

## 4. Methoden, Datenerhebung und -aufbereitung

---

### 4.1 Vorteile, Nachteile und Herausforderungen der (Online-)Befragung

Für die Datenerhebung mittels standardisiertem Fragebogen stehen zunächst drei grundlegende Möglichkeiten zur Wahl, deren Vor- und Nachteile je nach Ziel der Studie abgewogen werden müssen: das persönliche Interview (face-to-face), das telefonische Interview und die schriftliche Befragung (questionnaire).<sup>1</sup> Letztgenannte kann auf wiederum auf zwei Wegen erfolgen: postalisch oder online. Die Mitgliederbefragung 2019 wurde aus einer Reihe von Gründen bewusst als schriftliche Online-Befragung durchgeführt:

Erstens: Beide Befragungsgruppen der Mitgliederbefragung rekrutieren sich aus der Grundgesamtheit (Population) aller Mitglieder der Partei DIE LINKE, d.h. einem Personenkreis, der mit Stand Dezember 2018 etwa 62.000 potenzielle Teilnehmer<sup>2</sup> umfasst. Für mehr als die Hälfte ist in den Kontaktmöglichkeiten der parteiinternen Mitgliederdatenbank eine gültige Mailadresse hinterlegt. Eine Online-Befragung bietet hierbei die beste Möglichkeit, ressourcenschonend einen Großteil der Population in die Befragung einzubeziehen. Da auch der Vergleich zwischen unterschiedlichen Landesverbänden Gegenstand der Studie ist, wäre eine Erhebung mittels stichprobenbasierten persönlichen oder telefonischen Interviews bei den – zwingend erforderlich – hohen Fallzahlen (Rückläufen) weder effizient noch durchführbar gewesen.

Zweitens: Die Mitgliederbefragung stellt im Gegensatz zu offenen Web-Befragungen mit selbst rekrutierenden Panels<sup>3</sup> eine Feldstudie innerhalb einer speziellen Population<sup>4</sup> – der Mitglieder der Partei DIE LINKE – dar. Anders als bei den erstgenannten offenen Web-Befragungen, bei denen sich die Teilnehmer etwa aus beliebigen Besuchern einer bestimmten Internetseite rekrutieren, ist das Panel der Mitgliederbefragung geschlossen und bereits im Vorfeld hinsichtlich Größe und wesentlicher soziodemografischer

---

1 Vgl. Diekmann 2007, S. 437.

2 Vgl. DIE LINKE 2019.

3 Vgl. Schön 2004.

4 Vgl. Diekmann 2007, S. 528.

Merkmale (Geschlecht, Alter, Zugehörigkeit zu den Landesverbänden etc.) bekannt gewesen. Es bildet folglich eine listenbasierte Stichprobe aus einer Population mit hohem Abdeckungsgrad.<sup>5</sup> Der Coverage-Error, der bei offenen Web-Befragungen eine große Fehlerquelle darstellen kann<sup>6</sup>, gestaltet sich dabei weniger problematisch: Ein Overcoverage-Effect wird ausgeschlossen, da jeder Teilnehmer Teil der Population ist und sich – umgekehrt – im Panel keine Personen befinden, die nicht der Population angehören. Ein Undercoverage-Effect wird einerseits durch den hohen Abdeckungsgrad minimiert, da mehr als die Hälfte der Population zur Befragungsgesamtheit gehört. Andererseits kann die soziodemografische Struktur der Befragungsteilnehmer mit den bekannten Merkmalen der Population aus der Mitgliederdatenbank verglichen werden. Auf diesem Weg ist es möglich, zu überprüfen, ob etwa Unterschiede zwischen den Teilnehmern und der Population hinsichtlich Geschlechterverteilung und Altersstruktur bestehen oder bestimmte Landesverbände unter den Teilnehmern schwächer oder stärker vertreten sind als in der Population. Große Abweichungen können folglich in der Auswertung entsprechend mathematisch gewichtet werden.

Drittens: Mit Ausnahme der manuellen Quantifizierung von Freitextfeldern zeichnet sich die schriftliche Online-Befragung durch Schnelligkeit aus, denn jede einzelne Befragung erfolgt autonom ohne Moderation durch eine dritte Person und eingegebene Antworten werden sofort gespeichert. Ebenso wie bei computergestützten Telefoninterviews sind Datensätze dadurch bereits unmittelbar nach dem Ende der Feldphase – etwa in Form einer computergenerierten SPSS-Datei – verfügbar, sodass zwischen Befragung und Auswertung wenig Zeit vergeht.<sup>7</sup> Auch der Teilnehmer selbst profitiert von der Schnelligkeit, wodurch der Rücklauf positiv beeinflusst werden kann: Nach Abschluss des Fragebogens sind keine weiteren Handlungen, wie etwa der Gang zum Briefkasten bei postalischen Befragungen, notwendig.

Viertens: Jede Einzelbefragung verläuft gleich. Nebenfaktoren, die etwa bei Face-to-face-Befragungen in der Sympathie zwischen Teilnehmer und Interviewer (Geschlecht, Stimmlage, Kleidung, Alter) oder im Ort bzw. der Tageszeit der Befragung stark variieren können, gibt es nicht. Jeder Teilnehmer entscheidet selbst, wann er den Fragebogen ausfüllt und wie viel Zeit er dafür in Anspruch nehmen möchte. Antwortverzerrungen, die sich aus individuellen Fehlern des Interviewers ergeben können, sind nicht möglich; Hemmschwellen der Teilnehmer bei sensibleren Fragestellungen werden abgebaut.<sup>8</sup>

Fünftens: Der computergestützte Fragebogen beugt Fehlern beim Ausfüllen vor und bietet breite Möglichkeiten bei der Ausgestaltung des User-Interfaces. Im Gegensatz zu Fragebögen in Papierform ist es bei den meisten Frageformen für den Teilnehmer unmöglich, formale Fehler zu begehen, wodurch der Ausfall durch falsch ausgefüllte Bögen reduziert wird. So ist es etwa ausgeschlossen, dass bei Single-Choice-Fragen mehrere Antwortmöglichkeiten angekreuzt werden können.

5 Vgl. Couper 2000; Couper/Coutts 2006; Diekmann 2007, S. 524.

6 Vgl. Diekmann 2007, S. 521.

7 Vgl. Diekmann 2007, S. 522.

8 Vgl. Ebenda, S. 466–470.

Die Gestaltung der Benutzeroberfläche kann zudem durch grafische Elemente zur Verständlichkeit und Abwechslung innerhalb des Fragebogens positiv beitragen<sup>9</sup>: Für die Mitgliederbefragung wurde Weiß als Grundfarbe gewählt, wobei jedoch die aktuell mit dem Mauszeiger gewählte Antwortmöglichkeit im typischen, leicht ungesättigten Rot der Partei umrandet wurde. Auch der kompakte Einsatz des Parteilogos und der Befragungstitel in der Schriftart des Corporate-Designs der LINKEN zielten bei der Fragebogenkonstruktion darauf ab, eine vertraute Umgebung für Parteimitglieder zu schaffen. Durch variierende Eingabemodi in Form von Schiebereglern oder Drag-and-Drop-Feldern wurden zudem Alternativen zum einfachen Mausklick geschaffen, wodurch der Fragebogen auch medienergonomisch abwechslungsreicher gestaltet werden konnte.

