

# gesis

Leibniz-Institut  
für Sozialwissenschaften

German Longitudinal  
Election Study



## **GLES 2017**

### **Pretest**

ZA6821, Version 1.0.0

*Studienbeschreibung*

## Inhaltsverzeichnis

Vorbemerkungen .....	3
1. Studieninformationen .....	4
2. Konzeption .....	5
3. Stichprobenziehung.....	6
4. Erhebung .....	9
5. Datenaufbereitung.....	12
6. Hinweise und Anmerkungen.....	23
Links .....	24
Literaturverzeichnis.....	25

## 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](http://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](mailto: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](http://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](mailto:gles@gesis.org)

### Zitation bei Veröffentlichungen

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

Roßteutscher, Sigrid; Schmitt-Beck, Rüdiger; Schoen, Harald; Weßels, Bernhard, Wolf, Christof (2018): Pretest 2017 (GLES). GESIS Datenarchiv, Köln: ZA6821 Datenfile Version 1.0.0, doi: 10.4232/1.3057.

## **1. Studieninformationen**

### **1.1. Studiennummer**

ZA6821

### **1.2. Titel der Studie**

German Longitudinal Election Study, Pretest 2017

### **1.3. Version**

1.0.0, 04.07.2018, doi: 10.4232/1.3057

### **1.4. Erhebungszeitraum**

15.05.2017 – 22.05.2017

### **1.5. Primärforscher/innen**

Prof. Dr. Sigrid Roßteutscher	(Universität Frankfurt)
Prof. Dr. Rüdiger Schmitt-Beck	(Universität Mannheim)
Prof. Dr. Harald Schoen	(Universität Mannheim)
Prof. Dr. Bernhard Weßels	(Wissenschaftszentrum Berlin für Sozialforschung)
Prof. Dr. Christof Wolf	(GESIS – Leibniz-Institut für Sozialwissenschaften)

### **1.6. Finanzierende Stelle**

Deutsche Forschungsgemeinschaft e.V. (DFG)

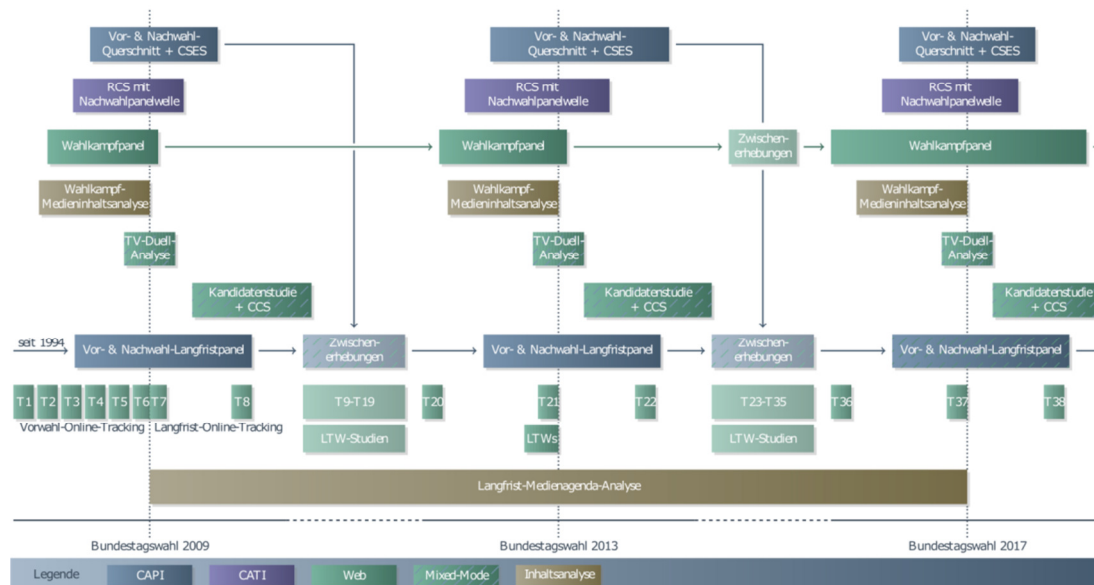
## 2. Konzeption

### 2.1. Hintergrund und Zielsetzung

Die German Longitudinal Election Study (GLES) ist die bislang größte nationale Wahlstudie in Deutschland. In dem von der DFG geförderten Projekt sollen die politischen Prädispositionen und Einstellungen sowie das politische Verhalten der wahlberechtigten Bürger bei vorerst drei aufeinanderfolgenden Wahlen beobachtet und analysiert werden. Das mit der Bundestagswahl 2009 gestartete Projekt wird ab dem Jahr 2018 von GESIS gestellt und in Zusammenarbeit mit der Fachgemeinschaft als institutionalisierte Wahlstudie durchgeführt.

Die GLES erlaubt die Analyse des Wahlverhaltens in quer- als auch längsschnittlicher Perspektive sowie in Hinblick auf kurzfristige Dynamiken während des Wahlkampfs und auch langfristige soziale Wandlungsprozesse über den Wahlzyklus hinweg (Schmitt-Beck et al. 2010).

Abbildung 1: Das Design der German Longitudinal Election Study (GLES)



### 2.2. Inhalt der Studie

Es handelt sich bei dem Datensatz um einen Pretest für die verschiedenen Komponenten, die zur Bundestagswahl 2017 durchgeführt werden.

