

gesis

Leibniz-Institut
für Sozialwissenschaften

German Longitudinal
Election Study



GLES 2017

Kandidatenstudie

ZA6814, Version 3.0.0

Studienbeschreibung

Inhaltsverzeichnis

Vorbemerkungen	3
1. Studieninformationen	4
2. Konzeption	5
3. Stichprobenziehung.....	7
4. Erhebung	8
5. Datenaufbereitung.....	9
6. Hinweise und Anmerkungen.....	20
Links	21
Literaturverzeichnis.....	22

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:

Roßteutscher, Sigrid; Schoen, Harald; Schmitt-Beck, Rüdiger; Weißels, Bernhard; Wolf, Christof; Giebler, Heiko; Melcher, Reinhold; Wagner, Aiko (2018): Kandidatenstudie (GLES 2017). GESIS Datenarchiv, Köln: ZA6814 Datenfile Version 3.0.0, doi: 10.4232/1.13089.

1. Studieninformationen

1.1. Studiennummer

ZA6814

1.2. Titel der Studie

German Longitudinal Election Study, Komponente 6, Kandidatenstudie (GLES 2017)

1.3. Version

3.0.0, 21.08.2018, doi: 10.4232/1.13089

1.4. Erhebungszeitraum

Versand Fragebögen und Zugangscodes: 25. Oktober 2017

Letzter Eingang Rücklauf: 07. Februar 2018

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)

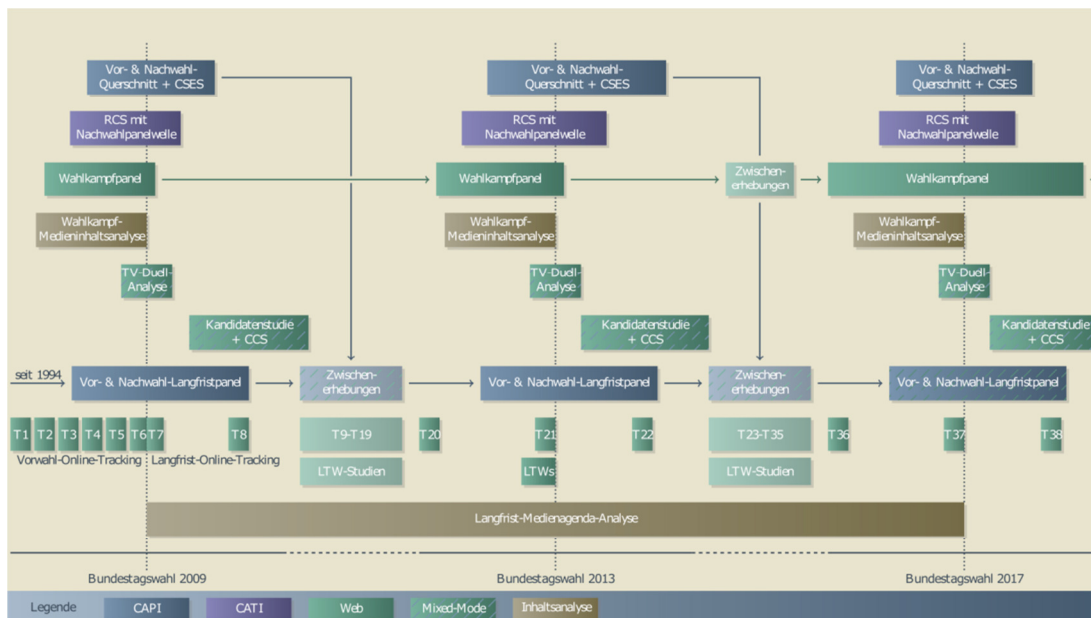
2. Konzeption

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 sowohl aus Quer- als auch Längsschnittsperspektive, in Hinblick auf kurzfristige Dynamiken während des Wahlkampfs sowie hinsichtlich langfristiger sozialer Wandlungsprozesse über den Wahlzyklus hinweg (Schmitt-Beck et al. 2010).

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



Die Deutsche Kandidatenstudie zur Bundestagswahl 2017 wurde als Teilprojekt (Komponente 6) im Rahmen der German Longitudinal Election Study 2017 (GLES) und in Kooperation mit dem Comparative Candidates Survey (CCS), einem international vergleichenden Forschungsprojekt zu Kandidaten¹ in Wahlkämpfen, durchgeführt. Die Befragung nutzt ein Dual-Mode-Design (postalisch und online). Zur Befragung wurden alle Direkt- und Listenkandidaten der im Bundestag vertretenden Parteien sowie die Direkt- und Listenkandidaten der FDP und AfD eingeladen.