Sechstens: Der computergestützte Fragebogen ermöglicht es, Daten zum Befragungsverhalten zu erheben.<sup>10</sup> Im Beispiel der Mitgliederbefragung 2019 wurde die Zeit, die jeder Teilnehmer für die komplette Befragung sowie für jede einzelne Seite benötigt hat, gemessen. Zusammen mit dem relativen Anteil unbeantworteter Fragen kann dadurch das Verhalten der Teilnehmer im Umgang mit der Befragung differenzierter betrachtet werden: Personen, die sich lediglich durch den Fragebogen »durchgeklickt« haben und den Großteil der Fragen entweder nicht oder zu schnell beantwortet haben (»Lurker«) konnten besser gefunden und ggf. von der Auswertung ausgeschlossen werden. Eine weitere hilfreiche Variable war die des Befragungszeitpunkts, d.h. zu welcher Tageszeit die meisten Fragebögen ausgefüllt wurden. So konnte mit den gewonnenen Daten nach der ersten Einladung der Zeitpunkt für den Versand der Erinnerungsmail nach dem Befragungsverhalten der Teilnehmer optimiert werden.

Neben der Rechtfertigung für die Datenerhebung mittels schriftlicher Online-Befragung ist die Auseinandersetzung mit den inhärenten Schwächen der Methode unabdingbar. Ihnen entgegenzuwirken bildete eine besondere Herausforderung bei der Konzeption der Mitgliederbefragung 2019.

Bereits die Vorauswahl der Befragungsgesamtheit aus der Population kann sich problematisch auf die Ergebnisse der Studie auswirken. Das entscheidende Auswahlkriterium war das Vorliegen einer E-Mailadresse in der Mitgliederdatenbank der LINKEN, da auch die Einladung zur Befragung ausschließlich schriftlich per E-Mail erfolgte. Durch diese Selektion werden internetaffine jüngere Personen, die E-Mails als Kommunikationsmedium regelmäßig nutzen und mit grundlegenden Computerkenntnissen vertraut sind, begünstigt. Für andere Personengruppen können hingegen Undercoverage-Effects vorliegen, in deren Folge valide Rückschlüsse von Befragungsergebnissen auf die Population erschwert werden.<sup>11</sup> Durch den o.g. hohen Abdeckungsgrad sowie die Möglichkeit, Teilnehmer und Population soziodemografisch zur vergleichen und ggf. entsprechende Gewichtungen vorzunehmen, wird dieser Schwäche in der Stichprobenauswahl jedoch innerhalb der Mitgliederbefragung entgegengewirkt (vgl. Abschnitt 4.4). Da der Hauptfokus der vorliegenden Arbeit zudem im besonderen Maße auf der Analyse der Neumitgliederstruktur – d.h. größtenteils jüngeren Personen – liegt, kann die internetgestützte Befragung positive Rücklauffekte in dieser weitestgehend mit dem Medium

9 Vgl. Diekmann 2007, S. 522.

10 Vgl. ebenda, S. 523.

11 Vgl. Diekmann 2007, S. 521.

vertrauten Gruppe entfalten: Die Datenerhebung mittels Online-Fragebogen erscheint als ein effektives, auf die Bedürfnisse der Hauptzielgruppe abgestimmtes Erhebungsinstrument. Die Alternative, postalische Einladungen für den Online-Fragebogen anstelle von E-Mails zu verwenden, hätte das Problem lediglich teilweise gelöst und neue Schwächen und Herausforderungen verursacht: Einerseits wären nach wie vor geringe EDV-Kenntnisse und ein unterstütztes Gerät für die Befragung notwendig gewesen, wobei das Aufrufen des Fragebogens durch den Medienbruch (postalische Einladung in Papierform/computergestützter Fragebogen im Internet) zusätzlich erschwert werden kann.<sup>12</sup> Andererseits wären postalische Einladungen zu kostenintensiv gewesen, um einen annähernd hohen Abdeckungsgrad der Befragungsgesamtheit im Vergleich zu E-Mails zu erhalten.

Eine weitere Herausforderung bestand darin, die Anonymität zu gewährleisten und gleichzeitig eine mehrfache Teilnahme durch Einzelpersonen zu verhindern. Die Glaubwürdigkeit der Anonymität wurde dadurch verstärkt, dass die Mitgliederbefragung in Kooperation mit der LINKEN, im Speziellen der Bundesgeschäftsstelle, durchgeführt wurde. Alle E-Mails wurden im Namen der LINKEN (mit einem Verweis auf das Dissertationsvorhaben des Verfassers) mit einem Einladungstext des Bundesgeschäftsführers Jörg Schindler verschickt. Damit wurde die Kooperation der Partei mit dem Verfasser der Studie sowie deren Unterstützung vermittelt, wodurch nicht zuletzt möglichem Misstrauen gegenüber der Befragung vorgebeugt wurde. Dagegen enthält der Fragebogen jedoch auch Items, bei denen kritische Äußerungen gegenüber der Partei als eine der durchführenden Institutionen möglich waren – sensible Fragen, die bei den Teilnehmern Misstrauen befeuern und Hemmschwellen hervorrufen können.<sup>13</sup> Aus diesem Grund wurde die Mitgliederbefragung 2019 als »personalisierte Umfrage«<sup>14</sup>, die eine der gebräuchlichsten Formen darstellt<sup>15</sup>, mit technischer Anonymisierung konstruiert. Kernstück dieser Variante bildet zunächst eine persönliche Einladung mit einem eindeutigen Zugang zur Befragung.<sup>16</sup> Dieser wurde in der Mitgliederbefragung durch einen einzigartigen, computergenerierten Link in jeder Einladung generiert, der den Befragten zur einmaligen Teilnahme berechtigte. Die Zuordnung des Links zur jeweiligen Person erfolgte dabei verschlüsselt, sodass der ausgefüllte Fragebogen technisch nicht auf den Einzelnen zurückverfolgt werden kann. Neben der Anonymität bietet die personalisierte Umfrage weitere Vorteile: So konnte die Befragung vom Teilnehmer an einer beliebigen Stelle pausiert und zu einem späteren Zeitpunkt fortgesetzt werden. Die Gefahr des »Mehrfachausfüllens« war technisch ebenso ausgeschlossen.

