
gesis

Leibniz-Institut
für Sozialwissenschaften

German Longitudinal
Election Study



GLES

Wahlkampf-Panel 2013

ZA5704, Version 3.2.0

Studienbeschreibung

Inhaltsverzeichnis

1.	Vorbemerkungen.....	4
2.	Studieninformationen	5
3.	Konzeption und Durchführung des Wahlkampf-Panels 2013.....	6
3.1.	Aufbau der Studie	6
3.2.	Fallauswahl	7
3.2.1	Grundgesamtheit – Das Online-Access-Panel der ResponDi AG	7
3.2.2	Auswahlverfahren und Quotierung.....	9
3.2.3	Zum Umgang mit fehlender Repräsentativität im Wahlkampfpanel	11
3.3	Einladung der Teilnehmer	11
3.4	Teilnehmerverifikation	13
3.5	Interviewdauer.....	15
3.6	Incentivierung.....	15
3.7	Teilnahmestatistiken	16
3.7.1	Teilnahmehäufigkeit.....	16
3.7.2	Verteilung der Teilnahmen über die Feldzeit	16
3.7.3	Ausschöpfung	17
4	Variablen im Datensatz	22
4.1	Vorbemerkung.....	22
4.2	GESIS Archiv-Variablen	23
4.3	Administrative Variablen.....	23
4.4	Kontextvariablen: Zuordnung von Wahlkreisen.....	24
4.5	Datenqualitätsvariablen.....	24
4.6	Gewichtungsvariablen.....	25
4.6.1	Querschnittsgewichte.....	25
4.6.2	Panelgewichte	28
4.7	Inhaltliche Variablen	32
4.8	Paradaten.....	42
4.9	Unveröffentlichte Variablen	43
5	Datenaufbereitung.....	45
5.1	Vorbemerkung.....	45
5.2	Codierung der fehlenden Werte	45
5.3	Codierung der Parteien	45
5.4	Codierung von offenen Fragen.....	47
5.5	Markierung der Wissensfragen	48
5.6	Umgang mit Nachfragen bei den Frageformaten „Checkbox“ und „Schieberegler“	48
5.7	Zusammenfassung von Zeitangaben	49
5.8	Zusammenfassung der Dependent Interviewing-Variablen.....	49

6	Datenqualität.....	52
6.1	Vorbemerkung	52
6.2	Qualitätsindikatoren	53
6.2.1	Selbsteinschätzung der Datenqualität durch die Befragten	53
6.2.2	Antwortzeiten	53
6.2.3	Straightlining	54
6.2.4	Item Non-Response.....	54
6.2.5	Weiß-nicht-Antworten	54
6.2.6	Kontrollitems	54
6.2.7	Ein kombinierter Index	55
6.2.8	Ein Abweichungsindex.....	55
6.3	Filterfehler	56
6.4	Rückmeldungen der Befragten	57
6.5	Versionshistorie	60
7.	Hinzuspielen anderer Datensätze.....	63
7.1	Hinzuspielen von Zeitvariablen.....	63
7.2	Hinzuspielen des GLES-Wahlkampfpanels 2009	63
7.3	Hinzuspielen der Kontrollquerschnitte	63
	Literaturverzeichnis	64
	Links.....	66

1. Vorbemerkungen

Arbeiten mit GLES Daten

Bei dieser Publikation und dem zugehörigen Datensatz handelt es sich um Daten der German Longitudinal Election Study (GLES), die von GESIS in Kooperation mit der Deutschen Gesellschaft für Wahlforschung (DGfW) herausgegeben werden. Vor der Veröffentlichung werden die Daten sorgfältig geprüft. Leider kann es trotz gründlicher Überprüfung der Daten passieren, dass Fehler in den Datensätzen unentdeckt bleiben. Fallen diese zu einem späteren Zeitpunkt auf, werden sie dokumentiert (Errata-Liste im Datenbestandskatalog, www.gesis.org/dbk) und zeitnah behoben.

Um hochqualitative Datensätze zur Verfügung stellen zu können, freuen wir uns über Ihre Mithilfe. Wenn Ihnen bei Ihrer Arbeit mit den GLES-Daten ein Fehler auffällt, helfen Sie uns sehr, wenn Sie eine kurze E-Mail an gles@gesis.org schicken. Bitte schicken Sie uns zusätzlich zu der Beschreibung des Fehlers auch die Studiennummer (ZA-Nummer) sowie die Versionsnummer des Datensatzes.

Wir empfehlen, stets mit der aktuellen Version der GLES-Daten zu arbeiten. Sie können diese unkompliziert über den Datenbestandskatalog herunterladen. Die Links zum direkten Download finden Sie auch auf den Seiten der GLES bei GESIS (www.gesis.org/gles).

Meldung von Veröffentlichungen

Um einen Überblick über die tatsächliche Nutzung der Daten zu erhalten, bitten wir Sie um eine kurze Mitteilung bei Veröffentlichungen, die Daten der GLES verwenden (bibliographische Angaben, Studiennummer des verwendeten Datensatzes). Veröffentlichungen, die vollständig oder teilweise auf Daten der GLES beruhen, werden in der offiziellen Bibliographie der GLES aufgeführt. Wenn es sich dabei um Konferenzpapiere o.ä. handelt, die nur schwer zugänglich sind, freuen wir uns über die Überlassung eines Exemplars bzw. eines PDF-Dokuments.

Kontakt

GESIS – Leibniz-Institut für Sozialwissenschaften

Postfach 122155

68072 Mannheim

E-Mail: gles@gesis.org

Zitation bei Veröffentlichungen

Wird in Publikationen auf Daten der GLES zurückgegriffen, bitten wir diese wie folgt zu zitieren:

Rattinger, Hans; Roßteutscher, Sigrid; Schmitt-Beck, Rüdiger; Weißels, Bernhard; Wolf, Christof; Plischke, Thomas; Wiegand, Elena (2016): Wahlkampf-Panel 2013 (GLES). GESIS Datenarchiv, Köln. ZA5704 Datenfile Version 3.2.0, doi: 10.4232/1.12561.

2. Studieninformationen

Studiennummer

ZA5704 (Version 3.2.0); doi: 10.4232/1.12561

Datensatzversionen

GLES Version: 3.2.0; Datei: ZA5704_v3-2-0.sav, ZA5704_v3-2-0.dta

GLES Version (Englisch): 3.2.0; File: ZA5704_v3-2-0_en.sav, ZA5704_v3-2-0_en.dta

Titel der Studie

German Longitudinal Election Study, Komponente 3: Wahlkampf-Panel

Primärforscher/innen

Prof. Dr. Hans Rattinger	(Universität Mannheim)
Prof. Dr. Sigrid Roßteutscher	(Goethe-Universität Frankfurt am Main)
Prof. Dr. Rüdiger Schmitt-Beck	(Universität Mannheim)
Prof. Dr. Bernhard Weißels	(Wissenschaftszentrum Berlin für Sozialforschung)
Prof. Dr. Christof Wolf	(GESIS – Leibniz-Institut für die Sozialwissenschaften)

Erhebungszeitraum

20. Juni 2013 bis 04. Oktober 2013

Welle 1 (20.06.-07.07.2013)
Welle 2 (18.07.-28.07.2013)
Welle 3 (01.08.-11.08.2013)
Welle 4 (15.08.-25.08.2013)
Welle 5 (02.09.-12.09.2013)
Welle 6 (16.09.-21.09.2013)
Welle 7 (24.09.-04.10.2013)

Erhebungsverfahren

Online-Befragung mit standardisiertem Fragebogen

Erhebungssoftware

GlobalPark, EFS Umfragecenter Version 7.1

Umfrageinstitut

Das Bamberger Centrum für Empirische Studien (BACES) der Otto-Friedrich-Universität Bamberg übernahm die Organisation und Durchführung der Datenerhebung. Zudem fungierte BACES als Schnittstelle zum Online-Access-Panel-Anbieter Respondi AG und übernahm die Programmierung des Fragebogens sowie das Hosting der Onlinebefragung.

Finanzierende Stelle

Deutsche Forschungsgemeinschaft e. V. (DFG)

3. Konzeption und Durchführung des Wahlkampf-Panels 2013

3.1. Aufbau der Studie

Das Wahlkampfpanel der German Longitudinal Election Study ermöglicht es, individuelle Veränderungen von politischen Einstellungen und Verhaltensweisen im Verlauf des Wahlkampfes zu untersuchen. Zu diesem Zweck wurde eine gleichbleibende Auswahl von Bürgern im Verlauf des Wahlkampfes in kurzen Abständen bis zu sieben Mal interviewt – sechs Mal vor der Wahl und einmal nach der Wahl.

Eine Neuerung im Design des Wahlkampfpanels 2013, im Vergleich zu 2009, besteht in der Einführung von Kontrollgruppen. Zusätzlich zu den Wiederholungsbefragten wurden zeitgleich zu der dritten, fünften und siebten Befragung drei unabhängige Stichproben mit etwa 1200 Personen gezogen, die dann anhand eines fast identischen¹ Fragebogens interviewt wurden (siehe Abbildung 1). Durch den Vergleich von Panel- und Querschnittsbefragten lässt sich untersuchen, ob und in welchem Ausmaß die wiederholte Befragung derselben Personen Einfluss auf die Messung von politischen Einstellungen und Verhalten nimmt. Die vorliegende Studienbeschreibung bezieht sich nur auf die sieben Wellen des Wahlkampfpanels. Die Dokumentation und die Daten der Kontrollquerschnitte wurden separat unter den ZA-Nummern 5753, 5754 und 5755 veröffentlicht.

Abbildung 1: Die Wellen und Kontrollquerschnitte des Wahlkampfpanels 2013

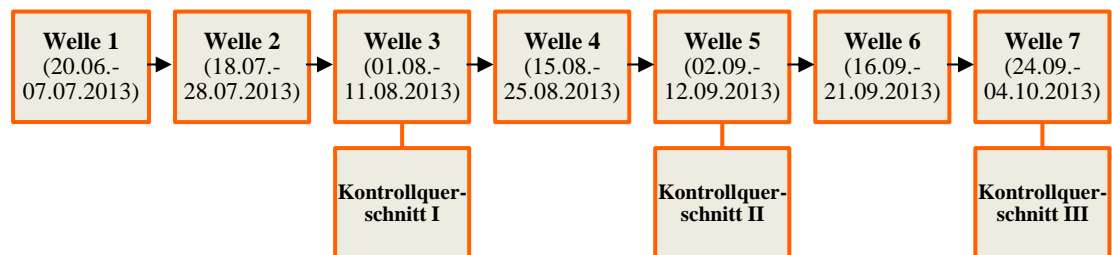


Tabelle 1 gibt einen Überblick über die inhaltlichen Schwerpunkte der sieben Befragungen. Als „Kernfragebogen“ wird der Bestand jener Fragen bezeichnet, der unverändert in jeder der sieben Befragungen gestellt wurde. Er umfasst etwa 35 Prozent des Fragebogens einer jeden Welle und enthält vor allem Fragen zum Wahl- und Kommunikationsverhalten. Darüber hinaus hat jede Welle bestimmte Schwerpunkte, wie z. B. Fragen zum strategischen Wählen, zu den Kandidaten oder zu Europa und der europäischen Schuldenkrise.

¹ Die Fragenkomplexe zur Soziodemographie wurden in drei Blöcken auf die Wellen aufgeteilt (siehe Tabelle 1). Der erste Block beinhaltete Fragen zum Haushalt, Familienstand, Religiosität sowie zu Beruf und Ausbildung. Der zweite Block befasste sich mit dem Migrationshintergrund der Befragten und der dritte Block erhob Informationen zum Beruf und Ausbildung des Partners. In den drei Kontrollquerschnitten wurden jeweils nur die Fragen des ersten Blocks gestellt.

Tabelle 1: Übersicht über die thematischen Schwerpunkte der sieben Panelwellen

Inhalte	Welle 1 (20.6.- 7.7.)	Welle 2 (18.7.- 28.7.)	Welle 3 (1.8.- 11.8.)	Welle 4 (15.8.- 25.8)	Welle 5 (2.9.- 12.9)	Welle 6 (16.9.- 21.9)	Welle 7 (24.9.- 4.10.)
Kernfragebogen	X	X	X	X	X	X	X
Wahlverhalten, retrospektiv	X						
Ideologie, Parteien	X		X		X		X
Positionsissues, Parteien		X		X		X	
Kandidaten		X		X	X		X
Strategisches Wählen		X		X		X	
Europa			X				X
Wirtschaft	X			X			X
Landtagswahl in Bayern		X	[X]	X	X	X	[X]
Landtagswahl in Hessen		X	[X]		X	X	X
Gerechtigkeit, allgemein	X						
Gerechtigkeit, Einkommen			X		X		X
Gerechtigkeit, Steuern		X		X		X	
TV-Duell, prospektiv				X			
TV-Duell, retrospektiv					X	X	X
Soziodemografie, Ego*	X						
Soziodemografie, Migrations- hintergrund		X	[X]	[X]			
Soziodemografie, Partner					X	[X]	[X]

[X] bedeutet, dass nur die Teilnehmer des Wahlkampfpanels die Fragen des Moduls erhalten haben, die bei der Abfrage dieses Themenblocks in einer der vorangegangenen Wellen nicht teilgenommen haben. Die Teilnehmer der Kontrollquerschnitte haben diese Fragemodule nicht erhalten.

*In den Kontrollquerschnitten zu dritten, fünften und siebten Welle wurde dieser Themenblock zur Soziodemografie ebenfalls erhoben (siehe Fußnote 1).

3.2. Fallauswahl

3.2.1 Grundgesamtheit – Das Online-Access-Panel der Respondi AG

Die Grundgesamtheit des Wahlkampfpanels bildet die zur Bundestagswahl 2013 wahlberechtigte Bevölkerung. Da es sich aber bei der vorliegenden Studie um eine Online-Befragung handelt, konnte keine Zufallsstichprobe realisiert werden. Die dieser Studie zugrunde liegende Auswahlgesamtheit umfasst somit nur die wahlberechtigten Teilnehmer des von der Respondi AG betriebenen Online-Access-Panels. Dieses Panel umfasste 2013 in Deutschland 96.445 aktive Personen. Als aktive Personen definiert Respondi diejenigen

Personen, die nach der (Double-opt-in-) Registrierung² die Stammdatenbefragung ausgefüllt haben und in den letzten zwölf Monaten an mindestens einer Umfrage teilgenommen haben.

Tabelle 2: Zusammensetzung des Respondi-Access-Panels nach soziodemographischen Merkmalen

	Anteil in Prozent
Geschlecht	
Weiblich	60,3
Männlich	39,7
Bildung	
Niedrig (d.h. kein Abschluss, Haupt- und Volksschule)	14,9
Mittel (d.h. Realschule, Mittlere Reife)	36,7
Hoch (d.h. Abitur, Fachhochschulreife, Studium)	47,2
Altersgruppen	
18-29 Jahre	32,6
30-39 Jahre	23,3
40-49 Jahre	21,6
50-59 Jahre	14,8
60 und älter	7,6

Fehlend zu 100%: keine Angabe

Die Mitglieder des Online-Access-Panels werden von Respondi auf unterschiedlichen Wegen rekrutiert. Zu diesem Zweck tritt Respondi unter dem Markennamen „Mingle“ an die Personen heran. Dies erfolgt überwiegend online, zu einem geringen Anteil aber auch offline. Die Mitglieder wirbt Respondi mehrheitlich über Umfragen auf themenspezifischen Portalen, in Foren und Communities an. Dabei ist das Motiv der Rekrutierung, dass sie ihre eigene Meinung kundtun dürfen und nicht die Aussicht auf eine monetäre Entlohnung.

Tabelle 3: Rekrutierungswege bei Respondi

	Anteil in Prozent
Mingle Trend Blog	2,6
Über Online-Werbung (davon beworbene Thementumfragen 50%; direkte Panelwerbung 50% inkl. Affiliate Marketing)	29,8
Selbsteinträge	13,5
Facebook Fanseite	26,2
Suchmaschinen	21,1
Empfehlungen (Freundschaftswerbung)	2,1
Kooperationen	4,6
Telefonakquise (CATI)	0,1

Respondi gibt an, ein wirkungsvolles Qualitätsmanagement zu betreiben. Über ein internes Bewertungssystem wird das Antwortverhalten der Mitglieder des Respondi-Panels kontinu-

² Double-opt-in bedeutet, dass ein Nutzer nach seiner Registrierung eine E-Mail erhält, die ihn dazu auffordert, die Mitgliedschaft zu bestätigen. Erst danach wird der Nutzer zur Stammdatenbefragung weitergeleitet.

ierlich evaluiert. Gründe für ein Löschen von Mitgliedern aus der Datenbank sind laut Respondi beispielsweise fehlende Teilnahme über einen Zeitraum von zwölf Monaten sowie Doppelanmeldungen und bewusste Falschangaben bei mehreren Befragungen.

Das Qualitätsmanagement sowie eine maßvolle Einladungsfrequenz sollen dazu beitragen, unerwünschte Effekte wie Panel-Verzerrungen oder Professionalisierung der Mitglieder zu vermeiden. Ein durchschnittliches Mitglied verweilt 18 Monate im Respondi-Panel. Innerhalb eines Jahres werden ca. 36 Prozent der Panelmitglieder aufgrund von Qualitätskontrollen und Panel-Mortalität aus der Datenbank entfernt.

Die durchschnittliche Teilnahmequote der Mitglieder liegt bei einer Umfrage mit fünf Feldtagen bei etwa 50 Prozent. Die Teilnahmequote wird von Respondi aus der Anzahl der begonnenen Interviews berechnet, also den beendeten Interviews, Screen-outs (Ausschluss bei Auswahl der zu befragenden Teilnehmer nach Thema), Quota-fulls (Ausschluss auf Grund erfüllter Quoten bei Quotenauswahl) und den abgebrochenen Interviews, die ins Verhältnis zur Gesamtzahl der Einladungen gesetzt werden.

Die Mitglieder des Access-Panels werden von Respondi für die Teilnahme an Befragungen incentiviert. Sie erhalten normalerweise fünf sogenannte Mingle-Punkte pro Minute, was 0,05 € entspricht (im Wahlkampfpanel 2013 wurde ein höherer Betrag bezahlt, siehe Kapitel 3.6). Sobald die Teilnehmer 1500 Mingle-Punkte gesammelt haben, kann dieser Betrag ausbezahlt, in Einkaufsgutscheine umgewandelt oder gespendet werden. Zusätzlich werden zur Panelpflege regelmäßig Verlosungen unter den Mitgliedern durchgeführt.

3.2.2 Auswahlverfahren und Quotierung

Die Teilnehmerschaft des Wahlkampfpanels 2013 setzt sich aus zwei Teilgruppen zusammen:

- Einer Stichprobe mit insgesamt 4.226 Personen, die auf der Grundlage einer Quotenauswahl gezogen wurde.
- Eine Gruppe bestehend aus 1.030 Personen, die bereits am Wahlkampfpanel 2009 teilgenommen hatten und für deren Wiedereinladung ins Wahlkampfpanel 2013 keine Quotenvorgabe galt. Durch die Wiederbefragung der Teilnehmer aus 2009 können die beiden Datensätze der Wahlkampfpanel 2009 und 2013 zusammengespielt werden (vgl. Abschnitt 7.2). Auf diese Weise erhält man einen 14-welligen Datensatz, der sowohl Kurz- und Langfristspektive vereint.

Die neu generierte Stichprobe wurde unter Berücksichtigung einer Quotenvorgabe aus der Auswahlgesamtheit der im Online-Access-Panel zur Verfügung stehenden Personen realisiert. Die Quoten beruhen auf den Merkmalen Geschlecht, Alter (in fünf Alterskategorien: 18-29, 30-39, 40-49, 50-59, 60+) und Bildung (in drei Kategorien: niedrig: kein Abschluss, Haupt- oder Volksschule; mittel: Realschule, mittlere Reife; hoch: Abitur, Fachhochschulreife). Um eine möglichst heterogene Stichprobe zu erhalten, sollte in ihr jede der dreißig möglichen Kombinationen der drei Quotenmerkmale in etwa gleichem Anteil (3,33%) repräsentiert sein. Geringfügige Abweichungen von diesem Quotenziel wurden in Kauf genommen, da Respondi eine ausreichende Repräsentation bestimmter Merkmalskombinationen nicht immer sicherstellen konnte. Beispielsweise konnten jüngere Männer mit niedriger Bildung nur sehr schwer zur Teilnahme motiviert werden.

Unabhängig davon wurden zudem alle Befragten des Wahlkampfpanels 2009 erneut eingeladen, die damals an mindestens vier Wellen teilgenommen hatten und 2013 noch immer aktive Respondi-Mitglieder waren. Insgesamt erfüllten 1.527 Personen diese Kriterien. Von ihnen nahmen 1.011 Personen die Einladung zur ersten Befragung an und wurden somit zu Teilnehmern des Wahlkampfpanels 2013. Darüber hinaus konnte für weitere 19 Befragte festgestellt werden, dass sie ebenfalls am Wahlkampfpanel 2009 teilgenommen hatten, je-

doch nicht die Kriterien erfüllten. Sie sollten ursprünglich durch eine „Blacklist“ von der Teilnahme abgehalten werden, die scheinbar nicht funktioniert hatte. Da diese Teilnehmer erst in Version 2 des Datensatzes identifiziert wurden, schließt die Quotierung noch diese 19 Fälle ein. Eine Nachberechnung zeigte jedoch, dass die Abweichungen in den Verteilungen nur geringfügig sind.

Tabelle 4: Quotierung im Wahlkampfpanel 2013

Quotierungsmerkmale	Soll in Prozent	Ist* in Prozent	
		N=4245 (ohne Teilnehmer des WKP 2009)	N=5256 (inkl. Teilnehmer des WKP 2009)
Geschlecht			
Weiblich	50,0	51,4	51,4
Männlich	50,0	48,6	48,6
Bildung			
Niedrig (d.h. kein Abschluss, Haupt- oder Volksschule)	33,3	35,2	30,8
Mittel (d.h. Realschule, Mittlere Reife)	33,4	32,1	34,3
Hoch (d.h. Abitur, Fachhochschulreife, Studium)	33,3	32,3	35,0
Altersgruppe			
18-29 Jahre	20,0	18,5	17,5
30-39 Jahre	20,0	17,8	18,0
40-49 Jahre	20,0	21,1	22,1
50-59 Jahre	20,0	21,8	19,8
60 Jahre und älter	20,0	20,7	19,8

* Tatsächliche Verteilung im Datensatz ohne Teilnehmer aus dem WKP2009, die bei der Quotierung nicht berücksichtigt wurden.

Ab den Befragungswellen zwei bis sieben trat das für Panelbefragungen typische Problem der „Panelmortalität“ auf: Nicht alle Personen aus der ersten Befragung nahmen auch an den nachfolgenden Wellen teil (vgl. hierzu auch Abschnitt 3.7). Wie jedoch aus Tabelle 5 ersichtlich wird, hatte Panelmortalität nur einen geringfügigen Einfluss auf die Verteilung der Quotenmerkmale. Der Anteil der Männer nahm im Verlauf des Wahlkampfpanels um etwa einen Prozentpunkt zu. Der Anteil der Niedriggebildeten nahm um etwa 1,5 Prozentpunkte ab, zu Gunsten des Anteils der Hochgebildeten. Der stärkste systematische Ausfall ist in Bezug auf die fünf verschiedenen Altersgruppen zu verzeichnen: Während der Anteil der jüngsten Altersgruppe im Verlauf des Wahlkampfpanels um fast drei Prozentpunkte abnahm, stieg der Anteil der ältesten Gruppe um fast zwei Prozentpunkte.

Das Ausmaß der systematischen Verzerrung erscheint jedoch etwas größer, wenn man sich die gemeinsame Verteilung der Quotierungsmerkmale betrachtet (nicht tabellarisch ausgewiesen). Durch den Panelausfall waren die 18-29-jährigen Männer mit niedriger Bildung am stärksten betroffen: In dieser Gruppe nahmen in der sechsten Welle nur noch 35 Prozent der Befragten aus der ersten Welle teil. Hingegen waren andere Gruppen fast überhaupt nicht von Panelmortalität betroffen: So lag die Teilnahmequote der über-60-jährigen Frauen mit hoher Bildung in jeder Befragung bei mindestens 88 Prozent.