Der Kern der Befragung besteht aus Daten zum Hintergrund der Befragten, ihrer Rekrutierung und Selektion, ihres Kampagnenverhaltens, ihrer Einstellungen sowie ihrer Policy-

¹ Aus Gründen der besseren Lesbarkeit wird auf die gleichzeitige Verwendung männlicher und weiblicher Sprachformen verzichtet. Sämtliche Personenbezeichnungen gelten gleichermaßen für beiderlei Geschlecht.

Positionen. Einige Fragen werden sowohl in der Kandidaten- als auch in der Querschnitts-Wählerbefragung (Komponente 1) gestellt, um so zum einen den Vergleich der Einstellungen von Wählern und Kandidaten zu ermöglichen und zum anderen die Bedeutung lokaler Kampagnen für das Wahlverhalten zu analysieren. Aggregat-Charakteristika der Wahlkreise aus der amtlichen Statistik erlauben zudem, den soziodemographischen Kontext mit einzu-beziehen. Im Rahmen der internationalen Kooperation mit dem Comparative Candidates Survey (CCS) enthält das Befragungsinstrument der Kandidatenstudie 2017 auch den CCS-Kernfragebogen und ermöglicht damit den internationalen Vergleich von Kandidaten, ihren Einstellungen und ihrem Verhalten, etwa im Wahlkampf.

2.1.1 Kernfragen

Selektion und Nominierung; politische Tätigkeiten und Ämter; Wahlkampftätigkeiten; Positionssues; Demokratie und politische Repräsentation.

2.1.2 Soziodemographie

Alter (Geburtsjahr); Geschlecht; höchster allgemeinbildender Schulabschluss; derzeitiger bzw. früherer Erwerbsstatus; derzeitiger bzw. früherer Beruf; beruflicher Status; Haushaltsgröße; Anzahl Kinder im Haushalt; Beschreibung des Wohnort; Familienstand; Konfession; Kirchgangshäufigkeit.

3. Stichprobenziehung

3.1. Untersuchungsgebiet

Deutschland (DE)

3.2. Grund- und Auswahlgesamtheit

Die Auswahlgesamtheit wurde auf Kandidaten derjenigen Parteien begrenzt, die vor der Wahl 2017 im Bundestag vertreten waren: SPD, CDU, CSU, Bündnis 90/Die Grünen und Die Linke. Diese wurden um Kandidaten der nach Umfragen im Vorfeld der Bundestagswahl größten, nicht im Parlament vertreten Parteien FDP und AfD ergänzt. Diese Beschränkung erfolgte aus der Überlegung heraus, vorwiegend Informationen über Kandidaten der im politischen Prozess relevanten Parteien zu erheben.

Die Auswahlgesamt betrug 2516 Kandidaten: 51% davon waren sowohl Wahlkreis- als auch Listenkandidaten (N=1276), 29% (N=737) reine Listenkandidaten und 20% entfielen auf Wahlkreiskandidaten (N=503).

4. Erhebung

4.1. Erhebungsverfahren

Postalische Befragung mit Option zur Beantwortung eines identischen Online-Fragebogens

4.2. Datenerhebung

Die Programmierung des Online-Fragebogens und die Durchführung der Datenerhebung wurden vom WZB in Berlin durchgeführt.

4.3. Erhebungssoftware

Die Programmierung des Online-Fragebogens erfolgte mittels der Software LimeSurvey.

4.4. Tests des programmierten Fragebogens

Der Online-Fragebogen wurde von GLES-Mitarbeitern intensiv geprüft. In den Tests gefundene Probleme oder Programmierfehler wurden bereinigt. Anschließend wurde der Fragebogen in der finalen Version von der Projektgruppe freigegeben.

4.5. Feldarbeit

Die Befragung erfolgte postalisch mit einer Option zur Teilnahme via Internet. Hierzu wurde neben einem schriftlichen Fragebogen auch ein identischer Online-Fragebogen zur Verfügung gestellt, der passwortgeschützt bzw. über per E-Mail versendete personalisierte Links zugänglich war. Für die Erstellung der Online-Version des Fragebogens wurde die Abteilung „Wissenschaftliche Information“ (WIN) des WZB beauftragt.