### 3. Stichprobenziehung

#### 3.1. Untersuchungsgebiet

Deutschland (DE), Nordrhein-Westfalen

#### 3.2. Grund- und Auswahlgesamtheit

Die Grundgesamtheit des Pretests bildet die zur Bundestagswahl 2017 wahlberechtigte Bevölkerung in Nordrhein-Westfalen. Da es sich aber bei der vorliegenden Studie um eine Online-Befragung handelt, konnte keine Zufallsstichprobe realisiert werden. Die dieser Studie zugrundeliegende Auswahlgesamtheit umfasst somit nur die wahlberechtigten Teilnehmer des von der Respondi AG betriebenen Online-Access-Panels. Das Respondi-Panel umfasste 2016 in Deutschland 65.000-70.000 aktive Personen. Als aktive Personen definiert Respondi diejenigen Personen, die nach der (Double-opt-in-) Registrierung<sup>1</sup> die Stammdatenbefragung ausgefüllt haben und in den letzten drei Monaten an mindestens einer Umfrage teilgenommen haben.

**Tabelle 1: Zusammensetzung des Respondi-Access-Panels nach soziodemographischen Merkmalen (Anteil in Prozent)**

	Respondi
<b>Geschlecht</b>	
Weiblich	62
Männlich	38
<b>Bildung</b>	
Niedrig (d.h. kein Abschluss, Haupt- und Volksschule)	19
Mittel (d.h. Realschule, Mittlere Reife)	39
Hoch (d.h. Abitur, Fachhochschulreife, Studium)	42
<b>Altersgruppen</b>	
18-29 Jahre	30
30-39 Jahre	21
40-49 Jahre	18
50-59 Jahre	18
60 und älter	13

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.

<sup>1</sup> 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.

**Tabelle 2: Rekrutierungswege bei Respondi**

	Anteil in Prozent
MingleTrend Blog	2,1
Über Online-Werbung (davon beworbene Themenumfragen 50%; direkte Panelwerbung 50% inkl. Affiliate Marketing)	38
Selbsteinträge	11
Facebook Fanseite	35
Suchmaschinen	10
Empfehlungen (Freundschaftswerbung)	0,9
Kooperationen	2,9
Telefonakquise (CATI)	0,1

Respondi gibt an, ein wirkungsvolles Qualitätsmanagement zu betreiben. Über ein internes Bewertungssystem wird das Antwortverhalten der Mitglieder des Respondi-Panels kontinuierlich evaluiert. Gründe für die Deaktivierung von Mitgliedern aus der Datenbank sind beispielsweise fehlende Teilnahme über einen Zeitraum von drei Monaten (Voraussetzung bei Respondi ist, dass die Mitglieder mindestens zehn Einladungen erhalten haben) 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 elf Monate im Respondi-Panel. Innerhalb eines Jahres werden ca. 40 Prozent der Panelmitglieder aufgrund von Qualitätskontrollen und Panel-Mortalität aus der Datenbank entfernt.

Die durchschnittliche Teilnahmequote der Respondi-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. Sobald die Teilnehmer 1.500 mingle-Punkte gesammelt haben, kann dieser Betrag in Einkaufsgutscheine umgewandelt oder gespendet werden. Ab 2.000 mingle-Punkten können diese auch ausgezahlt werden. Zusätzlich werden zur Panelpflege regelmäßig Verlosungen unter den Mitgliedern durchgeführt.

### 3.3. Auswahlverfahren

Die Stichprobe für den Pretest wird durch eine Quotenauswahl aus der Auswahlgesamtheit der bei Respondi zur Verfügung stehenden aktiven Panelteilnehmern generiert. Dabei werden nur volljährige Personen mit deutscher Staatsangehörigkeit berücksichtigt.

Die Teilnehmer werden nach den Merkmalen Geschlecht, Alter und Bildung quotiert.

**Tabelle 3: Quotierung**

<b>Quotierungsmerkmale</b>	<b>Soll in Prozent</b>	<b>Ist in Prozent</b>
<b>Geschlecht</b>		
Weiblich	50	53,1
Männlich	50	46,9
<b>Bildung</b>		
Niedrig (kein Abschluss, Haupt- oder Volksschule)	35	30,6
Mittel (Realschule, Mittlere Reife)	40	40,3
Hoch (Fachhochschulreife, Abitur)	25	29,1
<b>Altersgruppe</b>		
18-29 Jahre	25	19,0
30-39 Jahre	20	20,0
40-49 Jahre	25	22,4
50-59 Jahre	15	19,4
60 Jahre und älter	15	19,2



## **4. Erhebung**

### **4.1. Erhebungsverfahren**

Online-Befragung mit standardisiertem Fragebogen

### **4.2. Datenerhebung**

GESIS übernahm die Programmierung des Fragebogens und die Durchführung der Datenerhebung.

### **4.3. Erhebungssoftware**

Die Programmierung des Fragebogens und die Durchführung der Datenerhebung erfolgten mittels der Software QuestBack, EFS Survey.

### **4.4. Tests des programmierten Fragebogens**

Der programmierte Fragebogen wurde intensiv geprüft. In den Tests gefundene Probleme oder Programmierfehler wurden bereinigt. Anschließend wurde der Fragebogen in der finalen Version freigegeben.