Tabelle 5: Verteilung von Geschlecht, Bildung und Alter, getrennt nach Wellen in Prozent (inkl. Teilnehmer aus WKP2009)

Welle		W1	W2	W3	W4	W5	W6	W7
Geschlecht	Männlich	48,6	48,8	49,0	49,2	49,3	49,6	49,5
	Weiblich	51,4	51,2	51,0	50,8	50,7	50,4	50,5
Bildung	Niedrig	30,8	29,4	29,3	29,1	28,6	28,7	28,3
	Mittel	34,3	34,9	34,9	35,2	34,8	34,7	34,8
	Hoch	35,0	35,7	35,8	35,7	36,6	36,7	36,9
Altersgruppe	18-29 Jahre	17,5	15,9	15,2	15,0	15,0	14,5	14,7
	30-39 Jahre	18,0	17,4	16,9	17,1	16,7	16,7	16,8
	40-49 Jahre	22,1	22,1	22,1	22,2	22,4	22,4	22,4
	50-59 Jahre	22,7	23,5	23,9	23,8	23,9	24,2	24,1
	60 Jahre >	19,8	21,1	21,9	22,0	22,0	22,2	22,1
N		5256	4598	4432	4355	4257	4112	4231

3.2.3 Zum Umgang mit fehlender Repräsentativität im Wahlkampfpanel

Die Teilnehmer des Wahlkampfpanels sind nicht repräsentativ für die wahlberechtigte Bevölkerung. In Online-Access-Panels sind überdurchschnittlich viele junge, gebildete und internetaffine Menschen vertreten. Diese Gruppen unterscheiden sich in den erhobenen Einstellungen und Verhaltensweisen von der Gesamtbevölkerung. Daher ist eine Verallgemeinerung von *Anteilswerten*, die auf der Grundlage von Wahlkampfpanel-Daten ermittelt werden, nicht zulässig.

Der entscheidende Vorteil von Panel-Befragungen liegt in ihrem Potenzial, kausale Mechanismen zu untersuchen.³ Zwar kann nicht ausgeschlossen werden, dass fehlende Repräsentativität auch bei Panel-Kausalanalysen in einem gewissen Maße Verzerrungen hervorruft. Gleichwohl ist aber davon auszugehen, dass diese Verzerrungen wesentlich geringer ausfallen als bei Kausalanalysen mit repräsentativen Querschnittsdaten, deren Effektschätzwerte in einem erheblichen Ausmaß durch das Problem der „unbeobachteten Heterogenität“ (auch „omitted variable bias“ oder „Drittvariableneffekt“ genannt) betroffen sind (vgl. Green et al. 2001; Brüderl 2010). Anhand von Paneldaten und sogenannten „fixed-effects“-Panelmodellen kann unbeobachtete Heterogenität zu einem großen Anteil eliminiert werden (vgl. Brüderl 2010). Für den Zweck von Kausalanalysen liefern daher fixed-effects-Regressionen, die auf nicht-repräsentativen Paneldaten basieren, in der Regel weniger verzerrte Effektschätzwerte als gewöhnliche Querschnitts-Regressionsverfahren, die auf repräsentativen Daten beruhen.

3.3 Einladung der Teilnehmer

Zu Feldbeginn der ersten Welle wurden zunächst nur jene Personen eingeladen, die auch am Wahlkampfpanel 2009 teilgenommen hatten. Auf der Grundlage der Rücklaufquote konnte dann berechnet werden, wie viele Personen die neue Stichprobe enthalten sollte.

³ Dieser Abschnitt orientiert sich stark an den Ausführungen von Plischke (2014: 198-199).

Sechs Tage später wurden dann die ersten „neuen“ Befragten für das Wahlkampfpanel 2013 eingeladen. Um die Quotenvorgaben zu erreichen, wurden zunächst die vergleichsweise schwer zu erreichenden Personengruppen eingeladen (z.B. jüngere Personengruppen, Personen mit niedrigen Bildungsabschlüssen); die Einladungen für die anderen Personengruppen folgten einige Tage später. Je nach Bedarf oder bei noch nicht erreichten Quotenzielen wurden Teilnehmer, die die Umfrage noch nicht begonnen hatten, nach wenigen Tagen an die Einladung erinnert und erneut um Teilnahme gebeten (siehe Tabelle 6).

Tabelle 6: Teilnehmerrekrutierung in Welle 1

Datum	Gruppe	Gesamt	Geschlecht		Alter					Bildung		
			männ- lich	weib- lich	18- 29	30- 39	40- 49	50- 59	60+	nied- rig	mittel	hoch
20.6.	1: Teilnehmer WKP 2009 (E)	1527	718	809	260	305	412	336	214	183	641	703
22.6.	1: Teilnehmer WKP 2009 (R)	883	415	468	150	177	238	194	124	106	371	406
23.6.	2 (E)	3922	1961	1961	942	745	745	745	745	1530	1373	1019
25.6.	3 (E)	8000	3600	4400	4320	480	800	640	1760	2800	2400	2800
26.6.	1 (R)	594	279	315	101	119	160	131	83	71	250	273
26.6.	2 (R)	926	417	509	630	74	130	92		926		
1.7.	2 (R)	2084	1042	1042	584	500	500	500		813	729	542
1.7.	4 (E)	3000	1500	1500		1080	990	930			1800	1200
1.7.	3 (R)	834	417	417	667				167	834		
3.7.	4 (R)	477	239	238			239	238				477
6.7.	1, 2, 3, 4 (R)	512	307	205	512					512		
Einladungen Ge- samt		16449										

E: Einladung; R: Erinnerungsemail; WKP: Wahlkampfpanel

Ab der zweiten Welle wurden nur noch jene Personen eingeladen, die die erste Befragung vollständig absolviert hatten (Abbrecher in Welle eins wurden nicht mehr eingeladen). Jedoch wurde die Anzahl der versendeten Einladungen zunehmend kleiner, da zwischen jeder Welle eine kleine Anzahl von Personen aus dem Respondi-Panel ausgeschieden ist und dadurch für weitere Befragungen nicht mehr zur Verfügung stand. Ab der dritten Welle wurden die Teilnehmer in zwei gleich große Gruppen aufgeteilt und mit vier Stunden Abstand voneinander eingeladen, um eine Server-Überlastung zu vermeiden. Zusätzlich zu den Erinnerungsemails für die noch nicht erreichten Teilnehmern wurden jeweils einen Tag versetzt Erinnerungsemails an jene Personen versendet, die die Umfrage angefangen, aber abgebrochen hatten (siehe Tabelle 7).

Tabelle 7: Einladungen und Erinnerungen in den Wellen 2 bis 7

	Welle 2		Welle 3		Welle 4		Welle 5		Welle 6		Welle 7	
	Da- tum	N	Da- tum	N	Da- tum	N	Da- tum	N	Da- tum	N	Da- tum	N
Einladung	18.7	524 2	1.8	287 0	15.8	284 4	2.9	286 4	16.9	292 3	24.9	279 9
Einladung			1.8	234 3	15.8	236 0	2.9	232 4	16.9	225 7	24.9	236 8
Erste Erin- nerung der noch nicht erreichten Panelisten	21.7	136 7	4.8	155 6	18.8	170 3	5.9	150 7	18.9	165 3	27.9	138 6
Erste Erin- nerung der Unterbre- cher	22.7	244	5.8	206	19.8	166	6.9	186	19.9	136	27.9	232
Zweite Erinnerung der noch nicht er- reichten Panelisten	25.7	735	7.8	955	21.8	100 6	8.9	105 7	20.9	119 6	30.9	106 8
Zweite Erinnerung der Unter- brecher	26.7	179	8.8	174	22.8	125	9.9	150	21.9	114	1.10	120

Bei der Formulierung der Einladungs- und Reminder-E-Mails wurde großer Wert darauf gelegt, die Besonderheiten der Studie hervorzuheben. Die Befragung wurde den Panelisten unter dem Titel „GLES-Studie“ vorgestellt, und bei jeder Einladung wurde bereits im Betreff auf den Namen dieser Studie aufmerksam gemacht, um die Wiedererkennung zu gewährleisten (z. B. Betreff: „mingle – GLES-Studie Teil 2“). In Absprache mit Respondi wurde der Standardtext der E-Mails abgeändert und das GLES-Logo eingefügt. Das Layout und den genauen Wortlaut der Einladungstexte und Startseiten jeder Welle kann den Dokumenten „Bildschirmansichten“ entnommen werden, die den Daten beiliegt.

3.4 Teilnehmerverifikation

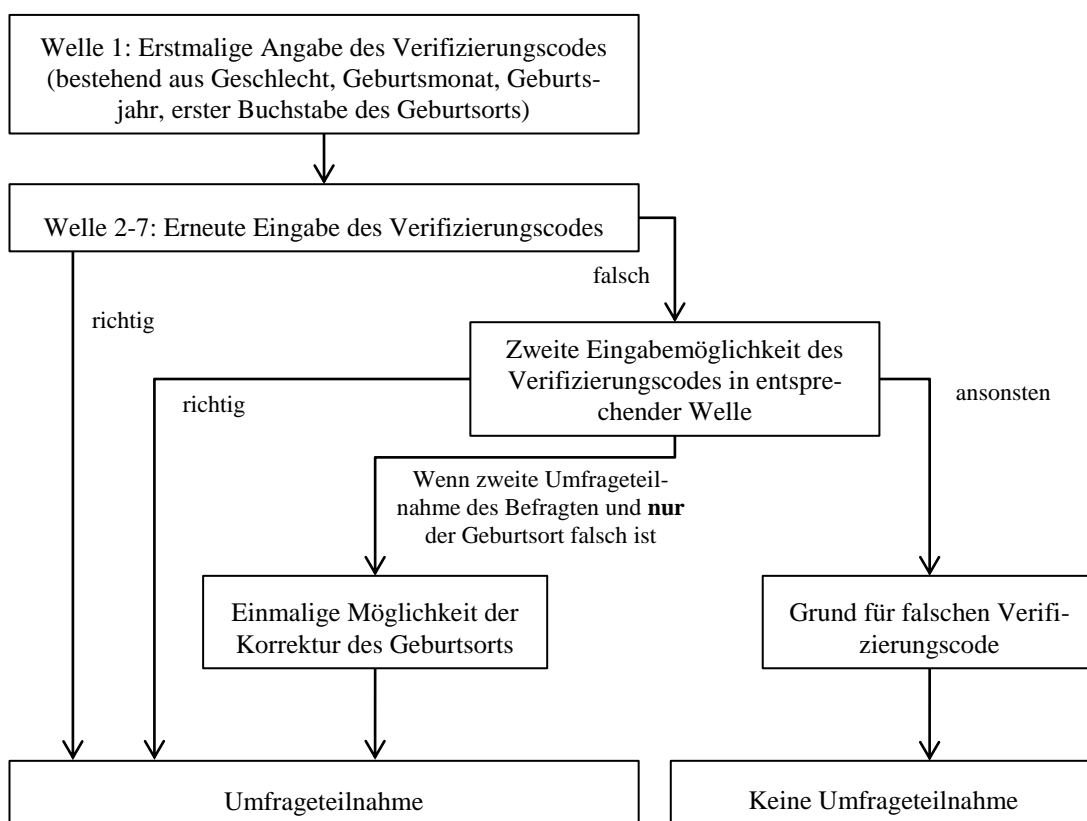
In online durchgeführten Panelstudien muss aufgrund der Anonymität der Teilnehmerschaft in einem besonderen Ausmaß darauf geachtet werden, dass an jeder Welle tatsächlich nur die jeweiligen Zielpersonen teilnehmen. Beispielsweise könnten Haushaltsangehörige, die Zugriff auf das E-Mail-Konto der Zielperson haben, anstatt der eigentlichen Zielperson teilnehmen. Um dies zu vermeiden, wurde in den Einladungs-E-Mails und auf der Startseite darum gebeten, von einer Umfrageteilnahme abzusehen, falls man nicht bereits an den vorangehenden Befragungen teilgenommen habe.

Darüber hinaus wurde zu Beginn jeder Welle eine Teilnehmerverifikation durchgeführt. Dies geschah über eine Abfrage von vier stabilen Merkmalen der Befragten: das Geschlecht, das Geburtsjahr, der Geburtsmonat und der erste Buchstabe des Geburtsortes. In der ersten Welle wurden die Angaben der Befragten zu den vier Merkmalen mit deren Stammdaten⁴

⁴ Stammdaten sind Personendaten, die die Teilnehmer des Wahlkampfpanels erstmalig bei ihrem Eintritt in das Respondi Online Access Panel über sich bereitstellten und die sie in regelmäßigen Ab-

abgeglichen. Ab der zweiten Welle wurde überprüft, ob die Angaben zu diesen vier Merkmalen mit jenen aus der ersten Welle übereinstimmten (vgl. Abbildung 2). Nur bei vollständiger Übereinstimmung wurden die Teilnehmer ohne Umwege zum Fragebogen weitergeleitet. Erwies sich mindestens eine Angabe als falsch, erhielten sie die Möglichkeit ein zweites Mal die vier Merkmale einzugeben. Wichen die Daten auch ein zweites Mal ab, wurde den Teilnehmern mitgeteilt, dass sie nicht an der Befragung teilnehmen könnten, dass das Problem aber untersucht und sie später nochmals eingeladen werden würden. Bevor die betroffenen Teilnehmer aussortiert wurden, konnten sie auf der letzten Seite angeben, weshalb die Verifikationsangaben ihrer Meinung nach von den Angaben aus der ersten Welle abwichen (vgl. „Grund für falschen Verifizierungscode“ in Abbildung 2). Die Zielpersonen wurden im Zuge der Reminder-E-Mails für Unterbrecher wieder zur Umfrage eingeladen und konnten dann erneut versuchen, sich zu verifizieren.

Abbildung 2: Ablauf des Verifizierungsprozesses



Eine einmalige Ausnahme im Verifizierungsprozess wurde bei dem Merkmal „Erster Buchstabe des Geburtsorts“ eingeräumt. Da es sich hierbei um eine offene Angabe handelte, wurde der Möglichkeit Rechnung getragen, dass in der ersten Welle ein Tippfehler gemacht wurde. Daher hatten die Teilnehmer einmalig die Möglichkeit, die in Welle 1 getätigte Angabe zu korrigieren, sofern *nur* der Buchstabe falsch war. 77 Personen korrigierten auf diese Weise den Anfangsbuchstaben ihres Geburtsortes ($n_{\text{Welle2}} = 72$; $n_{\text{Welle3}} = 2$; $n_{\text{Welle4}} = 2$; $n_{\text{Welle5}} = 0$; $n_{\text{Welle6}} = 0$; $n_{\text{Welle7}} = 1$).

ständen aufgefordert werden, zu aktualisieren. Diese Daten wurden somit bereits vor dem Beginn des Wahlkampfpanels durch Respondi erhoben und ermöglichen eine gesteuerte Einladung. Die von Respondi bereit gestellten Stammdaten (siehe 4.9) wurden im Wahlkampfpanel größtenteils nochmals erhoben, weil wenig über die Aktualität dieser Angaben bekannt ist.

3.5 Interviewdauer

Die Median-Befragungszeiten lagen im Durchschnitt über alle sieben Panelwellen bei fast genau 20 Minuten. Mit einer Medianzeit von 24:37 Minuten dauerte die erste Welle am längsten, während die fünfte Befragung mit 17:37 Minuten am kürzesten ausfiel (vgl. Tabelle 8).

Wie aus den Angaben zu den Minimal- und Maximalwerten ersichtlich wird, gab es in jeder Welle beträchtliche Variation in den Befragungszeiten. Relativ unkritisch sind dabei sehr lange Befragungszeiten, die vermutlich durch kurzzeitige Interview-Unterbrechungen hervorgerufen wurden (längere Unterbrechungen, die eine automatische Abmeldung von der Umfrage zur Folge hatten, sind in der Statistik nicht enthalten). Wesentlich problematischer sind hingegen sehr schnelle Interviews, weil diese darauf hindeuten, dass sich einige Befragte lediglich durch die Umfrage geklickt haben, ohne die Fragen sorgfältig gelesen zu haben. Der Umgang mit sehr schnellen Befragungszeiten wird in Abschnitt 6.2 thematisiert.

Tabelle 8: Interviewlänge

	N	Minimale Befragungs- dauer	Maximale Befra- gungsdauer	Arithmetisches Mittel	Median
Welle 1	5002	00:03:33	03:08:55	00:27:38	00:24:37
Welle 2	4236	00:02:18	02:09:10	00:24:31	00:21:07
Welle 3	4143	00:01:54	02:33:03	00:21:48	00:19:18
Welle 4	4089	00:01:55	02:47:13	00:21:34	00:18:52
Welle 5	4041	00:01:34	05:23:54	00:20:19	00:17:37
Welle 6	3906	00:01:39	02:17:08	00:20:46	00:18:19
Welle 7	3926	00:01:43	02:41:24	00:24:14	00:21:21

Die Bearbeitungsdauer wird nur für diejenigen Befragten ausgewiesen, die die Befragung weder unter- noch abgebrochen haben. Laut Erhebungssoftware hat eine Person in der zweiten Welle 105:44:24 für die Befragung ohne Unterbrechung aufgewendet. Diese Person wurde für die Darstellung der Bearbeitungsdauer in Welle nicht berücksichtigt. Darstellungsweise: hh:mm:ss

3.6 Incentivierung

Um eine möglichst hohe Wiederbefragungsbereitschaft zu erreichen, wurde die Teilnahme am Wahlkampfpanel stärker honoriert als bei gewöhnlichen Respondi-Umfragen. Mit Respondi wurde eine Basis-Incentivierung von 150 Mingle-Punkten (1,50€) für die Teilnahme an jeder einzelnen Welle vereinbart, während die übliche Bezahlung für ein 20-minütiges Interview bei 100 Mingle-Punkten liegt. Entgegen der Absprache hat Respondi für die Teilnahme an der ersten Welle nur 100 Mingle-Punkte ausgezahlt. Als Ausgleich wurden in der zweiten Welle einmalig 200 Mingle-Punkte ausgeschüttet (vgl. Tabelle 9).

Darüber hinaus wurde den Teilnehmern mitgeteilt, dass sie für eine häufige Teilnahme am Wahlkampfpanel bis zu 350 zusätzliche Mingle-Punkte erzielen können (vgl. Tabelle 10). Auf diese Weise hatten die Teilnehmer des Wahlkampfpanels die Möglichkeit, bei einer siebenmaligen Teilnahme am Wahlkampfpanel maximal 1.400 Mingle-Punkte ($7 \cdot 150 + 350$) zu erwerben.

Tabelle 9: Mingle-Punkte für die Teilnahme an den einzelnen Wellen

Teilnahme an...	Mingle-Punkte
Welle 1	100
Welle 2	200
Welle 3	150
Welle 4	150
Welle 5	150
Welle 6	150
Welle 7	150

Tabelle 10: Zusätzliche Mingle-Punkte für mindestens vier Teilnahmen

Teilnahme an ... Wellen	Mingle-Punkte
4	120
5	170
6	220
7	350

3.7 Teilnahme statistiken

3.7.1 Teilnahmehäufigkeit

Von den 5.256 Personen, die die erste Welle des Wahlkampfpanels beendet hatten, nahmen 3.487 Personen (66%) an allen sieben Befragungen teil (vgl. Tabelle 11). Lediglich rund 8 Prozent der Teilnehmer nahmen ausschließlich an der ersten Welle teil. Im Datensatz gibt die Variable „teilnahme“ detailliert Auskunft über den individuellen Teilnahmeverlauf über alle sieben Wellen hinweg (siehe auch Kapitel 4.3).

Tabelle 11: Anzahl abgeschlossener Befragungen

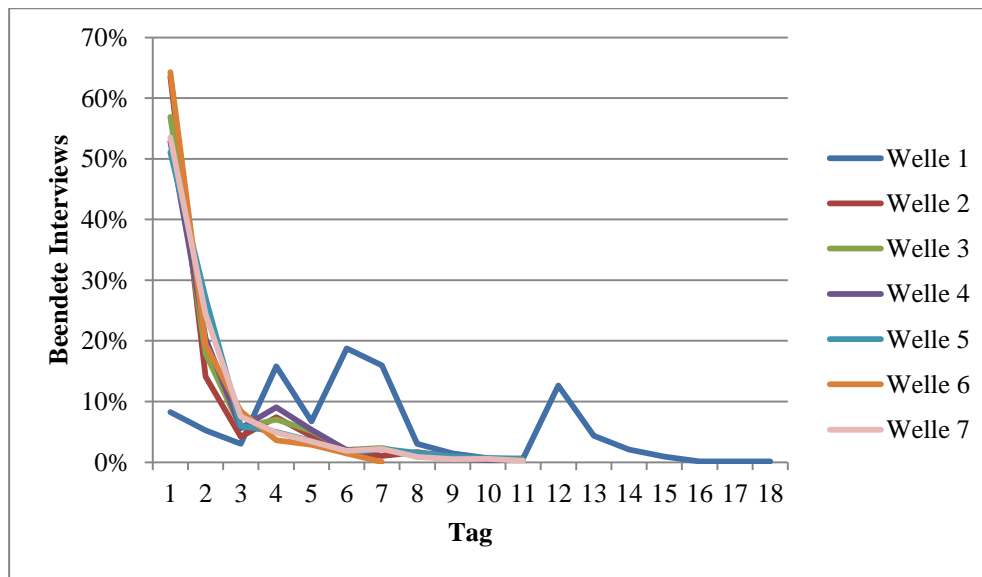
Wellenteilnahmen	absolut	%
1 Teilnahme	413	7,9
2 Teilnahmen	235	4,5
3 Teilnahmen	185	3,5
4 Teilnahmen	185	3,5
5 Teilnahmen	244	4,6
6 Teilnahmen	507	9,6
7 Teilnahmen	3487	66,3
Gesamt	5256	100,0

Abweichungen von 100 Prozent durch Runden der Zahlen.

3.7.2 Verteilung der Teilnahmen über die Feldzeit

In der ersten Welle fiel die Feldzeit aufgrund der gestaffelten Einladungen, die das Erreichen des Quoten-Solls garantieren sollten, um eine Woche länger aus. Die Umfrageteilnahmen verteilten sich dabei über die gesamte Länge der Feldzeit. Ab der zweiten Welle wurden alle Befragten am selben Tag eingeladen. Dies hatte eine sehr starke Konzentration der Umfragezugriffe auf die ersten zwei Tagen der Feldzeit zur Folge, in denen im Schnitt fast 78 Prozent aller Umfragen beendet wurden (vgl. Abbildung 3). Am vierten Feldtag ist in den meisten Wellen eine leichte Zunahme der Umfrageteilnahmen zu verzeichnen, die durch das Versenden der Erinnerungs-E-Mails erklärt werden kann.

Abbildung 3: Verteilung der beendeten Umfragen in der Feldzeit (in Prozent der Anzahl aller beendeten Umfragen pro Welle)



3.7.3 Ausschöpfung

Systematischer Unit Nonresponse ist eine Fehlerquelle in Umfragestudien, die die Qualität der erhobenen Daten erheblich mindern kann.⁵ Sind die in der Umfrage erhobenen Variablen mit der Teilnahmewahrscheinlichkeit korreliert, so sind Verzerrungen der Umfrageergebnisse („Nonresponse Bias“) die Folge. Die Angabe von Ausschöpfungsquoten („Response Rates“) erlaubt eine Einschätzung darüber, wie stark eine Umfragestudie von möglicherweise systematischer Nichtteilnahme betroffen ist. Die Angabe von standardisierten Response Rates und verwandten Indikatoren für das Teilnahmeverhalten der Personen in der Stichprobe sichert die Vergleichbarkeit dieser Maße über Studien hinweg. Für das Wahlkampf-Panel der GLES werden Indikatoren für das Teilnahmeverhalten nach den 2011 publizierten Standards der American Association for Public Opinion Research sowie Empfehlungen zur Berechnung der Ausschöpfung von Callegaro und DiSogra (2008) angegeben.

Bei Online-Befragungen ist die Interpretation von Response Rates in vielen Fällen problematisch, da in der Regel keine zufallsbasierten Stichprobenziehungsverfahren angewendet werden. Bei der Befragung von Teilnehmern aus selbstrekrutierten Online-Access-Panels, wie dem Online-Access-Panel von Respondi, ist es nicht sinnvoll, von einer Response Rate zu sprechen (vgl. The American Association for Public Opinion Research 2011). Die AAPOR empfiehlt daher für Befragungen von Teilnehmern aus selbstrekrutierten Online-Panels die Angabe einer „Participation Rate“, die sich als Anteil der verwertbaren Interviews an der absoluten Anzahl aller versendeten persönlichen Befragungseinladungen berechnet.