Das erste Anschreiben einschließlich Fragebogen, individualisiertem Zugangscode für den inhaltlich identischen Online-Fragebogen und adressiertem sowie frankiertem Rückumschlag wurde am 25. Oktober 2017 verschickt. Am 03. November folgte eine Einladung per E-Mail, die einen direkten Link zur Befragung beinhaltete. Eine erste Erinnerung wurde per E-Mail am 17. November versandt. Per Post wurde am 24. November eine zweite Erinnerung zusammen mit einer Postkarte verschickt, mit der bei Bedarf ein neuer Fragebogen angefordert werden konnte. Da die Kandidaten der CDU und CSU Ende November noch deutlich unterrepräsentiert waren, wurde am 05. Dezember per E-Mail eine parteispezifische Erinnerung versandt. Die letzte Erinnerung für 2017 wurde am 15. Dezember und eine abschließende Erinnerung am 12. Januar 2018 versandt (beide jeweils per E-Mail).

4.6. Ausschöpfung und Inklusionskriterium

Insgesamt haben von den 2516 kontaktierten Kandidaten 998 Personen die Befragung zumindest begonnen. Der vorliegende Datensatz beinhaltet aber nur jene Personen, bei denen für mindestens 75% der (numerischen) Variablen valide Information vorliegen. Somit werden Personen, welche die Befragung schon nach der Hälfte der Fragen abgebrochen haben, nicht berücksichtigt, wohl aber jene, die die Befragung erst kurz vor dem Ende abgebrochen haben.

Insgesamt weisen 803 Kandidaten die notwendige Anzahl valider Antworten auf. Dies entspricht einer Ausschöpfung von 32%, wobei davon 186 Personen bei der Bundestagswahl 2017 ein Mandat errungen haben.

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 wurden 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) dokumentiert und zeitnah behoben.

5.2. Einheiten im Datensatz

Der Datensatz umfasst 803 realisierte Interviews. Davon wurden 363 online und 440 postalisch durchgeführt.

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 5.9). Tabelle 1 gibt einen Überblick über die Anzahl der Variablen im Datensatz nach Variablentyp.

Tabelle 1: Anzahl der Variablen im Datensatz nach Variablentyp

Variablentyp	Anzahl
Metadaten	6
Paradaten	2
Substantielle Daten	310
Kandidatendaten des Bundeswahlleiters	8
Regionaldaten der amtlichen Statistik	249
Gewichtungsvariablen	2
Gesamt	577

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 2).

Tabelle 2: Metadaten

Variable	Label
study	Studiennummer
version	GESIS Archiv Version
doi	Digital Object Identifier
year	Erhebungsjahr
field	Erhebungszeitraum
glescomp	GLES-Komponente

5.5. Parادات

Parادات 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 Gerat 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 Parادات behandelt. Tabelle 3 gibt einen Uberblick Uber Parادات im Datensatz.

Tabelle 3: Parادات

Variable	Label
lfdn	Laufende Nummer
mode	Teilnahmeart

Laufende Nummer und Teilnahmeart

Die Variable *lfdn* „Laufende Nummer“ ist eine eindeutige Identifikationsnummer fur die Teilnehmer an der Befragung. Die Variable *mode* „Teilnahmeart“ gibt Auskunft daruber, ob der Befragte postalisch oder online teilgenommen hat.

5.6. Substantielle Daten

Substantielle Daten bilden den Hauptbestandteil des Datensatzes. Sie umfassen inhaltliche und soziodemographische Variablen. Der Datensatz umfasst Variablen zu funf verschiedenen Themengebieten. Die Variablenbezeichnungen lassen sich dabei den Themengebieten des Fragebogens zuordnen: Politischer Hintergrund und politische Aktivitaten (A), Wahlkampf (B), Politische Themen (C), Demokratie und Reprasentation (D), Angaben zur Person (E).