### **4.5. Incentivierung**

Um eine möglichst hohe Ausschöpfung zu erzielen, wurden Incentives eingesetzt. Für die Teilnahme an der ca. 25minütigen Studie erhielten die Befragten Bonuspunkte, die umgerechnet einem Wert von 2,50 Euro entsprechen.

### **4.6. Einladungen und Reminder**

Die Einladung der Teilnehmer erfolgte durch Respondi. Wenn ein Panelist die Einladung annahm, wurde er weitergeleitet und im Namen der für die Durchführung der Studie Verantwortlichen der GLES zur Befragung eingeladen.

Es wurde am Montagmorgen (15.05.2017) mit einem Soft Launch gestartet, ab dem Mittag erfolgte der Full Launch. Es wurden zwei Reminder versendet, einer am Mittwochmorgen (17.05.) und einer am Freitagmorgen (19.05.).

### **4.7. Feldverlauf**

Die nachfolgende Tabelle illustriert das Teilnahmeverhalten der zur Umfrage eingeladenen Teilnehmer des Respondi-Panels über den Erhebungszeitraum hinweg. Partielle und vollständige Interviews umfassen alle Interviews, die bis zur letzten Frage des Fragebogens durchgeführt wurden, unabhängig davon, ob die Umfrage unterbrochen wurde oder nicht.

Aufgrund von Probleme mit der Erfüllung der Quote von niedrig Gebildeten wurden ab dem 17.5. auch Personen über das Partnerpanel Gapfish eingeladen. Das Panel setzt sich im Endeffekt aus 89,3 Prozent Respondi Panelisten und 10,7 Prozent Gapfish Panelisten zusammen.

**Tabelle 4: Teilnahmestatistik nach Tagen**

Datum	Vollständige Interviews		
	N	Prozent	kum. Prozent
Mo 15.05.2017	1.081	58,5	58,5
Di 16.05.2017	343	18,6	77,1
Mi 17.05.2017	232	12,6	89,7
Do 18.05.2017	50	2,7	92,4
Fr 19.05.2017	99	5,4	97,7
Sa 20.05.2017	28	1,5	99,2
So 21.05.2017	14	0,8	100,0
<b>Gesamt</b>	<b>1.847</b>	<b>100,0</b>	

#### 4.8. Teilnehmerstatistik

Die Brutto-Stichprobe für die Online-Befragung wurde von Respondi aus ihrem Access-Panel gezogen. Die ausgewählten Panelisten wurden von Respondi zur Teilnahme eingeladen. Nur ein Teil der eingeladenen Panelisten hat die Einladung angenommen und die Befragung auch begonnen. Die Differenz zwischen begonnenen und beendeten Interviews ergibt sich aus drei Gruppen: Panelisten, die das Interview begannen, aber nicht beendeten, wurden der Kategorie „Abgebrochene Interviews“ zugeordnet. „Ausgescreent“ wurden jene Befragte, aufgrund von Alter, Bundesland und/oder Staatsangehörigkeit. Um den Quotenvorgaben zu genügen, wurden Panelisten mit bestimmten Merkmalen nach Erreichen der einzelnen Vorgaben von der Teilnahme ausgeschlossen. Die betroffenen Teilnehmer wurden „abgewiesen“.

Im Rahmen der Datenaufbereitung wurden zudem noch Zeitunterschreiter identifiziert.

**Tabelle 5: Teilnahme am Pretest**

Teilnehmerstatistik	Anzahl	Anteil in Prozent	
Von Respondi eingeladen	4.919	100,00%	
Interview nicht begonnen	208	4,23%	
Interview begonnen	4.711	95,77%	100,00%
Abgebrochene Interviews	462	9,81%	
Ausgescreent	215	4,56%	
Abgewiesen	2.186	46,40%	
Beendete Interviews	1.847	39,21%	100,00%
Zeitunterschreiter	177		9,58%
Beendete Interviews ohne Zeitunterschreiter	999		90,42%

\* Zur Definition von Zeitunterschreitern siehe Abschnitt 5.5.3.

## 4.9. Interviewdauer

Die Interviewdauer ist ein Indikator für den Umfang einer Befragung. Sie lässt Rückschlüsse darauf zu, wie groß der Aufwand war, den Befragte auf sich nehmen mussten, um den Fragebogen zu beantworten. Die nachfolgende Tabelle präsentiert deskriptive Statistiken zur Interviewdauer.

**Tabelle 6: Interviewdauer**

	<b>N</b>	<b>Min.</b>	<b>Max.</b>	<b>Arithmetisches Mittel</b>	<b>Median</b>
Interviewdauer	1.769	00:03:46	21:15:58	00:27:49	00:24:19

Anmerkungen: Die Interviewdauer wird nur für diejenigen Befragten ausgewiesen, die die Umfrage weder unter- noch abgebrochen haben. Wurde die Umfrage unterbrochen, ohne den Fragebogen im Browser zu schließen, wurde dies nicht als Unterbrechung gewertet.

## 5. Datenaufbereitung

### 5.1. Datenkontrollen

Die Daten dieser Studie wurden nach der Erhebung sorgfältig geprüft. Hierunter fiel die Überprüfung auf Vollständigkeit der Variablen im Datensatz, die Überprüfung auf Richtigkeit von Variablennamen und -labels, die Prüfung auf Vollständigkeit von Wertelabels, die Überprüfung der Codierung von Missing Values sowie eine Kontrolle der Randverteilung. Darüber hinaus erfolgte eine technische Kontrolle der Filterführungen. Alle gefundenen Fehler werden während der Datenaufbereitung dokumentiert und korrigiert. Trotz gründlicher Überprüfung der Daten kann es vorkommen, dass Fehler in den Datensätzen unentdeckt bleiben. Fallen diese zu einem späteren Zeitpunkt auf, werden sie in der Errataliste im GESIS Datenbestandskatalog ([www.gesis.org/dbk](http://www.gesis.org/dbk)) dokumentiert und zeitnah behoben.