Wie bei postalischen Erhebungen stellen Verständnisprobleme auch bei einem Online-Fragebogen eine grundsätzliche Schwäche dar, da die Teilnehmer innerhalb der schriftlichen Befragung nicht auf die Unterstützung eines Interviewers zurückgreifen können.<sup>17</sup> Daraus resultierend musste auch der Fragebogen der Mitgliederbefragung

12 Vgl. Gräf 2010, S. 21.

13 Vgl. Gräf, S. 22.

14 Ebenda, S. 21.

15 Vgl. ebenda.

16 Vgl. ebenda.

17 Vgl. Diekmann 2007, S. 514.

möglichst selbsterklärend und ohne weitverzweigte und komplexe Fragestellungen formuliert werden. Anders als bei Papierfragebögen bietet der Online-Fragebogen jedoch einen entscheidenden Vorteil bei der Ausgestaltung von Filterfragen: Der Teilnehmer muss nicht selbstständig den Punkt suchen, an dem er nach einer bestimmten Filterantwort die Befragung fortsetzen soll (Beispiel: »Wenn Sie diese Frage mit X beantworten, setzen Sie den Fragebogen bei Frage Y auf Seite Z fort.«). Dieser Schritt kann im Online-Fragebogen computergestützt automatisch erfolgen und kam bei der Mitgliederbefragung insofern zum Einsatz, als basierend auf der Wahl des eigenen Landesverbandes automatisch die Auswahl des Kreis-/Bezirksverbandes auf die entsprechend möglichen Optionen beschränkt wurde. Teilnehmer aus Sachsen konnten demnach etwa nicht versehentlich einen bayerischen Landkreis als ihren Kreisverband angeben.

Um den Befragten dennoch die Möglichkeit einer Kontaktaufnahme für technische oder inhaltliche Nachfragen zu bieten, wurde im Fragebogen der Mitgliederbefragung auf jeder Seite jeweils eine Kontaktmailadresse des Verfassers und der Bundesgeschäftsstelle verlinkt. Diese Möglichkeit wurde über den Befragungszeitraum in einem überschaubaren Maße und darüber hinaus oft für Feedback genutzt.

Die Erhebungsphase der Studie mittels standardisiertem Fragebogen erfolgte, nach einer vorherigen Informationsmail an alle Befragten, innerhalb von sechs Wochen zwischen dem 9. September und dem 20. Oktober 2019 über den Cloud-Service SoSci Survey<sup>18</sup>, einem Werkzeug zur professionellen Durchführung empirischer Studien. Es wurde 2003 und 2004 am Institut für Kommunikationswissenschaft der Universität München in Kooperation mit der Universität Zürich entwickelt und wird seit 2014 durch die SoSci Survey GmbH mit Serverstandort in Deutschland (München) über ein eigenständiges Unternehmen vertrieben.<sup>19</sup> Dabei erfreut sich das Tool einer wachsenden Beliebtheit innerhalb und außerhalb des akademischen Sektors: Wurden 2006 noch 50 Projekte mit SoSci Survey realisiert, stieg die Zahl auf über 13.500 Studien im Jahr 2014. Mit Ausnahme einer sog. Pro-Server-Lizenz, die auch bei der Mitgliederbefragung zum Einsatz kam, werden sämtliche Dienstleistungen für den akademischen Sektor kostenlos zur Verfügung gestellt. Die Pro-Server-Lizenz ermöglicht die Nutzung eines leistungsfähigeren Befragungsservers, der sich durch eine schnellere Verarbeitungsgeschwindigkeit für Studien ab einer höheren vierstelligen Teilnehmerzahl besonders eignet. Da bei der Mitgliederbefragung auch personenbezogene Daten erhoben wurden, war ein Auftragsverarbeitungsvertrag (AVV) mit der SoSci Survey GmbH im Sinne des Art. 28 DSGVO unabdingbar, um einen optimalen Datenschutz zu gewährleisten. Dieser konnte ausschließlich durch den Lizenzerwerb abgeschlossen werden, da lediglich der erweiterte Pro-Server eine BackUp-Vernichtung binnen eines Monats – die damit datenschutzrechtlich mit der DSGVO vereinbar ist – unterstützt.

Neben dem leistungsstarken Befragungsserver und den datenschutzrechtlich einwandfreien Bedingungen bietet SoSci Survey ein innovatives User-Interface (UI), dessen Nutzung lediglich an einigen wenigen Stellen Grundkenntnisse der Beschreibungssprachen HTML und PHP voraussetzt. Von der Fragebogenimplementierung über die intuitive und schnelle Paneladministration bis hin zum Serienmailversand und zur Ausgabe

18 Für diese Arbeit genutzte Version: SoSci Survey Version 3.1.06.

19 Vgl. SoSci Survey GmbH 2023.

erhobener Daten als SPSS-Datei bot der Cloud-Service daher eine optimale Grundlage zur Durchführung der Mitgliederbefragung 2019. Auch die besondere technische Herausforderung eines barrierefreien Zugangs für Menschen mit Sehbehinderungen konnte realisiert werden, sodass der Fragebogen mittels Screenreader und Braillezeile<sup>20</sup> bearbeitet werden konnte.

## 4.2 Pretests

Pretests zur medienergonomischen und inhaltlichen Optimierung des Erhebungsinstruments fanden im Vorfeld der Befragung zwischen Mai und Juli 2019 statt. Sie wurden mit verschiedenen Versionen in drei unterschiedlichen Testgruppen durchgeführt:

- Die erste vollständige Version des Fragebogens wurde im Rahmen des Doktorandenkreises von Eckhard Jesse, Tom Mannewitz und Isabelle-Christine Panreck im Mai 2019 in Meißen vorgestellt und von etwa 15 Doktoranden der Politikwissenschaft evaluiert.
- In den darauffolgenden Wochen wurde ein weiterer Pretest mit den hauptamtlichen Mitarbeitern der LINKEN-Bundesgeschäftsstelle durchgeführt.
- Der letzte Pretest fand schließlich im Juli 2019 unmittelbar in der Parteibasis durch Mitglieder des Ortsverbandes der Partei in Crimmitschau (Landesverband Sachsen) statt.

Die Wahl der Gruppen ermöglichte Pretests in einem breiten Personenspektrum. Dieses zeichnete sich durch Personen unterschiedlichen Alters sowie mit und ohne akademischen Hintergrund aus. Durch die Teilnahme der hauptamtlichen Partei-Mitarbeiter konnten zudem Erfahrungen aus vorangegangenen wissenschaftlichen Studien innerhalb der LINKEN in die Fragebogenkonstruktion einfließen, um auf diesem Weg etwa Missverständnissen bei der Formulierung der Fragen vorzubeugen und mögliche Hemmschwellen abzubauen. Der letzte Pretest innerhalb eines Ortsverbandes der Partei stellte schließlich einen praxisnahen Feldtest dar, da es sich hinsichtlich soziodemografischer Merkmale um die heterogenste Gruppe handelte. Hierbei wurden vor allem die Verständlichkeit der Fragen sowie ergonomische Komponenten wie Schriftgröße und Fragereihenfolge weiter optimiert.

