designeffekt	1,38	1,47	1,15	1,14	●●

Tabelle 74 CELLA 2: Geschlecht mit DSF-Schätzer nach Modell 2 (in %)

Geschlecht	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	MZ 2010
Männlich	50,3	50,7	50,5	50,4	48,3
Weiblich	49,7	49,3	49,5	49,6	51,7
Designeffekt	1,38	1,49	1,16	1,15	●●

Tabelle 75 CELLA 2: Geschlecht mit DSF-Schätzer nach Modell 3 (in %)

Geschlecht	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	MZ 2010
Männlich	50,3	50,8	50,5	50,5	48,3
Weiblich	49,7	49,2	49,5	49,5	51,7
Designeffekt	1,38	1,49	1,16	1,15	●●

Tabelle 76 CELLA 2: Migrationshintergrund mit DSF-Schätzer nach Modell 1 (in %)

Migrationshintergrund	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	MZ 2010
Ja (dt. Staatsbürgerschaft)	94,1	94,1	93,8	93,9	93,3
Nein	5,9	5,9	6,2	6,1	6,7
Designeffekt	1,38	1,47	1,15	1,14	●●

Tabelle 77 CELLA 2: Migrationshintergrund mit DSF-Schätzer nach Modell 2 (in %)

Migrationshintergrund	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	MZ 2010
Ja (dt. Staatsbürgerschaft)	94,1	94,2	93,8	93,9	93,3
Nein	5,9	5,8	6,2	6,1	6,7
Designeffekt	1,38	1,49	1,16	1,15	●●

Tabelle 78 CELLA 2: Migrationshintergrund mit DSF-Schätzer nach Modell 3 (in %)

Migrationshintergrund	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	MZ 2010
Ja (dt. Staatsbürgerschaft)	94,1	94,3	93,5	93,9	93,3
Nein	5,9	5,7	6,1	6,1	6,7
Designeffekt	1,38	1,47	1,15	1,14	●●

Tabelle 79 CELLA 2: Haushaltsgröße mit DSF-Schätzer nach Modell 1 (in %)

Haushaltsgröße	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	MZ 2010
0 Personen	0,1	0,1	0,1	0,1	●●
1 Person	20,9	19,9	19,6	20,2	19,2
2 Personen	34,6	34,2	34,3	34,3	34,0
3 Personen	19,2	19,4	19,8	19,6	18,3
4 Personen	15,6	16,3	16,3	16,1	18,6
5+ Personen	9,7	10,2	9,9	9,7	9,9
Designeffekt	1,38	1,47	1,15	1,14	●●

Tabelle 80 CELLA 2: Haushaltsgröße mit DSF-Schätzer nach Modell 2 (in %)

Haushaltsgröße	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	MZ 2010
0 Personen	0,1	0,1	0,1	0,1	●●
1 Person	20,9	20,0	19,6	20,1	19,2
2 Personen	34,6	34,6	34,5	34,5	34,0
3 Personen	19,2	19,2	19,7	19,5	18,3
4 Personen	15,6	16,1	16,3	16,1	18,6
5+ Personen	9,7	10,0	9,8	9,6	9,9
Designeffekt	1,38	1,49	1,16	1,15	●●

Tabelle 81 CELLA 2: Haushaltsgröße mit DSF-Schätzer nach Modell 3 (in %)

Haushaltsgröße	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	MZ 2010
0 Personen	0,1	0,1	0,1	0,1	●●
1 Person	20,9	19,8	19,6	20,1	19,2
2 Personen	34,6	34,6	34,5	34,5	34,0
3 Personen	19,2	19,4	19,7	19,6	18,3
4 Personen	15,6	16,1	16,3	16,1	18,6
5+ Personen	9,7	10,1	9,8	9,6	9,9
Designeffekt	1,38	1,49	1,16	1,15	●●

Tabelle 82 CELLA 2: Familienstand mit DSF-Schätzer nach Modell 1 (in %)