Bei Variablen wie etwa dem Geburtsjahr oder Geschlecht, fur die offizielle Informationen des Bundeswahlleiters vorlagen, wurden diese auch verwendet. Die Informationen aus der Befragung finden sich ebenfalls im Datensatz und sind mit einem „_org“ gekennzeichnet. Zusatzlich wurden Jahresangaben zu Jahreszahlen recodiert, falls dies eigentlich gefragt war. Schlielich erfolgte auch eine Reprasentierung der Wahlkampfmittelquellen (b6a bis b6c), soweit die Prozentangaben nicht in Summe 100 ergaben oder Geldbetrage angegeben wurden.

Tabelle 4: Inhaltliche und soziodemographische Variablen

Variable	Label
a1	Kandidat: Antritt Partei
a2	Kandidat: Jahr Parteieintritt
a3a-e	Kandidatur: 2013, 2009, 2005, 2002, 1998
a4a-j	Kandidat, politische Laufbahn
a5	Kandidatur: Zeit für Parteiaktivitäten
a6	Nominierung: Ermutigung Kandidatur
a7a-b,as-bs	Einfluss auf Nominierung: Wahlkreiskandidatur/ Listenkandidatur (andere)
a8a-b	Nominierung umkämpft: Wahlkreiskandidatur/ Listenkandidatur
a9a-b	Wahlkampf, Wohnsitz im Wahlkreis, Jahr
a10aa-ag,gs	Wahlkampf, Mitgliedschaft: Gewerkschaft, Wirtschafts-/Berufsvereinigung, kirchliche Organisation, Umweltorganisation, Menschen/Bürgerrechtsorganisation, Verein (andere)
a10ba-bg,gs	Wahlkampf, Unterstützung: Gewerkschaft, Wirtschafts-/Berufsvereinigung, kirchliche Organisation, Umweltorganisation, Menschen/Bürgerrechtsorganisation, Verein (andere)
a11a-h	Kontaktierung durch Bürger
b1a-b	Wahlkampf: Beginn, Beginn Vollzeit
b2	Wahlkampf: Zeitaufwand
b3a-c	Wahlkampf: Teamumfang, insgesamt/ aus Wahlkampfbudget/ unbezahlt
b4	Wahlkampf: professionelle Beratung
b5	Wahlkampf: Budget
b6a-c	Wahlkampf: Budget, Anteil Parteimittel/ Spenden/ eigene private Mittel
b7a-u	Wahlkampf, Aktivitäten
b8aa-ae	Wahlkampf: Nutzungsgründe Facebook
b8ba-be	Wahlkampf: Nutzungsgründe Twitter
b9	Wahlkampf: Fokus der Aufmerksamkeit
b10,s	Wahlkampf: lokale Themen (andere)
b11a-i	Wahlkampf, Themenfokus
b12a-e	Wahlkampf, Negative Campaigning
b13a-b	Wahlkampf, Einschätzung Gewinnchancen Beginn: Wahlkreis-/ Listenkandidatur
b14	Wahlkampf, Medienberichterstattung: Häufigkeit
b15	Wahlkampf, Medienberichterstattung: Schwierigkeit
b16a-b	Wahlkampf, Medienberichterstattung: Partei/ Person
b17a-d	Wahlkampf, Medienberichterstattung: Einfluss auf Wahlprogrammthemen/ -kampfführung/ -verhalten/ politisches Wissen
b18a-c	Wahlkampf, Parteiprogramm
b19a-h	Wahlkampf, Faktoren für Wahlergebnis

b20a-h	Wahlkampf, Wählerwettbewerb: CDU, CSU, SPD, Die Linke, Bündnis 90/Die Grünen, FDP, AfD, Nichtwähler
c1as	Wichtigstes Problem (offen)
c1bs	Zweitwichtigstes Problem (offen)
c1cs	Drittwichtigstes Problem (offen)
c2a-n	Einstellungen
c3	Links-Rechts: Selbsteinstufung
c4a-g	Links-Rechts, Einstufung Partei
c5	Links-Rechts: Einstufung eigene Wähler
c6	Libertär-Autoritärer Issue Ausländer: Ego
c7	Libertär-Autoritärer Issue Ausländer: eigene Partei
c8	Libertär-Autoritärer Issue Ausländer: eigene Wähler
c9	Sozio-Ökonomischer Issue: Ego
c10	Sozio-Ökonomischer Issue: eigene Partei
c11	Sozio-Ökonomischer Issue: eigene Wähler
c12	Umweltissue Klima: Ego
c13	Umweltissue Klima: eigene Partei
c14	Umweltissue Klima: eigene Wähler
c15	Europa: Bewertung Mitgliedschaft
c16	Europa, Einigung: Ego
c17	Europa: Demokratiezufriedenheit
c18a-i	Staatsausgaben
c19a-f	Bewältigung internationaler Probleme
c20	BRD: Demokratiezufriedenheit
d1	Repräsentationsstil: Wähler vs. Partei
d2	Repräsentationsstil: eigene Meinung vs. Wähler
d3	Repräsentationsstil: eigene Meinung vs. Partei
d4	Repräsentationsstil: Wähler vs. Partei vs. eigene Meinung
d5a-e	Repräsentationsnorm
d6a-h	Aussagen Demokratie
d7a-e	Aussagen innerparteiliche Demokratie
d8a-g	Qualität politischer Entscheidungen
e1a-j	Psychologische Konstrukte
e2	Wohnort: Größe
e3	Wohnort: Dauer
e4	Wohnort: Beschreibung