---

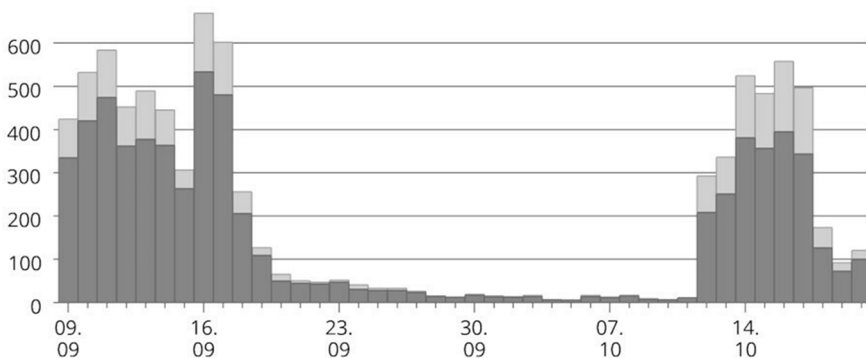
20 Bei Screenreadern handelt es sich um Software-Anwendungen, die Menschen mit Sehbehinderungen alternative Schnittstellen zu Textmodi oder grafischen Oberflächen bieten. Auf diesem Weg werden etwa Texte mittels Sprachausgabe vorgelesen. Eine Braillezeile stellt hingegen ein Ausgabegerät dar, das dem Nutzer Texte in Brailleschrift-Zeichen ausgibt. Beides wird oftmals in direkter Kombination eingesetzt.

### 4.3 Gesamtrücklauf und Abbruchquote

Der ursprüngliche Gesamtumfang des Befragungspool betrug 34.887 Mitglieder der Partei DIE LINKE, bei denen im parteiinternen Mitgliederprogramm eine Mailadresse hinterlegt war. 92 Personen hatten im Vorfeld der Teilnahme an der Befragung widersprochen und sich elektronisch selbst aus dem Panel ausgetragen. Für 1293 Mailadressen lagen technische Problemberichte vor, sodass der Fragebogen den jeweiligen Empfänger nicht erreicht hat. Die Ursachen hierfür können vielfältig sein (z.B. automatische Versandblockierung durch Abwesenheitsnotizen), es ist jedoch davon auszugehen, dass ein Großteil der nicht erreichbaren Mailadressen zum Zeitpunkt der Befragung nicht mehr existent war. Drei Personen konnten wegen sonstiger technischer Probleme nicht erreicht werden. Der tatsächliche Befragungspool umfasste daher 33.499 Personen.

Mit einer Teilnehmerzahl von 6597 Personen erzielte die Befragung einen zufriedenstellenden Gesamtrücklauf von 19,7 Prozent, der im Zeitverlauf in Abbildung 6 dargestellt wird. Gemessen an der Gesamtmitgliederstärke der Partei zum Ende des Jahres 2019 (60.862 Mitglieder<sup>21</sup>), haben sich somit 10,8 Prozent aller Parteimitglieder an der Umfrage beteiligt. Eingerechnet sind hierbei ausschließlich Teilnehmer, die den Fragebogen vollständig bis zur letzten Seite ausgefüllt und abgeschickt haben.

Abbildung 6: Rücklauf im Zeitverlauf (Diagramm erstellt mit SoSci Survey). Dunkelgraue Balken kennzeichnen vollständig ausgefüllte Fragebögen, hellgraue Bereiche begonnene Fragebögen



Die Haupteinladung wurde zwischen dem 9. und 17. September jeweils an etwa 5000 Personen pro Tag verschickt. Mit Ausnahme des 15. und 16. September erfolgte der Versand an allen Tagen am späten Nachmittag und frühen Abend zwischen 17 Uhr und 19 Uhr. Die tägliche Versandmenge wurde bewusst reduziert, um technisch zu vermeiden, dass Mails von Spamfiltern aussortiert werden und so der Rücklauf geschmälert wird. Bis zur Erinnerungsmail am 12. Oktober wurden etwa 4200 Bögen ausgefüllt. Mit der Erinnerungsmail, die an alle Teilnehmer versandt wurde, die die Befragung bis zu diesem Zeitpunkt bislang nicht oder nur unvollständig ausgefüllt haben, konnten bis zum Ende der Befragung schließlich fast 2400 weitere Rückläufe generiert werden. Für den

21 Die Zahl entstammt der Mitgliederdatenbank der Partei DIE LINKE zum Stichtag 31.12.2019.

Zeitpunkt der Erinnerungsmail wurde jeweils der Vormittag zwischen 9 und 12 Uhr gewählt, um eine zeitliche Varianz zur Haupteinladung zu erzeugen.

Die Abbruchquote, d.h. der Anteil derer, die den Fragebogen begonnen, aber nicht beendet haben, liegt bei 22,2 Prozent. Die detaillierte Aufschlüsselung nach der zuletzt bearbeiteten Seite (Abb. 7) zeigt, dass bei 1271 von 1886 nicht vollständig ausgefüllten Fragebögen (67,4 Prozent) die Befragung direkt nach der Hinweisseite (Seite 1) abgebrochen wurde.

Abbildung 7: Letzte bearbeitete Seiten (Tabelle erstellt mit SoSci Survey)

| Letzte bearbeitete Seite | Datensätze abgeschlossen / Interviews gesamt / kumulativ |      |      |                                                                                    |
|--------------------------|----------------------------------------------------------|------|------|------------------------------------------------------------------------------------|
| Seite 7                  | 6597                                                     | 6602 | 6602 |  |
| Seite 6                  | 0                                                        | 11   | 6613 |  |
| Seite 5                  | 0                                                        | 64   | 6677 |  |
| Seite 4                  | 0                                                        | 143  | 6820 |  |
| Seite 3                  | 0                                                        | 373  | 7193 |  |
| Seite 2                  | 0                                                        | 1271 | 8464 |  |
| Seite 1                  | 0                                                        | 19   | 8483 |  |
| Gesamt                   | 6597                                                     | 8483 |      |                                                                                    |

## 4.4 Analysegruppen, gruppenspezifischer Rücklauf und Gewichtung

### 4.4.1 Begriffsdefinition und Einordnung der Analysegruppen

Alle Teilnehmer der Mitgliederbefragung 2019 haben unabhängig von ihrem individuellen Beitrittsjahr den gleichen Fragebogen erhalten und sind bei der Datenerhebung in einen gemeinsamen Datensatz eingeflossen. Dieser allumfassende und alle Beitrittsjahre einschließende Datensatz wird bei der Auswertung dazu genutzt, um die Gesamtpartei abzubilden und den Vergleich dazu zu ermöglichen, wie sich die Gruppe der Neumitglieder im Kontext zur Gesamtpartei strukturiert und positioniert.

Die Gruppe der Neumitglieder stellt die Hauptzielgruppe der Mitgliederbefragung 2019 dar und wird durch alle im Zeitraum Januar 2014 bis August 2019<sup>22</sup> eingetretenen Personen definiert. Sie ist demnach eine Teilmenge, die bedingt durch das Eintrittsjahr alle Mitglieder, die der Partei vor 2014 beigetreten sind, aus dem ursprünglichen Datensatz exkludiert. In der grafischen und schriftlichen Datenauswertung wird an verschiedenen Stellen ein vergleichender Bezug zwischen beiden Gruppen hergestellt, wobei sich der Begriff »Gesamtpartei« auf den allumfassenden Datensatz bezieht und der Begriff »Neumitglieder« auf die ausgegliederte Befragungsgruppe der Neumitglieder.