Zur Berechnung der Ausschöpfungsquoten wird zwischen ausreichenden und unzureichenden Interviews unterschieden. Als ausreichendes Interview werden alle Befragungen gewertet, die bis zum Ende von teilnahmeberechtigten Personen durchgeführt wurden. Dies schließt vollständige (I) und partielle (P) Interviews ein, die dadurch entstehen können, dass Teilnehmer zwar nicht alle Fragen beantwortet, die Befragung aber dennoch beendet haben. Durch eine Screening-Frage zu Beginn der Befragung in der ersten Welle wurden nur jene Personen zugelassen, die in Deutschland wahlberechtigt waren und die bestimmte Quotie-

⁵ Dieser Abschnitt orientiert sich stark an den Ausführungen zur Ausschöpfung im Technical Report zum Wahlkampf-Panel 2009 (siehe Steinbrecher et al. 2013: 25f., sowie Blumenstiel/Gummer 2012).

rungsmerkmale aufwiesen (Geschlecht, Alter, Bildung). Die übrigen Personen wurden ausgescreent bzw. abgewiesen, wenn die festgelegten Quoten bereits erfüllt waren (NE) und gehen nicht mit in die Berechnung der Ausschöpfungsquoten ein.

Die unzureichenden Interviews definieren sich zum einen durch ein abgebrochenes Interview oder dadurch, dass überhaupt kein Interview erfolgte. In die Kategorie des Abbruchs (RBreakoff) fallen die Teilnehmer, die vor dem Ende der Befragung den Fragebogen verlassen und auch nicht mehr zu einem späteren Zeitpunkt zu ihm zurückkehren. Ab der zweiten Welle treten jedoch zwei unterschiedliche Typen von Abbrechern auf: Teilnehmer können entweder vor oder nach dem Verifizierungsprozess die Umfrage vorzeitig verlassen. Während die Teilnahmeberechtigung nach erfolgreicher Verifizierung gewährleistet ist, so bleibt sie bei einem Abbruch vor der Verifizierung unsicher. Die Teilnahmeberechtigung ist jedoch relevant für die Berechnung einer Abbruchrate, um sie möglichst vergleichbar zu machen. Daher sollten diese zwei Typen nicht undifferenziert verwendet werden. Durch das Paneldesign ist die Eignung bereits in der ersten Welle ermittelt worden, sodass es sich nur dann um einen nicht teilnahmeberechtigten Befragten handeln kann, wenn eine dritte unbekannte Person den Respondi-Account ebenfalls nutzt. Da also Befragte, die vor der Verifizierung abbrechen, nicht zweifelsfrei als teilnahmeunberechtigt eingestuft werden können, wurde in Folge dessen entschieden, dass zwei Abbruchraten berechnet werden. Eine, die nur Abbrüche durch Teilnehmer mit gültiger Teilnahmeberechtigung (R1Breakoff) beinhaltet und eine zweite, die zusätzlich auch Abbrüche durch Teilnehmer mit unklarer Teilnahmeberechtigung (R2Breakoff) umfasst.

Bei Respondi-Mitgliedern, die bereits in der ersten Welle nicht auf die Einladung zur Panelteilnahme reagiert haben und damit kein Interview gaben, kann keine Aussage über ihre Teilnahmeberechtigung gemacht werden, sodass sie als unklar teilnahmeberechtigt definiert werden müssen (UH). Entsprechend gehen sie nur für die Berechnung der Ausschöpfungsquoten der ersten Welle ein, da für die folgenden Wellen nur die Teilnehmer als teilnahmeberechtigt galten, die das Interview der ersten Welle vollständig absolviert hatten. Reagierten diese teilnahmeberechtigten Respondi-Mitglieder in den folgenden Wellen nicht auf die jeweilig versendeten Einladungen, so wurden sie als kein Kontakt gewertet (NC). Als Sonstige werden schließlich die Respondi-Mitglieder definiert, die das Respondi-Access-Panel in der laufenden Feldphase ganz verlassen haben und somit der Befragung nicht mehr zur Verfügung standen (O). In den Berechnungen der Participation Rate gehen diese Fälle mit ein, da sie potenziell ein Interview hätten geben können.

Ergänzt werden diese Kategorien abschließend durch die Kategorie der Sonstigen mit unklarer Teilnahmeberechtigung. Sie schließt die Teilnehmer ein, die aufgrund falscher Angaben im Verifizierungsprozess von der Befragung ausgeschlossen wurden. Nach üblichen AAPOR Empfehlungen würden diese Fälle als „ausgescreent“ gelten und nicht in die Berechnung der Participation Rate eingehen, da sie als nicht teilnahmeberechtigt gewertet werden würden. Da jedoch die grundsätzliche Teilnahmeberechtigung in der ersten Welle bereits ermittelt wurde und damit der Adressat der Einladung ab Welle zwei in jedem Fall teilnahmeberechtigt ist, muss bei einem Scheitern an der Verifizierung von einer unbekannten dritten Person ausgegangen werden, die den Respondi-Account ebenfalls oder zu Unrecht benutzt hat. Da Teilnehmern, die den Verifizierungscode falsch eingaben, mehrmals durch erneutes Einladen die Gelegenheit gegeben wurde, den richtigen Code einzugeben (siehe Abschnitt 3.4), ist davon auszugehen, dass bei einer erfolglosen Verifizierung der adressierte Teilnehmer nicht in Kontakt mit der Umfrage kam. Würde man diesen Teilnehmer nun ausscreenen, so würde man unterstellen, dass der adressierte Teilnehmer sich so verändert habe, dass er keine Teilnahmeberechtigung mehr hätte und auch für die folgenden Wellen nicht mehr teilnahmeberechtigt wäre. Dies ist jedoch nicht der Fall. Jeder berechtigter Teilnehmer, der das Respondi-Panel nicht verlassen hat, bekam für jede Welle erneut eine Einladung. Folglich handelt es sich bei einem eingegebenen falschen Verifizierungscode nicht um einen fal-

schen Adressaten, sondern das teilnahmeberechtigte Respondi-Mitglied reagierte lediglich nicht auf die Einladung und wäre daher ebenfalls der Kategorie „kein Kontakt“ zuzuweisen. Diese Unterscheidung ist für die Vergleichbarkeit der Participation Rate höchst relevant, da eine Klassifikation als nicht teilnahmeberechtigt dazu führen würde, dass die Bezugsgröße „alle verschickten Einladungen“ variiert, obwohl die Anzahl der Einladungen von teilnahmeberechtigten Personen gleich bliebe. Aus diesem Grund sollen auch die Interviews mit in die Berechnung eingehen, die an der Verifizierung scheiterten. Da aber unbekannt ist, worin die Gründe genau liegen, werden sie als Sonstige mit unklarer Teilnahmeberechtigung definiert (UO). Die Berechnung der Participation Rates ergibt sich dann wie folgt:

$$\text{Participation Rate (PR)} = \frac{I + P}{(I + P) + (R1_{\text{Breakoff}} + R2_{\text{Breakoff}} + NC + O) + (UH + UO)}$$

Als weiteren informativen Indikator berichten wir den Anteil der Befragungsabbrüche in der Online-Befragung. Hierfür wird die „Breakoff Rate“ (Callegaro/DiSogra 2008) berechnet. Wie zuvor erläutert, wird zwischen dem Abbruch vor und nach der Verifizierung unterschieden und damit auch zwei Abbruchraten berichtet. R1Breakoff setzt sich wie oben beschrieben aus allen Abbrüchen zusammen, die nach der Verifizierung erfolgten. R2Breakoff umfasst hingegen die Abbrecher, die bereits vor der Verifizierung die Umfrage verließen. I und P bilden die Anzahl der ausreichenden Interviews. Die Breakoff Rates geben mit anderen Worten den Anteil der Befragungsabbrüche an allen begonnen Interviews wieder.

$$\begin{aligned} \text{Breakoff Rate 1 (BR1)} &= \frac{R1_{\text{Breakoff}}}{(I + P) + (R1_{\text{Breakoff}})} \\ \text{Breakoff Rate 2 (BR2)} &= \frac{(R1_{\text{Breakoff}} + R2_{\text{Breakoff}})}{(I + P) + (R1_{\text{Breakoff}} + R2_{\text{Breakoff}})} \end{aligned}$$

Die Brutto-Stichprobe für die erste Welle des Wahlkampfpanels wurde von Respondi aus ihrem Online-Access-Panel gezogen. Die ausgewählten Panelisten wurden von Respondi zur Teilnahme eingeladen. Die Brutto-Stichprobe für die zweite bis siebte Welle des Wahlkampf-Panels bildeten die Panelisten, welche die erste Welle des Panels beendeten. Sie bilden somit die Berechnungsgrundlage der Participation und Breakoff 1 & 2 Rate. Tabelle 12 gibt einen Überblick über die Raten pro Welle.

Tabelle 12: Ausschöpfungsstatistik des Wahlkampfpanels

Code	Beschreibung	W1	W2	W3	W4	W5	W6	W7
I + P	Vollständige und partiell vollständige Interviews	5256	4517	4332	4285	4205	4062	4192
R1 _{Breakoff}	Abbruch nach der Verifizierung	271	81	100	70	52	50	39
R2 _{Breakoff}	Abbruch vor der Verifizierung	/	34	51	32	34	27	21
NC	Kein Kontakt	/	596	729	811	890	1037	912
O	Sonstige	/	14	43	52	68	76	89
UH	Unklare Teilnahmeberechtigung	8841	/	/	/	/	/	/
UO	Sonstige mit unklare Teilnahmeberechtigung	/	14	1	6	7	4	3
NE	Quote voll bzw. in Welle 1 ausgescreent	2081	/	/	/	/	/	/
Anzahl der von Respondi versandten Befragungseinladungen		16449	5242	5213	5204	5188	5180	5167
PR	Participation Rate in %	36,6	85,9	82,4	81,5	80,0	77,3	79,8
BR1	Breakoff Rate 1 in %	4,9	1,8	2,3	1,6	1,2	1,2	0,9
BR2	Breakoff Rate 2 in %	/	2,5	3,4	2,3	2,0	1,9	1,4

Im Datensatz enthält die Variable „kpX_compl“ den Teilnahmestatus der Panelisten gemäß AAPOR-Standards. Anhand der Variablen soll die Kalkulation von Nonresponse-Raten in Forschungsarbeiten erleichtert werden. Tabelle 13 gibt einen Überblick über die Bedeutung der Codes.

Tabelle 13: Ausprägungen der Variablen kpX_compl

Code	Wertelabel	Erläuterung
<u>1 Teilnahmeberechtigte Personen, ausgefüllter Fragebogen</u>		
1.10	Vollständig ausgefüllt	Personen, die den kompletten Fragebogen absolviert haben. Wenn eine „weiß nicht“-Kategorie vorgegeben wurde und diese genutzt wurde, so gilt auch dies als eine vollständige Antwort. Wenn hingegen mindestens einmal weitergeklickt wurde, ohne eine Antwort gegeben zu haben, so gilt der Fall als unvollständig.
1.20	abgebrochen/teilweise ausgefüllt	Die Bedeutung dieser Kategorie unterscheidet sich je nach Welle. In Welle 1 sind dies Personen, die mindestens einmal keine Angabe gemacht haben, aber das Interview beendet haben, so dass sie weiterhin Teil der Panelstudie bleiben (Abbrecher in der ersten Welle wurden nicht wieder eingeladen und sind im veröffentlichten Datensatz nicht enthalten). In den Wellen 2-7 sind dies Personen, die sich korrekt identifiziert haben, mindestens eine Frage nach der Identifizierung beantwortet haben, aber dann entweder a) abgebrochen haben und die Umfrage nicht wieder aufgenommen haben oder b) mindestens einmal ohne gültige Angabe „weiter“ geklickt haben.
<u>2 Teilnahmeberechtigte Personen, kein Interview</u>		
2.11	Implizite Verweigerung bei Personen mit bekannter Teilnahmeberechtigung	Personen, die sich mit ihrem korrekten Identifizierungscode eingeloggt hatten (Abfrage zur Identifizierung der Zielperson), aber sofort danach abbrechen und keine einzige Frage beantworteten (nur in den Wellen 2-7).
2.20	kein Kontakt	Personen, die nicht auf die Einladung reagieren (nur in den Wellen 2-7).
2.30	kein Kontakt, Panelmitgliedschaft beendet	Personen, die nicht auf die Einladung reagierten, weil sie aus dem Respondi-Panel ausgeschieden sind (nur in den Wellen 2-7).
<u>3 Unsichere Teilnahmeberechtigung, kein Interview</u>		
3.93	Implizite Verweigerung bei Personen mit unbekannter Teilnahmeberechtigung	Personen, die die Umfrage abgebrochen hatten, noch bevor sie den Verifizierungscode eingaben (nur in Welle 2-7).
3.94	Abbruch aufgrund eines falschen Verifizierungscodes	Personen, die einen falschen Verifizierungscode eingegeben hatten und daher nicht zur Umfrage weitergeleitet wurden.

Erläuterung: „X“ steht für die jeweilige Befragungswelle

4 Variablen im Datensatz

4.1 Vorbemerkung

Der Datensatz des Wahlkampf-Panels enthält verschiedene Variablen-Typen:

- *GESIS Archiv-Variablen* beschreiben den Datensatz und seine Erstellung. Sie umfassen Variablen, die für die Archivierung und Distribution der Studie benötigt werden, z.B. die Studiennummer, die Versionskennung und die Feldzeit der Erhebung.
- *Administrative Variablen*: Variablen zur Teilnahme der Befragten im Wahlkampfpanel (z.B. AAPOR-Codes oder Dummies, die Auskunft über die Teilnahme an bestimmten Wellen geben).
- *Kontextvariablen* geben Auskünfte über den regionalen Kontext der Befragten.
- *Datenqualitätsvariablen* sind Indikatoren, die potenziell dazu verwendet werden könnten, um die Qualität der gegebenen Antworten eines Befragten zu beurteilen (vgl. hierzu ausführlich Kapitel 6).
- *Gewichtungsvariablen* umfassen Querschnitts- und Panelgewichte (vgl. Abschnitt 4.6).
- *Inhaltliche Variablen*: Die eigentlich interessierenden Merkmale der Befragten, die anhand des Fragebogens erhoben wurden.
- *Paradaten* dokumentieren sowohl den technischen Ablauf des tatsächlichen Interviewprozesses als auch Informationen zu Umfrageaktivitäten der Respondentmitglieder vor und während der Feldzeit. Hierzu zählen beispielsweise die Anzahl an erhaltenen Befragungseinladungen und Umfrageteilnahmen sowie die technische Ausstattung der Befragten (z.B. Browser-Version, Javascript-Version).

Einen Überblick über die Gesamtzahl der Variablen in den einzelnen Wellen gibt Tabelle 14. In den folgenden Unterkapiteln wird der Inhalt der verschiedenen Variablentypen genauer beschrieben.

Tabelle 14: Anzahl der Variablen Gesamt und nach Wellen

	W1	W2	W3	W4	W5	W6	W7	Wellen- über- grei- fend	Gesamt
GESIS-Archiv-Variablen	1	1	1	1	1	1	1	6	13
Administrative Variablen	6	6	6	6	6	6	6	4	46
Kontextvariablen	/	/	/	/	/	/	/	9	9
Datenqualitätsvariablen	17	13	13	13	13	13	15	1	98
Gewichtungsvariablen	/	/	/	/	/	/	/	8	8
Inhaltliche Variablen	216	263	303	298	338	325	330	76	2149
Paradaten	8	7	7	7	7	7	7	13	63
Zeitvariablen	280	222	252	268	312	308	356	/	1998
Gesamt	528	512	582	593	677	660	715	120	4384

4.2 GESIS Archiv-Variablen

Tabelle 15: Archivvariablen von GESIS

Variablenname	Erläuterung
study	Gibt die (ZA-)Studiennummer des Datensatzes im vierstelligen Format an, unter der die Studie im Archivsystem der GESIS abgelegt und zu finden ist
version	Version des Datensatzes, beginnend mit 1.0.0
year	Erhebungsjahr der Studie
field	Feldzeit der Studie. Aufgrund des Paneldesigns finden sich im Datensatz zusätzlich die Variablen kpX_field, die jeweils die Feldzeit der entsprechenden Welle wiedergeben
glescomp	Zuteilung zu einer der elf Komponenten der GLES. Hier Komponente 3 für das Wahlkampfpanel
survey	Differenzierung zwischen mehreren Erhebungen, die innerhalb einer GLES-Komponente durchgeführt wurden. Im Wahlkampfpanel 2013 sind dies der kumulierte Datensatz aller sieben Wellen (codiert mit 1) und die Kontrollquerschnitte zu Welle 3,5 und 7 (codiert mit 2-4)

Erläuterung: „X“ steht für die jeweilige Befragungswelle

4.3 Administrative Variablen

Tabelle 16: Administrative Variablen

Variablenname	Erläuterung
wkp2009	Teilnahme am Wahlkampfpanel 2009 (0=nein; 1=ja)
mut09	Markierung der Fälle, die in 2009 als Panelmutanten identifiziert worden sind
kpX_compl	Teilnahmestatus der Panelisten gemäß AAPOR-Standards
kpX_dispcode	Dispositionscode der Teilnehmer, der seinen Status im Feld angibt, also ob der Teilnehmer z.B. die Umfrage bereits begonnen oder bereits fertiggestellt hat.
kpX_dropout	Vollständigkeit des Interviews (0=vollständig; 1=unvollständig)
kpX_interrupt	Unterbrechung der Befragung (0=keine Unterbrechung; 1=mit Unterbrechung)
wXa	Information, ob der Befragte den jeweiligen Welle X begonnen hat oder nicht (0=nicht teilgenommen; 1=Interview abgebrochen/beendet (nach Verifizierung)).
wXb	Information, ob der Befragte den jeweiligen Befragungsteil X beendet hat oder nicht (0=abgebrochen/nicht teilgenommen; 1=Interview beendet).
n_teilnahmen	Anzahl der Befragungen, die ein Teilnehmer beim Wahlkampfpanel beendet hat.
teilnahme	siebenstelliger Zahlencode, der angibt, an welchen Befragungen des Wahlkampfpanels 2013 eine Teilnahme stattfand. 0 = nicht teilgenommen; 1 = teilgenommen

Erläuterung: „X“ steht für die jeweilige Befragungswelle

4.4 Kontextvariablen: Zuordnung von Wahlkreisen

Die einzige Kontextvariable im Wahlkampfpanel ist der Wahlkreis der Befragten. Die Zuordnung der Befragten zu einem Wahlkreis erfolgte auf der Grundlage der von ihnen genannten Postleitzahl des Hauptwohnsitzes. Zu diesem Zweck stand eine Korrespondenztabelle des Bundeswahlleiters zur Verfügung.

Bei der Verwendung der Wahlkreisinformationen ist zu beachten, dass eine Postleitzahl nicht immer eine eindeutige Zuordnung zu einem Wahlkreis erlaubt. Vor allem in großstädtischen Gebieten kann einer Postleitzahl im Extremfall bis zu fünf Wahlkreisen zugeordnet werden. Sofern eine eindeutige Zuordnung möglich war – in 4.438 Fällen (84,4%) – wurde der Wahlkreis in der Variable „elecdist“ gespeichert. Waren mehrere Einordnungen möglich – in 694 Fällen (13,2%) – so sind die bis zu fünf möglichen Wahlkreise in den Variablen „elecdist1“–„elecdist5“ enthalten. Die restlichen 124 Befragten (2,4%) gaben keine gültige Postleitzahl an oder verweigerten die Angabe.

Zusätzlich wurde ein Abgleich zwischen dem genanntem Bundesland und der angegebenen Postleitzahl durchgeführt. Bei 15 Befragten stimmt die Postleitzahl nicht mit dem Bundesland überein. Diese Fälle wurden in der Variable „kpx_info“ markiert.

4.5 Datenqualitätsvariablen

In dem veröffentlichten Datensatz sind Variablen enthalten, die den Nutzern eine Einschätzung der Datenqualität ermöglichen sollen. Ausführliche Erläuterungen der Qualitätsindikatoren sind in Kapitel 6 zu finden.

Tabelle 17: Datenqualitätsvariablen

Variablenname	Erläuterung	W1	W2	W3	W4	W5	W6	W7
kpX_speederindex	Setzt die Beantwortungsdauer eines Befragten ins Verhältnis zu allen anderen Befragten.	X	X	X	X	X	X	X
kpX_dev	Abweichendes Antwortverhalten Index	X	X	X	X	X	X	X
kpX_itnrp	Anteil der Antworten mit „keine Angabe“ (Item Non-response)	X	X	X	X	X	X	X
kpX_strl	Straightlining-Index	X	X	X	X	X	X	X
kpX_mtime	Durchschnittliche Antwortdauer pro Seite in Sekunden	X	X	X	X	X	X	X
kpX_mtimex	Durchschnittliche Antwortdauer pro Seite in Sekunden (bereinigt für Wellenvergleich)	X	X	X	X	X	X	X
kpX_dokn	Anteil der Antworten mit „weiß nicht“	X	X	X	X	X	X	X
kpX_qual	Gesamtqualitätsindikator	X	X	X	X	X	X	X
kpX_4210	Aufmerksamkeit Umfrageteilnahme	X	X	X	X	X	X	X
kpX_4220	Sorgfalt Umfrageteilnahme	X	X	X	X	X	X	X

kpX_4230, s	Grund für die Umfrageteilnahme	X							X
kpX_4240	Bewertung der Umfrage	X	X	X	X	X	X	X	X
kpX_4250	Teilnahme an Umfragen, Anzahl der Mitgliedschaften in Online-Panels	X							
kpX_4260	Teilnahme an Umfragen, Anzahl der Umfrageteilnahmen im letzten Monat	X							X
kpX_4270s	Kommentarfeld, in das Befragte Ihre Anmerkungen zur Umfrage eintragen konnten	X	X	X	X	X	X	X	X
kpX_	Kontrollitems	050q; 1571q	2900q	050q	040q	2910q	2940q	050q	

Erläuterung: „X“ steht für die jeweilige Befragungswelle.

4.6 GewichtungsvARIABLEN

4.6.1 Querschnittsgewichte

Anhand von Querschnittsgewichten kann die Verteilung bestimmter Variablen eines Datensatzes an bekannte Verteilungen in der Grundgesamtheit angepasst werden. Dabei wird davon ausgegangen, dass zumindest ein schwacher Zusammenhang zwischen den Anpassungsvariablen und den inhaltlichen Untersuchungsvariablen besteht.

Zur Berechnung der Querschnittsgewichte der GLES wurden sowohl sozial- als auch regionalstrukturelle Merkmale ausgewählt. Bei der Berechnung der Querschnittsgewichte für die GLES wurde auf das iterative Anpassungsverfahren („iterative proportional fitting“, IPF) zurückgegriffen (Deming/Stephan 1940). Bei der IPF-Gewichtung wird die Ist-Verteilung der einzelnen Zellen schrittweise an die jeweilige Soll-Verteilung der GewichtungsvARIABLEN angepasst. Bei diesem schrittweisen Prozess der Anpassung (Iteration) bilden die nach jedem Anpassungsschritt berechneten Gewichtungsfaktoren die Ausgangsbasis für die Anpassung des nächsten Merkmals. Der Anpassungsprozess endet, wenn die Differenz zwischen der gewichteten Randverteilung aller Faktoren und der Soll-Verteilung das Abbruchkriterium von 0,05 unterschreitet.⁶ Um sehr große Gewichtungsfaktoren zu vermeiden, werden die Faktoren (nach jedem Iterationsschritt) auf den fünffachen Mittelwert der Gewichtungsvariable (also fünf) getrimmt.⁷ Allerdings war eine solche Trimmung im Wahlkampfpanel weder bei den Querschnitts- noch bei den Panelgewichten notwendig, da die Abweichungen von Soll- und Ist-Werte nicht groß ausfielen.

⁶ Die Querschnittsgewichte wurden mit dem Statistikprogramm Stata berechnet, wobei auf das Ado „ipfweight“ von Michael Bergmann zurückgegriffen wurde.

⁷ Dieses Vorgehen findet auch bei der Berechnung der Gewichte der American National Election Study (ANES) Anwendung (siehe dazu: DeBell et al. 2009).

Tabelle 18: Übersicht Gewichte

Gewicht	Variablenname
Sozial- und regionalstrukturelles Gewicht (Anpassung MZ2012)	wei_mz
Sozial- und regionalstrukturelles Gewicht (Anpassung Onliner)	wei_on

Bei der Erstellung der Querschnittsgewichte für das Wahlkampfpanel wurde, basierend auf den Ist-Verteilungen der ersten Panelwelle, zum einen an die Randverteilungen des (N)Onliner Atlas 2012⁸ und zum anderen an die Verteilungen des Mikrozensus 2012 angepasst. Bei der Erstellung der Sollverteilungen des Mikrozensus 2012 ging nur die wahlberechtigte Bevölkerung in Privathaushalten am Ort der Hauptwohnung ein.⁹

Angepasst wurde an die soziodemographischen und regionalstrukturellen Merkmale: Geschlecht, Alter, Bildung und alte bzw. neue Bundesländer (inkl. Berlin). Das Alter wurde in vier Gruppen kategorisiert: „18 bis unter 30 Jahre“, „30 bis unter 45 Jahre“, „45 bis unter 60 Jahre“ und „60 Jahre und älter“. Das Merkmal Bildung wurden in drei Gruppen unterteilt:

- niedrige Bildung: Schule beendet ohne Abschluss, Hauptschulabschluss, Volksschulabschluss, bin noch Schüler/in
- mittlere Bildung: Realschulabschluss, Mittlere Reife, Fachschulreife oder Abschluss der polytechnischen Oberschule 10. Klasse
- hohe Bildung: Fachhochschulreife (Abschluss einer Fachoberschule etc.), Abitur bzw. erweiterte Oberschule mit Abschluss 12. Klasse (Hochschulreife)

⁸ Die Randverteilungen (für Personen ab 18 Jahren) aus dem (N)Onliner Atlas wurden für die Gewichtung freundlicherweise von TNS Infratest zur Verfügung gestellt.