### 5.2. Einheiten im Datensatz

Der Datensatz umfasst 1.847 realisierte Interviews.

### 5.3. Variablen im Datensatz

Der Datensatz setzt sich aus verschiedenen Typen von Variablen bzw. Daten zusammen. Diese umfassen Meta-, Para- und substantielle Daten sowie Kontext- und GewichtungsvARIABLEN (siehe auch die Abschnitte 5.4 bis 5.7). Die Tabelle 8 gibt einen Überblick über die Anzahl der Variablen im Datensatz nach Variablentyp.

**Tabelle 7: Anzahl der Variablen im Datensatz nach Variablentyp**

Variablentyp	Anzahl
Metadaten	23
Paradaten	7
Substantielle Daten	800
Gewichtungsvariablen	4
<b>Gesamt</b>	<b>834</b>

### 5.4. Metadaten

Metadaten beschreiben den Datensatz und seine Erstellung (Callegaro 2013). Sie umfassen Variablen, die für die Archivierung und Distribution der Studie benötigt werden. In der GLES umfasst dies die Studiennummer des Datensatzes im vierstelligen Format, unter welcher die Studie in den Archivsystemen von GESIS abgelegt und zu finden ist, die Versionskennung des Datensatzes, Angaben zum Erhebungsjahr und -zeitraum, die Zuordnung der Studie zu den Komponenten der GLES sowie die Kennung der Erhebung innerhalb der Komponente der GLES (siehe Tabelle 9).

**Tabelle 8: Metadaten**

Variable	Label
study	Studiennummer
version	GESIS Archiv Version
doi	Digital Object Identifier
year	Erhebungsjahr
field	Erhebungszeitraum
glescomp	GLÉS-Komponente
survey	Erhebung/Welle
c_0001-0016	Split 1 – Split 16

## 5.5. Paradata

Paradata sind Informationen, die aus dem Erhebungsprozess resultieren (Couper 2000), z.B. das Datum und die Uhrzeit des Beginns und des Endes der Interviewteilnahme, die Interviewdauer, Antwortzeitmessungen oder Informationen zum verwendeten Gerät und dem verwendeten Webbrowser (Callegaro 2013). Auch administrative Informationen wie Identifikationsnummern, Informationen zum Teilnahmestatus von Befragungsteilnehmern und Informationen zur Zuweisung von Teilnehmern zu Gruppen auf Grund von inhaltlichen Angaben oder zufallsbasierten Verfahren werden hier als Paradata behandelt. Die Tabelle 10 gibt einen Überblick über Paradata im Datensatz.

**Tabelle 9: Paradata**

Variable	Label
ifdn	Laufende Nummer
partstat	Teilnahmestatus
datetime	Datum/Zeit
duration	Interviewdauer in Sekunden
speederindex	Zeitunterschreiter-Index
speederflag	Zeitunterschreiter (10% der Befragten mit den niedrigsten Indexwerten)
javascript	Installation Java

### 5.5.1 Laufende Nummer und Teilnahmestatus

Die Variable *ifdn* „Laufende Nummer“ ist eine eindeutige Identifikationsnummer für die Teilnehmer an der Befragung. Die Variable *partstat* „Teilnahmestatus“ gibt Auskunft darüber, ob Teilnehmer die Befragung vollständig absolviert oder abgebrochen haben, oder ob sie auf Grund bereits erfüllter Quoten abgewiesen wurden.

### 5.5.2 Zeitvariablen

Es wurde Zeitvariable (Variable *duration*) serverseitig und sekundengenau als Paradata erhoben. Diese beinhaltet die Zeit, die ein Befragter zwischen dem Betrachten und Bearbeiten der Begrüßungsseite bis zum Schließen der letzten Seite der Befragung benötigt hat. Bei der Analyse der Interviewdauer ist zu beachten, dass Befragte den Fragebogen verlassen und die Befragung zu einem späteren Zeitpunkt fortsetzen konnten.

### 5.5.3 Zeitunterschreiter

Aufgrund der fehlenden Kontrolle der Befragten durch einen Interviewer in Online-Befragung und der besonderen Anreiz- und Belohnungsstruktur durch die Incentivierung der Befragten ist diese Studie mit dem Problem außergewöhnlich schneller Antwortzeiten („Zeitunterschreitung“) konfrontiert. Zeitunterschreitung bedeutet, dass einige Befragte einzelne Fragen oder auch die gesamte Umfrage erheblich schneller beantworten als der Großteil der Teilnehmer. Eine schnelle Beantwortung der Fragen in der Umfrage ist noch kein Problem an sich, da es aufgrund bestimmter sozialstruktureller und persönlicher Merkmale (z.B. Bildung, Alter, Intelligenz, Reaktionsgeschwindigkeit) deutliche Unterschiede in der Beantwortungsgeschwindigkeit zwischen Befragten geben kann. Dennoch ist davon auszugehen, dass bei einer erheblichen Unterschreitung der mittleren Antwortzeit die Antwortqualität leidet, da sich Befragte mutmaßlich wenig bei der Beantwortung der Fragen bemühen und daher oberflächliche oder willkürliche Antworten abgeben, keine Angabe machen oder „weiß nicht“ antworten, obwohl sie eine substantielle Antworten hätten abgeben können (vgl. Krosnick 1991).