Mit Blick auf den Vergleich beider Analysegruppen (Tab. 17) fällt in der Differenzierung auf, dass die Gruppe der Neumitglieder mit 23.028 von 60.862 Parteimitgliedern über ein Drittel (37,8 Prozent) der Grundgesamtheit umfasst. Fast 20.000 von ihnen konnten in der Mitgliederbefragung durch das Vorliegen einer gültigen Mailadresse

22 Die Befragung fand im September 2019 statt, sodass nicht alle Eintritte aus 2019 berücksichtigt werden.

erreicht werden, sodass etwa zwei Drittel des Befragungspools der Gruppe der Neumitglieder zuzuordnen sind. In der Schlussfolgerung erscheint es nachvollziehbar, dass die Zahl der Mitglieder mit ungültiger oder fehlender Mailadresse in der Mitgliederdatenbank steigt, je weiter der Parteieintritt zurückliegt. Weiterhin legt der Blick auf die altersspezifische Nutzung von Internet und E-Mails in Deutschland eine weitere Begründung nah: Das Internet wird von älteren Personen – besonders ab 70 Jahren – generell weniger genutzt<sup>23</sup> und darüber hinaus greifen ältere Internetnutzer im Vergleich zu jüngeren Altersgruppen weniger stark auf E-Mails als Kommunikationsmedien zurück<sup>24</sup>.

Einerseits konnte die Hauptzielgruppe der Studie mit dem Erhebungsinstrument demnach umfassend erreicht werden, andererseits wird durch den naheliegenden Undercoverage-Effekt, insbesondere mit Blick auf den Anteil älterer Teilnehmer, eine nachträgliche Gewichtung der Daten für die Gesamtpartei notwendig (Abschnitt 4.4.3).

Tabelle 17: Analysegruppen

|                 | Gesamtpartei                                                                              | Neumitglieder<br>(Teilgruppe von Gesamtpartei)                                                                                     |
|-----------------|-------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|
| Definition      | Alle Mitglieder der Partei DIE LINKE mit gültiger Mailadresse in der Mitgliederdatenbank. | Alle Mitglieder der Partei DIE LINKE mit gültiger Mailadresse in der Mitgliederdatenbank und einem Eintrittsjahr 2013 oder später. |
| Grundgesamtheit | 60.862 Personen                                                                           | 23.028 Personen                                                                                                                    |
| Teilnehmer      | 6597 Teilnehmer (10,8 Prozent)                                                            | 3800 Teilnehmer (16,5 Prozent)                                                                                                     |

4.4.2 Datensatz Neumitglieder

Die Untersuchungsgruppe der Neumitglieder umfasst in ihrer Grundgesamtheit insgesamt 23.028 Personen, die zwischen 2014 und 2019 in die Partei eingetreten sind. Im Rahmen der Mitgliederbefragung 2019 konnten fast 20.000 Personen (über 80,0 Prozent der Grundgesamtheit) befragt werden, von denen 3800 den Fragebogen ausgefüllt haben (vgl. vorheriger Abschnitt). Um die Qualität der erhobenen Daten hinsichtlich ihrer repräsentativen Aussagekraft für die Grundgesamtheit zu beurteilen, ist ein Abgleich mit der Mitgliederdatenbank der Partei DIE LINKE für die Merkmale Geschlecht<sup>25</sup> (Abb. 8), Alter (Abb. 9) und Landesverband (Abb. 10) möglich. Die Grundlage für die Daten der Grundgesamtheit bieten dabei parteiinterne Auszüge aus der Mitgliederdatenbank sowie der interne Jahresbericht zur Mitgliederentwicklung für 2019.

23 Vgl. Statistisches Bundesamt 2022, S. 13.  
24 Vgl. ebenda, S. 16.  
25 Die Mitgliederdatenbank der Partei DIE LINKE unterschied zum Zeitpunkt der Erhebung lediglich zwischen dem männlichen und weiblichen Geschlecht. Die Option zur Angabe eines diversen Geschlechts war für Parteimitglieder erstmals bei der Datenerhebung für die vorliegende Studie möglich, wodurch Vergleichsdaten fehlen.

Abbildung 8: Geschlechterverteilung Neumitglieder (Grundgesamtheit versus MB 2019)

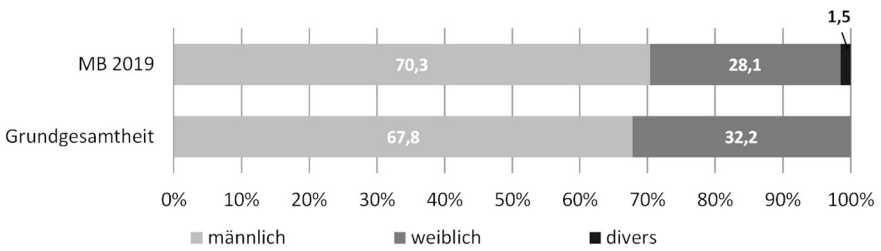


Abbildung 9: Altersverteilung Neumitglieder (Grundgesamtheit versus Daten MB 2019)