⁹ Bei den vom Mikrozensus ausgewiesenen Werten handelt es sich um absolute Personenzahlen nach gebundener Hochrechnung.

Tabelle 19: Ist- und Soll-Verteilungen der zur Erstellung der Gewichte verwendeten Variablen

Merkmal	Ist (in Prozent)		Soll (in Prozent)	
	N=5256		Mikrozensus 2012	(N)Onliner Atlas 2012
Geschlecht				
männlich	48,6		48,6	52,5
weiblich	51,4		51,4	47,6
Altersgruppe (gekreuzt mit Bildung)				
18 bis unter 30 Jahre	niedrig	4,1	4,2	
	mittel	17,5 5,6	5,4	22,6
	hoch	7,7	6,8	
30 bis unter 45 Jahre	niedrig	8,5	5,3	
	mittel	29,8 10,9	8,3	28,9
	hoch	10,3	8,5	
45 bis unter 60 Jahre	niedrig	11,7	10,1	
	mittel	33,0 11,3	10,2	30,9
	hoch	10,0	7,8	
60 Jahre und älter	niedrig	6,4	22,1	
	mittel	19,8 6,4	5,9	17,6
	hoch	7,0	5,3	
Bildung				
niedrig	30,8			34,6
mittel	34,3			32,1
hoch	35,0			33,3
Bundesland				
Alte Bundesländer	78,2		78,6	79,9
Neue Bundesländer (inkl. Berlin)	21,8		21,4	20,1

Da für den Mikrozensus zusätzlich die gemeinsame Verteilung der Merkmale Alter und Bildung vorliegt, konnte differenzierter an die Grundgesamtheit angepasst werden. Vorteil dieses Vorgehens ist, dass in jeder Altersgruppe an die einzelnen Bildungskategorien angeglichen werden kann und somit eventuelle Verzerrungen durch besonders stark oder schwach besetzte Gruppen entgegengewirkt wird. Dies ist jedoch nur für die Anpassung an den Mikrozensus möglich, da der (N)Onliner Atlas über keine Information zur gemeinsamen Verteilung der beiden Merkmale verfügt. Entsprechend wurden für die Anpassung an die Onlinebevölkerung die Merkmale einzeln mit aufgenommen (siehe Tabelle 19).

Die Berechnung der beiden Gewichtungsvariablen wurde jeweils nach der dritten Iteration abgebrochen, da die Differenz zwischen den gewichteten Istwerten in der Stichprobe und den Sollwerten des Mikrozensus bzw. (N)Onliner Atlas das Abbruchkriterium von 0,05 unterschritt. In beiden Fällen war keine Trimmung notwendig. Eine Übersicht über die berechneten Gewichte bietet die nachfolgende Tabelle.

Tabelle 20: Übersicht über die Gewichtungsfaktoren

	N	Mean	Std.Dev	Min	Max	1.Q	Median	3.Q	Max./Min. ¹⁰
wei_mz	5256	1	0,65	0,61	3,65	0,77	0,85	0,90	5,98
wei_on	5256	1	0,19	0,71	1,67	0,85	0,96	1,12	2,35

4.6.2 Panelgewichte

Panelgewichte korrigieren die systematischen Verzerrungen, die in Panelstudien durch unregelmäßige Teilnahmen (Panelattrition) an den Panelwellen entstehen. Hierbei handelt es sich in aller Regel um Propensity Score Gewichte. In Folge werden die Berechnungsmethoden, die verwendeten Modelle sowie die erstellten Gewichte erläutert.

Bei der Propensity Score Gewichtung werden für alle Befragten individuelle Teilnahmewahrscheinlichkeiten geschätzt. Im Falle eines Panels ist von Bleibewahrscheinlichkeiten zu sprechen. Hier wird für alle Befragten der Verbleib in den Panelwellen geschätzt (außer für die erste Welle, an der alle Befragten teilgenommen haben). Üblicherweise verwendet man zur Modellierung des Ausfallprozesses logistische Regressionen. Das geschätzte Modell liefert eine Erklärung des Ausfallprozesses und erlaubt für jeden Befragten die individuelle Bleibewahrscheinlichkeit zu schätzen. Die Inverse der Bleibewahrscheinlichkeit bildet den individuellen Gewichtungsfaktor für einen Befragten, welcher eingesetzt werden kann, um für Panelattrition zu kontrollieren. Konkret bedeutet das: Befragte mit einer niedrigen Bleibewahrscheinlichkeit erhalten ein höheres Gewicht, während Befragte mit einer höheren Bleibewahrscheinlichkeit ein niedrigeres Gewicht erhalten. Für jede Panelwelle wird ein neues Modell geschätzt, im Falle des Wahlkampfpanels für die Teilnahme in Welle zwei, drei, vier, fünf, sechs und sieben. Dabei werden immer die Informationen der letzten Welle verwendet, da davon auszugehen ist, dass die gemachten Erfahrungen in der letzten Befragung die Teilnahme in der darauffolgenden Welle beeinflussen sollten. Das bedeutet, die Bleibewahrscheinlichkeiten für Welle drei werden mit Informationen aus Welle zwei geschätzt. Da einige Merkmale als zeitlich invariant anzunehmen sind, bspw. das Geschlecht oder das Geburtsjahr, können in diesen Fällen die Informationen aus der ersten Welle verwendet werden.

Bei der Berechnung der Gewichte ergeben sich zwei Probleme. Erstens ist es möglich, dass Befragte fehlende Werte in den im Modell verwendeten Variablen aufweisen. Folglich ist es nicht möglich, für diese Befragte Bleibewahrscheinlichkeiten zu schätzen. In solchen Fällen wird die mittlere Bleibewahrscheinlichkeit der anderen Befragten zugewiesen. Zweitens kann es sein, dass Befragte in der vorangehenden Welle nicht teilgenommen haben. Auch in diesem Fall kann keine Bleibewahrscheinlichkeit geschätzt werden. Liegen aus früheren Wellen Bleibewahrscheinlichkeiten für den Befragten vor, werden diese zur Berechnung des Gewichtungsfaktors verwendet. Beispielsweise nimmt ein Fall in den Wellen eins, zwei, drei und fünf teil. Da in Welle vier keine Informationen zum Befragten vorliegen, ist es nicht mög-

¹⁰ Der Wert Max/Min gibt das Verhältnis zwischen dem höchsten und niedrigsten Gewichtungsfaktor an. Im Idealfall werden die Gewichte weder besonders groß noch sehr klein, so dass ein niedriger Wert positiv zu betrachten ist.

lich die Bleibewahrscheinlichkeit für Welle fünf zu bestimmen. Daher wird die Bleibewahrscheinlichkeit aus Welle drei verwendet, um den Gewichtungsfaktor für Welle fünf zu berechnen. Ist dies nicht möglich, weil eine Teilnahme bisher nur in Welle eins stattfand, wird die mittlere Bleibewahrscheinlichkeit der anderen Befragten imputiert.

Die zur Modellierung verwendeten Variablen sind in Tabelle 21 aufgeführt. Die Auswahl orientiert sich an einigen theoretischen Ansätzen zur Erklärung des Teilnahmeverhaltens in Umfragen, welche schon der Panelgewichtung in der GLES 2009 zu Grunde lagen (vgl. Blumenberg/Gummer 2013: S. 16ff). Bedingt durch den Modus, das Frageprogramm sowie zusätzlich verfügbare Informationen, wurde das Erklärungsmodell für das aktuelle Wahlkampfpanel angepasst. Die auf dieser Basis berechneten Modelle sind Tabelle 22 zu entnehmen.

Tabelle 21: Verwendete Variablen

Erläuterung der verwendeten Variablen	Codierung
Alter (und Alter ²)	18-99
Geschlecht (männlich / weiblich)	0/1
Bildung (niedrig / mittel / hoch)	0/1 & 0/1 & 0/1
Lebenspartner	0/1
Erwerbsstatus (Erwerbsperson / Hausfrau/mann / RentnerIn)	0/1 & 0/1 & 0/1
Haushaltsgröße	1-8
Ostdeutschland	0/1
Politisches Interesse	0-4
Häufige Gespräche über Politik	0/1
Stärke der Parteiidentifikation	0-5
Zufriedenheit mit Demokratie	0/1
Kanzlerpräferenz	0/1
Wahlbeteiligungsabsicht	0/1
Politisches Wissen: Korrekte Zuordnung der Spitzenkandidaten der großen Parteien	0/1
Internetnutzung (Intensität)	0-7
Teilnahmemotivation: Index aus subjektiver Sorgfalt und Genauigkeit der Beantwortung der Fragen sowie Bewertung der Umfrage	factor scores
Kontrollitems (korrekt erkannt)	W2: 0-2, W3-7: 0/1
Big 5	factor scores
Monetäre Motivation: Grund der Teilnahme „Mingle Punkte“:	0/1
Erfahrung mit Umfragen: Anzahl an bisher teilgenommenen Umfragen (0; 1-4; 5-10; 11+)	0-3
Anzahl bisherige Teilnahmen am WKP	1-6 (je nach Welle)
Häufigkeit Item Nonresponse (Antwortverweigerung)	0-122
Dauer des letzten Interview (und Dauer ²)	Zeit in Sekunden
Verwendetes Eingabegerät (PC / Smartphone / Tablet)	0/1 & 0/1 & 0/1

Tabelle 22: Logistische Regressionen zur Erklärung der Teilnahme in den Wellen 2-7

	Teilnahme in					
	Welle 2	Welle 3	Welle 4	Welle 5	Welle 6	Welle 7
	b (se)	b (se)	b (se)	b (se)	b (se)	b (se)
Alter	0,039 ^a (0,021)	-0,050 (0,033)	-0,000 (0,035)	0,017 (0,033)	0,039 (0,031)	0,037 (0,044)
Alter ²	-0,000 (0,000)	0,001* (0,000)	0,000 (0,000)	0,000 (0,000)	-0,000 (0,000)	-0,000 (0,000)
Bildung: mittel	0,430*** (0,117)	0,131 (0,163)	-0,047 (0,187)	0,093 (0,179)	-0,183 (0,179)	0,041 (0,250)
Bildung: hoch	0,353** (0,126)	0,155 (0,175)	-0,032 (0,204)	0,564** (0,213)	-0,068 (0,198)	0,460 (0,288)
Lebenspartner	-0,042 (0,110)	-0,316* (0,157)	0,048 (0,173)	0,206 (0,172)	-0,159 (0,168)	-0,420 (0,262)
Erwerb: Hausfrau / -mann	0,491* (0,215)	0,118 (0,297)	0,327 (0,324)	-0,315 (0,284)	0,087 (0,282)	-0,389 (0,386)
Erwerb: Rentner	0,116 (0,187)	-0,059 (0,253)	0,185 (0,306)	0,278 (0,339)	0,029 (0,288)	-0,191 (0,378)
Haushaltsgröße	-0,097* (0,040)	-0,088 (0,056)	-0,109 a (0,066)	-0,091 (0,067)	0,029 (0,067)	-0,053 (0,093)
Ostdeutschland	-0,032 (0,114)	0,073 (0,158)	0,253 (0,186)	-0,012 (0,179)	0,356* (0,179)	0,512 a (0,272)
Politisches Interesse	0,093 (0,059)	0,085 (0,078)	0,128 (0,091)	0,337*** (0,090)	0,076 (0,086)	0,084 (0,126)
Häufigkeit pol. Gespräche	-0,111 (0,106)	-0,176 (0,145)	-0,134 (0,170)	-0,432* (0,170)	-0,105 (0,165)	0,089 (0,240)
Stärke der Parteiidentifikation	-0,029 (0,028)	-0,062 (0,039)	0,016 (0,045)	0,030 (0,045)	0,026 (0,043)	0,050 (0,063)
Zufriedenheit mit Demokratie	0,141* (0,056)	-0,079 (0,077)	0,038 (0,087)	0,086 (0,087)	0,183* (0,083)	0,063 (0,123)
Kanzlerpräferenz	0,078 (0,101)	0,185 (0,140)	-0,070 (0,162)	-0,004 (0,162)	-0,202 (0,165)	0,019 (0,235)
Wahlbeteiligungsabsicht	-0,040 (0,139)	0,200 (0,185)	0,290 (0,208)	-0,480* (0,227)	0,395 a (0,209)	0,067 (0,317)
Politisches Wissen	0,198 a (0,114)	0,090 (0,150)	0,161 (0,177)	-0,267 (0,176)	0,124 (0,168)	-0,180 (0,245)
Internetnutzung	0,057 (0,043)	0,137* (0,056)	0,055 (0,068)	0,094 (0,065)	0,125* (0,060)	0,227** (0,074)
Teilnahmemotivation	0,248*** (0,061)	0,162* (0,072)	0,171* (0,085)	0,309*** (0,082)	0,330*** (0,074)	0,167 (0,113)
Kontrollfragen	0,160* (0,078)	-0,159 (0,209)	-0,114 (0,248)	0,006 (0,200)	0,141 (0,219)	0,074 (0,389)
Big 5: Extraversion	-0,198*** (0,059)	-0,102 (0,078)	-0,173 a (0,092)	-0,003 (0,092)	-0,023 (0,088)	0,006 (0,130)
Big 5: Offenheit	-0,060 (0,067)	-0,020 (0,088)	-0,084 (0,104)	-0,023 (0,102)	-0,130 (0,098)	-0,008 (0,150)
Big 5: Gewissenhaft.	0,097 (0,067)	-0,016 (0,089)	-0,010 (0,106)	-0,116 (0,104)	-0,016 (0,099)	0,247 a (0,148)

Fortsetzung
22

	Teilnahme in					
	Welle 2	Welle 3	Welle 4	Welle 5	Welle 6	Welle 7
	b (se)	b (se)	b (se)	b (se)	b (se)	b (se)
Big 5: Neurotizismus	0,112 a (0,064)	-0,115 (0,086)	0,108 (0,100)	-0,114 (0,103)	0,035 (0,097)	-0,028 (0,145)
Big 5: Verträglichkeit	0,079 (0,067)	0,006 (0,090)	0,293** (0,108)	-0,148 (0,106)	-0,008 (0,103)	0,085 (0,153)
Monetäre Motivation	0,189 a (0,114)	0,022 (0,151)	-0,066 (0,173)	0,238 (0,181)	0,092 (0,168)	-0,210 (0,238)
Erfahrung mit Umfragen	0,341*** (0,058)	0,351*** (0,080)	0,185* (0,091)	0,109 (0,090)	0,100 (0,086)	0,016 (0,126)
Bisherige Teilnahmen	---	---	1,910*** (0,201)	1,269*** (0,140)	1,130*** (0,100)	1,103*** (0,098)
Item Nonresponse	-0,038*** (0,008)	-0,007 (0,012)	-0,012 (0,009)	-0,023** (0,007)	0,000 (0,013)	-0,014 a (0,009)
Dauer letztes Interview	-0,000** (0,000)	0,000 (0,000)	-0,000 (0,000)	-0,001 a (0,000)	-0,000 (0,000)	-0,000 (0,000)
Dauer letztes Interview ²	0,000a (0,000)	-0,000 (0,000)	-0,000 (0,000)	0,000 (0,000)	0,000 (0,000)	0,000 (0,000)
Device: Smartphone	-0,228 (0,228)	0,215 (0,389)	0,452 (0,517)	-0,431 (0,384)	0,918 (0,572)	-0,626 (0,521)
Device: Tablet	-0,211 (0,258)	0,245 (0,438)	0,311 (0,477)	1,055 a (0,610)	0,016 (0,437)	-0,354 (0,569)
Konstante	-0,321 (0,596)	1,691* (0,827)	-3,893*** (1,118)	-3,401** (1,106)	-5,267*** (1,002)	-5,147*** (1,388)
N	4810	4142	3959	3914	3823	3719
Nagelkerke R ²	0,12	0,08	0,11	0,13	0,14	0,19

a p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Tabelle 23 fasst die berechneten Panelgewichte zusammen. Sie können eingesetzt werden, wenn Variablen aus späteren Wellen untersucht werden sollen, für welche anzunehmen ist, dass sie durch Panelattrition verzerrt sind. Für jede Welle liegt ein Panelgewicht „wei_wX“ vor, wobei X die jeweilige Welle kennzeichnet. Mittels Multiplikation können die Panelgewichte mit den Querschnittsgewichten kombiniert werden, um weiterhin noch an die Grundgesamtheit der wahlberechtigten Personen in Deutschland (Mikrozensus) oder der Online-Population ((N)Onliner-Atlas) anzupassen.

Tabelle 23: Panelgewichte für die Wellen 2-7

	N	Mean	Std.Dev	Min	Max	1.Q	Median	3.Q	Max./Min. ¹¹
wei_w2	4598	1	0,13	0,88	2,90	0,93	0,97	1,02	3,30
wei_w3	4432	1	0,05	0,93	1,64	0,96	0,99	1,02	1,76
wei_w4	4355	1	0,07	0,95	2,19	0,97	0,99	1,00	2,31
wei_w5	4257	1	0,10	0,94	4,13	0,97	0,99	1,00	4,39
wei_w6	4112	1	0,09	0,94	3,11	0,97	0,98	1,00	3,31
wei_w7	4231	1	0,01	0,97	2,78	0,98	0,99	1,00	2,87

4.7 Inhaltliche Variablen

Die Vergabe der Variablennamen erfolgte nach einem einfachen Schema. Die ersten drei Stellen des Variablennamens sind für die jeweiligen Befragungswellen reserviert, z.B. „kp1“ für die erste Welle und „kp2“ für die zweite Welle¹². Anschließend folgt die Item-Nummer. Die Variable „Politische Interesse“ hat beispielsweise die Item-Nummer „010“. Dementsprechend ist die Variable „Politisches Interesse“, die in Welle 1 erhoben wurde, unter der Bezeichnung „kp1_010“ zu finden. Sofern ein Item unverändert sowohl im Wahlkampfpanel 2009 als auch im Wahlkampfpanel 2013 erhoben wurde, hat es in beiden Datensätzen dieselbe Item-Nummer. Wurden jedoch größere Veränderungen im Fragetext oder der Antwortskala vorgenommen, wurde die letzte Ziffer der Itemnummer um eins erhöht.

Tabelle 24: Inhaltliche Variablen

Variablenstamm	Variablenlabel	W1	W2	W3	W4	W5	W6	W7
kpX_screen 1	Wahlberechtigung	X						
kpX_screen 2	Wahlberechtigung, Nachfrage	X						
kpX_2280	Geschlecht	X	X	X	X	X	X	X
kpX_2291	Geburtsmonat	X	X	X	X	X	X	X
kpX_2290	Geburtsjahr	X	X	X	X	X	X	X
	Wahlverhalten	W1	W2	W3	W4	W5	W6	W7
kpX_170	Wahlbeteiligungsabsicht	X	X	X	X	X	X	
kpX_180	Tatsächliche Wahlbeteiligung							X
kpX_190	Beabsichtigte Stimmabgabe	X	X	X	X	X	X	
kpX_191	Stimmabgabe Briefwähler			X	X	X	X	X
kpX_200	Tatsächliches Wahlverhalten							X
kpX_270	Sicherheit der Wahlentscheidung	X	X	X	X	X	X	
kpX_252	Gründe für Nichtwahl, gestützt	X	X	X	X	X	X	X
kpX_260s	Gründe für Wahlentscheidung, offen	X	X	X	X	X	X	

¹¹ Siehe Fußnote 10

¹² Eine Ausnahme von dieser Regel, stellen die sozialstrukturellen Merkmale der Befragten dar. Weil diese über die Feldzeit hinweg stabil sein sollten, beginnen die Variablennamen - unabhängig davon in welcher Welle sie erhoben wurden - mit dem Kürzel kpX.

Variablenstamm	Variablenlabel	W1	W2	W3	W4	W5	W6	W7
kpX_261	Gründe für Wahlentscheidung, gestützt			[X]	[X]	[X]	[X]	X
kpX_211	Consideration Set, hypothetisch	X		X			X	X
kpX_221	Consideration Set	X		X			X	X
kpX_2751	Consideration Set am Tag der Briefwahl			X	X	X	X	X
kpX_2760	Grund Briefwahl, gestützt			X	X	X	X	X
kpX_320	Zeitpunkt der Wahlentscheidung							X
kpX_321	Zeitpunkt der Nichtwahlentscheidung							X
kpX_331	Schwierigkeit der Wahlbeteiligung			[X]	[X]	[X]	[X]	X
kpX_330	Schwierigkeit der Wahlbeteiligung			[X]	[X]	[X]	[X]	X
kpX_340	Zufriedenheit Wahlergebnis							X
kpX_2770	Hypothetische Stimmabgabe nach Briefwahl				X	X	X	X
kpX_280	Hypothetische Wahlentscheidung							X
kpX_290	Hypothetische Stimmabgabe nach Bundestagswahl							X
kpX_341	Wahlergebnis: Gewinner und Verlierer							X
Wahlverhalten, retrospektiv		W1	W2	W3	W4	W5	W6	W7
kpX_2780	Wahlbeteiligung BTW 2009	X						
kpX_2790	Sicherheit Rückerinnerung BTW 2009	X						
kpX_350a	Recall vorangegangene BTW	X						
Ideologie		W1	W2	W3	W4	W5	W6	W7
kpX_1490	Links-Rechts-Einstufung, Parteien	X		X		X		X
kpX_680	Links-Rechts-Einstufung, Spitzenkandidaten		X		X	X		X
kpX_1500	Links-Rechts-Selbsteinstufung	X	X	X	X	X	X	X
Einstellungen zu Parteien und Regierung		W1	W2	W3	W4	W5	W6	W7
kpX_430	Skalometer Parteien	X	X	X	X	X	X	X
kpX_730	Skalometer Regierung	X	X	X	X	X	X	X
kpX_2800	Ambivalenz Parteien, negative Gefühle	X		X			X	
kpX_2801	Ambivalenz Parteien, positive Gefühle	X		X			X	
Valenzissues		W1	W2	W3	W4	W5	W6	W7
kpX_840s	Wichtigstes Problem	X	X	X	X	X	X	X
kpX_850	Lösungskompetenz: Wichtigstes Problem	X	X	X	X	X	X	X
kpX_851	Lösungskompetenz: Wichtigstes Problem, andere Partei	X	X	X	X	X	X	X
kpX_860s	Zweitwichtigstes Problem	X	X	X	X	X	X	X
kpX_870	Lösungskompetenz: Zweitwichtigstes Problem	X	X	X	X	X	X	X
kpX_871	Lösungskompetenz: Zweitwichtigstes Problem, andere Partei	X	X	X	X	X	X	X