Familienstand	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	MZ 2010
Verheiratet	49,1	50,0	49,5	49,3	46,7
Eingetragene Lebenspartnerschaft	0,1	0,1	0,1	0,1	0,1
Geschieden	9,2	8,8	9,0	9,1	6,4
Verwitwet	5,1	4,8	4,5	4,6	7,7
Ledig	36,5	36,2	36,9	36,9	39,0
Designeffekt	1,38	1,47	1,15	1,14	●●

Tabelle 83 CELLA 2: Familienstand mit DSF-Schätzer nach Modell 2 (in %)

Familienstand	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	MZ 2010
Verheiratet	49,1	50,7	49,8	49,6	46,7
Eingetragene Lebenspartnerschaft	0,1	0,1	0,1	0,1	0,1
Geschieden	9,2	8,8	9,0	9,1	6,4
Verwitwet	5,1	5,0	4,5	4,6	7,7
Ledig	36,5	35,4	36,6	36,6	39,0
Designeffekt	1,38	1,49	1,16	1,15	●●

Tabelle 84 CELLA 2: Familienstand mit DSF-Schätzer nach Modell 3 (in %)

Familienstand	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	MZ 2010
Verheiratet	49,1	50,8	49,9	49,7	46,7
Eingetragene Lebenspartnerschaft	0,1	0,1	0,1	0,1	0,1
Geschieden	9,2	8,8	9,0	9,1	6,4
Verwitwet	5,1	4,9	4,5	4,6	7,7
Ledig	36,5	35,4	36,6	36,6	39,0
Designeffekt	1,38	1,49	1,16	1,15	●●

Tabelle 85 CELLA 2: Privates Nettoeinkommen mit DSF-Schätzer nach Modell 1 (in %)

Nettoeinkommen in Euro	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	MZ 2010
b.u. 2000	64,3	63,9	63,6	63,7	60,5
b.u. 4500	24,4	24,6	25,1	25,1	13,3
b.u. 10000	2,1	2,1	2,2	2,2	1,3
10000+	0,5	0,5	0,6	0,6	0,2
Kein eigenes Einkommen	8,7	8,7	8,5	8,4	18,7
Andere/Keine Angaben	●●	●●	●●	●●	6,0
Designeffekt	1,38	1,47	1,15	1,14	●●

Tabelle 86 CELLA 2: Privates Nettoeinkommen mit DSF-Schätzer nach Modell 2 (in %)

Nettoeinkommen in Euro	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	MZ 2010
b.u. 2000	64,3	63,6	63,6	63,6	60,5
b.u. 4500	24,4	25,0	25,0	25,1	13,3
b.u. 10000	2,1	2,2	2,2	2,3	1,3
10000+	0,5	0,5	0,6	0,6	0,2
Kein eigenes Einkommen	8,7	8,7	8,5	8,4	18,7
Andere/Keine Angaben	●●	●●	●●	●●	6,0
Designeffekt	1,38	1,49	1,16	1,15	●●

Tabelle 87 CELLA 2: Privates Nettoeinkommen mit DSF-Schätzer nach Modell 3 (in %)

Nettoeinkommen in Euro	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	MZ 2010
b.u. 2000	64,3	63,4	63,5	63,6	60,5
b.u. 4500	24,4	25,2	25,1	25,1	13,3
b.u. 10000	2,1	2,2	2,2	2,2	1,3
10000+	0,5	0,5	0,6	0,6	0,2
Kein eigenes Einkommen	8,7	8,7	8,5	8,4	18,7
Andere/Keine Angaben	••	••	••	••	6,0
Designeffekt	1,38	1,49	1,16	1,15	••

A.6 Schätzergebnisse DSF für Influenza 2014

Tabelle 88 Influenza 2014: Bildungsabschluss mit DSF-Schätzer nach Modell 2 (in %)