In der Fachliteratur gibt es keine etablierten Standards für die Identifikation von Zeitunterschreitern. 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 (vgl. Mayerl und Urban 2008). Diese Befragten werden dann entweder aus dem Datensatz ausgeschlossen oder durch Markervariablen gekennzeichnet.

In dieser Studie werden Zeitunterschreiter anhand einer modifizierten Routine des von Roßmann (2010) weiterentwickelten Algorithmus identifiziert. Hierfür wird ein Zeitunterschreiter-Index gebildet (Variable *speederindex*), der sowohl die Antwortdauer der Befragten auf allen Bildschirmseiten der Befragung als auch die Interviewdauer je Befragtem einbezieht und Indexwerte zwischen größer 0 und kleiner 2 annimmt. 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 anzeigen. Als Zeitunterschreiter werden alle Befragten im unteren 10%-Perzentil der Verteilung des Summenindex gekennzeichnet. Die Kennzeichnung von Zeitunterschreitern ist im Datensatz in der Variable *speederflag* enthalten. Diese Markervariable kann zum Ausschluss der Fälle aus Analysen als auch zur Durchführung von Robustheitsanalysen verwendet werden.

In dieser Studie wurden 177 Fälle (10,01%) als Zeitunterschreiter markiert.

**Tabelle 10: Statistiken zur Stichprobe mit und ohne Zeitunterschreiter**

Merkmal	Mit Zeitunterschreibern	Ohne Zeitunterschreiter
<b>Interviewdauer (hh:mm:ss)</b>		
Min.	00:03:46	00:12:25
Mittelwert	00:27:49	00:29:42
Median	00:24:19	00:24:23
<b>Geschlecht (in %)</b>		
Männlich	46,89	47,17
Weiblich	53,11	52,83
<b>Altersgruppe (in %)</b>		
18 bis unter 30 Jahre	19,95	17,46
30 bis unter 45 Jahre	30,10	28,45
45 bis unter 60 Jahre	31,73	32,79
60 Jahre und älter	19,22	21,29
<b>Bildung (in %)</b>		
Niedrig	30,59	30,84
Mittel	40,34	39,70
Hoch	29,07	29,46
<b>N</b>	1.769	1.592

## 5.6. Substantielle Daten

Substantielle Daten bilden den Hauptbestandteil des Datensatzes. Sie umfassen inhaltliche und soziodemographische Variablen.

**Tabelle 11: Inhaltliche und soziodemographische Variablen**

Variable	Label
v1	Alter
v2	Geschlecht
v3	Schulabschluss
v5	Bundesland
v112	Staatsbürgerschaft
v6	Politisches Interesse
v7	Politik in den Medien
v8	Internal Efficacy
v9	Einstellungen geg. Eliten
v10	Einstellungen zu out-groups
v11	Nationale Identität
v12	Korruption
v13	Issues
v14	Leistung Bundesregierung
v15	Leistung Landesregierung
v16	Interessenvertretung Partei
v17	Welche Partei Interessenvertretung
v18	Allgemeine wirtschaftliche Lage, retrospektiv
v19	Wahlbeteiligung LTW
v20	Recall LTW

---

v21	Einschaetzung: Parteiwahl
v22	Protestwahl: Wahlentscheidung
v23	Entscheidung bereit: Wahlenthaltung
v24	Protestwahl: Wahlbeteiligung
v25	Ueberzeugungskommunikation: Ego - Familie/Freunde
v26	Ueberzeugungskommunikation: Einfluss auf Entscheidung von Familie/Freunde
v27	Ueberzeugungskommunikation: Familie/Freunde - Ego
v28	Ueberzeugungskommunikation: Einfluss Familie/Freunde auf Ego
v29	Wahlberechtigung vorangegangene LTW
v30	Wahlteilnahme letzte LTW
v31	Recall vorangegangene LTW
v32	Unterschied Regierung
v33	Unterschied Politik
v34	Skalometer Parteien
v35	Skalometer Politiker
v36	Links-Rechts-Einstufung Parteien
v37	Links-Rechts-Selbsteinstufung
v38	Demokratiezufriedenheit
v39	Parteineigung
v40	Nachfrage Parteineigung
v41	Welche Parteineigung
v42	Staerke Parteineigung
v43	Propensity to Vote
v44	Leistung Bundesregierung
v45	Leistung Landesregierung
v46	Wichtigste Informationsquelle
v47a	Zweitwichtigste Informationsquelle
v47b	Drittwichtigste Informationsquelle
v48	Allgemeine Nutzung soziale Medien
v49	Politische Nutzung soziale Medien
v50/v51	Unkonventionelle Offline-Partizipation
v52	Soziale Medien, Nachfrage
v53	Mediennutzung Internet – Nachrichten
v54	Mediennutzung Internet – Social Media
v55	Mediennutzung Internet – Bias positiv
v56	Mediennutzung Internet – Bias positiv Partei
v57	Mediennutzung Internet - Bias negativ Partei
v58	Nutzung Tageszeitung
v59	Nutzung Tageszeitung – Haeufigkeit
v60	Parteien-Bias Tageszeitung - Bias positiv
v61	Parteien-Bias Tageszeitung – Bias positiv Partei
v62	Parteien-Bias Tageszeitung – Bias negativ Partei
v63	Rezeption Tagespresse: offline - online
v64	Nutzung TV-Nachrichten
v65	Parteien-Bias TV-Nachrichten - Bias positiv