e5,s	Geburtsland: Deutschland / Ausland (offen)
e6,s	Deutsche Staatsbürgerschaft seit Geburt / andere (offen)
e7a-b,as,bs	Geburtsland Eltern: Deutschland, Mutter/Vater, anderes Land (offen)
e8a-b,as,bs	Staatsbürgerschaft Eltern bei Geburt: Mutter/ Vater, andere (offen)
e9,s	Schulabschluss, anderer (offen)
e10	Berufliche Bildung
e10as-cs	Berufliche Bildung, Hochschulabschluss Bachelor/ Master/ anderer (offen)
e11,as,bs	Erwerbsstatus, früherer, arbeitslos/ Rente, Pension, Vorruhestand (offen)
e12s	Beruf (offen)
e13	Kirchgangshäufigkeit
e14,s	Religion, andere (offen)
e15	Familienstand
e16a-f	Haushalt, weitere Personen, Geburtsjahr, 1.-6. Person
e17a-f	Kinder, nicht im Haushalt, 1.-6. Kind
e18a-e	Häufigkeit Tätigkeit
e21as-bs	Bedeutung: Links/ Rechts (offen)

Die offene Abfrage der drei wichtigsten Probleme in Deutschland (c1as-cs), der lokalen Wahlkampfthemen (b10s) sowie der offenen Berufsnennungen (e12s) wurden vom Bamberger Centrum für Empirische Studien (BACES) manuell codiert. Die Codierung der drei wichtigsten Probleme in Deutschland sowie der lokalen Wahlkampfthemen erfolgte unter Verwendung von Codierschemata, die von der GLES-Projektgruppe entwickelt und bereitgestellt wurden. Die offenen Berufsnennungen wurden nach ISCO 2008 codiert. Diese können auf den Seiten der GLES abgerufen werden (<http://www.gesis.org/gles>).

Darüber hinaus werden die offenen Fragen als String-Variablen in einer externen Datei bereitgestellt. Hierfür wurden die offen erhobenen Antworten intensiv auf datenschutzrechtlich bedenkliche Nennungen geprüft und nötigenfalls zum Schutz der Befragten zensiert.

5.7. Kandidateninformationen des Bundeswahlleiters

Zuzüglich zu den Antworten der Befragten wurden Kandidateninformationen herangespielt, die vom Bundeswahlleiter bezogen wurden. Die Angaben zum Kandidaturtyp dienen bei der Datenaufbereitung der schriftlichen Befragungen zur Überprüfung der Filter für Fragen, die nur an Listen- oder nur an Wahlkreiskandidaten gerichtet waren.

Tabelle 5: Kandidateninformationen Bundeswahlleiter

Variable	Label
kandidaturtyp	Kandidaturtyp
wkname	Wahlkreisname
wknr	Wahlkreisnummer
bundesland	Bundesland

listenplatz	Listenplatz
gewaehlt	Erfolgreiche Kandidatur?
geschlecht	Geschlecht
geburtsjahr	Geburtsjahr

5.8. Regionaldaten der amtlichen Statistik

Neben den Antworten aus der Befragung wurden dem Datensatz auch Regionaldaten der amtlichen Statistik auf Wahlkreis- und Bundeslandebene zugespielt. Die Informationen zu den Wahlkreisen selbst sind in den Variablen ohne Suffix zu finden. Informationen auf Bundeslandebene sind hingegen durch das Suffix „_l“ und auf nationaler Ebene durch „_ges“ gekennzeichnet. Eine detaillierte Erklärung der einzelnen Strukturdaten findet sich auf den Seiten des Bundeswahlleiters². Die zugespielten Daten sind:

Tabelle 6: Regionaldaten

Variable	Label
Wahlstatistik	
wk_gewinner	Wahlkreisgewinner 2017
kand1st17(_l)/-13(_l) kand2st17(_l)/-13(_l) kand1st17/-13_ges kand2st17/-13_ges	Erst- und Zweitstimme des Kandidaten 2017 und 2013 pro Wahlkreis, Bundesland und gesamt
wahl_ber17(_l) wahl_ber13(_l)	Wahlberechtigte 2017 und 2013 Erst- und Zweitstimmen pro Wahlkreis und Bundesland
wbet17(_l) wbet13(_l)	Wahlbeteiligung 2017 und 2013 in Prozent pro Wahlkreis und Bundesland
wa1st17(_l) wa1st13(_l)	Wähler 2017 und 2013 pro Wahlkreis und Bundesland
un1s17(_l)/-13(_l) un2st17(_l)/-13(_l) gu1st17(_l)/-13(_l) gu2st17(_l)/-13(_l)	Gültige und ungültige Erst- und Zweitstimmen 2017 und 2013 pro Wahlkreis und Bundesland
cdu-, spd-, fdp, lin-, gru-, csu-, afd1st(_l)(_p) und -2st(_l)(_p)	Erst- und Zweitstimmen der Parteien (CDU, SPD, FDP, Die Linke, Bündnis 90/Die Grünen, CSU, AfD) pro Wahlkreis und Bundesland (absolut und in Prozent)
Regionaldaten	
gemeinden(_l)	Anzahl der Gemeinden pro Wahlkreis und Bundesland (am 31.12.2015)
flaeche(_l)	Fläche in km ² pro Wahlkreis und Bundesland (am 31.12.2015)
bev(_l) bev_deutsch(_l) bev_ausl(_l)	Bevölkerungszahl in 1000 (insgesamt, deutsche Staatsbürgerschaft, Ausländer) pro Wahlkreis und Bundesland (am 31.12.2015)
bev_dichte(_l)	Bevölkerungsdichte pro Wahlkreis und Bundesland (am 31.12.2015)

² <https://www.bundeswahlleiter.de/bundestagswahlen/2017/strukturdaten.html>

bev_gebsaldo(_I) bev_wandsaldo(_I)	Geburten- und Wanderungssaldo 2015 je 1000 Einwohner pro Wahlkreis und Bundesland
bev_om(_I) bev_mm(_I)	Zensus 2011, Bevölkerung ohne/mit Migrationshintergrund in Prozent pro Wahlkreis und Bundesland (09.05.2011)
alter_u18(_I) alter_18_24(_I) alter_25_34(_I) alter_35_59(_I) alter_60_74(_I) alter_75plus(_I)	Altersgruppen in Prozent, unter 18 Jahre, 18- 24 Jahre, 25-34 Jahre, 35-59 Jahre, 60-74 Jahre und 75 Jahre und älter (am 31.12.2015)
rel_kath(_I) rel_evang(_I) rel_sonstige(_I)	Zensus 2011, Religionszugehörigkeit (katholisch, evangelisch, sonstige) in Prozent pro Wahlkreis und Bundesland (am 09.05.2011)
eigenq(_I)	Zensus 2011, Wohnungen in Wohngebäuden Eigentümerquote pro Wahlkreis und Bundesland (am 09.05.2011)
abschluss(_I) abschluss_kein(_I) abschluss_hs(_I) abschluss_rs(_I) abschluss_brfs(_I) abschluss_gy(_I)	Angaben zu Bildungsabschlüssen (alle Absolventen, keine Abschlüsse, Hauptschul-, Realschulabschluss, Hochschulreife, berufliche Schulen) je 1000 Einwohner pro Wahlkreis und Bundesland (2015)
bip_pp(_I)	Bruttoinlandsprodukt 2014 in Euro je Einwohner pro Wahlkreis und Bundesland
einkm(_I)	Verfügbares Einkommen privater Haushalte 2014 in Euro je Einwohner pro Wahlkreis und Bundesland
kndbetr(_I)	Kindertagesbetreuung, betreute Kinder je 1000 Einwohner pro Wahlkreis und Bundesland (am 01.03.2016)
kfz_anz(_I)	Kraftfahrzeugbestand pro 1000 Einwohner pro Wahlkreis und Bundesland (am 01.01.2016)
firmen_gesamt(_I) firmen_hndwrk(_I)	Unternehmensregister 2014, Unternehmen und Handwerksunternehmen je 1000 Einwohner pro Wahlkreis und Bundesland
wohnung_bau(_I)	Fertiggestellte Wohnungen je 1000 Einwohner 2014 pro Wahlkreis und Bundesland
wohnung_bestand(_I)	Bestand an Wohnungen je 1000 Einwohner pro Wahlkreis und Bundesland (am 31.12.2015)
svpfl_besch_gesamt(_I) svpfl_besch_forst(_I) svpfl_besch_prod(_I) svpfl_besch_handel(_I) svpfl_besch_oeffentlich(_I) svpfl_besch_sonstige(_I)	Sozialversicherungspflichtige Beschäftigte (insgesamt, Land-/Forstwirtschaft und Fischerei, Produzierendes Gewerbe, Handel/Gastgewerbe/Verkehr, öffentliche und private Dienstleister, sonstige) pro Wahlkreis und Bundesland (am 30.06.2016)
alq(_I) alq_mann(_I) alq_frau(_I) alq_alter_15_19(_I) alq_alter_55_65(_I)	Arbeitslosenquote Ende März 2017 (insgesamt, Männer, Frauen, 15-19-Jährige, 55-65-Jährige) pro Wahlkreis und Bundesland
hartziv(_I) hartziv_ne(_I) hartziv_ausl(_I)	SGB II Empfänger je 1000 Einwohner (insgesamt, nicht erwerbsfähige Hilfsbedürftige, Ausländer) pro Wahlkreis und Bundesland (am 31.12.2016)