Variablenstamm	Variablenlabel	W1	W2	W3	W4	W5	W6	W7
	Positionsisues	W1	W2	W3	W4	W5	W6	W7
kpX_1070	Sozio-ökonomische Dimension, Parteien		X		X		X	
kpX_1080	Sozio-ökonomische Dimension, Kanzlerkandidaten		X		X	X		X
kpX_1090	Sozio-ökonomische Dimension, Ego	X	X	X	X	X	X	X
kpX_1110	Zuzugsmöglichkeiten Ausländer, Parteien		X		X		X	
kpX_1120	Zuzugsmöglichkeiten Ausländer, Kanzlerkandidaten		X		X	X		X
kpX_1130	Zuzugsmöglichkeiten Ausländer, Ego	X	X	X	X	X	X	X
kpX_1270	Klimaschutz, Parteien		X		X		X	
kpX_1280	Klimaschutz, Kanzlerkandidaten		X		X	X		X
kpX_1290	Klimaschutz, Ego	X	X	X	X	X	X	X
kpX_1100	Sozio-ökonomische Dimension, Wichtigkeit		X		X		X	
kpX_1140	Zuzugsmöglichkeiten Ausländer, Wichtigkeit		X		X		X	
kpX_1300	Klimaschutz, Wichtigkeit		X		X		X	
kpX_2810	Europäische Hilfe, Parteien			X				X
kpX_2820	Europäische Hilfe, Kanzlerkandidaten			X		X		X
kpX_2830	Europäische Hilfe, Ego			X		X		X
kpX_2840	Europäische Hilfe, Wichtigkeit			X				X
kpX_1210	Integration, Ego			X			X	
kpX_1250	Europäische Integration, Ego			X			X	
kpX_1482	Außenpolitik, Ego			X			X	
kpX_2850	Ego-Positionen I	X			X			
kpX_2860	Ego-Positionen II		X			X		
kpX_2870	Ego-Positionen III; Aktuelle Fragen			X			X	
	Kandidaten	W1	W2	W3	W4	W5	W6	W7
kpX_650	Skalometer Politiker	X	X	X	X	X	X	X
kpX_650	Skalometer Politiker		X		X		X	
kpX_661	Eigenschaften der Kanzlerkandidaten, Merkel		X		X	X		X
kpX_661	Eigenschaften der Kanzlerkandidaten, Steinbrück		X		X	X		X
kpX_671	Kanzlerpräferenz	X	X	X	X	X	X	X
kpX_3000	Bewertungsmaßstäbe Kandidatenbewertung		X		X	X		X
	Strategisches Wählen	W1	W2	W3	W4	W5	W6	W7
kpX_910	Koalitionsskalometer		X		X		X	X
kpX_961	Koalitionserwartungen		X		X		X	
kpX_1040	Wahlkreisgewinner		X		X		X	
kpX_1050	Wahrnehmung Meinungsumfragen		X		X		X	
kpX_1051	Glaubwürdigkeit Meinungsumfragen		X		X		X	
kpX_3010	Einzugswahrscheinlichkeit kleine Parteien		X		X		X	
kpX_3020	Bewertung der Wahlkreiskandidaten		X				X	

Variablenstamm	Variablenlabel	W1	W2	W3	W4	W5	W6	W7
Europa		W1	W2	W3	W4	W5	W6	W7
kpX_3030	Angst, europäische Schuldenkrise			X				X
kpX_3040	Betroffenheit Ego, europäische Schuldenkrise			X				X
kpX_3050	Betroffenheit Deutschland, europäische Schuldenkrise			X				X
kpX_3060	Performanz Bundesregierung, Bekämpfung europäische Schuldenkrise			X				X
kpX_3070	Performanz Kanzlerkandidaten, europäische Schuldenkrise			X		X		X
kpX_3080	Hauptschuld europäische Schuldenkrise			X				X
kpX_3090	Rückkehr zur D-Mark			X				X
Wirtschaft		W1	W2	W3	W4	W5	W6	W7
kpX_780	Eigene wirtschaftliche Lage, aktuell	X			X			X
kpX_760	Eigene wirtschaftliche Lage, retrospektiv	X			X			X
kpX_770	Verantwortlichkeit eigene wirtschaftliche Lage	X			X			X
kpX_790	Eigene wirtschaftliche Lage, prospektiv	X			X			X
kpX_820	Allgemeine wirtschaftliche Lage, aktuell	X			X			X
kpX_800	Allgemeine wirtschaftliche Lage, retrospektiv	X			X			X
kpX_810	Verantwortlichkeit allgemeine wirtschaftliche Lage	X			X			X
kpX_830	Allgemeine wirtschaftliche Lage, prospektiv	X			X			X
kpX_3110	Europäische wirtschaftliche Lage, aktuell			X				X
Gerechtigkeit		W1	W2	W3	W4	W5	W6	W7
kpX_2250	Gerechte Gesellschaftsordnung	X						
kpX_2270	Gerechtigkeit, Ego	X						
kpX_3330	Gerechtigkeit, Entwicklung	X						
kpX_3340	Wichtigkeit Gerechtigkeit	X			X			
kpX_3350	Gerechtigkeit, Verständnis	X						
kpX_3360	Gerechtigkeit, Merkel	X			X	X		
kpX_3360	Gerechtigkeit, Steinbrück	X			X	X		
kpX_3370	Gerechtigkeit, Einkommen			X		[X]		[X]
kpX_3380	Wichtigkeit Gerechtigkeit, Einkommen			X		X		X
kpX_3390	Einkommensgerechtigkeit, Parteien			X		X		X
kpX_3400	Gerechtigkeit, Steuern		X		[X]		[X]	
kpX_3410	Wichtigkeit Steuergerechtigkeit		X		X		X	
kpX_3420	Steuergerechtigkeit, Parteien		X		X		X	
kpX_2700	Alt vs. Jung Aussagen			X				
Mediennutzung		W1	W2	W3	W4	W5	W6	W7
kpX_1620	Wichtigste Informationsquelle, gewöhnlich	X						

Variablenstamm	Variablenlabel	W1	W2	W3	W4	W5	W6	W7
kpX_1631	Allgemeine Internetnutzung, Durchschnitt	X						
kpX_1640	Internetnutzung, politisch gewöhnlich	X						
kpX_3540	Dauer Internetnutzung, gewöhnlich	X						
kpX_1651s	Internetnutzung, Seite gewöhnlich, offen	X						
kpX_1721	Printmediennutzung, politisch gewöhnlich	X						
kpX_1741	Fernsehnutzung, Nachrichten gewöhnlich	X						
kpX_1760	Nachrichtenmagazine, gewöhnlich	X						
kpX_1600	Internetnutzung, politisch aktuell		X	X	X	X	X	X
kpX_1661	Printmediennutzung, politisch aktuell		X	X	X	X	X	X
kpX_1681	Fernsehnutzung, Nachrichten aktuell		X	X	X	X	X	X
kpX_1700	Nachrichtenmagazine, aktuell		X	X	X	X	X	X
kpX_396	Media Reliance							X
Politische Gespräche		W1	W2	W3	W4	W5	W6	W7
kpX_1931	Politische Gespräche, allgemein	X						
kpX_1951	Gesprächspartner 1, Verständnis von Politik	X						
kpX_1961	Gesprächspartner 1, Meinungsverschiedenheiten	X						
kpX_1991	Gesprächsnetzwerk, Parteipräferenz allgemein	X						
kpX_1932	Politische Gespräche, aktuell		X	X	X	X	X	X
kpX_1970	Gesprächspartner 1, Wahlentscheidung		X	X	X	X	X	X
Wissen und Involvierung		W1	W2	W3	W4	W5	W6	W7
kpX_090	Politisches Wissen: 5%-Klausel		X			X	X	X
kpX_110	Politisches Wissen: Erst-/Zweitstimme		X		X		X	X
kpX_130	Politisches Wissen: Wahlrecht D		X		X		X	X
kpX_3430	Politisches Wissen: Zuordnung Politiker/Parteien	X		X				X
kpX_3430	Politisches Wissen: Zuordnung Politiker/Parteien		X		X		X	
kpX_3440	Wirtschaftliches Wissen: Arbeitslosenquote	X			X			X
kpX_010	Politisches Interesse, allgemein	X	X	X	X	X	X	X
kpX_3450	Wirtschaftliches Interesse	X			X			X
kpX_380	Wichtigkeit Wahlausgang		X	X	X	X	X	
kpX_390	Interesse am Wahlkampf		X	X	X	X	X	
kpX_391	Interessanter Wahlkampf							X
kpX_392	Aufmerksamkeit Wahlkampf		X	X	X	X	X	X
kpX_395	Wahlkampf hilfreich			[X]	[X]	[X]	[X]	X
kpX_421	Parteikontakte I		X	X	X	X	X	X
kpX_421	Parteikontakte II		X	X	X	X	X	X
kpX_070	Andere Partizipationsformen, retrospektiv		X					

Variablenstamm	Variablenlabel	W1	W2	W3	W4	W5	W6	W7
Wahlkampf im Internet		W1	W2	W3	W4	W5	W6	W7
kpX_3490	Wahl-O-Mat, Nutzung					X	[X]	[X]
kpX_3500	Wahl-O-Mat, Grund der Nutzung					X	[X]	[X]
kpX_3510	Wahl-O-Mat, Übereinstimmung					X	[X]	[X]
TV-Duell		W1	W2	W3	W4	W5	W6	W7
kpX_1780	TV-Duell: Rezeptionswahrscheinlichkeit				X			
kpX_1790	TV-Duell: erwarteter Ausgang				X			
kpX_1800	TV-Duell: Rezeption					X	[X]	[X]
kpX_1810	TV-Duell: Wahrgenommener Ausgang Merkel					X	X	X
kpX_1820	TV-Duell: Wahrgenommener Ausgang Steinbrück					X	X	X
kpX_1870	TV-Duell: Rezeption des Medienurteils					X	X	X
kpX_1880	TV-Duell: wahrgenommenes Medienurteil Merkel					X	X	X
kpX_1890	TV-Duell: wahrgenommenes Medienurteil Steinbrück					X	X	X
kpX_1900	TV-Duell: Rezeption des Urteils des sozialen Umfelds					X	X	X
kpX_1910	TV-Duell: Wahrgenommenes Urteil des sozialen Umfelds zu Angela Merkel					X	X	X
kpX_1920	TV-Duell: Wahrgenommenes Urteil des sozialen Umfelds zu Peer Steinbrück					X	X	X
kpX_3460	TV-Duell: Anzahl Mitseher					X		
kpX_3470	TV-Duell: Emotionale Nähe zu Mitsehern					X		
kpX_3480	TV-Duell: Verständnis von Politik der Mitseher					X		
kpX_3520	TV-Duell: Interaktionsintensität					X		
kpX_3530	TV-Duell: Opinion Leadership 1					X		
kpX_3531	TV-Duell: Opinion Leadership 2					X		
kpX_3550	TV-Duell: Wahrgenommener Ausgang von Mitsehern					X		
kpX_3560	TV-Duell: Kanzlerpräferenz von Mitsehern					X		
kpX_3570	TV-Duell: Aufmerksamkeit					X	[X]	[X]
kpX_3580	TV-Duell: Spezifische Debattenleistung Merkel/Steinbrück					X	[X]	[X]
kpX_3600	TV-Duell: Aussagen Merkel/Steinbrück					X	[X]	[X]
Landtagswahl in Bayern		W1	W2	W3	W4	W5	W6	W7
kpX_3120	Wahlbeteiligungsabsicht, Bayern		X	[X]	X	X		
kpX_3130	Beabsichtigte Stimmabgabe, Bayern		X	[X]	X	X		
kpX_3140	Zufriedenheit Regierung, Bayern		X	[X]	X	X		
kpX_650	Skalometer Politiker: Christian Ude		[X]	X	X	X		
kpX_3160	Aufmerksamkeit Wahlkampf, Bayern		X	[X]	X	X		
kpX_3170	Antizipierte Mehrheitsverhältnisse, Bayern		X	[X]	X	X		

Variablenstamm	Variablenlabel	W1	W2	W3	W4	W5	W6	W7
kpX_3180	Wahlbeteiligung, Bayern						X	[X]
kpX_3190	Tatsächliche Stimmabgabe, Bayern						X	[X]
kpX_3200	Zeitpunkt der Wahlentscheidung, Bayern						X	[X]
kpX_3210	Kenntnis Wahlergebnis: Bayern						X	[X]
kpX_3220	Wahlergebnis: Gewinner und Verlierer, Bayern retrospektiv						X	[X]
Landtagswahl in Hessen		W1	W2	W3	W4	W5	W6	W7
kpX_3230	Wahlbeteiligungsabsicht, Hessen		X	[X]		X	X	
kpX_3240	Beabsichtigte Stimmabgabe, Hessen		X	[X]		X	X	
kpX_3250	Zufriedenheit Regierung, Hessen		X	[X]		X	X	
kpX_650h, i	Skalometer Politiker: Volker Bouffier/Thorsten Schäfer-Gümbel		[X]	X		X	X	
kpX_3270	Aufmerksamkeit Wahlkampf, Hessen		X	[X]		X	X	
kpX_3280	Antizipierte Mehrheitsverhältnisse, Hessen		X	[X]		X	X	
kpX_3290	Wahlbeteiligung, Hessen							X
kpX_3300	Tatsächliche Stimmabgabe, Hessen							X
kpX_3310	Zeitpunkt der Wahlentscheidung, Hessen							X
Aktuelle Fragen		W1	W2	W3	W4	W5	W6	W7
kpX_2900	Meinungen zum Abhörskandal		X	X	X	X	X	X
kpX_2910	Aktuelle sozioökonomische Fragen; Meinung zum Strafprozess gegen Wulff		X			X		
kpX_2920	Meinungen zum Drohnen-Projekt			X				
kpX_2930	Spekulationen zu Rot-Rot-Grün			X		X	X	
kpX_2940	Meinungen zu PKW-Maut, Veggie-Day, Mietpreisbremse				X		X	X
kpX_2950	Gutes Wahlplakat				X			
kpX_2960	Gutes Wahlplakat: Partei				X			
kpX_2961	Gutes Wahlplakat, andere Partei				X			
kpX_2970	Meinungen zum Syrien-Konflikt					X	X	X
kpX_2980	Wahlarena: Rezeption						X	
kpX_2990	Wahlarena: Wahrgenommener Ausgang Merkel/Steinbrück						X	
kpX_2991	Wahlarena: Aussagen, Kenntnis						X	
kpX_2992	Wahlarena: Aussagen, Bewertung						X	
kpX_3590	Kenntnis von den Steuerplänen der Grünen						X	X
kpX_3591	Meinung zu den Steuerplänen der Grünen						X	X
kpX_4010	Guter Wahlwerbespot						X	
kpX_4020	Guter Wahlwerbespot: Partei						X	
kpX_4021	Guter Wahlwerbespot: andere Partei						X	
kpX_4030	Kenntnis von dem Mittelfinger-Titelbild von Steinbrück						X	X
kpX_4031	Meinung zu dem Mittelfinger-Titelbild von Steinbrück						X	X

Variablenstamm	Variablenlabel	W1	W2	W3	W4	W5	W6	W7
kpX_4040	Meinung zu Jürgen Trittins Pädophilie-Affäre							X
kpX_4050	Bewertung der Wahlkampf-Endphase							X
kpX_4060	Koalitionspräferenz							X
	Persönlichkeit, Werte, Grundorientierungen	W1	W2	W3	W4	W5	W6	W7
kpX_1570	Needs	X						
kpX_2180	Big 5	X						
kpX_2090	Parteiidentifikation	X		X		X		X
kpX_2091	Parteiidentifikation, andere Partei	X		X		X		X
kpX_2100	Parteiidentifikation, Stärke	X		X		X		X
kpX_2110	Parteiidentifikation, Dauer	X						X
kpX_2121	Parteiidentifikation Art	X						X
kpX_2081	Inglehart-Items				X	[X]	[X]	
kpX_2170	Allgemeine Lebenszufriedenheit			X				
kpX_2200	Verbundenheitsbatterie			X				
kpX_020	Demokratiezufriedenheit	X						X
kpX_030	Idee Demokratie			X				
kpX_040	Politikverdrossenheit			X	X			
kpX_050	Efficacy und Wahlnorm	X		X				X
kpX_050	Demokratieprinzipien			X				
kpX_060	Extremismusbatterie			X				
kpX_160	Institutionenvertrauen			X				
	Soziodemografie	W1	W2	W3	W4	W5	W6	W7
kpX_2441	Haushaltsgröße	X						
kpX_2450	Anzahl Haushaltsmitglieder unter 18 Jahre	X						
kpX_2591	Nettoeinkommen HH, Kategorien	X						
kpX_2461	Organisationsmitgliedschaft	X						
kpX_2470	Gewerkschaftsmitgliedschaft HH	X						
kpX_4000	Parteimitgliedschaft	X						
kpX_4110	Parteimitgliedschaft, andere Partei	X						
kpX_3910	Religiosität	X						
kpX_2480	Religionszugehörigkeit	X						
kpX_2491	Häufigkeit Gottesdienst	X						
kpX_2580	Schichtzugehörigkeit, subjektiv	X						
kpX_2601	Bundesland	X						
kpX_2600	Wohnort	X						
kpX_2602	Postleitzahl	X						
kpX_2320	Schulabschluss	X						
kpX_2330	Berufliche Bildung	X						
kpX_2340	Erwerbstätigkeit	X						
kpX_2350	Frühere Erwerbstätigkeit	X						
kpX_2360	Arbeitslosigkeit vergangene 10 Jahre	X						
kpX_2370	Dauer Arbeitslosigkeit	X						
kpX_2380	Beruf	X						

Variablen- stamm	Variablenlabel	W1	W2	W3	W4	W5	W6	W7
kpX_3610	Angestellter - Differenzierung	X						
kpX_3620	Arbeiter - Differenzierung	X						
kpX_3630	Akademisch freier Beruf, Selbständiger - Differenzierung	X						
kpX_3640	Beamter - Differenzierung	X						
kpX_3650	Beschäftigungssektor	X						
kpX_3660	Wirtschaftssektor	X						
kpX_3670	Zeit/Leiharbeit	X						
kpX_3680	Angst Stellenverlust	X						
kpX_3690	Angst Betriebsverlust	X						
kpX_2390	Früherer Beruf	X						
kpX_3710	Früherer Beruf Angestellter - Differenzierung	X						
kpX_3720	Früherer Beruf Arbeiter - Differenzierung	X						
kpX_3730	Früherer Beruf Akademisch freier Beruf, Selbständiger - Differenzierung	X						
kpX_3740	Früherer Beruf Beamter - Differenzierung	X						
kpX_3750	Früherer Beruf - Beschäftigungssektor	X						
kpX_3760	Früherer Beruf – Wirtschaftssektor	X						
kpX_2541	Geburtsland	X						
kpX_3920	Geburtsland, Bundesland	X						
kpX_3930	Geburtsland, anderes Land	X						
kpX_2551	Alter Zuzug	X						
kpX_3940	Alter Zuzug, Ostdeutschland	X						
kpX_3950	Alter Zuzug, Westdeutschland	X						
kpX_3960	Alter Zuzug, Berlin	X						
kpX_2520	Deutsche Staatsbürgerschaft seit Geburt		X	[X]	[X]			
kpX_2571	Geburtsland, Mutter		X	[X]	[X]			
kpX_2571	Geburtsland, Vater		X	[X]	[X]			
kpX_2572	Geburtsland, Mutter, anderes Land		X	[X]	[X]			
kpX_2572	Geburtsland, Vater, anderes Land		X	[X]	[X]			
kpX_3980	Zuzug Mutter		X	[X]	[X]			
kpX_3990	Zuzug Vater		X	[X]	[X]			
kpX_3970	Erfassung von Migration in 3. Generation		X	[X]	[X]			
kpX_4120	Im HH gesprochene Sprache		X	[X]	[X]			
kpX_4130	Im HH gesprochene Sprache, andere Sprache		X	[X]	[X]			
kpX_4140	Identifikation von Aussiedlern		X	[X]	[X]			
kpX_4150	Identifikation von Asylbewerbern		X	[X]	[X]			
kpX_2301	Familienstand	X				X	[X]	[X]
kpX_2311	Partner vorhanden	X				X	[X]	X
kpX_2312	Partner im Haushalt	X				X	[X]	X
kpX_3890	Wahlabsicht Partner	X				X	[X]	X

Variablen- stamm	Variablenlabel	W1	W2	W3	W4	W5	W6	W7
kpX_3900	Wahlabsicht Partner, andere Partei	X				X	[X]	X
kpX_2391	Schulabschluss Partner					X	[X]	[X]
kpX_2400	Erwerbstätigkeit Partner					X	[X]	[X]
kpX_2410	Frühere Erwerbstätigkeit Partner					X	[X]	[X]
kpX_2420	Beruf, Partner					X	[X]	[X]
kpX_2430	Früherer Beruf, Partner					X	[X]	[X]
kpX_3700	Berufliche Bildung Partner					X	[X]	[X]
kpX_3770	Angestellter – Differenzierung, Partner					X	[X]	[X]
kpX_3780	Arbeiter – Differenzierung, Partner					X	[X]	[X]
kpX_3790	Akademisch freier Beruf, Selbständiger – Differenzierung, Partner					X	[X]	[X]
kpX_3800	Beamter – Differenzierung, Partner					X	[X]	[X]
kpX_3810	Beschäftigungssektor, Partner					X	[X]	[X]
kpX_3820	Wirtschaftssektor, Partner					X	[X]	[X]
kpX_3830	Früherer Beruf Angestellter – Differenzierung, Partner					X	[X]	[X]
kpX_3840	Früherer Beruf Arbeiter – Differenzierung, Partner					X	[X]	[X]
kpX_3850	Früherer Beruf Akademisch freier Beruf, Selbständiger – Differenzierung, Partner					X	[X]	[X]
kpX_3860	Früherer Beruf Beamter – Differenzierung, Partner					X	[X]	[X]
kpX_3870	Früherer Beruf Beschäftigungssektor, Partner					X	[X]	[X]
kpX_3880	Früherer Beruf Wirtschaftssektor, Partner					X	[X]	[X]
kpX_4200	Aktuelles Befinden			X	X		X	X

Erläuterung: „X“ steht für die jeweilige Befragungswelle. [X] bedeutet, dass nur die Teilnehmer des Wahlkampfpanels die Fragen des Moduls erhalten haben, die bei der Abfrage dieses Themenblocks in einer der vorangegangenen Wellen nicht teilgenommen haben (siehe hierzu auch Kapitel 5.8).

4.8 Paradata

Im veröffentlichten Datensatz sind zwei verschiedene Arten von Paradata enthalten. Die eine Gruppe von Variablen gibt Auskunft über den technischen Ablauf des Interviewprozesses. Diese werden von der Erhebungssoftware EFS von Global Park automatisch gespeichert (Siehe Tabelle 25). Die zweite Gruppe von Variablen enthält Informationen zur Umfrageaktivität der Panelisten und wurde von Respondi zur Verfügung gestellt (Siehe Tabelle 26).

Tabelle 25: Paradata zum technischen Interviewverlauf

Variablenname	Erläuterung
kpX_browser	Gibt den Namen des Browsers (User Agent) so an, wie er an den EFS Survey-Server übertragen wurde. Beispiele finden Sie z.B. unter http://de.wikipedia.org/wiki/User_Agent .
kpX_javascript	Enthält das Resultat der Überprüfung auf JavaScript, die optional bei Umfragebeginn ausgeführt werden kann (0=JavaScript abgeschaltet; sonst=Versionsnummer, z.B. 10 für JavaScript 1.0)
kpX_flash	Enthält das Resultat der Überprüfung auf das Adobe Flash-Plugin, die optional bei Umfragebeginn ausgeführt werden kann und insbesondere für die Flash-Fragetypen benötigt wird (0=Flash-Plugin nicht vorhanden; sonst=Versionsnummer, z.B. 800 für Flash-Plugin 8.0)
kpX_datetime	Datum und Uhrzeit des Umfragebeginns, d.h. des Aufrufs der ersten Fragebogenseite. Die Angaben zur Uhrzeit im Datensatz entsprechen der Greenwich Mean Time (GMT).
kpX_date_of_last_access	Datum und Uhrzeit des letzten Zugriffs auf die Umfrage. Die Angaben zur Uhrzeit im Datensatz entsprechen der Greenwich Mean Time (GMT).
lfdn	Jeder Teilnehmer des Fragebogens bekommt eine laufende Nummer.
lfdn09	Laufende Nummer aus dem Wahlkampfpanel 2009
kp1_quota	Enthält die ID der zugewiesenen Quote (d.h. die per Zuweisungsmodus ausgewählte Quote). Sie wird verwendet, um ungeeignete Teilnehmer auszuscreenen
kpX_lastpage	Gibt die letzte vom Teilnehmer abgeschickte Seite an. D.h. wenn der Teilnehmer die Umfrage abgebrochen hat, erfolgte der Abbruch auf der Folgeseite.
kpX_duration	Die Bearbeitungsdauer, d.h. die Zeit, die zwischen dem ersten Zugriff und dem letzten Zugriff des Probanden auf den Fragebogen abgelaufen ist. Unterbricht ein Proband den Fragebogen und nimmt die Bearbeitung später wieder auf (Dispositionscodes 23, 32), so wird kpX_duration= -99 gesetzt, da eine sinnvolle Berechnung nicht möglich ist.

Erläuterung: „X“ steht für die jeweilige Befragungswelle.

Die von Respondi angeforderten Paradata geben Aufschluss über den Eintritt in das Panel, den Rekrutierungsweg sowie das Umfrageverhalten der Panelisten a) in den letzten zwölf Monaten und b) in den letzten drei Monaten. Stichtag für die Datenabfrage der zeitlichen Informationen war der 21. September 2013.