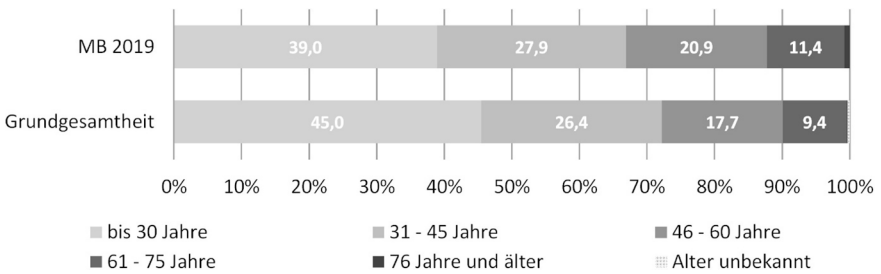
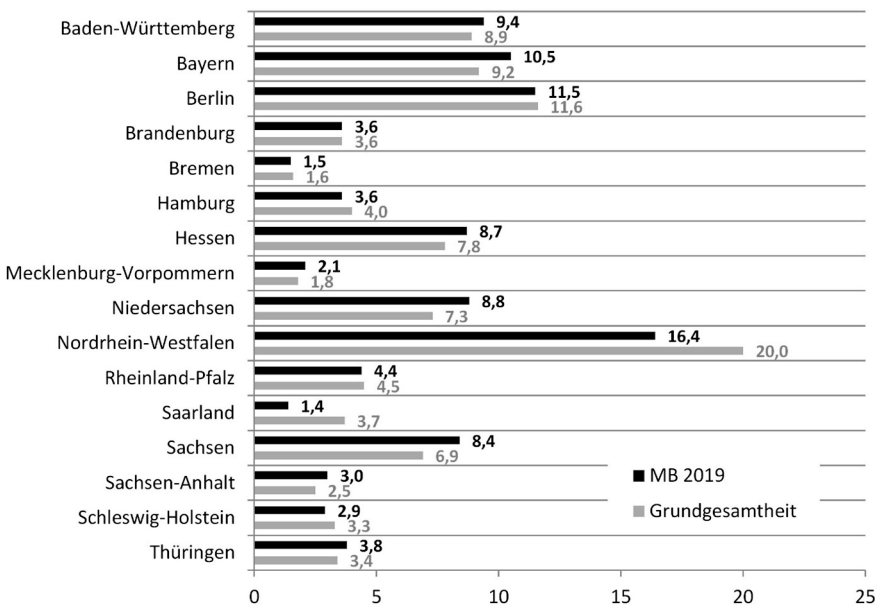
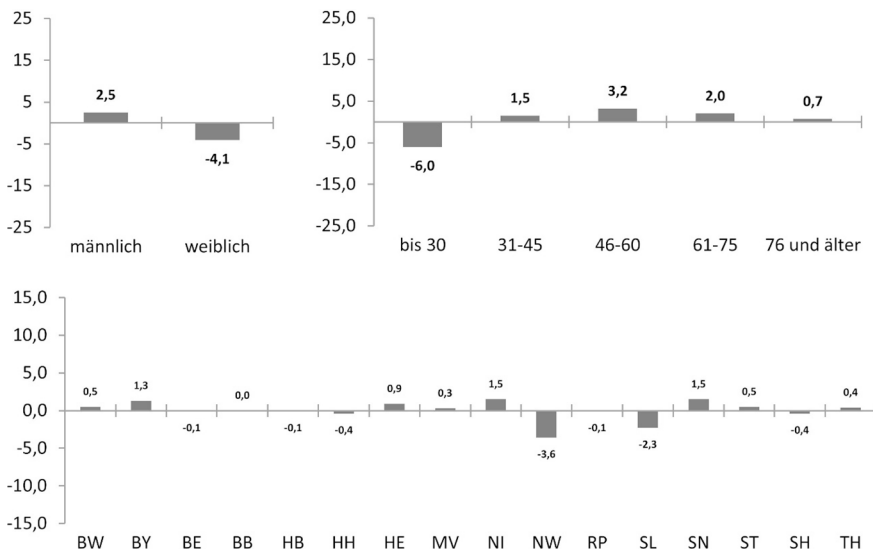


Abbildung 10: Neumitglieder nach Landesverband (Grundgesamtheit versus Daten MB 2019; relative Häufigkeiten in Prozent)



Aus der Gegenüberstellung der im Rahmen der Mitgliederbefragung 2019 erhobenen Daten mit jenen aus der Mitgliederdatenbank der Partei lässt sich zunächst folgern, dass die Erhebungsmethode für die Hauptuntersuchungsgruppe der Neumitglieder effizient funktioniert hat. Durch die hohe Anzahl gültiger Mailadressen konnte der überwiegende Teil der Grundgesamtheit in die Befragung einbezogen werden. Die ausgeprägte Teilnahmebereitschaft – etwa jeder Fünfte hat den Fragebogen ausgefüllt – bewirkt einerseits eine große Fallzahl und andererseits eher geringe Abweichungen hinsichtlich der für die Grundgesamtheit vorliegenden Merkmalsausprägungen aus der Mitgliederdatenbank (Abb. 11).

Abbildung 11: Abweichungen der Teilnehmerstruktur von der Grundgesamtheit (Neumitglieder; Abweichungen in Prozentpunkten)



Neben einem leichten Unterschied in der Geschlechterverteilung fällt auf, dass jüngere Mitglieder bis 30 Jahre in der Teilnehmerstruktur zugunsten von Teilnehmern im Altersbereich 46–75 in geringerem Maße unterrepräsentiert sind. Da der Bericht aus der Mitgliederdatenbank eine weitere Differenzierung der Altersgruppe »bis 30 Jahre« zulässt<sup>26</sup>, ist festzustellen, dass die beiden Gruppen 21–25 und 26–30 Jahre um jeweils etwas mehr als drei Prozentpunkte weniger im Teilnehmerfeld der Mitgliederbefragung 2019 abgebildet werden. Die Verteilung der Landesverbandszugehörigkeit spiegelt weiterhin die tatsächliche Struktur der Grundgesamtheit treffend wider, wobei sich die größten Abweichungen (Nordrhein-Westfalen und Saarland) unterhalb einer Marke von 4 Prozentpunkten bewegen. Die Ergebnisse der vorliegenden Arbeit müssen folglich

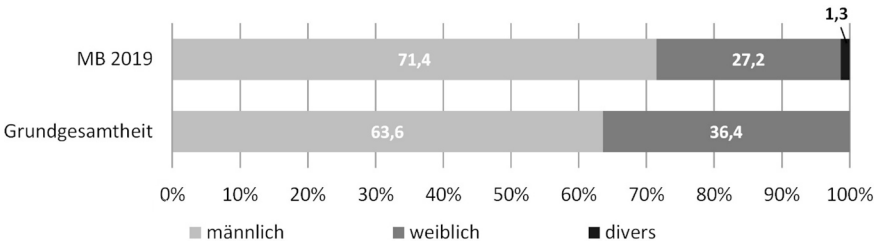
26 Der Auszug fasst beginnend mit der Altersgruppe »16-20 Jahre« jeweils 5 Jahre als eine Altersgruppe zusammen.

auf der Basis lediglich leichter Abweichungen zwischen Teilnehmern und tatsächlicher Mitgliederstruktur betrachtet werden.<sup>27</sup>

#### 4.4.3 Datensatz Gesamtpartei

Die Untersuchungsgruppe der Gesamtpartei umfasst in ihrer Grundgesamtheit insgesamt 60.863 Personen, die zum Stichtag 31. Dezember 2019 Mitglied der Partei DIE LINKE waren. Im Rahmen der Mitgliederbefragung 2019 konnten 33.499 Personen (55,0 Prozent der Grundgesamtheit) befragt werden, von denen 6.597 den Fragebogen ausgefüllt haben (vgl. Abschnitt 4.4.1). Der Datensatz der Neumitglieder ist dem der Gesamtpartei inhärent und wird ergänzt durch Teilnehmer, die vor 2014 der Partei beigetreten sind. Auch mit Blick auf alle Parteimitglieder kann ein Abgleich der Teilnehmerstruktur mit der Grundgesamtheit durch die vorliegenden Daten zu Geschlecht (Abb. 12), Alter (Abb. 13) und Landesverband (Abb. 14) erfolgen. Die folgenden Abbildungen beziehen sich ebenfalls auf den parteiinternen Jahresbericht zur Mitgliederentwicklung für 2019.