---



---

v66	Parteien-Bias TV-Nachrichten – Bias positiv Partei
v67	Parteien-Bias TV-Nachrichten - Bias negativ
v68	Parteien-Bias TV-Nachrichten – Bias negativ Partei
v69	Rezeption TV-Nachrichten: offline - online
v70	TV-Sender: Bevorzugter Sender
v71	Gespraechе ueber Politik
v72	Gespraechspartner: Wahlentscheidung
v73	Gespraechspartner: wuerde Partei niemals waehlen
v74	Gespraechspartner: weitere Parteien, die GP waehlen wuerde
v75	Gespraechspartner: Wahlabsicht BTW
v76	Autoritarismus (ANES)
v77	Efficacy/Populismus/electoral Integrity
v78	Globalisierung
v79	Meinungsklima-Wahrnehmungen: Globalisierung
v80	Globalisierung, Ego
v81	Globalisierung, Salienz
v82	Politisches Wissen: Erst-/Zweitstimme
v83	Politisches Wissen: 5%-Huerde
v84a	Politisches Wissen: Zuordnung Politiker/Parteien
v85	Politisches Wissen: Haushaltsdefizit
v86	Politisches Wissen: Arbeitslosenquote
v87	Familienstand
v88	Partner vorhanden
v89	Organisationsmitgliedschaft
v4	Berufliche Bildung
v90	Erwerbstaetigkeit
v92	Fruehere Erwerbstaetigkeit
v94	Beruf
v95	Frueherer Beruf
v96	Beschaefigungssektor
v97	Nettoeinkommen HH, mit Kategorien
v98	Religionszugehoerigkeit
v99	Haeufigkeit Gottesdienst
v100	Geburtsland
v101	Geburtsland: anderes Land
v102	Geburtsland Eltern
v103	Im HH gesprochene Sprache
v104	Im HH gesprochene Sprache: andere Sprache
v105	Wohngegend
v106	Haushaltsgroesse
v107	Vertrauen, Institutionen
v108	Ego-Positionen
v109	Survey Evaluation
v110	Antwortprozess
v111	Gesamtbewertung

---

#### 5.6.4 Codierung von halboffene Fragen

Halboffene Fragen wurden codiert. Die Codierung erfolgte in der Regel nach Vorgaben und anhand von Codierschemata der GLES-Projektgruppe. Im Datensatz werden in der Regel nur die vercodeten Variablen veröffentlicht. Bei Interesse an den nicht vercodeten offenen Antworten, melden Sie sich bitte unter [gles@gesis.org](mailto:gles@gesis.org).

#### 5.6.5 Codierung der Parteien

Die Codierung der Parteien erfolgte bei allen geschlossenen, halboffenen und offenen Fragen nach dem einheitlichen Codierschema der GLES. Das Codierschema kann auf den Seiten der GLES bei GESIS abgerufen werden (<http://www.gesis.org/gles>).

In der Regel werden die Angaben der Befragten zu Parteien in zwei Variablen veröffentlicht, die mit Version A und Version B bezeichnet werden. Version A enthält für Parteienfragen stets die Parteien, die auf dem verwendeten Musterstimmzettel aufgeführt sind, sowie eine Kategorie für „andere“ Parteien. Version B unterscheidet sich von Version A ausschließlich durch eine detaillierte Ausweisung der von den Befragten genannten „anderen“ Parteien.

### 5.7. GewichtungsvARIABLEN

Mit Hilfe von Anpassungsgewichten kann die Verteilung von 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. Die für die GLES-Daten erstellten Gewichte werden mittels einer stets gleichen Vorgehensweise berechnet, so dass sie über die verschiedenen GLES-Komponenten hinweg vergleichbar sind.

Zur Berechnung der Anpassungsgewichte der GLES wurden sowohl sozial- als auch regionalstrukturelle Merkmale ausgewählt. Bei der gleichzeitigen Anpassung an mehrere Merkmale ergeben sich häufig Nullzellen oder sehr kleine Zellbesetzungen, was bei einer einfachen Soll/Ist-Gewichtung zu Problemen führt. Daher wurde bei der Berechnung der Anpassungsgewichte für die GLES auf das iterative Anpassungsverfahren („iterative proportional fitting“, IPF)<sup>2</sup> zurückgegriffen. 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.<sup>3</sup> Um sehr große Gewichtungsfaktoren zu vermeiden wurden die Faktoren (nach jedem Iterationsschritt) auf den fünffachen Mittelwert der Gewichtungsvariable (also fünf) getrimmt.<sup>4</sup>

---

<sup>2</sup> Siehe zum IPF-Verfahren Deming und Stephan (1940).

<sup>3</sup> Die Anpassungsgewichte wurden mit dem Statistikprogramm Stata berechnet, wobei auf das frei verfügbare Ado „ipfweight“ (Bergmann 2011) zurückgegriffen wurde. Zur Installation muss lediglich „ssc install ipfweight“ in die Kommandozeile von Stata eingegeben werden.