Tabelle 26: Von Respondi erworbene Paradata

Variablenname	Erläuterung
p_enter_date	Eintrittsdatum ins Respondi-Panel
p_recruit	Rekrutierungsart
p_recruit1	Allgemeine Rekrutierungsart
p_numinv2	Anzahl der Umfrageeinladungen (letzten 12 Monate; Stichtag: 21.09.2013)
p_numcpl2	Anzahl der Umfrageteilnahmen (letzten 12 Monate; Stichtag: 21.09.2013)
p_numstr2	Anzahl begonnener Umfragen (letzten 12 Monate; Stichtag: 21.09.2013)
p_numinc2	Anzahl unterbrochener Umfragen (letzten 12 Monate; Stichtag: 21.09.2013)
p_numinv3	Anzahl der Umfrageeinladungen (letzten 3 Monate; Stichtag: 21.09.2013)
p_numcpl3	Anzahl der Umfrageteilnahmen (letzten 3 Monate; Stichtag: 21.09.2013)
p_numstr3	Anzahl begonnener Umfragen (letzten 3 Monate; Stichtag: 21.09.2013)
p_numinc3	Anzahl unterbrochener Umfragen (letzten 3 Monate; Stichtag: 21.09.2013)

4.9 Unveröffentlichte Variablen

Bei der Durchführung von Online-Umfragen werden Variablen erhoben oder an GESIS übermittelt, die zwar für die Durchführung der Online-Umfrage relevant sind, aber keine inhaltliche Aussagekraft haben. Diese Variablen werden nicht veröffentlicht, können aber bei Bedarf von GESIS (gles@gesis.org) bezogen werden. Ebenso können aus datenschutzrechtlichen Gründen einige erhobenen Variablen nicht zum freien Download zur Verfügung gestellt werden (wie z.B. die Postleitzahl). Bei diesen Variablen ist ein Bezug unter bestimmten Bedingungen möglich.

Tabelle 27: Übersicht über die nicht veröffentlichten Variablen und die Bezugsmöglichkeit

Variablenname	Erläuterung	Bezugsmöglichkeit
kpx_ats	Absoluter Zeitstempel	auf Anfrage erhältlich
kpx_page_history	Seitenhistorie	auf Anfrage erhältlich
kpx_lfdpagenr	Laufende Seitennummer	auf Anfrage erhältlich
kpx_invcode	Persönlicher Einladungscode für jeden Teilnehmer	auf Anfrage erhältlich
kpx_codelink	Persönlicher Link für den Befragten zur Umfrage. Besteht aus URL der Umfrage und persönlichem Einladungscode.	auf Anfrage erhältlich
kpx_2293s	Geburtsort (erster Buchstabe)	auf Anfrage erhältlich
kpx_2061	Postleitzahl	auf Anfrage erhältlich
kpx_xxxx_org bzw. _c1-2	Sämtliche Originalvariablen und Nachfragevariablen, die bei der Zusammenfassung der Nachfragvariablen, der Parteiversionen und der Dependent-Interviewing Variablen zur Sicherung der ursprünglichen Daten erstellt wurden bzw. nicht mehr benötigt werden	auf Anfrage erhältlich
kpX_2280_c1	Geschlecht	auf Anfrage erhältlich
kpX_2291_c1	Geburtsmonat	auf Anfrage erhältlich
kpX_2290_c1	Geburtsjahr	auf Anfrage erhältlich
kpX_2293s_c1	Geburtsort	auf Anfrage erhältlich
kpX_2280_c2	Geschlecht	auf Anfrage erhältlich
kpX_2291_c2	Geburtsmonat	auf Anfrage erhältlich
kpX_2290_c2	Geburtsjahr	auf Anfrage erhältlich
kpX_2293s_c2	Geburtsort	auf Anfrage erhältlich

kpX_2293_c3	Korrektur der Eingabe des Geburtsortes	auf Anfrage erhältlich
kpX_4280	Grund für falschen Verifizierungscode	auf Anfrage erhältlich

Stammdaten

s0	Geschlecht	auf Anfrage erhältlich
s1	Geburtstag	auf Anfrage erhältlich
s2	Geburtsmonat	auf Anfrage erhältlich
s3	Geburtsjahr	auf Anfrage erhältlich
s4	Familienstand	auf Anfrage erhältlich
s5	Schulische Ausbildung	auf Anfrage erhältlich
s6	Berufliche Ausbildung	auf Anfrage erhältlich
s7	Berufstätigkeit	auf Anfrage erhältlich
s8	Berufsbranche	auf Anfrage erhältlich
s9	Beruf	auf Anfrage erhältlich
s10	Abteilung im Beruf	auf Anfrage erhältlich
s11	Nettoeinkommen	auf Anfrage erhältlich
s12	Haushaltseinkommen	auf Anfrage erhältlich
sHH	Anzahl der Haushaltsmitglieder	auf Anfrage erhältlich
sHHK	Anzahl der Kinder im Haushalt	auf Anfrage erhältlich

Erläuterung: „X“ steht für die jeweilige Befragungswelle.

5 Datenaufbereitung

5.1 Vorbemerkung

Die Datenaufbereitung des Wahlkampf-Panels erfolgte in einem ersten Schritt wellenweise. Anschließend wurden die Datensätze der einzelnen Wellen in einen gemeinsamen Datensatz überführt. Alle Aufbereitungsschritte folgten der Richtlinie einer minimal-invasiven Bearbeitung. Fehler in den Daten wurden mit Flaggenvariablen gekennzeichnet. Die Aufbereitung wurde mittels Syntax durchgeführt und lässt sich somit in all ihren Schritten nachvollziehen.

5.2 Codierung der fehlenden Werte

Fehlende Werte wurden bei allen Variablen nach dem einheitlichen Codierschema der GLES vergeben, welches im SPSS-Format negative Werte im Bereich -71 bis -99 und in Stata-Codes von .a bis .p vorsieht. Den Datensätzen liegt ein Do-File bei, anhand dessen die SPSS Missing Codes im Stata-Datensatz in die vorgesehenen Stata-Codes transformiert werden können.

Tabelle 28: Kurzübersicht über die Missing-Value-Codes der GLES

SPSS-Code	Stata-Code	Label
-71	.p	nicht bekannt
-84	.k	keine Erst-/Zweitstimme abgeben
-85	.j	nicht wählen
-92	.h	Fehler in Daten
-93	.g	Interview abgebrochen
-95	.e	nicht teilgenommen
-97	.c	trifft nicht zu
-98	.b	weiß nicht
-99	.a	keine Angabe

5.3 Codierung der Parteien

Zur besseren Vergleichbarkeit der GLES Komponenten untereinander erfolgt die Codierung der politischen Parteien bei allen Fragen, die eine geschlossene oder offene Abfrage zu politischen Parteien beinhalteten, nach einem einheitlichen Codierschema. Das Codierschema kann auf den Seiten der GLES bei GESIS abgerufen werden (<http://www.gesis.org/gles>). Alle Angaben zu politischen Parteien werden in zwei Variablen veröffentlicht, die mit Version A und Version B gekennzeichnet sind. Version A enthält nur die im Bundestag vertretenen Parteien der Legislaturperiode 2009 bis 2013 sowie die Kategorie „andere Partei“. Version B unterscheidet sich von Version A in einer detaillierten Ausweisung auch kleinerer Parteien (AfD, Die Grauen, Freie Wähler, NPD, oedp, Piraten, REP).¹³ Diese Parteien konnten von den Befragten im Rahmen einer Nachfrage nur dann

¹³ Ursprünglich war in der Liste der Antwortoptionen auch die Partei „DVU“ enthalten. Da sich die DVU der NPD angeschlossen hat, wurden die DVU-Nennungen der NPD hinzugefügt.

ausgewählt werden, wenn sie zuvor auf die Frage nach einer Partei – zur Auswahl standen zunächst nur die im Bundestag vertretenen Parteien – auf „andere Partei“ klickten.

Einige Befragte haben bei der Nachfrage, bei denen sie zwischen den kleineren Parteien auswählen sollten, keine Angabe gemacht oder die Umfrage abgebrochen. Diese Personen erhielten sowohl in Version A als auch in Version B den Code -99 (keine Angabe) bzw. -93 (Umfrage abgebrochen), obwohl in diesen Fällen eigentlich eine partielle inhaltliche Information vorlag: Sie hatten auf der ersten Seite eine „andere Partei“ angeklickt. Falls man diese ursprüngliche Information wieder herstellen möchte, sind in Tabelle 29 die laufenden Nummern der Befragten im Datensatz dokumentiert (Variable: „lfdn“), bei denen die hier beschriebene Veränderung vorgenommen wurde.

Da bei der Frage nach der Stärke, Dauer und Art der Parteiidentifikation („kpX_2100“, „kpX_2110“, „kpX_2121“) nur danach gefiltert wurde, ob sie bei der Parteiidentifikation („kpX_2090“) eine Angabe gemacht haben, tritt nun für die Befragten, bei denen der Wert in -99 umkodiert wurde der Fall auf, dass sie Werte bei „kpX_2100“, „kpX_2110“ und „kpX_2121“ aufweisen. Um die Filterung konsistent zu gestalten, werden die Angaben dieser Fälle in -97 „trifft nicht zu“ umkodiert. Die betroffenen Fälle lassen sich Tabelle 29 in den Zeilen zur Parteiidentifikation entnehmen.

Tabelle 29: Befragte, bei denen aufgrund von Antwortverweigerung oder Abbruch nachträgliche Änderungen bei Parteivariablen vorgenommen wurden

Variable	Label	Laufende Nummer der Befragten (lfdn)
Welle 1		
kp1_190ba-b	Beabsichtigte Stimmabgabe, Zweistimme	1278, 2686, 3465, 3745, 3761, 4519
kp1_850a-b	Lösungskompetenz: Wichtigstes Problem	1845, 4424, 4988
kp1_870a-b	Lösungskompetenz: Zweitwichtigstes Problem	104
kp1_350aa-b	Recall vorangegangene BTW Erststimme	2598
kp1_350ba-b	Recall vorangegangene BTW Zweitstimme	718, 973, 1513, 4323
kp1_2090a-b	Parteiidentifikation	95, 261, 326
Welle 2		
kp2_190ba-b	Beabsichtigte Stimmabgabe: Zweitstimme	190, 1227, 1311, 1980, 2021, 2183, 2536, 2892, 3758,
kp2_850a-b	Lösungskompetenz: Wichtigstes Problem	1778, 2640, 4582
kp2_870a-b	Lösungskompetenz: Wichtigstes Problem	4630, 5125
Welle3		
kp3_190ba-b	Beabsichtigte Stimmabgabe: Zweitstimme	255, 1227, 1289, 3136
kp3_850a-b	Lösungskompetenz: Wichtigstes Problem	1090, 3251
kp3_870a-b	Lösungskompetenz: Zweitwichtigstes Problem	1138, 3251, 4873
kp3_2090a-b	Parteiidentifikation	3753
Welle 4		
kp4_190ba-b	Beabsichtigte Stimmabgabe: Zweitstimme	2519, 2536, 3787, 4308
kp4_850a-b	Lösungskompetenz: Wichtigstes Problem	4346
kp4_870a-b	Lösungskompetenz: Zweitwichtigstes Problem	3251

Welle 5		
kp5_190ba-b	Beabsichtigte Stimmabgabe: Zweitstimme	707, 2825
kp5_850a-b	Lösungskompetenz. Wichtigstes Problem	907
kp5_870a-b	Lösungskompetenz. Zweitwichtigstes Problem	2341, 4364
kp5_2090a-b	Parteiidentifikation	1032
kp5_3890a-b	Wahlabsicht Partner	4381
Welle 6		
kp6_190ba-b	Beabsichtigte Stimmabgabe: Zweitstimme	89, 1593, 2023, 2091, 2175, 3096, 3286, 3473, 4330
kp6_2770ba-b	Hypothetische Zweitstimme nach Briefwahl	1311
kp6_870a-b	Lösungskompetenz: Zweitwichtigstes Problem	1140, 2255, 2661
Welle 7		
kp7_200ba-b	Tatsächliches Wahlverhalten nach BTW: Zweitstimme	4423, 4906
kp7_870a-b	Lösungskompetenz: Zweitwichtigstes Problem	284, 1182, 3595
kp7_2090a-b	Parteiidentifikation	4982, 5038
kp7_3890	Wahlabsicht Partner	568, 2656

5.4 Codierung von offenen Fragen

Offene Fragen (Fragen ohne vorgegebene Antwortmöglichkeiten) wurden durch das Umfrageinstitut BACES codiert. Die Codierung erfolgte unter Verwendung von Codierschemata, die von der GLES-Projektgruppe entwickelt wurden. Dies betrifft die Variablen „Gründe für die Wahlentscheidung“ („kpx_260“), die Fragen nach dem wichtigsten und zweitwichtigsten Problem in Deutschland (Agendafragen, „kpx_840“, „kpx_860“), die Frage nach der am häufigsten genutzten politischen Internetseite („kp1_1651s“) sowie der Grund für die Umfragebeteiligung („kp1_4230s“).

Mit Abschluss der Codierarbeiten stehen nun auch die Codierungen der offenen Angaben zur Verfügung. Sie sind durch das Suffix _c1 gekennzeichnet, wobei die Zahl 1 die Position der Nennung angibt. Es wurden bis zu fünf Nennungen codiert, sodass Suffix von _c1 (z.B.: „kp1_4230_c1“) bis _c5 vergeben wurden. Die offenen Angaben selbst bleiben weiterhin im Datensatz enthalten. Diese wurden vor der Veröffentlichung auf datenschutz sensible Inhalte geprüft und gegebenenfalls zum Schutz der Befragten zensiert.

In Folge der Codierung wurden Angaben, die keine inhaltliche Mitteilung darstellen mit -99 kodiert. Diese Angaben wurden von der Befragungssoftware jedoch natürlich als Angabe gewertet, sodass diese Fälle auch jeweils Angaben zu weiterführenden Fragen wie der Lösungskompetenz „kpX_850a,b“ gemacht haben. Es bleibt dem Nutzer überlassen, ob er diese Angaben für seine Analyse verwendet oder sie ausschließt.

Um mit den offenen Antworten zu arbeiten empfehlen wir die Nutzung des SPSS-Datensatzes, da aufgrund von Programmrestriktionen alle offenen Angaben im Stata-Datensatz bei 244 Zeichen abgeschnitten werden.

Die entsprechenden Codierschemata können auf den Seiten der GLES bei GESIS abgerufen werden (<http://www.gesis.org/gles>).

5.5 Markierung der Wissensfragen

Anhand von mehreren Fragen wurde im Wahlkampfpanel das politische und wirtschaftliche Wissen der Umfrage-Teilnehmer getestet. Die richtige Antwort auf eine Wissensfragen wurden im entsprechenden Wertelabel mit einem Sternchen (*) versehen:

Tabelle 30: Wissensfragen mit markierten Antworten

Variable	Label	Welle
kpx_110	Politisches Wissen: Erst- /Zweitstimme	2, 4, 6, 7
kpx_130	Politisches Wissen: Wahlrecht D	2, 4, 6, 7
kpx_3430a-f	Politisches Wissen: Zuordnung Politiker/Parteien	1, 3, 7
kpx_3430j-o	Politisches Wissen: Zuordnung Politiker/Parteien	2, 4, 6
kpx_3440	Wirtschaftliches Wissen: Arbeitslosenquote	1, 4, 7

Eine Ausnahme ist die Frage nach der 5%-Hürde („kpx_090“), die in den Wellen zwei, fünf, sechs und sieben gestellt wurde. Da die Frage offen gestellt wurde, wurden die Antworten in eine neue Dummyvariable „kpx_090_v1“ überführt, die angibt, ob der Befragte die richtige Antwort gegeben hat.

5.6 Umgang mit Nachfragen bei den Frageformaten „Checkbox“ und „Schieberegler“

Bei bestimmten Frageformaten war es notwendig, im Falle einer nicht erfolgten Antwort eine Nachfrage zu stellen, um die Bedeutung der nicht erfolgten Antwort zu klären. Dies betraf die beiden Frageformate „Checkbox“ und „Schieberegler“, die jeweils einen voreingestellten Wert (meistens von Null) aufwiesen. Wenn die Befragten bei diesen Frageformaten weiterklickten, ohne eine Veränderung auf der Seite vorgenommen zu haben, hätte man in diesen Fällen nicht zwischen einer möglichen substanziellen Antwort, wie z.B. „Null Tage“, oder einer nicht-substanziellen Antwortverweigerung unterscheiden können. Daher wurde nach dem Grund der nicht erfolgten Antwort gefragt – ob sie entweder keine Angaben machen wollten oder ob der Wert von „Null“ eine substanziell zu interpretierende Antwort darstellt. Zusätzlich wurde diesen Befragten die Möglichkeit gegeben, ihre Antwort zu korrigieren und Werte anzugeben, die von Null abweichen. Im veröffentlichten Datensatz wurden die sofort getätigten und die durch Nachfrage gewonnenen Antworten in der Stammvariable zusammengefasst. Ob die Antwort erst auf Nachfrage gegeben wurde, lässt sich über eine Flagvariable herausfinden, die aus dem Variablenstammmamen und dem Suffix „flag“ besteht. Dies betrifft die in Tabelle 31 aufgeführten Variablen.

Tabelle 31: Variablen, deren Nachfragen zusammengefasst wurden

Variablenname	Variablenlabel	Frageformat	Welle
kp1_1721a-g	Printmediennutzung, politisch gewöhnlich	Schieberegler	1
kp1_1741a-e	Fernsehnutzung, Nachrichten gewöhnlich	Schieberegler	1
kp1_1931a-e	Politische Gespräche, allgemein	Schieberegler	1
kpX_421aa-lf	Parteikontakte I und II	Checkbox	2-7
kpX_1661a-h	Printmediennutzung, politisch aktuell	Schieberegler	2-7
kpX_1681a-f	Fernsehnutzung, Nachrichten aktuell	Schieberegler	2-7
kpX_1932a-f	Politische Gespräche, aktuell	Schieberegler	2-7

Erläuterung: „X“ steht für die jeweilige Befragungswelle.

5.7 Zusammenfassung von Zeitangaben

Im Wahlkampfpanel wurde die mittlere Dauer der täglichen Internetnutzung anhand zweier Antwortlisten in Stunden und Minuten erhoben. Auch bei der Erfassung der Arbeitslosigkeit in den letzten 10 Jahren wurde die Antwort anhand zweier Listen in sowohl Jahren und Monaten erfasst. In beiden Fällen wurden die zwei Zeitangaben zu einer Zeitvariablen zusammengefasst: Die Dauer der täglichen Internetnutzung wird in Minuten („kp1_3450“), die Dauer der Arbeitslosigkeit in Monaten ausgewiesen („kp1_2370“).

5.8 Zusammenfassung der Dependent Interviewing-Variablen

Der Begriff „Dependent Interviewing“ (DI) bezeichnet in dieser Studie eine spezielle Form der Filterführung, bei der das Erhalten einer Frage oder ihr Wortlaut von Angaben aus *vorangegangenen* Wellen abhängt. Im Wahlkampfpanel wurde diese Technik in zwei unterschiedlichen Kontexten eingesetzt:

1. *Um Doppelerhebungen zu vermeiden.* Ein passendes Beispiel ist die Briefwahl: Sofern Befragte in einer Welle angegeben hatten, dass sie bereits Briefwahl gemacht hatten, wurde ihnen in den folgenden Wellen die Frage nach der Wahlteilnahme nicht mehr gestellt. Stattdessen erhielten sie Fragen, die auf Briefwähler zugeschnitten waren.
2. *Um eine Messung nachzuholen.* Bei vergleichsweise stabilen Merkmalen (wie z. B. sozialstrukturelle Merkmale, Wertorientierungen, Erinnerung an vergangenes Verhalten) erschien es zweckmäßig, diese im Wahlkampfpanel nur ein einziges Mal zu erheben – dafür jedoch möglichst bei *allen* Befragten. Da es aufgrund des begrenzten Fragenbogenplatzes nicht möglich war, solche stabilen Merkmale allesamt in der ersten Welle zu erheben (an der alle Befragten teilnahmen), wurde ihre Abfrage teilweise auf spätere Wellen verschoben. Wenn Befragte an der entsprechenden Welle nicht teilgenommen hatten, wurde bei ihnen die Messung in späteren Wellen nachgeholt. Auf diese Weise gelang es, trotz des Panelausfalls für fast alle Befragten Messwerte zu erhalten.

Die hier beschriebene Zusammenfassung trifft nur auf jene DI-Variablen zu, die aus einer nachgeholten Messung (Punkt 2) resultierten. Um das Arbeiten mit den nachgeholten Messwerten zu vereinfachen, wurden diese nachträglich in der Variablen aus jener Welle abgespeichert, in der das entsprechende Merkmal zum ersten Mal abgefragt wurde.

Ein Beispiel: In der zweiten Welle wurden Einstellungen zur Steuergerechtigkeit erhoben („kp2_3400a-d“). Nahm ein Befragter an der zweiten Welle nicht teil, so erhielt er die Fragen in der vierten Welle („kp4_3400a-d“), ansonsten in der sechsten Welle („kp6_3400a-d“). Im Zuge der Datenaufbereitung wurden die Messwerte aus der vierten und der sechsten Welle in die Variablen „kp2_3400a-d“ verschoben. Die Variablen „kp4_3400a-d“ und „kp6_3400a-d“ wurden danach gelöscht. Über eine Flag-Variable im Datensatz („kp2_3400flag“) kann nachvollzogen werden, aus welcher Welle die jeweiligen Angaben stammen.¹⁴¹⁵

¹⁴ Eine Ausnahme stellen die Variablen „kpx_2301“, „kpx_2311“ und „kpx_2312“ dar. Sie wurden jeweils einmal in der ersten Welle abgefragt und dann erneut in Welle fünf, da sichergestellt werden musste, ob in der Zwischenzeit eine Änderung im Familienstand eingetreten ist. Eine Nichtteilnahme in Welle fünf führte dazu, dass die Panelisten die Frage „kpx_2301“ in Welle sechs und sieben erneut bekommen konnten. Entsprechend wurden die Informationen aus Welle sechs und sieben in fünf zusammengefasst und die gegebenen Antworten in Welle eins beibehalten. Aus demselben Grund – um mögliche Veränderungen im Beziehungsstatus zu berücksichtigen – wurden die Fragen „kpx_2311“ und „kpx_2312“ in Welle sieben auch denjenigen erneut gezeigt, die bereits in Welle fünf eine Antwort gegeben hatten. Zusammengefasst wurden nur die Informationen derjenigen Befragten, die an Welle fünf und sechs nicht teilnahmen. Entsprechend verbleiben auch „kp7_2311“ und „kp7_2312“ im Datensatz.

Es ist somit in Ausnahmefällen möglich, dass eine Variable aus einer bestimmten Welle inhaltlich interpretierbare Messwerte von Befragten enthält, die an dieser Welle nicht teilgenommen haben. Tabelle 32 liefert eine Auflistung der betroffenen Variablen.

Tabelle 32: Liste der Variablen, bei denen eine nachträgliche Zusammenfassung nachgeholter Messwerte erfolgte

Variablenname	Variablenlabel	Erstes Auftreten in Welle...	Nachgeholt in Welle...
kpx_3400a-d	Gerechtigkeit, Steuern	2	4, 6
kpx_2520	Deutsche Staatsbürgerschaft seit Geburt	2	3, 4
kpx_2570a	Geburtsland, Mutter, anderes Land	2	3, 4
kpx_2570b	Geburtsland, Vater, anderes Land	2	3, 4
kpx_2571a,b	Geburtsland Mutter/Vater	2	3, 4
kpx_3970	Erfassung von Migration in 3. Generation	2	3, 4
kpx_3980	Zuzug Mutter	2	3, 4
kpx_3990	Zuzug Vater	2	3, 4
kpx_4120	Im HH gesprochene Sprache	2	3, 4
kpx_4130	Im HH gesprochene Sprache, andere Sprache	2	3, 4
kpx_4140	Identifikation von Aussiedlern	2	3, 4
kpx_4150	Identifikation von Asylbewerbern	2	3, 4
kpx_3370a-d	Gerechtigkeit, Einkommen	3	5, 7
kpx_2081a-d	Inglehart-Items	4	5, 6
kpx_1800	TV-Duell: Rezeption	5	6, 7
kpx_3570	TV-Duell: Aufmerksamkeit	5	6, 7
kpx_3580a-l	TV-Duell: Spezifische Debattenleistung Merkel/Steinbrück	5	6, 7
kpx_3600a-d	TV-Duell: Aussagen Merkel/Steinbrück	5	6, 7
kpx_2301	Familienstand	1/5*	6, 7
kpx_2311	Partner vorhanden	1/5*	6, (7)**
kpx_2312	Partner im Haushalt	1/5*	6, (7)**
kpx_2391	Schulabschluss Partner	5	6, 7
kpx_2400	Erwerbstätigkeit Partner	5	6, 7
kpx_2410	Frühere Erwerbstätigkeit Partner	5	6, 7
kpx_2420	Beruf, Partner	5	6, 7
kpx_2430	Früherer Beruf, Partner	5	6, 7
kpx_3700	Berufliche Bildung Partner	5	6, 7

¹⁵ Um die DI-Filterführungen nicht komplex werden zu lassen, wurde in Kauf genommen, dass in Ausnahmefällen auch eine Doppelerhebung der entsprechenden Merkmale erfolgen konnte (diese Ausnahmen waren „Abbrecher“ aus vergangenen Wellen, bei denen nicht klar war, ob sie vor oder nach der Erhebung des entsprechenden Merkmals abgebrochen hatten). In diesem Fall wurde bei der Zusammenfassung der Variablen die *letzte* Antwort des Befragten verwendet.

