

Teilnahmeverhalten und Stichprobenverzerrung in der deutschen Stichprobe des European Social Survey

Michael Weinhardt & Stefan Liebig

Universität Bielefeld

Einleitung

Teilnahmeverhalten (Nonresponse) und Stichprobenverzerrungen (Bias) zählen vermutlich zu den beiden wichtigsten Forschungsgebieten der Surveymethodologie. Beide Bereiche hängen eng zusammen: Die systematische, nicht zufällige Verweigerung oder Nichtteilnahme an einer Studie führen zu Ergebnissen, die bei Verallgemeinerungen auf die Grundgesamtheit diese nur verzerrt wiedergeben (Groves und Peytcheva 2008). Dies stellt auch eine Herausforderung für den European Social Survey (ESS) dar. Bereits Billiet et al. (2007) vermuten den Zusammenhang zentraler ESS-Inhalte mit dem Prozess der Studienteilnahme: „In the ESS we can expect substantial differences between respondents and non-respondents. Many of the variables covered in the first round of the ESS survey (social participation, political interest and involvement, civic duties) have previously been found or are believed to correlate substantially with survey participation (Voogt und Saris 2003; Groves und Couper 1998). Accordingly, we expect that non-participation in the ESS will be highly likely to cause biased estimates and limit the generalisability to each population.“ Gleichzeitig wird in der Literatur seit einiger Zeit der Einfluss des Studienthemas auf die Teilnahmebereitschaft diskutiert (Groves et al. 2004; Marcus et al. 2007; Zillmann et al. 2013). Die Abhängigkeit der Teilnahmewahrscheinlichkeit von

Inhalten und Themen der Befragung muss bei der Beurteilung der Stichprobenqualität demnach berücksichtigt werden.

Die systematische Teilnahmeverweigerung ist zudem nicht die einzige Quelle verzerrter Stichproben. Schon ein defizitärer Stichprobenrahmen oder Probleme bei der Umsetzung eines Ziehungsdesigns im Feld können sich negativ auf die Stichprobenqualität auswirken. Im European Social Survey wird daher besonderer Wert auf eine hohe methodische Qualität bei der Durchführung gelegt, die auch durch externe Überprüfungen bestätigt wird (Kohler 2008). Der hohe methodische Anspruch äußert sich auch in der Wahl des Ziehungsdesigns: In Deutschland wird eine registerbasierte Einwohnermeldeamtsstichprobe verwendet (Neller 2005), die in der Literatur zur Gewinnung von Personenstichproben im Vergleich zu anderen Verfahren als überlegen angesehen wird (z.B. Menold 2014). Der Beitrag widmet sich deshalb zunächst der Überprüfung dieses Anspruchs, mittels einer registerbasierten Stichprobe die Grundgesamtheit möglichst getreu abzubilden.

Die in der Literatur diskutierten Verfahren zur Identifikation von Stichprobenverzerrungen basieren meist auf dem Abgleich mit externen Kriterien, die als Vergleichsstandards herangezogen werden (vgl. Kohler 2007; Schouten et al. 2009). Diese Verfahren erlauben jedoch lediglich einen Abgleich mit „objektiven“, „harten“ strukturellen Merkmalen; die Identifizierung von Verzerrungen bei der Messung „weicher“, stärker subjektiver Maße wie Meinungen und Einstellungen ist wesentlich schwieriger, da hier praktisch keine „objektiven“ Referenzdaten zur externen Evaluation vorliegen. Gleichzeitig bieten sich subjektive Maße nicht auf gleiche Weise für die Konstruktion „interner Repräsentativitätskriterien“ (Kohler 2007) an wie Strukturmerkmale der Geschlechterverteilung oder Haushaltszusammensetzung. Aus diesen Gründen sind Studien zu Stichprobenverzerrungen bei Einstellungsmaßen nur selten in der Literatur zu finden. Da nicht davon auszugehen ist, dass Verzerrungen von „harten“ und „weichen“ Kriterien denselben Mechanismen folgen, liegen hierzu kaum erprobte Verfahren vor. Deshalb widmet sich dieser Beitrag der Betrachtung einer im ESS durchgeführten Kurzbefragung von Personen, die die Teilnahme an der eigentlichen ESS-Befragung verweigert haben (vgl. Billiet et al. 2007).