5.9. GewichtungsvARIABLEN

Für die Kandidatenstudie (ZA5716) wurden zwei Anpassungsgewichte erstellt. Die Datensätze enthalten die folgenden Gewichte:

Tabelle 7: Übersicht über die Gewichte

Gewicht	Variablenname
Grundgewicht Kandidaten	wei_kandi
Grundgewicht MandatsträgerInnen	wei_mdb

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.

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)³ 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.⁴ Um sehr große Gewichtungsfaktoren zu vermeiden, wurden die Faktoren (nach jedem Iterationsschritt) auf den fünffachen Mittelwert der Gewichtungsvariable (also fünf) getrimmt.⁵

³ Siehe zum IPF-Verfahren Deming und Stephan (1940).

⁴ 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.

⁵ Dieses Vorgehen findet auch bei der Berechnung der Gewichte der American National Election Study (ANES) Anwendung (siehe DeBell et al. 2009).

Tabelle 8: Soll- und Ist-Verteilungen (in Prozent)

Merkmal	Kandidaten		Mandatsträger	
	Soll	Ist	Soll	Ist
Parteizugehörigkeit				
SPD	19,04	18,68	21,58	30,11
CDU	19,96	12,83	28,21	24,19
CSU	3,58	1,74	6,49	2,15
FDP	14,59	16,44	11,28	10,75
Bündnis 90/Die Grünen	14,31	20,67	9,45	11,29
Die Linke	14,11	16,31	9,73	11,29
AfD	15,42	13,33	13,26	10,22
Geschlecht				
männlich	67,85	69,24	69,25	63,44
weiblich	32,15	30,76	30,75	36,56
Altersgruppe				
unter 30 Jahre	10,65	11,21	1,69	1,61
30 bis unter 45 Jahre	29,09	21,92	29,48	24,73
45 bis unter 60 Jahre	43,48	46,82	52,33	56,99
60 Jahre und älter	16,77	20,05	16,50	16,67
Kandidaturtyp				
Wahlkreis	19,99	23,29	10,01	7,53
Liste	29,29	21,42	3,10	2,69
Wahlkreis und Liste	50,72	55,29	86,88	89,78

Bei der Berechnung der Gewichtungsfaktoren konnte das Abbruchkriterium von 0,05 für die Anpassung an die Kandidaten/-innen nach vier Schritten, für die Anpassung an die Mandatsträger/-innen nach sechs Schritten erreicht werden. Eine Übersicht über die Anzahl der Iterationen und die maximale Abweichung bietet die nachfolgende Tabelle.