<sup>4</sup> Dieses Vorgehen findet auch bei der Berechnung der Gewichte der American National Election Study (ANES) Anwendung (siehe DeBell et al. 2009).

Bei der Erstellung der Anpassungsgewichte für den Pretest wurde auf die aktuellsten verfügbaren Randverteilungen des Mikrozensus<sup>5</sup> und des (N)Onliner Atlas<sup>6</sup> zurückgegriffen. Bei der Erstellung der Sollverteilungen des Mikrozensus ging nur die Bevölkerung über 18 Jahren mit deutscher Staatsbürgerschaft in Privathaushalten am Ort der Hauptwohnung ein. Die Sollverteilungen des (N)Onliner Atlas beruhen auf den Angaben zu Personen mit deutscher Staatsbürgerschaft ab einem Alter von 18 Jahren.

Für diese Studie wurden insgesamt vier Anpassungsgewichte erstellt, wobei die Gewichte einmal mit und einmal ohne Zeitunterschreiter berechnet wurden.

**Tabelle 12: Übersicht über die sozial- und regionalstrukturellen GewichtungsvARIABLEN**

Gewichtungsvariable	Anpassung an Randverteilungen des ...	Berücksichtigung von Zeitunterschreitern
wei_mzz	Mikrozensus 2013	ja
wei_mzoz	Mikrozensus 2013	nein
wei_onz	(N)Onliner Atlas 2014	ja
wei_onoz	(N)Onliner Atlas 2014	nein

Angepasst wurde an sozio-demographische und regionalstrukturelle Merkmale: Geschlecht, Alter, Bildung.

Die Variable Alter wurde kategorisiert und umfasst nun vier Gruppen: „18 bis unter 30 Jahre“, „30 bis unter 45 Jahre“, „45 bis unter 60 Jahre“ und „60 Jahre und älter“.

Bei der Kategorisierung der Variable Bildung wurden die drei folgenden Gruppen gebildet:

- 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).

Die BIK-Regionsgrößenklassen wurden zu zwei bzw. drei Kategorien zusammengefasst (siehe nachfolgende Tabelle). Die unterschiedliche Kategorisierung ist den verfügbaren Randverteilungen geschuldet. Für die Anpassung an den Mikrozensus wurde an drei Kategorien angepasst, während beim (N)Onliner Atlas lediglich die Anpassung an zwei Kategorien möglich war.<sup>7</sup> Fehlende Werte wurden für die Erstellung der GewichtungsvARIABLEN auf den Modalwert der Verteilung gesetzt.

<sup>5</sup> Bei den vom Mikrozensus ausgewiesenen Werten handelt es sich um absolute Personenzahlen nach gebundener Hochrechnung.

<sup>6</sup> Die Randverteilungen (für Personen ab 18 Jahren) aus dem (N)Onliner Atlas wurden für die Gewichtung freundlicherweise von TNS Infratest zur Verfügung gestellt.

<sup>7</sup> Die Zusammenfassung der Variablen erfolgte entsprechend dem Vorgehen in anderen GLES-Komponenten.

**Tabelle 13: Soll- und Ist-Verteilungen der zur Erstellung der Gewichte verwendeten Variablen**

Merkmal	Ist in Prozent		Soll in Prozent	
	Mit Zeitunter- schreibern (N=1.847)	Ohne Zeitun- terschreiber (N=1.592)	Mikrozensus	(N)Onliner Atlas
<b>Geschlecht</b>				
Männlich	46,89	47,17	48,4	52,6
Weiblich	53,11	52,83	51,6	47,4
<b>Altersgruppe</b>				
18 bis unter 30 Jahre	18,95	17,46	16,0	21,3
30 bis unter 45 Jahre	30,10	28,45	21,7	27,3
45 bis unter 60 Jahre	31,73	32,79	28,4	32,9
60 Jahre und älter	19,22	21,29	33,9	18,5
<b>Bildung</b>				
Niedrig	30,59	30,84	41,2	34,9
Mittel	40,34	39,70	29,7	31,3
Hoch	29,07	29,46	29,1	33,8

Die Tabelle 18 gibt die Anzahl der zur Berechnung notwendigen Iterationen sowie deskriptive Statistiken zu den resultierenden Gewichtungsfaktoren wieder.

**Tabelle 14: Deskriptive Statistiken zu den Gewichtungsfaktoren**

Statistik	wei_mzz	wei_mzoz	wei_onz	wei_onoz
Anzahl Iterationen <sup>8</sup>	4	4	4	4
Mittelwert	1,00	1,00	1,00	1,00
Standardabweichung	0,43	0,37	0,22	0,23
Minimum	0,55	0,59	0,64	0,58
Maximum	2,20	2,00	1,41	1,51
1. Quartil	0,67	0,73	0,79	0,82
Median	0,89	0,94	1,00	1,01
3. Quartil	1,07	1,11	1,16	1,22
Max./Min. <sup>9</sup>	4,00	3,39	2,20	2,60
N	1.847	1.592	1.847	1.592

<sup>8</sup> Die Anzahl der Iterationen gibt an, wie viele Anpassungsschritte bei der Gewichtung vorgenommen wurden bis die Abweichung zwischen den gewichteten Istwerten in der Stichprobe und den Sollwerten aus der Referenzverteilung weniger als 0,05 betrug. Bei Erreichen von 20 Iterationsschritten wird die Anpassung ebenfalls beendet.