Trotz der offensichtlichen Problematik eines solchen Vorgehens zur Beurteilung von Nonresponse – besonders in der Tatsache begründet, dass auch eine solche Verweigererbefragung selbst von Nonresponse betroffen

ist – werden die daraus resultierenden Daten hier zur Analyse herangezogen, da sonst kaum Informationen oder Studien zu dieser Thematik vorliegen. Zentraler Bestandteil dieses Verweigererfragebogens im deutschen ESS ist ein subjektives Item zur Selbsteinschätzung des politischen Interesses, das in derselben Form im Hauptfragebogen enthalten ist. Damit lässt sich der Zusammenhang des politischen Interesses mit der Teilnahmewahrscheinlichkeit betrachten. Gleichzeitig rückt das Problem des Zusammenhangs von Studieninteresse und Teilnahmebereitschaft in den Fokus: Hängt die Teilnahmebereitschaft der Befragten von zentralen Befragungsinhalten ab, führt dies zu verzerrten Messergebnissen bei diesen Inhalten und damit korrelierten Variablen (vgl. Groves et al. 2004). Da politische Einstellungen und politische Teilhabe zentrale Befragungsinhalte des ESS sind, ist für die Stichprobenqualität des ESS entscheidend, ob die Bereitschaft zur Teilnahme vom politischen Interesse der Zielpersonen abhängt.

Der Aufbau dieses Beitrags ist wie folgt: Zunächst wird eine kurze Einführung über den deutschen ESS und sein Stichprobendesign gegeben. Dabei wird auch über die Entwicklung von Ausschöpfung, Verweigerung und Nichtkontaktierung seit der ersten Erhebung des ESS im Jahr 2002 berichtet. Im zweiten Abschnitt wird die Güte der Brutto- wie Nettostichprobe durch Vergleiche mit externen Daten basierend auf dem Zensus 2011 und der Bundeswahlstatistik evaluiert. Im dritten Abschnitt wird auf das politische Interesse der potenziellen Befragten als Teilnahmegrund eingegangen. Hierfür wird auf Daten des ESS-Verweigerungsfragebogens zurückgegriffen und auch thematisiert, inwieweit ein solches Instrument hilfreiche Informationen über Nichtteilnehmer liefern kann und auch andere Variablen von einer Stichprobenverzerrung betroffen sind. In der abschließenden Diskussion der Ergebnisse wird auf mögliche Konsequenzen für das Studiendesign und die Feldarbeit des ESS in Deutschland eingegangen.

Die deutsche ESS-Teilstichprobe als Einwohnermeldeamtsstichprobe

Der ESS ist eine der führenden Erhebungen für die europavergleichende Politik- und Sozialforschung und wird seit 2002 im Abstand von zwei Jahren wiederholt durchgeführt (vgl. Keil und van Deth 2012); Ende des Jahres 2014 bereits zum siebten Mal. Deutschland hat bisher an allen sieben Welten teilgenommen. Die Grundgesamtheit bildet hier die in Privathaushal-

ten lebende Wohnbevölkerung ab 15 Jahren. Die Befragung ist nicht auf deutsche Staatsbürger begrenzt, allerdings findet die Befragung nur auf Deutsch statt, sodass gute Deutschkenntnisse Voraussetzung sind, um an der Studie teilnehmen zu können (Übersetzungen etwa durch Familienmitglieder sind nicht gestattet).

Traditionell nutzt der ESS in Deutschland ein disproportionalen Stichprobendesign: Um Ost-West Vergleiche zu ermöglichen, werden 2.000 Interviews auf dem Gebiet der ehemaligen Bundesrepublik und 1.000 Interviews auf dem Gebiet der ehemaligen DDR angestrebt. Das Auswahlverfahren ist zweistufig angelegt. Zunächst werden die Gemeinden nach einem speziellen Verfahren als primary sampling units (PSUs) ausgewählt. Dafür wird aus Bundesländern, Kreisen und BIK-Gemeindegrößenklassen ein Schichtungstableau erstellt, aus dessen Zellen die einzelnen Gemeinden anschließend mit einer Wahrscheinlichkeit proportional zu ihrer Größe gezogen werden.¹ Im zweiten Schritt werden die Einwohnermeldeämter der Gemeinden angeschrieben und gebeten, aus ihrem Register eine Personenstichprobe nach einem vorgegebenen Verfahren zu ziehen, aus denen wiederum die Einsatzstichprobe für das Feld durch das Umfrageinstitut gezogen wird. Pro Sample-Point wird dabei die gleiche Zahl an Personenadressen ausgewählt. Durch dieses Verfahren entsteht eine nach regionalen Gesichtspunkten geschichtete, selbstgewichtete Stichprobe. Das bedeutet, dass innerhalb Ost- und Westdeutschlands alle Zielpersonen die gleiche theoretische Wahrscheinlichkeit besitzen, in die Stichprobe zu gelangen.