Bildungsabschluss	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	Zensus 2011
Höchstens 7 Jahre Schulbesuch	0,7	0,6	0,6	0,6	5,0
Haupt-/Volksschulabschluss	15,7	15,4	13,9	14,0	36,6
Mittlere Reife	33,1	33,1	34,3	34,2	29,0
Fachhochschulreife	9,6	9,7	9,4	9,3	8,3
Abitur	40,9	41,2	41,9	41,9	21,1
Designeffekt	1,43	1,45	1,29	1,27	●●

Tabelle 89 Influenza 2014: Geschlecht mit DSF-Schätzer nach Modell 1 (in %)

Geschlecht	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	Zensus 2011
Männlich	46,4	46,3	45,5	45,5	48,3
Weiblich	53,6	53,7	54,5	54,5	51,7
Designeffekt	1,43	1,45	1,29	1,27	●●

Tabelle 90 Influenza 2014: Geschlecht mit DSF-Schätzer nach Modell 2 (in %)

Geschlecht	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	Zensus 2011
Männlich	46,4	46,3	45,5	45,5	48,3
Weiblich	53,6	53,7	54,5	54,5	51,7
Designeffekt	1,43	1,45	1,29	1,27	●●

Tabelle 91 Influenza 2014: Migrationshintergrund mit DSF-Schätzer nach Modell 1 (in %)

Migrationshintergrund	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	Zensus 2011
Ja (in Dtl. geboren)	87,7	88,1	87,7	87,8	82,3
Nein	12,3	11,9	12,3	12,2	17,7
Designeffekt	1,43	1,45	1,29	1,27	••

Tabelle 92 Influenza 2014: Migrationshintergrund mit DSF-Schätzer nach Modell 2 (in %)

Migrationshintergrund	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	Zensus 2011
Ja (in Dtl. geboren)	87,7	88,1	87,7	87,7	82,3
Nein	12,3	11,9	12,3	12,3	17,7
Designeffekt	1,43	1,45	1,29	1,27	••

Tabelle 93 Influenza 2014: Haushaltgröße mit DSF-Schätzer nach Modell 1 (in %)

Haushaltsgröße	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	Zensus 2011
1 Person	21,8	20,7	18,4	18,9	20,4
2 Personen	34,3	34,4	34,3	34,5	35,9
3 Personen	19,3	18,8	20,6	20,3	19,5
4 Personen	16,1	17,9	18,1	17,8	15,3
5 Personen	5,9	5,6	6,0	6,0	5,5
6+ Personen	2,6	2,7	2,6	2,5	3,4
Designeffekt	1,43	1,45	1,29	1,27	••

Tabelle 94 Influenza 2014: Haushaltsgröße mit DSF-Schätzer nach Modell 2 (in %)

Haushaltsgröße	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	Zensus 2011
1 Person	21,8	20,8	18,4	18,9	20,4
2 Personen	34,3	34,5	34,3	34,5	35,9
3 Personen	19,3	18,6	20,6	20,3	19,5
4 Personen	16,1	18,0	18,1	17,8	15,3
5 Personen	5,9	5,5	6,0	6,0	5,5
6+ Personen	2,6	2,6	2,6	2,5	3,4
Designeffekt	1,43	1,45	1,29	1,27	●●

Tabelle 95 Influenza 2014: Alter mit DSF-Schätzer nach Modell 1 (in %)

Altersgruppe	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	Zensus 2011
18-29	19,8	19,6	19,3	19,1	17,0
30-39	17,8	17,6	17,9	17,9	14,2
40-49	21,2	22,1	23,0	22,8	19,9
50-59	24,3	23,8	25,4	25,2	17,3
60-69	8,8	9,3	7,9	8,2	13,2
70-79	6,8	6,6	5,6	5,8	12,1
80+	1,2	1,1	0,9	0,9	6,3
Designeffekt	1,43	1,45	1,29	1,27	●●

Tabelle 96 Influenza 2014: Alter mit DSF-Schätzer nach Modell 2 (in %)