Abbildung 12: Geschlechterverteilung Gesamtpartei (Grundgesamtheit versus MB 2019)



Stärker noch als in der Gruppe der Neumitglieder sind Männer in der Teilnehmerstruktur der Mitgliederbefragung 2019 überrepräsentiert. Die größte Abweichung wird jedoch mit Blick auf die Altersverteilung erkennbar: Zu Gunsten des breiten Altersbereichs von Personen bis 60 Jahre sind ältere Personen ab 76 unter den Studienteilneh-

27 Aufgrund der zwar vorhandenen, jedoch überschaubaren Unterschiede zwischen Teilnehmerstruktur und Grundgesamtheit wird von einer nachträglichen Gewichtung des Datensatzes abgesehen. Diese wäre nicht zuletzt in Form einer Poststratifizierung durch das Fehlen einer vollständigen, alle Merkmale umfassenden Kontingenztafel für die Analysegruppe der Neumitglieder nicht möglich, wobei bestenfalls auf alternative Gewichtungsverfahren, die lediglich Randverteilungen berücksichtigen – wie etwa RIM-Weighting bzw. Raking – zurückgegriffen werden könnte. Letztgenannte sind jedoch dadurch gekennzeichnet, dass sie anfälliger für Verzerrungen in der statistischen Auswertung sind und die Gefahr von Fehlschlüssen zugunsten eines durch die Gewichtung eher geringen Zugewinns an (soziodemografischer) Präzision erhöhen. Eine Gewichtung des Alters hätte zudem mit Bezug auf die Ausprägung »bis 30 Jahre« zur Folge, dass zwar die Altersgruppen »21-25« und »26-30« durch ein höheres Gewicht der Grundgesamtheit angeglichen wären, alle Personen unter 21 jedoch gleichzeitig mit einem zu hohen Gewicht versehen werden würden, da diese sowohl in der Grundgesamtheit als auch in der Teilnehmerstruktur zu nahezu gleichen Anteilen vertreten sind. Weiterhin wäre etwa eine Gewichtung nach dem Geschlecht durch die in der Mitgliederdatenbank fehlenden Daten zur Merkmalsausprägung »divers« problematisch.

mern stark unterrepräsentiert. Hier zeigt sich eine mögliche Kehrseite der Erhebungsmethode. Diese konnte einerseits die Hauptuntersuchungsgruppe der Neumitglieder durch die hohe Verfügbarkeit von gültigen Mailadressen und die Befragung mittels internetgestütztem Fragebogen gut erreichen, andererseits konnten die ältesten Mitglieder der Partei – die einen nicht zu verachtenden Teil der Gesamtmitgliedschaft ausmachen – nicht ausreichend gut angesprochen werden. Das Ungleichgewicht zwischen Teilnehmern und Grundgesamtheit erstreckt sich dadurch auch auf die Repräsentation der Landesverbände: Durch die geringere Teilnahme der ältesten Parteimitglieder sind insbesondere die ostdeutschen (älteren) Landesverbände unter den Teilnehmern weniger stark vertreten (Abb. 14).

Abbildung 13: Altersverteilung Gesamtpartei (Grundgesamtheit versus MB 2019)

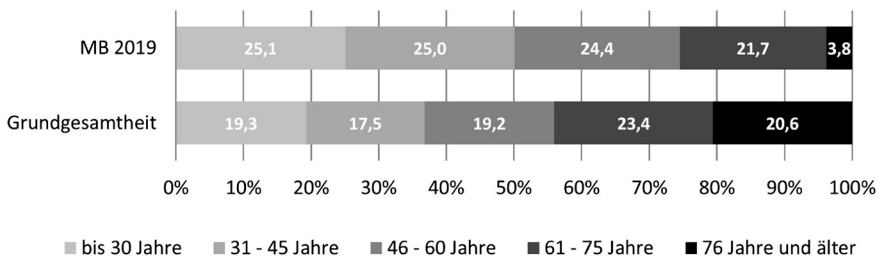
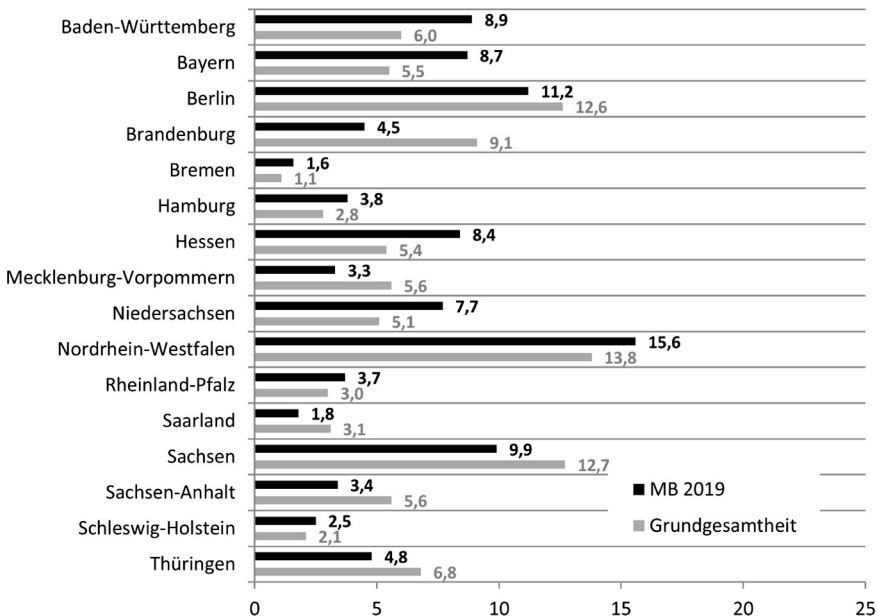


Abbildung 14: Gesamtpartei nach Landesverband (Grundgesamtheit versus Daten MB 2019; relative Häufigkeiten in Prozent)



Um Verzerrungen in der Datenauswertung entgegenzuwirken, erscheint eine Anpassungsgewichtung bzw. Kalibrierung der erhobenen Daten daher empfehlenswert.<sup>28</sup> Durch das Vorliegen einer vollständigen Kreuztabelle für die Merkmale Geschlecht<sup>29</sup>, Alter und Landesverband in der Mitgliederdatenbank bieten sich hierfür mehrere Optionen mittels Postratifikation (nachträgliche Schichtung)<sup>30</sup> an, die eine Anpassungsgewichtung nach einem, zwei oder allen drei Merkmalen ermöglichen. Je mehr Merkmalsvariablen dabei in die Gewichtung einbezogen werden, umso mehr Zellen enthält die Kreuztabelle. Mehr Zellen bedeuten zudem eine geringere Fallzahl pro Zelle, sodass die Gefahr entsteht, dass einzelne Zellen mit nur wenigen Teilnehmern besetzt werden.<sup>31</sup> In der Folge können durch die Integration aller drei Variablen in den Gewichtungsprozess zwar die soziodemografischen Unterschiede zwischen Teilnehmern und Grundgesamtheit kosmetisch kaschiert werden, jedoch entstehen hierdurch schnell hohe Gewichtungsfaktoren, mit denen einzelne Teilnehmer schnell um ein Vielfaches stärker bzw. häufiger in die Auswertung einbezogen werden, wodurch infrage gestellt werden könnte, ob einzelne Ergebnisse der Studie noch als realitätsnah angesehen werden können.<sup>32</sup> Zudem würden durch die Gewichtung im Beispiel der Mitgliederbefragung 2019 Personen, die etwa ein diverses Geschlecht oder kein Alter angegeben haben, mit einem modellierten Gewichtungsfaktor »1« versehen, da für sie mathematisch kein Gewichtungsfaktor ermittelt werden kann.<sup>33</sup>