Ein Beispiel: In der sechsten Welle des ESS wurden in der ersten Auswahlstufe für Ost- und Westdeutschland zusammen 168 Sample-Points gezogen, die sich auf 155 Gemeinden verteilten (infas Methodenbericht 2013). Es wurden zwei Stichprobenstranchen mit zusammen 8.904 Adressen im Feld eingesetzt, 53 pro Sample-Point. Dieser Stichprobenansatz resultierte in 2.993 Interviews, von denen nach Datenprüfung und Interviewerkontrolle 2.975 als auswertbar eingestuft wurden.

Bei Einwohnermeldeamtsstichproben handelt es sich um reine Personenstichproben, in denen alle Personen der Grundgesamtheit eine im Prinzip gleiche Auswahlwahrscheinlichkeit besitzen. Der Umweg über eine Haushaltsstichprobe entfällt. Gleichzeitig reduziert diese Art der Stichproben-

1 Für eine genaue Beschreibung der technischen Details dieses Vorgehens siehe von der Heyde (2014).

ziehung, im Vergleich etwa zum Random-Route-Verfahren, den Einfluss der Interviewer auf die Auswahl der Befragten. Zudem sind bei einer registerbasierten Stichprobe Name und Adresse der Befragten bekannt, was es erlaubt, ein personalisiertes Anschreiben zu versenden (vgl. Groves et al. 2004). Aufgrund dieser und anderer Vorteile gilt eine Einwohnermeldeamtsstichprobe anderen Formen der Stichprobenziehung als qualitativ überlegen.

Es gibt allerdings auch Nachteile: So benötigt eine Einwohnermeldeamtsstichprobe einen sehr langen Vorlauf von etwa sechs Monaten, da zunächst die Einwohnermeldeämter der gezogenen Gemeinden kontaktiert und für eine Teilnahme geworben werden müssen.² Dies bedeutet einen sehr hohen administrativen Aufwand, zusätzlich fallen zum Teil beträchtliche Kosten in Form von administrativen Gebühren an, die regional variieren. Zusätzlich besteht zumindest theoretisch das Problem, dass sich die Grundgesamtheit zwischen Stichprobenziehung und Datenerhebung verändert haben kann. Zudem weisen die amtlichen Register trotz Meldepflicht eine relativ hohe Fehlerquote auf. Diese wurde insbesondere durch den Zensus 2011 deutlich, in dessen Verlauf sich zeigte, dass die Bevölkerung vieler Städte und Gemeinden tatsächlich niedriger war, als die auf den amtlichen Daten basierenden Bevölkerungsfortschreibungen vermuten ließen. Insbesondere Personen, die ins Ausland verziehen, melden sich demnach häufig nicht ab und verbleiben so fälschlicherweise in den Registern.

Auch aus anderen Gründen können die Register fehlerhaft sein, etwa wenn Jungerwachsene noch bei den Eltern gemeldet sind, inzwischen aber für ein Studium oder aus einem anderen Grund verzogen sind. Außerdem ist die Datenqualität der Melderegister selbst Schwankungen unterworfen. Regionen mit höherer regionaler Mobilität (z.B. Berlin) sind davon besonders betroffen. Hier kann es zu systematischen Verzerrungen kommen, die in unmittelbaren Zusammenhang mit dem Thema einer Studie stehen können: Gruppen mit höherer regionaler Mobilität haben eine geringere Chance in die Stichprobe zu gelangen und unterscheiden sich gleichzeitig vom Rest der Bevölkerung z.B. durch Alter und Bildungsgrad.

2 Hier kommt es zudem vereinzelt zu Verweigerungen auf der Ebene der Kommunen, die dann strukturell ersetzt werden müssen.