Tabelle 9: Anzahl der Iterationen und maximale Abweichung

Gewicht	Anzahl Iterationen	Maximum deviation (in percentage points)
Grundgewicht Kandidaten/-innen	4	0,012
Grundgewicht Mandatsträger/-innen	6	0,018

Eine Begrenzung der Gewichtungsfaktoren auf den Faktor 5 – wie teilweise in den anderen Komponenten der GLES – war nicht notwendig. Eine Übersicht über die Gewichtungsfaktoren bietet die untenstehende Tabelle.

Tabelle 10: Übersicht über die Gewichtungsfaktoren

	Obs	Mean	Std.Dev	Min	Max	1.Q	Median	3.Q	max/min ⁶
Grundgewicht Kandidaten/- innen	803	1	0,35	0,510	2,651	0,766	0,915	1,153	6,896
Grundgewicht Mandatsträger/- innen	186	1	0,39	0,542	3,740	0,726	0,917	1,146	5,199

5.10. 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.

In den Interviews, die schriftlich durchgeführt wurden, traten in manchen Fällen Eingabefehler auf, da Befragte Filterführungen ignorierten. Diese wurden mit dem Code -92 kenntlich gemacht. Wenn aufgrund der ignorierten Filterführung fälschlicherweise offene Angaben erfolgten, wurden diese jedoch nicht gelöscht.

Tabelle 11: 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

⁶ 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, sodass ein niedriger Wert positiv zu betrachten ist.

-81	.n	noch nicht entschieden
-72	.o	nicht einzuschätzen
-71	.p	nicht bekannt

5.11. Analyse-Systeme

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

Tabelle 12: 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) abgerufen werden. Dort werden auch alle Änderungen zwischen den einzelnen Versionen der Datensätze dokumentiert.

6.2. Versionshistorie

Veränderung von Version 1.0.0 zu Version 2.0.0

- Das Anpassungsgewicht für die Mandatsträger *wei_mdb* war fehlerhaft und wurde korrigiert.
- Die Codierungen für die offenen Angaben zu den wichtigsten politischen Problemen (c1as, c1bs, c1cs), den lokalen Wahlkampfthemen (b10s) und den Berufsbezeichnungen (e12s) wurden ergänzt.
- Die Variablenwerte der Variable *gewählt* „Erfolgreiche Kandidatur?“ wurden angepasst.
- Die Buchstaben-Suffixe für die Variablen b7, b11, b20, e1 und e12 wurden in der Reihenfolge vergeben, wie sie im Fragebogen abgefragt wurden. In Version 1.0.0 wurden die Buchstaben-Suffixe noch analog zu den äquivalenten Variablen in der Kandidatenstudie 2013 benannt.
- Die Benennung der Variablen b11i (in Version 1.0.0 „b11d“), e11as und e11bs wurden korrigiert bzw. konkretisiert.

Veränderung von Version 2.0.0 zu Version 3.0.0

- Die Variablen *wk_gewinner13* und *wk_gewinner17* waren fehlerhaft und wurden gelöscht. Diese beiden Variablen indizierten fälschlicherweise, welche Partei im jeweiligen Wahlkreis gewonnen hat, und nicht, ob der Kandidat im Wahlkreis gewonnen hat oder nicht. Es wurde eine neue Variable *wk_gewinner* „Wahlkreisgewinner 2017“ generiert, die angibt, ob ein Kandidat den Wahlkreis gewonnen hat.
- Es wurden 12 neue Variablen hinzugespielt, die die Erst- und Zweitstimmenanteile des Kandidaten im Wahlkreis, im Bundesland und insgesamt für 2017 und 2013 enthalten.

Links

Projektseite der GLES: www.gles.eu

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

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

Comparative Candidates Survey (CCS): <http://www.comparativecandidates.org/>

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

- Bergmann, Michael (2011): IPFWEIGHT: Stata module to create adjustment weights for surveys. Statistical Software Components S457353, Boston: Boston College Department of Economics.
- Callegaro, Mario (2013): Paradata in Web Surveys, in: Kreuter, Frauke (Hrsg.), Improving Surveys with Paradata: Analytic Uses of Process Information. Hoboken, NJ: Wiley, 261-280.
- 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.
- 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.