<sup>9</sup> 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.

## 5.8. Unveröffentlichte Variablen

Aus datenschutzrechtlichen Gründen können nicht alle erhobenen Variablen zum freien Download zur Verfügung gestellt werden. Generell müssen frei verfügbare Datensätze so aufgebaut sein, dass ein unverhältnismäßiger Aufwand zur De-Anonymisierung betrieben werden müsste. Das heißt, die Datensätze müssen faktisch anonym sein.

In diesem Datensatz wurden Systemvariablen aus dem veröffentlichten Datensatz entfernt, die lediglich zur Durchführung der Online-Befragung relevant sind, aber keine inhaltliche Aussagekraft haben. Darüber hinaus werden offen erhobene Antworten im Normalfall nur als vercodete Variablen veröffentlicht. Bei Interesse an den entfernten Systemvariablen oder den nicht vercodeten offenen Antworten, melden Sie sich bitte ebenfalls unter [gles@gesis.org](mailto:gles@gesis.org).

## 5.9. Fehlende Werte

Fehlende Werte wurden bei allen Variablen nach dem einheitlichen Codierschema der GLES vergeben, welches negative Werte im Bereich -71 bis -99 und in Stata Codes von .a bis .p vorsieht. In den SPSS-Datensätzen sind diese Werte standardmäßig als fehlende Werte definiert. Den Stata -Datensätzen hingegen liegt ein Do-File bei, anhand dessen die fehlenden Werte in die vorgesehenen Missing-Value-Codes recodiert werden können.

**Tabelle 15: Übersicht über die Missing-Value-Codes der GLES**

Code	Stata Missing-Value-Code	Label
-99	.a	keine Angabe
-98	.b	weiß nicht
-97	.c	trifft nicht zu
-96	.d	Split
-95	.e	nicht teilgenommen
-94	.f	nicht in Auswahlgesamtheit
-93	.g	Interview abgebrochen
-92	.h	Fehler in Daten
-86	.i	nicht wahlberechtigt
-85	.j	nicht wählen
-84	.k	keine Erst-/Zweitstimme abgeben
-83	.l	ungültig wählen
-82	.m	keine andere Partei wählen
-81	.n	noch nicht entschieden
-72	.o	nicht einzuschätzen
-71	.p	nicht bekannt

## 5.10. Analyse-Systeme

Der Datensatz ist für die Analyse-Systeme SPSS (IBM) und Stata (StataCorp LP) verfügbar.

**Tabelle 16: Analysesysteme**

Analyse-System	Dateiformat	Versionsbeschränkungen
SPSS	.sav	Version 18 oder höher
Stata	.dta	Version 12 oder höher

## **6. Hinweise und Anmerkungen**

### **6.1. Errata**

Eine aktuelle Errataliste kann über den GESIS Datenkatalog ([www.gesis.org/dbk](http://www.gesis.org/dbk)) abgerufen werden. Dort werden auch alle Änderungen zwischen den einzelnen Versionen der Datensätze dokumentiert.

## Links

Projektseite der GLES: [www.gles.eu](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>



## Literaturverzeichnis

- AAPOR (2011): Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 7. Aufl. The American Association for Public Opinion Research.
- Bergmann, Michael (2011): IPFWEIGHT: Stata module to create adjustment weights for surveys. Statistical Software Components S457353, Boston: Boston College Department of Economics.
- BIK Aschpurwis+Behrens (2001): BIK Regionen: Ballungsräume, Stadtregionen, Mittel-/Unterzentrengebiete. Methodenbeschreibung zur Aktualisierung 2000, [www.bik-gmbh.de/texte/BIK-Regionen2000.pdf](http://www.bik-gmbh.de/texte/BIK-Regionen2000.pdf) [05.11.2011].
- Callegaro, Mario (2013): Paradata in Web Surveys, in: Kreuter, Frauke (Hrsg.), Improving Surveys with Paradata: Analytic Uses of Process Information. (Aufl.) Hoboken, NJ: Wiley, 261-280.
- Callegaro, Mario und Charles Disogra (2008): Computing Response Metrics for Online Panels, in: Public Opinion Quarterly 72, 1008-1032.
- Couper, Mick P. (2000): Usability Evaluation of Computer-Assisted Survey Instruments, in: Social Science Computer Review 18, 384-396.
- 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 und 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: 427-444.
- Kaczmirek, Lars (2008): Human-Survey Interaction. Usability and Nonresponse in Online Surveys. Universität Mannheim, Mannheim.
- Krosnick, Jon A. (1991): Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys, in: Applied Cognitive Psychology 5, 213-236.
- Mayerl, Jochen und Dieter Urban (2008): Antwortreaktionszeiten in Survey-Analysen. Messung, Auswertung und Anwendung. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Roßmann, Joss (2010): Data Quality in Web Surveys of the German Longitudinal Election Study 2009. 3rd ECPR Graduate Conference. Dublin.
- Schmitt-Beck, Rüdiger, Hans Rattinger, Sigrid Roßteutscher und Bernhard Weßels (2010): Die deutsche Wahlforschung und die German Longitudinal Election Study (GLES), in: Faulbaum, Frank & Wolf, Christof (Hrsg.), Gesellschaftliche Entwicklungen im Spiegel der empirischen Sozialforschung. Wiesbaden: VS Verlag für Sozialwissenschaften, 141-172.