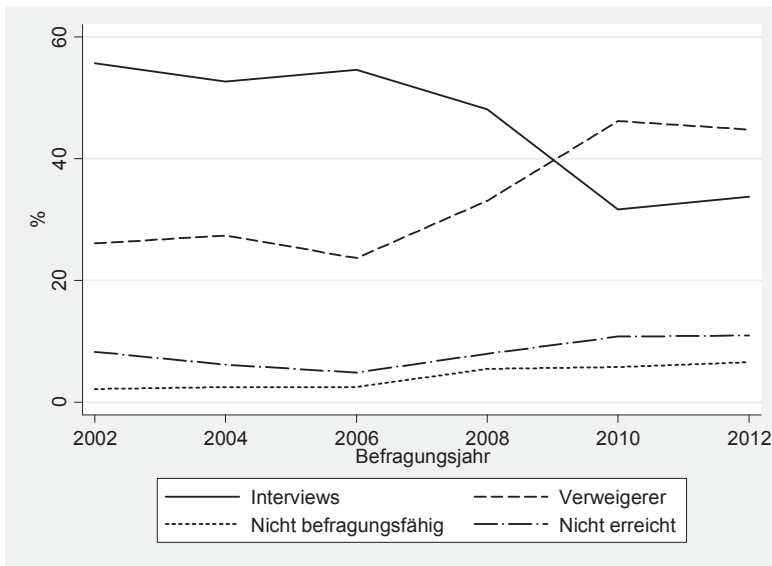


Abbildung 1: Anteil von Interviews, Verweigerungen, Nichterreichten sowie nicht befragungsfähiger Personen der deutschen ESS-Teilstichprobe in den ersten 6 Wellen an der Bruttostichprobe (ohne stichprobenneutrale Ausfälle).

Wie Abbildung 1 zeigt, betrug die Ausschöpfungsquote (berechnet nach ESS-Standard, Anteil der Interviews an der um stichprobenneutrale Ausfälle gereinigten Stichprobe, vgl. Stoop et al. 2010) in Deutschland für die ersten drei Wellen jeweils über 50 Prozent, fiel danach jedoch auf 32 bzw. 34 Prozent in den Jahren 2010 und 2012 ab. Gleichzeitig ist zu sehen, dass die schlechtere Ausschöpfung vor allem an ansteigenden Verweigerungsraten festzumachen ist. Während der Anteil von nicht erreichten Personen und Personen, die zur Teilnahme nicht in der Lage sind, über die Zeit innerhalb von leichten Schwankungen konstant bleibt, steigt die Verweigerungsrate zeitlich parallel zum Abfall der Ausschöpfungsquote an. Niedrigere Ausschöpfungsquoten sind also nicht Folge einer nachlassenden Qualität der Bruttostichprobe oder veränderter Kontaktierungsbemühungen, sondern sind vor allem auf ein verändertes Verhalten der potenziellen Befragten zurückzuführen. Wie die Güte von Brutto- und Nettostichprobe zu beurteilen

ist und wie sich das Befragtenverhalten generell auf die Repräsentativität der Stichprobe auswirkt, wird in den folgenden Abschnitten erörtert.

Repräsentativität der Brutto- und Nettostichprobe

Zur Beurteilung der Repräsentativität der ESS-Daten wird zunächst ein Abgleich der Bruttostichprobe mit der Grundgesamtheit vorgenommen und dann die realisierte Nettostichprobe mit der Bruttostichprobe verglichen. Beides geschieht auf Basis sozio-demografischer Merkmale, die über die Melderegister verfügbar sind. Danach wird die Nettostichprobe mit weiteren Daten aus dem Zensus 2011 und der Bundeswahlstatistik verglichen, um Anhaltspunkte einer möglichen Verzerrung der ESS-Stichprobe zu überprüfen.

Abgleich der Bruttostichprobe mit der Grundgesamtheit

Tabelle 1 zeigt einen Vergleich zwischen den Bruttostichproben der bisherigen ESS-Wellen mit der Grundgesamtheit für Bundesländer und BIK-Gemeindegrößenklassen³ getrennt nach West- und Ostdeutschland (die zugrundeliegenden Werte sind in Tabelle 7 im Appendix ersichtlich, für Welle 5 sind die entsprechenden Daten der Bruttostichprobe nicht verfügbar). Eingetragen sind Maximal- und Mittelwert der Prozentpunktdifferenzen zwischen den einzelnen Merkmalskategorien, um das Ausmaß der Abweichungen abschätzen zu können. Insgesamt fallen die Abweichungen zwischen Ziehungsrahmen und Bruttostichprobe sehr gering aus, die mittlere Differenz liegt für beide Variablen in allen dargestellten Wellen unter einem Prozentpunkt.