Altersgruppe	SF-Schätzer	DSC-Schätzer	$\widehat{DSF}_1$	$\widehat{DSF}_2$	Zensus 2011
18-29	19,8	19,5	19,3	19,1	17,0
30-39	17,8	17,7	18,0	17,9	14,2
40-49	21,2	22,1	23,0	22,8	19,9
50-59	24,3	23,8	25,4	25,3	17,3
60-69	8,8	9,2	7,8	8,2	13,2
70-79	6,8	6,6	5,5	5,8	12,1
80+	1,2	1,1	0,9	0,9	6,3
Designeffekt	1,43	1,45	1,29	1,27	●●

Die Vorzüge von Dual-Frame-Telefonerhebungen in Deutschland wurden bereits im Zuge der beiden CELLA-Erhebungen ausgiebig untersucht. Beide Erhebungen wurden durch GESIS, das Leibniz-Institut für Sozialwissenschaften, gemeinsam mit der Technischen Universität Dresden 2007/2008 und 2010/2011 veranlasst. Neben einer Vielzahl an Erkenntnissen über die Only-Population und den Möglichkeiten und Limitationen solcher Erhebungen, war ein entscheidendes Ergebnis die Entwicklung eines Gewichtungsansatzes, der auch heutzutage noch immer Verwendung findet. Dieser basiert auf einem Horvitz-Thompson-Schätzer zur Berücksichtigung ungleicher Auswahlwahrscheinlichkeiten sowie einem GREG-Schätzer zur Kalibrierung und Anpassungsgewichtung bei Non-response. Jedoch berücksichtigt dieses Vorgehen keine Antwortausfälle, die sich aufgrund der Auswahlgrundlage ergeben. Eine weitere Untersuchung der Daten der Erhebung CELLA 2 lässt indes vermuten, dass weitere Faktoren, fernab der soziodemografischen Charakteristika die Teilnahmewahrscheinlichkeit in solchen Erhebungen beeinflussen kann. Da die Kontaktergebnisse nach verwendetem Telekommunikationsmedium innerhalb dieser Erhebung erheblich voneinander abweichen, besteht die Vermutung, dass eine weitere Quelle des Nonresponse durch sich unterscheidende Nutzungsgewohnheiten von bestimmten Telekommunikationsmedien beschrieben werden kann. Um dieser Möglichkeit nachzugehen, führt die vorliegende Arbeit einen Gewichtungsansatz ein, der gerade diese von der Auswahlgrundlage und dem Telekommunikationsmedium abhängigen Quellen stärker berücksichtigt. Dazu wird zuvor der Bereich der sog. Dual-User in mehrere disjunkte Nutzungskategorien unterteilt und anschließend ein modifizierter Composite-Gewichtungsansatz vorgestellt. Dieses Vorgehen wird anschließend anhand von geeigneten Erhebungen ausgiebig untersucht.

The benefits of dual-frame telephone surveys in Germany have been extensively examined within both CELLA studies that GESIS, the Leibniz Institute for the Social Sciences, conducted in cooperation with the Dresden University of Technology in 2007/2008 and 2010/2011. Besides numerous insight regarding the population of only-households and the potentials of dual-frame surveys, a key-development has been a weighting approach that is nowadays commonly practised when weighting dual-frame surveys in Germany. The weighting approach adjusts for unequal inclusion probabilities by employing the Horvitz-Thompson-Estimator and for frame-independent nonresponse caused by socio-demographic characteristics using the GREG-Estimator.

However, this weighting approach does not consider nonresponse in dual-frame surveys that is frame-dependent. A further investigation of CELLA 2's data suggests that there might be factors apart from socio-demographic characteristics that can affect individual response behavior. Since CELLA 2's call-outcomes of the landline and mobile sample differ considerably, another source of nonresponse might be determined by varying usage-patterns of specific means of communications within the overlap domain of a dual-frame survey. Therefore, this book will introduce a weighting approach that takes such frame-dependent, device-specific Response behavior into account by subdividing the overlap domain into several, disjoint categories determined by usage-patterns and employing a composite design weighting approach. This newly introduced weighting approach will then be evaluated by application on appropriate dual-frame surveys.