Variablenname	Variablenlabel	Erstes Auftreten in Welle...	Nachgeholt in Welle...
kpx_3770	Angestellter – Differenzierung, Partner	5	6, 7
kpx_3780	Arbeiter – Differenzierung, Partner	5	6, 7
kpx_3790	Akademisch freier Beruf, Selbständiger – Differenzierung, Partner	5	6, 7
kpx_3800	Beamter – Differenzierung, Partner	5	6, 7
kpx_3810	Beschäftigungssektor, Partner	5	6, 7
kpx_3820	Wirtschaftssektor, Partner	5	6, 7
kpx_3830	Früherer Beruf Angestellter – Differenzierung, Partner	5	6, 7
kpx_3840	Früherer Beruf Arbeiter – Differenzierung, Partner	5	6, 7
kpx_3860	Früherer Beruf Beamter – Differenzierung, Partner	5	6, 7
kpx_3850	Früherer Beruf Akademisch freier Beruf, Selbständiger – Differenzierung, Partner	5	6, 7
kpx_3870	Früherer Beruf Beschäftigungssektor, Partner	5	6, 7
kpx_3880	Früherer Beruf Wirtschaftssektor, Partner	5	6, 7
kpx_3180	Wahlbeteiligung, Bayern	6	7
kpx_3190a,b	Tatsächliche Stimmabgabe, Bayern	6	7
kpx_3200	Zeitpunkt der Wahlentscheidung, Bayern	6	7
kpx_3210	Kenntnis Wahlergebnis, Bayern	6	7

* Nachfrage aus Welle 6/7 in 5 zusammengefasst.

** Information aus Welle 7 nur dann in Welle 5 zusammengefasst, wenn an Welle 5 und 6 nicht teilgenommen wurde.

6 Datenqualität

6.1 Vorbemerkung

Da bei Online-Umfragen keine Interviewer anwesend sind, ist die Interviewsituation durch einen hohen Grad an Anonymität gekennzeichnet.¹⁶ Anonymität kann ein Vorteil sein, wenn es um die Erhebung von sozial unerwünschten Einstellungen oder Verhaltensweisen geht, da Befragte in selbst-administrierten Umfragen in der Regel ehrlichere Antworten geben (vgl. Joinson, 1999: 435; Mühlenfeld, 2004; Taddicken, 2009). Sie hat jedoch den Nachteil, dass niemand die Sorgfalt und Ernsthaftigkeit im Antwortverhalten der Befragten kontrolliert (Gräf/Heidingsfelder, 1999: 120). Dadurch entsteht ein Anreiz für wenig motivierte Befragte, die Fragen nur oberflächlich oder unpräzise zu beantworten („satisficing“, vgl. Kroshnick, 1991, 1999). Sie können sich auf diese Weise die Incentives sichern, ohne sich mit der Umfrage Mühe zu machen. Satisficing kann sich in sehr unterschiedlichen Weisen im Antwortverhalten äußern: etwa in einer sehr schnellen Beantwortungszeit, in einem hohen Anteil von Antwortverweigerung, zufälligen Antworten oder nicht-differenzierenden Antworten („straight-lining“). Darüber hinaus ist bekannt, dass es in Online-Access-Panels einen geringen Anteil von Befragten gibt, der absichtlich falsche Angaben macht (Downes-Le Guin et al., 2006; Baker/Downes-Le Guin, 2007).

Trotz zahlreicher Maßnahmen zur Förderung guter Datenqualität ist das Wahlkampfpanel 2013 nicht frei von diesen Problemen. Gleichzeitig sollte jedoch berücksichtigt werden, dass auch bei telefonisch oder persönlich durchgeführten Studien Datenqualitätsprobleme auftreten (wenngleich anderer Art, wie z. B. Schwierigkeiten mit dem Umgang nicht visualisierter Antwortskalen bei CATI-Umfragen), die jedoch in den Daten weniger auffallen und daher auch seltener thematisiert und zu Bewusstsein gebracht werden. Der vergleichsweise schlechte Ruf von Online-Studien rührt zum Teil daher, dass aufgrund des sehr stark technisierten Ablaufs der Datenerhebung zahlreiche Messwerte anfallen, die das Ausmaß fehlender Datenqualität erst transparent machen (z. B. durch die automatische Messung von Antwortzeiten).

Diese Transparenz ist ein großer Vorteil von Online-Studien: Zwar mag der Anteil der problematischen Daten in Online-Studien größer sein als in persönlich durchgeführten Interviews. Gleichzeitig schärft die Beobachtung von schlechten Daten das Bewusstsein, dass das Problem existiert. Sie zwingt die Forschung auf diese Weise, sich stärker mit diesem Thema zu befassen. Darüber hinaus können Datenqualitäts-Messwerte nützlich sein, um die Sensitivität von Analyse-Ergebnissen im Hinblick auf unterschiedliche Grade der Datenqualität zu testen.

Das Wahlkampfpanel 2013 hält eine große Menge von verschiedenen Indikatoren bereit, die zur Evaluation von Datenqualität eingesetzt werden können (und sollten). Es wird empfohlen, bei der Analyse von Wahlkampfpanel-Daten zu testen, welche Auswirkungen der Ein- bzw. Ausschluss von Daten mit unterschiedlicher Qualität auf die Ergebnisse hat. Der Nutzer sollte auf der Grundlage eigener Erfahrung mit den Qualitätsindikatoren selbst darüber entscheiden (und dies in Forschungsarbeiten adäquat ausweisen), ab welchen Schwellenwerten ein Ausschluss der Befragten sinnvoll erscheint. Eine ausführliche Beschreibung dieser Indikatoren erfolgt demnächst in einem gesonderten Bericht. Eine Kurzzusammenfassung der Indikatoren liefert der Abschnitt 6.2.

Datenqualitätsprobleme in Online-Studien resultieren nicht nur aufgrund mangelnder Motivation der Befragten, sondern auch – wenngleich in selteneren Fällen – durch technische Prob-

¹⁶ Dieser Abschnitt orientiert sich stark an den Ausführungen von Plischke (2014: 197-198).

leme. Filterfehler werden in Abschnitt 6.3 beschrieben. In Abschnitt 6.4 erfolgt eine Auflistung von weiteren Problemen, auf die uns die Befragten aufmerksam gemacht haben.

6.2 Qualitätsindikatoren

6.2.1 Selbsteinschätzung der Datenqualität durch die Befragten

Am Ende jeder Umfrage wurden die Teilnehmer des Wahlkampfpanels gefragt, wie aufmerksam („kpX_4210“) und wie sorgfältig („kpX_4220“) sie die Fragen beantwortet haben. Obwohl zu bezweifeln ist, dass diese Fragen immer wahrheitsgemäß beantwortet wurden, gaben einige wenige Befragte fehlende Sorgfalt und Aufmerksamkeit zu. Ersten Analysen zufolge weisen diese Befragte auch bei objektiveren Datenqualitäts-Indikatoren schlechtere Werte auf.

6.2.2 Antwortzeiten

Wenn Befragte für die Beantwortung der Umfrage nur sehr wenig Zeit benötigen, ist dies nicht notwendigerweise ein Indiz für mangelnde Datenqualität. Es gibt viele Personen, die Fragetexte sehr schnell lesen und aufgrund ihrer vorgefestigten Meinungen auch schnell beantworten können. Gleichzeitig ist jedoch bekannt, dass es bei allen Online-Umfragen auch solche Befragte gibt, die sich einfach nur durch die Umfrage durchklicken, ohne die Fragen und Antworten überhaupt gelesen zu haben. Diese Personen weisen in der Regel sehr geringe Antwortzeiten auf.

In der Fachliteratur gibt es keine etablierten Standards für die Identifikation von „zu schnellen“ Antworten. In der Regel beziehen die Maße für deren Identifikation den Median bzw. Mittelwert der Verteilung und die Streuung mit ein und wählen auf dieser Basis ein Abschneidekriterium, das nicht unterschritten werden darf (Mayerl/Urban 2008, 58ff). Diese Befragten werden dann entweder aus dem Datensatz ausgeschlossen oder durch Markervariablen gekennzeichnet.

Im Wahlkampfpanel-Datensatz sind drei Indikatoren für schnelle Antwortzeiten enthalten, die unterschiedliche Aspekte messen, aber empirisch stark miteinander korrelieren:

- Der Indikator von Roßmann (2010; Variable „kpX_speederindex“) bezieht sowohl die Antwortdauer der Befragten auf allen Bildschirmseiten der Befragung als auch die Gesamtdauer je Befragtem ein. Diese Indexwerte nehmen einen Wertebereich zwischen 0 und 2 ein. Ein Indexwert von 1 zeigt eine mittlere Antwortgeschwindigkeit an, während Werte gegen 0 im Mittel sehr schnelle und Werte gegen 2 im Mittel sehr langsame Antwortzeiten indizieren.
- Die Variable „kpX_mtime“ enthält die durchschnittliche Dauer (in Sekunden) für die Beantwortung einer Umfrageseite. Bei der Berechnung dieses Mittelwertes wurden Verweildauern von mehr als 20 Sekunden pro Seite auf 20 Sekunden getrimmt, um sicherzustellen, dass keine Ausreißer in die Mittelwerte einfließen.
- Die Variable „kpX_mtimex“ ist eine leicht modifizierte Variante von „kpX_mtime“. Sie wurde erstellt, um mittlere individuelle Antwortzeiten zwischen den Wellen des Wahlkampfpanels vergleichbar zu machen. Hierfür wurden die individuellen Beantwortungszeiten pro Umfrageseite um die durchschnittliche Beantwortungszeit für diese Seite bereinigt, um den Umstand zu berücksichtigen, dass verschiedene Umfrageseiten unterschiedlich viel Aufwand erforderten. Anschließend wurde für jedes Individuum über alle bereinigten Bearbeitungszeiten der Mittelwert gebildet. Interpretationsbeispiel: Ein Wert von -2 zeigt an, dass eine Person pro Seite um 2 Sekunden schneller geantwortet hat als der durchschnittliche Befragte.

6.2.3 Straightlining

Unter „Straightlining“ wird ein spezielles Antwortmuster bei der Bearbeitung von Matrixfragen verstanden, das in der Regel auf eine nicht sorgfältige Bearbeitung des Fragebogens schließen lässt. Straightlining ist dadurch definiert, dass Befragte in der Item-Batterie stets den gleichen Wert innerhalb einer Spalte anklicken (meistens die Mittelkategorie), so dass die Antworten eine senkrechte Linie repräsentieren. Natürlich kann ein solches Muster auch inhaltlich begründet sein, wenn die „wahren“ Antworten zufälligerweise tatsächlich auf einer Linie liegen. Doch wenn Straightlining bei mehreren Matrixfragen wiederholt auftritt, ist dies wahrscheinlich eher auf eine nicht sorgfältige Durchführung des Interviews zurückzuführen.

Für die Berechnung eines Straightlining-Index wurden alle Matrixfragen mit mindestens drei Zeilen herangezogen. Anschließend wurde für jede Person bestimmt, bei wie viel Prozent aller Matrixfragen Straightlining vorliegt. Diese Prozentzahlen sind in der Variablen „kpX_strl“ enthalten.

6.2.4 Item Non-Response

Item Non-Response tritt auf, wenn Befragte eine Antwort auf eine Frage verweigern (Antwortcode -99). Ein hoher Anteil an Item Non-Response kann darauf hindeuten, dass sich die Befragten nur durch den Fragebogen durchgeklickt haben, ohne Frage- und Antwortstimuli beachtet zu haben. Die Variable „kpX_itnrp“ enthält den Anteil aller Fragen, die die Befragten in einer Welle erhalten haben, bei denen die Antwort verweigert wurde.

6.2.5 Weiß-nicht-Antworten

In politischen Umfragen deuten hohe Anteile an Weiß-nicht-Antworten in der Regel auf einen geringen politischen Wissensstand hin. Allerdings werden Weiß-nicht-Antworten in der Literatur auch als Indikator für Satisficing angesehen (Krosnick 1991). Aus diesem Grund wurde ein Indikator für das Ausmaß an Weiß-nicht-Antworten in einer Welle gebildet: Die Variable „kpX_dokn“ enthält den Anteil aller Weiß-nicht-Antworten, gemessen an der Anzahl aller Weiß-nicht-Optionen, die der Befragte in einer Welle erhalten hat.

6.2.6 Kontrollitems

„Satisficing“ äußert sich in Online-Umfragen oftmals dadurch, dass Personen wahllos durch die Antwortkategorien klicken oder – im speziellen Fall von Matrixfragen – immer die Antwortoption aus derselben Spalte wählen („straightlining“), ohne die Fragen richtig gelesen zu haben. Um ein solches Antwortverhalten zu identifizieren, wurde pro Welle mindestens einmal ein „Kontrollitem“ in eine Matrixfrage eingebaut. Bei diesem Item wurden die Personen nicht um ihre politische Meinung gebeten, sondern dazu angewiesen, eine bestimmte Antwortkategorie anzuklicken (vgl. Abbildung 4). „Falsche“ Antworten sind nicht notwendigerweise ein Zeichen für schlechte Datenqualität; Befragte können die falsche Angabe auch absichtlich gemacht haben (z. B. um gegen die Kontrolle zu protestieren). Dennoch haben vorläufige Analysen bereits gezeigt, dass falsche Antworten beim Kontrollitem sehr stark mit Straightlining und schnellen Antworten assoziiert sind. Die Kontrollitems weisen im Datensatz jeweils das Suffix „q“ auf (z. B. „kp1_050q“).

Abbildung 4: Beispiel für eine Kontrollfrage (Screenshot aus Welle 1)

	stimme überhaupt nicht zu	stimme eher nicht zu	teils/ teils	stimme eher zu	stimme voll und ganz zu
Die Politiker kümmern sich darum, was einfache Leute denken.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Die Politiker bemühen sich um einen engen Kontakt zur Bevölkerung.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich traue mir zu, mich an einem Gespräch über politische Fragen aktiv zu beteiligen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wichtige politische Fragen kann ich gut verstehen und einschätzen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bitte klicken Sie hier zur Überprüfung der Funktionsweise unseres Fragebogens „stimme eher zu“ an.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
In der Demokratie ist es die Pflicht jedes Bürgers, sich regelmäßig an Wahlen zu beteiligen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Anmerkung: Rote Umrandung dient der Hervorhebung und wurde nachträglich hinzugefügt.

6.2.7 Ein kombinierter Index

Schließlich wurde ein Gesamtqualitätsindikator (kpX_qual) gebildet, in dem einige der oben aufgeführten Einzelindikatoren enthalten sind. Dieser Indikator verfügt über einen theoretischen Wertebereich von 0 bis 1 und gibt die Wahrscheinlichkeit an, mit der Befragte bei einem Kontrollitem (siehe 6.2.6) eine falsche Antwort gegeben haben. Die Antwortqualität ist demnach umso geringer einzuschätzen, je höher der Index-Wert ausfällt.

Zur Berechnung des Index wurde eine Fixed-Effects-Panelregression gerechnet. Als abhängige Variable fungierten die Antworten bei den Kontrollitems (0=richtige Antwort; 1=falsche Antwort). Als potentielle Erklärungsvariablen wurden alle die in 6.2.1 bis 6.2.5 beschriebenen Qualitätsindikatoren getestet. Beibehalten wurde jedoch am Ende ein Modell, bei dem nur die Bearbeitungszeiten (dritter Index in 6.2.2) und der Straightlining-Indikator (6.2.3) enthalten waren sowie eine komplexe, nicht-lineare Interaktionen zwischen diesen beiden Indikatoren. Schließlich wurden auf der Grundlage dieses Modells die vorhergesagten Wahrscheinlichkeiten berechnet, deren Werte den kombinierten Index bilden. Eine ausführliche Beschreibung dieses Index erfolgt in einem separaten Bericht.

6.2.8 Ein Abweichungsindex

Der Index „kpX_dev“ zeigt an, wie stark sich die Antworten eines Teilnehmers in einer Welle von seinen Antworten in allen anderen Wellen unterscheiden. Er beruht auf den Antworten zu allen Fragen, die in den sieben Wellen des Wahlkampfpanels mindestens zweimal gestellt wurden.

Hintergrund dieses Index ist die Problematik, dass es sich bei Online-Umfragen nur sehr schwer feststellen lässt, ob tatsächlich die ausgewählte Zielperson an der Befragung teilgenommen hat (siehe Abschnitt 3.4). Anhand des Index sollen mögliche Nicht-Zielpersonen identifiziert werden: Je höher der Indexwert, desto wahrscheinlicher ist es, dass es sich bei der befragten Person nicht um die eigentliche Zielperson gehandelt hat. Allerdings ist aufgrund des in Abschnitt 3.4 beschriebenen Verifizierungsmechanismus davon auszugehen, dass dies nur sehr selten vorgekommen sein sollte.

Die Logik des Index lässt sich durch ein Beispiel verdeutlichen: Angenommen, an der ersten Welle habe die ausgewählte Zielperson teilgenommen, in der zweiten Welle der Ehepartner der Zielperson, und in allen anderen Wellen wieder die Zielperson. In diesem Fall wären für die zweite Welle deutlich andere Antworten zu erwarten als bei denselben Fragen in den anderen Wellen. Dementsprechend verzeichnet dann der Abweichungsindex in der zweiten Welle einen höheren Wert. Nach einer groben Faustregel können Indexwerte größer als 2

als „verdächtig“ gelten. Eine genaue Beschreibung des Index erfolgt in einem separaten Bericht.

6.3 Filterfehler

Filterfehler in Welle 1

- Der Befragte mit der lfdn 2313 hat aus ungeklärten Gründen trotz korrekter Filterbedingung die Frage „kp1_2470“ (Gewerkschaftsmitglied HH) nicht gesehen. Wir vermuten Probleme bei der Internetverbindung.
- Der Wert der Variable „kp1_190a“ (Beabsichtigte Stimmabgabe, Erststimme) wurde für den Befragten mit der lfdn 60 von „-97 - trifft nicht zu“ auf „801 - andere Partei“ nachträglich geändert, da auf dem Nachfragescreen zur anderen Partei eine gültige Antwort gegeben wurde. Des Weiteren wurden ebenfalls für den Fall lfdn 60 die Antworten bei „kp1_350a-b“ (Recall vorangegangene BTW: Erst- /Zweitstimme) von „-97 - trifft nicht zu“ auf „801 - andere Partei“ nachträglich geändert, da auf den Nachfragescreens zur anderen Partei gültige Antworten gegeben wurden. In beiden Fällen lässt sich die Ursache des Fehlers nicht rekonstruieren. Wir vermuten Probleme bei der Internetverbindung.
- Eine Person (lfdn 3356) gab zwar an, zur Wahl gehen zu wollen, wurde aber aus nicht nachvollziehbaren Gründen trotzdem über das Item „kp1_190a-b“ (Beabsichtigte Stimmabgabe) gefiltert. Des Weiteren hat der Befragte aus unerklärlichen Gründen die Items „kp1_650o,h,j“ (Skalometer Politiker) nicht gesehen (die Variablen erhielten ein system missing), obwohl er die ersten drei Items „kp1_650a,k,c“ der Matrix beantwortet hat. Nachträglich wurden die system missing der Items „kp1_650o,h,j“ in -92 „Fehler in Daten“ umkodiert.
- Welle 1: Eine Person (lfdn 700) hat bei „kp1_1931a-f“ (Politische Gespräche, allgemein) widersprüchliche Angabe gemacht. Zum einen gab sie an, zwei Tage in der Woche mit Bekannten politische Gespräche geführt zu haben und zum anderen, dass sie keine Unterhaltung führe. Aufgrund eines Plausibilitätschecks, der aus ungeklärten Gründen hier nicht funktionierte, hätte eine solche Antwortkombination gar nicht auftreten dürfen. Dieser Fall ist in „kpx_info“ markiert.

Filterfehler in Welle 2

- Bei der Variable „kp2_3990“ weist ein Befragter (lfdn 2820) einen systemfehlenden Wert auf, obwohl diese Person diese Seite tatsächlich gesehen hat. Der „system missing“ wurde in -92 „Fehler in Daten“ umkodiert.

Filterfehler in Welle 4

- Ein Befragter (lfdn 5233) hat aufgrund eines Filterfehlers die Frage zur Wahlbeteiligungsabsicht („kp4_170“) nicht gesehen, obwohl er in „kp3_170“ angegeben hat, „vielleicht zur Wahl“ zu gehen. Der Fall ist in der Variable „kpx_info“ dokumentiert.
- 24 Befragte haben aufgrund eines Filterfehlers die Frage „Gründe für Nichtwahl, gestützt“ (Variable kp4_252) erhalten, obwohl sie bei der Wahlbeteiligungsabsicht (Variable „kp4_170“) angegeben haben, „bestimmt zur Wahl“ zu gehen. Diese Fälle sind in der Variable „kpx_info“ dokumentiert.
- Zwei Befragte (lfdn 553 und 1800) haben aufgrund eines Filterfehlers die Frage „Gründe für Nichtwahl, gestützt“ (Variable „kp4_252“) erhalten, obwohl sie bei der

Wahlbeteiligungsabsicht (Variable „kp4_170“) angegeben haben, „bereits Briefwahl gemacht“ zu haben. Diese Fälle sind in der Variable „kpx_info“ dokumentiert.

- Zehn Befragte haben aufgrund eines Filterfehlers die Frage „Gründe für Nichtwahl, gestützt“ (Variable „kp4_252“) erhalten, obwohl sie bei der Wahlbeteiligungsabsicht (Variable „kp4_170“) angegeben haben, „wahrscheinlich zur Wahl“ zu gehen. Diese Fälle sind in der Variable „kpx_info“ dokumentiert.
- Vier Befragte (lfdn 590, 1099, 2883, 4713) haben aufgrund eines Filterfehlers die Frage „Gründe für Nichtwahl, gestützt“ (Variable „kp4_252“) erhalten, obwohl sie in Welle 1 (Variable „kp1_252“) bereits einen Grund angegeben und an Welle 2 und 3 nicht teilgenommen hatten. Diese Fälle sind in der Variable „kpx_info“ dokumentiert.

Filterfehler in Welle 5

- Ein Befragter (lfdn 3677) hat aufgrund eines Filterfehlers fälschlicherweise die Frage „kp5_170“ (Wahlbeteiligungsabsicht) erhalten, obwohl er zuvor in Welle 4 angegeben hatte, Briefwahl gemacht zu haben.
- Ein Befragter (lfdn 3996) hat aufgrund eines Filterfehlers fälschlicherweise die Fragen „kp5_170“, „kp5_191a-b“, „kp5_2751a-g“, „kp5_2760“, „kp5_330“, „kp5_331“ und „kp5_395“ (Wahlbeteiligungsabsicht, Briefwahlstimmen, Consideration Set, Schwierigkeit der Wahlentscheidung und –beteiligung, Wahlkampf hilfreich) erhalten, obwohl er zuvor in Welle 4 angegeben hatte, Briefwahl gemacht zu haben. Dafür hat er fälschlicherweise die Frage „kp5_2770a-b“ (Hypothetische Stimmabgabe nach Briefwahl) nicht erhalten. Die gemachten Angaben zur Briefwahl zwischen Welle 4 und 5 sind teilweise inkonsistent.

6.4 Rückmeldungen der Befragten

Am Ende jeder Umfrage wurde den Befragten die Gelegenheit gegeben, Feedback zu geben. Neben häufigen positiven Rückmeldungen wurden vereinzelt auch Kommentare gesammelt, die über technische Probleme oder fehlerhafte Antworten informieren. Alle berichteten Probleme werden im Folgenden aufgelistet.

Technische Probleme

Ein Befragter (lfdn in Welle 2: 3666) berichtete, dass anstelle der Politikerbilder (Items „kpX_3430j-o“) rote Kreuze zu sehen waren. Die Werte der Variablen „kp2_3430j-o“ wurden in -92 „Fehler in Daten“ umkodiert. Ein anderer Studienteilnehmer (lfdn in Welle 6: 5157) konnte die unteren Bilder der Politiker nicht im Browserfenster sehen. Entsprechend wurden die Variablen, die eine -99 „keine Angabe“ aufwiesen („kp6_3430j, l, m“) in -92 „Fehler in Daten“ umkodiert.