In Welle sechs weist die Einsatzstichprobe bezüglich der genannten Merkmale in den alten Bundesländern einen hohen Grad an Übereinstimmung auf, in den neuen Bundesländern gibt es etwas größere Abweichungen. Derartige Differenzen auf Bundeslandebene können bei einer Zufallsstichprobe mit einer vergleichsweise kleinen Anzahl von Sampling Points wie in Ostdeutschland auftreten. In den vorherigen Wellen ist eine solche Diskrepanz nicht zu beobachten. Insgesamt garantiert das Ziehungsdesign

3 Die BIK-Gemeindegrößenklassen beschreiben den siedlungsstrukturellen Typ der Gemeinde, insbesondere ihre Zugehörigkeit zu Ballungsgebieten.

aus den Registern der Meldeämter, wie zu erwarten, eine sehr genaue Entsprechung der Bruttostichprobe mit der Grundgesamtheit.

Tabelle 1: Maximum und Mittelwert der Abweichungen in Prozentpunkten zwischen Grundgesamtheit (Ziehungsrahmen) und Bruttostichprobe im deutschen ESS für Bundesland und BIK-Gemeindegrößenklasse getrennt nach Ost und West

ESS	Bundesland (West)		Bundesland (Ost)		BIK (West)		BIK (Ost)	
	Max.	Mittelw.	Max.	Mittelw.	Max.	Mittelw.	Max.	Mittelw.
2002	-1,2	0,65	-0,90	0,62	-1,00	0,39	-1,00	0,47
2004	0,5	0,35	-0,80	0,62	0,50	0,29	-1,10	0,53
2006	-0,6	0,25	-0,70	0,38	-0,60	0,20	-0,90	0,39
2008	-0,5	0,19	-0,60	0,40	-0,50	0,25	-1,70	0,67
2012	-0,7	0,25	-1,60	0,77	-0,50	0,26	1,00	0,32

Quelle: ESS Methodenberichte der Wellen 1-6 (infas 2003, 2005, 2007, 2009, 2013) und eigene Berechnungen.

Abgleich von Brutto- und Nettostichprobe

Ein Vorteil von Einwohnermeldeamtsstichproben besteht darin, dass einige Merkmale der ausgewählten Personen bereits vorab bekannt sind. Dazu gehören neben dem Wohnort auch das Geschlecht und das Alter der Person sowie deren Staatsangehörigkeit. Diese können für einen Vergleich von Bruttostichprobe und realisierter Stichprobe herangezogen werden, um auf diese Weise Stichprobenverzerrungen zu ermitteln.

Bei einem Vergleich anhand der Merkmale Ost/West, Bundesland, Gemeindegrößenklasse, Geschlecht und Alter für die sechste Welle zeigen sich insgesamt nur sehr geringe Abweichungen zwischen der realisierten Stichprobe und der Bruttostichprobe (Tabelle 2: Brutto-, Nettostichprobe und Nichtteilnehmer im Vergleich sozio-demografischer Merkmale). Bei den regionalen Merkmalen sind die beiden Flächenländer Bayern und Niedersachsen in der realisierten Stichprobe leicht überrepräsentiert, während Nordrhein-Westfalen, Berlin und in geringerem Maße Hamburg leicht unterrepräsentiert sind. Dies könnte ein Effekt des Verstärkerungsgrades sein:

Die Verteilung nach BIK-Gemeindegrößenklassen deutet darauf hin, dass Großstädte mit mehr als 500.000 Einwohnern in der realisierten Stichprobe vergleichsweise unterrepräsentiert sind. Effekte der sozio-ökonomischen Struktur dürften ebenfalls eine Rolle spielen.

Bei der Betrachtung individueller Merkmale zeigt sich, dass Frauen in der realisierten Stichprobe unterrepräsentiert sind: Ihr Anteil liegt um zwei Prozentpunkte niedriger als in der Bruttostichprobe. Differenzen zeigen sich auch, wie bereits in den vorhergehenden ESS-Wellen, bei unterschiedlichen Altersgruppen: Eine geringere Ausschöpfung bei Senioren und Rentnern (65-Jährige und älter) und den 30- bis 39-Jährigen steht eine erhöhte Ausschöpfung insbesondere bei Personen zwischen 50 und 64 Jahren gegenüber. Die Struktur lässt einen Zusammenhang mit den jeweiligen Lebensabschnitten vermuten, die sich auf individuelle Zeitkontingente und damit auch auf Kontaktierungswahrscheinlichkeit und Teilnahmebereitschaft auswirken. Insgesamt weist die Nettostichprobe anhand dieser Kriterien jedoch eine gute Passung an die Bruttostichprobe auf.

Tabelle 2: Brutto-, Nettostichprobe und Nichtteilnehmer im Vergleich sozio-demografischer Merkmale

Wohnbevölkerungsanteil ab 15 Jahren	Brutto- stichprobe		Netto- stichprobe		Differenz Brutto-Netto	Nicht- teilnehmer	
Verteilungsmerkmal	N	%	N	%	%-Punkte	N	%
<i>Gesamt:</i>	8.904	100,0	2.975	100,0	0	5.929	100,0
Westdeutschland	5.883	66,1	1.969	66,2	0,1	3.914	66,0
Ostdeutschland	3.021	33,9	1.006	33,8	-0,1	2.015	34,0
<i>Bundesland:</i>							
Schleswig-Holstein	265	3,0	98	3,3	0,3	167	2,8
Hamburg	159	1,8	40	1,3	-0,5	119	2,0
Niedersachsen	689	7,7	272	9,1	1,4	417	7,0
Bremen	53	0,6	15	0,5	-0,1	38	0,6
Nordrhein-Westfalen	1.590	17,9	476	16,0	-1,9	1.114	18,8
Hessen	530	6,0	166	5,6	-0,4	364	6,1
Rheinland-Pfalz	371	4,2	135	4,5	0,3	236	4,0
Baden-Württemberg	954	10,7	316	10,6	-0,1	638	10,8
Bayern	1.060	11,9	401	13,5	1,6	659	11,1
Saarland	53	0,6	11	0,4	-0,2	42	0,7

Fortsetzung Tabelle 2

Wohnbevölkerungsanteil ab 15 Jahren	Brutto- stichprobe		Netto- stichprobe		Differenz Brutto-Netto	Nicht- teilnehmer	
Verteilungsmerkmal	N	%	N	%	%-Punkte	N	%
Berlin	477	5,4	120	4,0	-1,4	357	6,0
Brandenburg	530	6,0	176	5,9	-0,1	354	6,0
Mecklenburg-Vorpommern	371	4,2	129	4,3	0,1	242	4,1
Sachsen	848	9,5	306	10,3	0,8	536	9,0
Sachsen-Anhalt	530	6,0	177	6,0	0	353	6,0
Thüringen	424	4,8	137	4,6	-0,2	287	4,8
<i>Gemeindegrößenklasse:</i>							
1: u. 2.000	212	2,4	55	1,9	-0,5	157	2,6
2: 2.000 – u. 5.000	318	3,6	126	4,4	0,8	192	3,2
3: 5.000 – u. 20.000	742	8,3	279	9,4	1,1	463	7,8
4: 20.000 – u. 50.000	1.113	12,5	389	13,1	0,6	724	12,2
5: 50.000 – u. 100.000, Rest	795	8,9	302	10,2	1,3	493	8,3
6: 50.000 – u. 100.000, Kern	265	3,0	83	2,8	-0,2	182	3,1
7: 100.000 – u. 500.000, Rest	1.219	13,7	448	15,1	1,4	771	13,0
8: 100.000 – u. 500.000, Kern	1.325	14,9	427	14,4	-0,5	898	15,1
9: 500.000 u. größer, Rest	795	8,9	266	8,9	0	529	8,9
10: 500.000 u. größer, Kern	2.120	23,8	600	20,2	-3,6	1.520	25,6
<i>Geschlecht:</i>							
Männlich	4.304	48,3	1.500	50,4	2,1	2.804	47,3
Weiblich	4.589	51,5	1.472	49,5	-2,0	3.117	52,6
Fehlende	11	0,1	3	0,1	0	8	0,1
<i>Altersgruppen:</i>							
15 bis 29 Jahre	1.528	17,2	563	18,9	1,7	965	16,3
30 bis 39 Jahre	1.123	12,6	329	11,1	-1,5	794	13,4
40 bis 49 Jahre	1.477	16,6	510	17,1	0,5	967	16,3
50 bis 64 Jahre	2.087	23,4	746	25,1	1,7	1.341	22,6
65 Jahre und älter	2.106	23,7	617	20,7	-3,0	1.489	25,1
Fehlende	583	6,6	210	7,1	0,5	373	6,3

Quelle: Methodenbericht ESS 6 (infas 2013) und eigene Berechnungen.

Nonresponse Bias

Qualitätssicherung sozialwissenschaftlicher Umfragen

Schupp, J.; Wolf, C. (Hrsg.)

2015, X, 478 S. 32 Abb., Softcover

ISBN: 978-3-658-10458-0