Mehrere Personen berichteten über Probleme bei der Darstellung der Fragetexte (lfdn in Welle 3: 5216; in Welle 4: 3206, 1821, 5216; in Welle 5: 1432, 1436, 1821, 2790; in Welle 6: 2790; in Welle 7: 428, 2790, 5202). Es wurden keine Änderungen in den Daten vorgenommen.

Eine Person gab an, dass die Namen der Politiker nicht dargestellt wurden. Dabei ist jedoch unklar, auf welche Fragen er sich bezieht (lfdn in Welle 4: 5006). Es wurden keine Änderungen in den Daten vorgenommen.

Ein Befragter berichtete von technischen Problemen bei der Durchführung der Umfrage (lfdn in Welle 6: 5066). Es wurden keine Änderungen in den Daten vorgenommen.

Eine Person berichtete, dass die Schieberegler in der Umfrage nicht funktioniert haben (lfdn in Welle 2: 649). Es wurden keine Änderungen in den Daten vorgenommen.

Eine Person (lfdn in Welle 6: 5040) berichtet, dass die Schieberegler bei der Frage nach der aktuellen politischen Printmediennutzung nicht funktioniert haben, sodass die Werte der Variablen „kp6_1661a-h“ in -92 „Fehler in Daten“ umkodiert wurden.

Mehrere Personen berichteten, dass die Ladezeiten der Umfrage lange gedauert haben (lfdn in Welle 2: 1199, 4690; in Welle 6: 1700, 4009). Es wurden keine Änderungen in den Daten vorgenommen.

Mehrere Personen gaben an, dass sie Probleme bei der Bedienung der Umfrage mit Smartphone oder Tablet hatten (lfdn in Welle 1: 153, 1752; Welle 4: 2906, 3549). Es wurden keine Änderungen vorgenommen.

Befragte, die auf Fehler in ihren Antworten aufmerksam machten

Mehrere Befragte gaben an, dass sie bei der Wahlabsichtsfrage die Antwort verweigern wollten und aufgrund des Fehlens einer passenden Kategorie die Option „weiß nicht“ anklickten (lfdn in den Wellen 1 bis 6: 2436; in Welle 1: 1804, 2436, 4424; in Welle 3: 1804). Die Werte dieser Fälle wurden in -99 „keine Angabe“ umkodiert.

Zwei Befragte haben die Erst- und Zweitstimme verwechselt („kp2_190aa-b, ba-b“: lfdn 720; „kp3_190aa-b, ba-b“: 2332). Die Angaben wurden entsprechend korrigiert.

In Welle 1 wies eine Person darauf hin, dass sie eine falsche Angabe bei der Postleitzahl gemacht habe (lfdn 872). Die Postleitzahl („kp1_2602“) wurde nachträglich in -99 „keine Angabe“ umkodiert.

In Welle 1 wies eine Person darauf hin, dass sie eine falsche Angabe bei der Wohnortsgröße („kp1_2600“) gemacht habe (lfdn 2462). Entsprechend der im Kommentar genannten Einwohnerzahl wurde die Variable umkodiert.

In Welle 1 lieferte eine Person die Information nach, dass sie in einer kirchlichen Gruppe aktiv sei (lfdn 3024). Diese Angabe wurde in „kp1_2461e“ entsprechend eingegeben.

In Welle 1 haben vier Personen bei der Frage nach dem Einkommen („kp1_2591“) nach eigener Aussage keine wahre Auskunft gegeben (lfdn 872, 1956, 3325, 3629). Die Werte wurden in -99 „keine Angabe“ umkodiert.

In Welle 1 haben sich acht Personen über die Frage nach den Einkommen („kp1_2591“) beschwert, haben dort aber dennoch eine Antwort gegeben. Es könnte sein, dass ihre Angaben falsch sind (lfdn 1738, 1884, 2321, 3345, 3378, 4031, 4285, 4685). Es wurden keine Änderungen in den Daten vorgenommen.

In Welle 2 hat eine Person bei den Variablen „kp2_661a-c“ die Antwortmöglichkeiten „trifft eher zu“ und „trifft eher nicht zu“ verwechselt (lfdn 4555). Die Werte wurden entsprechend umkodiert.

In Welle 2 hat eine Person ihre Angabe korrigiert, wonach ihre Mutter im Ausland geboren sei (lfdn 44). Daher wurden die Angaben in den Variablen „kp2_2571a“ (Geburtsland, Mutter) auf „Ja“ für Deutschland gesetzt und die übrigen Variablen „kp2_2570a“, „kp2_3980“, „kp2_4120“, „kp2_4140“, „kp2_4150“ entsprechend in -97 „trifft nicht zu“ umkodiert.

In Welle 3 hat ein Befragter bei den Variablen „kp3_190ba-b“ aus Versehen SPD anstatt FDP angeklickt (lfdn 1539). Dies wurde nachträglich korrigiert.

In Welle 4 verweist ein Befragter darauf, aufgrund des Wechsels von Skalenbeschriftungen häufiger Werte angeklickt habe, die nicht seiner Meinung entsprechen (lfdn 2111). Es wurde keine Änderung in den Daten vorgenommen.

In Welle 4 gab ein Befragter an, dass er bei der Links-Rechts-Einstufung von Angela Merkel und Peer Steinbrück („kp4_680a-b“) die Pole „Links“ und „Rechts“ vertauscht habe (lfdn 1784). Seine Angaben wurden jeweils spiegelbildlich umkodiert.

In Welle 5 gab ein Befragter an, bei der Links-Rechts Einstufung der Parteien („kp5_1490a-i“) die Pole verwechselt zu haben (lfdn 4928). Seine Angaben wurden entsprechend spiegelbildlich umkodiert.

In Welle 5 gab ein Befragter an, dass er bei der Frage nach seiner Meinung zum Zuzug von Ausländern („kp5_1130“) in die falsche Richtung geantwortet hat und eigentlich genau die entgegengesetzte Meinung vertritt (lfdn 2518). Die Eingabe wurde entsprechend umkodiert.

In Welle 6 gaben zwei Befragte an, dass sie bei der Skalierung ihrer negativen Gefühle zu den Parteien („kp6_2800a-f“) aus Versehen genau wie bei den positiven Gefühlen geantwortet hatten, da sie die veränderte Skalierung nicht beachtet hatten (lfdn 211, 4337). Da die intendierten Antworten nicht rekonstruiert werden können, werden die Angaben in -92 „Fehler in Daten“ umkodiert.

In Welle 7 gab eine Person an, zwei Fragen („kp7_820“ und „kp7_3030“) übersprungen zu haben und reichte im Kommentarfeld seine Antworten nach (lfdn 5045). Die zwei Variablen wurden seinen Angaben entsprechend nachträglich umkodiert.

In Welle 7 gab eine Person an, doch die Wahlplakate aller Parteien gesehen zu haben („kp7_421ga-f“; lfdn 832). Die Variablen wurden entsprechend umkodiert.

In Welle 7 revidierte ein Befragter im Kommentarfeld seine Zuordnung, dass Angela Merkel zur CSU gehöre und ordnete sie der CDU zu (lfdn 1385). Entsprechend wurde der Wert in „kp7_3430a“ umkodiert.

Sonstiges

Welle 1: Obwohl eine Person (lfdn 970) im Kommentarfeld angab, dass sie ungültig wählen würde, hat sie trotzdem eine Angabe bei „kp1_190aa-b, ba-b“ gemacht. Es wurden keine Änderungen vorgenommen.

Wellen 2 und 3: Eine Person (lfdn 2315) hat beim offenen Kommentarfeld in den Wellen 2 und 3 angegeben, dass sie ungültig wählen wird. Dennoch hat sie bei den Parteiwahl-Variablen „kp2/kp3_190aa-b, ba-b“ gültige Parteiangaben gemacht. Eine weitere Person (lfdn 1298) hat dieselbe Anmerkung in Welle 3 für „kp3_190aa-b, ba-b“ gemacht. Es wurden keine Änderungen vorgenommen.

Welle 5: Eine Person (lfdn 2127) gab an, dass sie aufgrund einer fehlenden „Weiß nicht“ Option häufiger die Mittelkategorie gewählt hat. Es wurden keine Änderungen vorgenommen.

Welle 6: Bei dem Befragten mit der lfdn 3717 handelt es sich möglicherweise nicht um die Zielperson, da dessen Angaben zum TV-Duell widersprüchlich sind. In Welle 5 hat er angegeben, dass TV-Duell gesehen zu haben und in Welle 6 gab er im Kommentarfeld an, Fragen zum TV-Duell beantworten zu müssen, obwohl er es nicht gesehen hat. Es wurden keine Änderungen vorgenommen.

6.5 Versionshistorie

Veränderung von Version 1.0.0 zu Version 2.0.0

Systemfehlende Werte in den Variablen „wei_w2-7“ (Gewichtungsvariablen) und „kp1_2180a-l“ (Big 5) wurden in die einheitlichen GLES-Missing-Codes umgewandelt.

Durch einen Abgleich der IDs aus den Wahlkampfpaneln 2009 und 2013 wurden Fälle identifiziert, die fälschlicherweise in der Variable „wkp2009“ als *teilgenommen* klassifiziert wurden. Gleichzeitig waren manche Personen in der Variable nicht markiert, obwohl sie in dem Wahlkampfpanel 2009 teilgenommen hatten. Dies wurde nun korrigiert.

Die Variable „mut09“ zur Identifikation von Panelmutanten im Wahlkampfpanel 2009 wurde ergänzt.

Die Variable „kpX_1230“ (Europäische Integration, Ego) und „kp4_1781“ (TV-Duell: Rezeptionswahrscheinlichkeit) liefen im Wahlkampfpanel 2009 als „kpX_1250“ bzw. „kpX_1780“. Zur Vereinheitlichung der Bezeichnung von Variablen in beiden Datensätze wurden diese beiden Variablen umbenannt und erhielten die Bezeichnung aus 2009.

Die Berechnung der Variable „kp4_speederindex“ war fehlerhaft und wurde korrigiert.

Durch die Erstellung der Versionen A und B bei allen Parteiabfragen ist es möglich, dass Befragte fehlende Werte bei der Parteiidentifikation aufweisen, aber trotzdem eine inhaltliche Antwort bei „kpX_2100“ (Stärke Parteiidentifikation) gegeben haben. Dieser Umstand beruht auf der zweistufigen Abfrage der Parteiidentifikation (siehe Kapitel 5.3). Diese Antworten bei „kpX_2100“, die mit keiner inhaltlichen Angabe verknüpft sind, wurden in -97 „trifft nicht zu“ umkodiert.

Zur Vervollständigung der Liste und Verwendung bei Analysen wurden die Variable „w1a“ (Wellenteilnahme Welle 1 angefangen) und „w1b“ (Wellenteilnahme Welle 1 beendet) ergänzt.

Es wurden die Codierungen der offenen Angaben hinzugespielt (siehe Kapitel 5.4).

Auch wurden Korrekturen von Rechtschreibfehlern in Labels und in der Fragebogendokumentation vorgenommen.

Veränderung von Version 2.0.0 zu Version 3.0.0

Der Großteil der Veränderungen in Version 3 stellt eine bessere Vergleichbarkeit der Wahlkampfpanel 2009 und 2013 her. Dazu wurden folgende Veränderungen vorgenommen:

Aufnahme der Variable „lfdn09“, die die laufende Nummer des Befragten aus dem Wahlkampfpanel 2009 angibt und anhand derer die Datensätze selbst zusammengespielt werden können.

Die Codes der Variable „kpX_260“ (Gründe für Wahlentscheidung) wurden an das Codierschema aus dem Wahlkampfpanel 2009 angepasst.

Die Codierung der Variable „kp1_2480“ (Religionszugehörigkeit) wurde an das allgemeine GLES Schema angepasst und ist damit auch identisch zu 2009.

Variablen, die bisher denselben Variablennamen wie 2009 hatten, sich aber im Wortlaut der Frage oder Antwortmöglichkeiten unterschieden, wurden umbenannt. Deren Variablenstamm, wurde um eine Ziffer erhöht. Gleichzeitig gab es Fragen, die genau wie in 2009 erhoben wurden, aber fälschlicherweise einen neuen Variablennamen erhalten hatten. Diese wurden in denselben Variablennamen wie 2009 umbenannt. Dies betrifft folgende Variablen:

„kpX_2570a, b“ (Geburtsland Vater, Mutter)	→	„kpX_2572a,b“
„kpX_250“ (Grund Nichtwahl, gestützt)	→	„kpX_252“
„kpX_660a-u“ (Eigenschaften der Kandidaten)	→	„kpX_661a-u“
„kpX_960“ (Erwartungen Koalitionen)	→	„kpX_961“
„kpX_1650c1-3“ (Internetnutzung Seiten)	→	„kpX_1651c1-3“
„kpX_1720a-h“ (Printmediennutzung)	→	„kpX_1721a-h“
„kpX_1740a-f“ (Fernsehnutzung)	→	„kpX_1741a-f“
„kpX_1660a-h“ (Printmedien, aktuell)	→	„kpX_1661a-h“
„kpX_1680a-f“ (Fernsehnutzung, aktuell)	→	„kpX_1681a-f“
„kpX_2080a-d“ (Inglehart Items)	→	„kpX_2081a-d“
„kpX_2460a-i“ (Organisationsmitgliedschaft)	→	„kpX_2461a-i“
„kpX_2540“ (Geburtsland)	→	„kpX_2541“
„kpX_2550“ (Alter Zuzug)	→	„kpX_2551“
„kpX_2590“ (Nettoeinkommen)	→	„kpX_2591“
„kpX_2181a-m“ (Big 5)	→	„kpX_2180a-m“
„kpX_1571a-q“ (Needs)	→	„kpX_1570a-q“

Darüber hinaus wurden teilweise die Suffixe der Variablen an die Vergabe im Wahlkampfpanel 2009 angepasst (nicht genannte Suffixe bleiben unverändert):

„kpX_661“	„kpX_261“	„kpX_650“
i → t	b → d2	b → k
k → u	c → e2	d → o
	d → b	e → h
	e → h	f → j
	f → c	g → q
	g → f	h → r
	h → g2	i → z
	i → k2	j → t
	j → i2	o → u
	k → j	k → v
		l → w
		n → x
		m → y

In den Variablen „kpX_1070a“, „kpX_1110a“, „kpX_1270a“ und „kp7_341a“ wurden das jeweilige Suffix „a“ durch das Suffix „j“ ersetzt, da CDU/CSU im Wahlkampfpanel 2009 getrennt abgefragt wurden.

Die unter 6.2 erläuterten Qualitätsindikatoren „kpX_dev“, „kpX_itnrp“, „kpX_strl“, „kpX_mtime“, „kpX_mtimex“, „kpX_dokn“ und „kpX_qual“ wurden dem Datensatz hinzugefügt.

Die Paradata „p_numinv“, „p_numcpl“, „p_numstr“ und „p_numinc“ wurden aus dem Datensatz entfernt, da sie nicht wie angenommen Informationen über das Panelverhalten seit dem Eintritt der Befragten in das Respondi-Panel enthalten, sondern nur der letzten 372 Tage und damit sich kaum von den Variablen „p_numinv2“ bis „p_numinc2“ unterscheiden.

Neben der deutschen Version wird nun auch eine englische Version des Datensatzes, des Fragebogens und der Studienbeschreibung angeboten.

Es wurden Korrekturen in der Schreibweise der Variablen- und Wertelabels gemäß der GLES-Standards vorgenommen.

Veränderung von Version 3.0.0 zu Version 3.1.0

Korrektur von Variablen- und Wertelabels.

Die Skala der Variable „kp7_3591“ wurde falschherum kodiert und wurde korrigiert.

Anders als im Fragebogen festgestellt, unterschied sich die Frageformulierung bei „kpX_2820a,b“ in Welle 5 im Vergleich zu den Versionen, die in Welle 3 und 7 liefen. Dies wurde nun entsprechend im Fragebogen dokumentiert.

Veränderung von Version 3.1.0 zu Version 3.2.0

Die Label der Werte 0 und 1 der Variablenbatterie „kpX_421aa-kf,y“ (Parteikontakte) waren im englischen Datensatz vertauscht. Dies ist nun korrigiert.

7. Hinzuspielen anderer Datensätze

7.1 Hinzuspielen von Zeitvariablen

Die zum Wahlkampfpanel zugehörigen Zeitvariablen werden in einem separaten Datensatz zum Herunterladen angeboten und können mittels der Variable „lfdn“ (Laufende Nummer) an das Wahlkampfpanel herangespielt werden. Alternativ kann auch ein Do-File für Stata oder eine SPSS-Syntax heruntergeladen werden, welche die Zusammenführung der beiden Datensätze übernehmen.

Der Datensatz enthält zwei Arten von Zeitvariablen: Die erste Gruppe von Zeitvariablen gibt die Antwortzeit eines Befragten auf einer speziellen Umfrageseite wieder. Diese Variablen sind nach dem Schema T_Variablenname benannt (Beispiel: „T_kp1_010“ für das Politische Interesse). Wenn mehrere Items auf einer Umfrageseite abgefragt wurden, was beispielsweise bei Matrixfragen der Fall ist, so ist die Zeitvariable der Umfrageseite nur mit dem jeweiligen Variablenstamm benannt (Beispiel „T_kp1_430“ und nicht „T_kp1_430a-f“).

Die zweite Gruppe von Zeitvariablen enthält die kumulierte Zeit eines Befragten, die benötigt wurde, um bis an die jeweilige Stelle der Umfrage zu gelangen. Diese kumulierten Zeitvariablen sind entsprechend der Reihenfolge der Fragen innerhalb des Fragebogens durchnummeriert, wobei die Zahl in der Bezeichnung nach dem T die Stelle innerhalb der Umfragesequenz angibt. Zum Beispiel wurde das politische Interesse auf dem 8. Bildschirm in der 1. Welle abgefragt und die kumulierte Zeitvariable wurde demnach mit „T8_kp1_010“ benannt.

7.2 Hinzuspielen des GLES-Wahlkampfpanels 2009

Für das Hinzuspielen der Daten des Wahlkampfpanels 2009 (ZA-Nr. 5305) gibt es eine Variable „wkp2009“, die angibt, ob der Befragte im Wahlkampfpanel 2009 teilgenommen hat. Um die Daten anhand der Variable „lfdn09“ hinzuzuspielen, müssen aber zuvor die Variablennamen in einer der beiden Datensätze geändert werden, da andernfalls eine nicht zulässige Doppelung von Variablen auftritt (beispielsweise existiert die Variable „kp2_010“ in beiden Datensätzen). Dem Nutzer steht eine bereits kumulierte Version der Datensätze zur Verfügung (ZA5757) und kann bei GESIS heruntergeladen werden.

7.3 Hinzuspielen der Kontrollquerschnitte

Wie in Kapitel 3.1 beschrieben, wurden zeitgleich zur dritten, fünften und siebten Befragung des Wahlkampfpanels drei Querschnittsbefragungen mit jeweils etwa 1200 Personen und einem fast identischen Fragebogen durchgeführt. Die Daten der Kontrollgruppen wurden bei GESIS unter den ZA-Nummern 5753, 5754 und 5755 veröffentlicht. Da die Variablen im Wahlkampfpanel und in den Kontrollquerschnitten dieselben Namen haben, können alle Datensätze problemlos aneinandergesetzt werden. Bei GESIS steht dem Nutzer ein Stata-Do-File zum Zusammenfügen der Datensätze bereit.

Literaturverzeichnis

- Baker, Reg und Theo Downes-Le Guin (2007): Separating the Wheat from the Chaff: Ensuring Data Quality in Internet Samples. Konferenzpapier vorgestellt auf: The Challenges of a Changing World: Proceedings of the Fifth International Conference of the Association for Survey Computing in Southampton, 12. bis 14. September.
- Blumenberg, Manuela S. und Tobias Gummer (2013): Gewichtung in der German Longitudinal Election Study 2009. In: GESIS – Technical Reports 2013|19.
- Blumenstiel, Jan Eric und Tobias Gummer (2012): Langfrist-Panels der German Longitudinal Election Study (GLES): Konzeption, Durchführung, Aufbereitung und Archivierung. GESIS - Technical Reports, 2012/11.
- Brüderl, Josef (2010). Kausalanalyse mit Paneldaten. in: Christof Wolf und Henning Best (Hrsg.). *Handbuch der sozialwissenschaftlichen Datenanalyse*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 963-994.
- Callegaro, Mario und Charles DiSogra (2008): Computing Response Metrics for Online Panels, in: *Public Opinion Quarterly* 72, 1008-1032.
- DeBell, Matthew, Jon A. Krosnick, Arthur Lupia und Caroline Roberts (2009): User's Guide to the Advance Release of the 2008-2009 ANES Panel Study. Palo Alto, CA and Ann Arbor, MI: Stanford University and University of Michigan.
- Deming, Edwards W. und Frederick F. Stephan (1940): On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known. In: *The Annals of Mathematical Statistics* 11 (4): 427-444.
- Downes-Le Guin, Theo; Joanne Mechling und Reg Baker (2006): Great Results from Ambiguous Sources: Cleaning Internet Panel Data. Konferenzpapier vorgestellt auf: ESOMAR World Research Conference: Panel Research 2006 in Barcelona, 27. bis 29. November.
- Gräf, Lorenz und Martin Heidingsfelder (1999): Bessere Datenqualität bei WWW-Umfragen – Erfahrungen aus einem Methodenexperiment mit dem Internet-Rogator. in: Bernad Batinic; Andreas Werner; Lorenz Gräf und Wolfgang Bandilla (Hrsg.). *Online Research. Methoden, Anwendungen und Ergebnisse*. Göttingen: Hogrefe, S. 113-126.
- Green, Donald P.; Soo Yeon Kim und David H. Yoon (2001): Dirty Pool. In: *International Organization* 55, S. 441-468.
- Joinson, Adam (1999): Social Desirability, Anonymity, and Internet-Based Questionnaires. In: *Behavior Research Methods, Instruments, & Computers* 31, S. 433-438.
- Krosnick, Jon A. (1991): Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. In: *Applied Cognitive Psychology* 5, S. 213-236.
- Krosnick, Jon A. (1999): Survey research. In: *Annual Review of Psychology* 50, S. 537-567.
- Mayerl, Jochen und Urban, Dieter (2008): Antwortreaktionszeiten in Survey-Analysen. Messung, Auswertung und Anwendungen. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Mühlenfeld, Hans U. (2004): Der Mensch in der Online-Kommunikation: Zum Einfluss webbasierter, audiovisueller Fernkommunikation auf das Verhalten von Befragten. Wiesbaden: Deutscher Universitäts Verlag.
- Plischke, Thomas (2014): Wann Wähler entscheiden. Abläufe von Entscheidungsprozessen und der Zeitpunkt der Wahlentscheidung. Nomos: Baden-Baden.
- Roßmann, Joss (2010): Data Quality in Web Surveys of the German Longitudinal Election Study 2009. 3rd ECPR Graduate Conference. Dublin.
- Steinbrecher, Markus/ Roßmann, Joss und Michael Bergmann (2013): Das Wahlkampf-Panel der German Longitudinal Election Study 2009. Konzeption, Durchführung, Aufbereitung und Archivierung, GESIS Technical Reports 2013/17.
- Taddicken, Monika (2009): Die Bedeutung von Methodeneffekten der Online-Befragung: Zusammenhänge zwischen computervermittelter Kommunikation und erreichbarer Datengüte. In: Nikolaus Jackob; Harald Schoen und Thomas Zerback (Hrsg.). *Sozialforschung im Internet*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 91-107.

The American Association for Public Opinion Research (Hrsg.) (2011): Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. Siebte Auflage. o.O.

Links

Projektseite der GLES: <http://www.gles.eu>

Deutsche Gesellschaft für Wahlforschung e. V. (DGfW): <http://www.dgfw.info>

GESIS – Leibniz-Institut für Sozialwissenschaften: <http://www.gesis.org/gles>

Universität Frankfurt: <http://www.uni-frankfurt.de>

Universität Mannheim: <http://www.uni-mannheim.de>

Wissenschaftszentrum Berlin für Sozialforschung gGmbH (WZB): <http://www.wzb.eu>

Bamberger Centrum für Empirische Studien (BACES) <http://www.uni-bamberg.de/zentren/baces/>

Erstellt wurde die Studienbeschreibung von GESIS – Leibniz-Institut für Sozialwissenschaften und der Universität Mannheim in Zusammenarbeit mit BACES.

Mannheim, Juni 2016