Mit Blick auf die tatsächlichen Abweichungen der Teilnehmer von der Grundgesamtheit (Abb. 15) sticht die Abweichung in der ältesten Altersgruppe besonders hervor. Eine Gewichtung nach ausschließlich diesem Merkmal für jede Altersgruppe mittels

$$\text{Gewichtungsfaktor } w = \frac{\%N_{\text{Grundgesamtheit}}}{\%N_{\text{Teilnehmer}}}$$

ergibt folglich die nachfolgenden Gewichte für die einzelnen Altersgruppen (Tab. 18):

Tabelle 18: Anpassungsgewichte durch das Merkmal »Alter«

|                               | Bis 30<br>Jahre | 31–45<br>Jahre | 46–60<br>Jahre | 61–75<br>Jahre | 76 und<br>älter | Kein<br>Alter |
|-------------------------------|-----------------|----------------|----------------|----------------|-----------------|---------------|
| %N <sub>Grundgesamtheit</sub> | 19,3            | 17,5           | 19,2           | 23,4           | 20,6            | -             |
| %N <sub>Teilnehmer</sub>      | 21,1            | 25,0           | 24,4           | 21,7           | 3,8             | -             |
| w                             | <b>0,7689</b>   | <b>0,7000</b>  | <b>0,7869</b>  | <b>1,0783</b>  | <b>5,4211</b>   | <b>1,000</b>  |

Abbildung 15 zeigt die Auswirkungen einer einfachen Gewichtung nach Altersklassen in Bezug auf die von der Grundgesamtheit bekannten Merkmale. Zunächst werden

28 Vgl. Gabler 2012, S. 147; Sand/Kunz 2020, S. 1.

29 Mit Ausnahme der Merkmalsausprägung »divers«.

30 Vgl. Sand/Kunz 2020, S. 13.

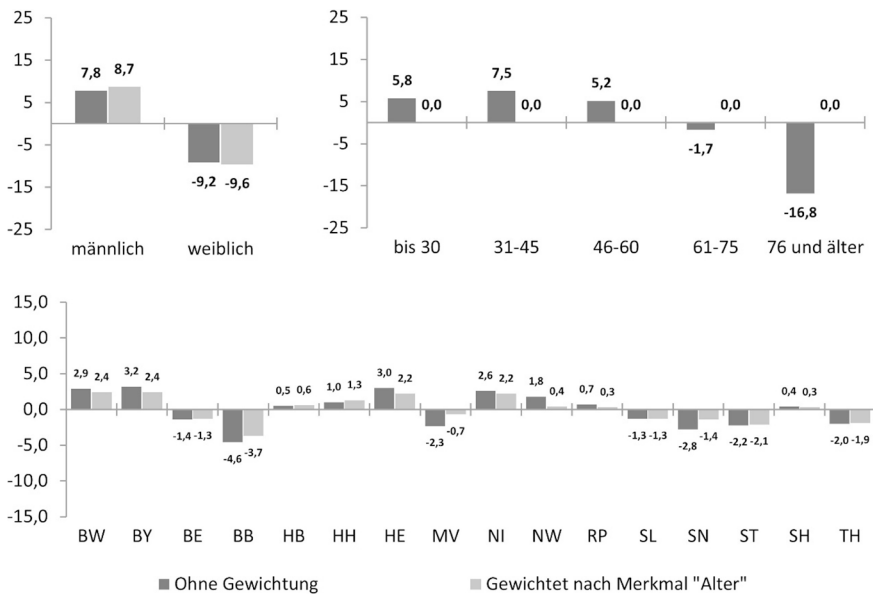
31 Vgl. ebenda.

32 Vgl. Riepl 2012.

33 Vgl. Reinecke 2005, S. 8.

die Abweichungen der Variable Alter auf 0,0 Prozentpunkte in allen Altersklassen reduziert, da die berechneten Gewichte auf ihnen basieren. Die Abweichungen der Variable Geschlecht werden leicht verstärkt, während sich die Differenzen mit Blick auf den Landesverband verringern. Letztgenannter Effekt lässt sich auf den Zusammenhang zwischen Alter und Landesverband zurückführen, da die o.g. ostdeutschen Landesverbände einen Großteil der älteren Parteimitglieder umfassen und nun aufgrund der höheren Gewichtung ebendieser Personen stärker in die Auswertung einfließen. Wenngleich die Abweichungen nicht vollständig geglättet werden, wird zumindest eine Annäherung an die Grundgesamtheit – mit Ausnahme der Stadtverbände Bremen und Hamburg – erreicht.

Abbildung 15: Abweichungen der Teilnehmerstruktur von der Grundgesamtheit (Gesamtpartei mit und ohne Gewichtung nach Alter; Abweichungen in Prozentpunkten)



Trotz der mathematischen Möglichkeit, Geschlecht und Landesverband bei der Berechnung der Gewichtungsfaktoren zu berücksichtigen (und damit alle Abweichungen in Abbildung 12 auf 0,0 Prozentpunkte zu senken), wäre dies mit zwei entscheidenden Problemen verbunden, die nicht zuletzt einen maßvollen und sparsamen Einsatz von Gewichtungen infrage stellen würden:

- 1) Die Variable »Geschlecht« kann nicht vollständig in den Gewichtungsprozess einfließen, da die parteiinterne Mitgliederdatenbank keine Vergleichswerte zur Merkmalsausprägung »divers« enthält. In der Folge müsste ein bestimmter Personenkreis aufgrund seines Geschlechts entweder vollständig aus der Analyse ausgeschlossen oder allenfalls auch mit einem modellierten Gewichtungsfaktor versehen werden, da mathematisch keine Gewichtungswerte für ihn ermittelt werden könnten. Durch die

Abweichungen von Männern und Frauen zugunsten eines höheren männlichen Anteils würden Frauen in der Altersgruppe »76 und älter« einen noch höheren Gewichtungsfaktor oberhalb von 5,4211 erhalten, wodurch die einzelne Befragte in dieser Gruppe noch häufiger in die Auswertung einfließen würde.

- 2) Die Berücksichtigung der Variable »Landesverband« führt zu einer Kreuztabelle mit sehr hoher Zellenzahl<sup>34</sup>, wodurch sich die Zahl der Merkmalsträger pro Zelle stark reduziert. Die Folge wären mitunter enorm hohe Gewichtungsfaktoren für geringe Teilnehmergruppen in schwach besetzten Zellen.<sup>35</sup>

Aus diesen Gründen erfolgt die Datengewichtung für die Analysegruppe der Gesamtpartei im Auswertungsteil der vorliegenden Studie lediglich durch das Altersmerkmal. Ergebnisse für diese Untersuchungsgruppe müssen dadurch auf der Grundlage betrachtet werden, dass (a) eine Gewichtung hinsichtlich des Alters stattfindet und dass (b) die in Abbildung 15 dargestellten Abweichungen der Teilnehmerstruktur von der Grundgesamtheit bezüglich Geschlecht und Landesverband (hellgraue Balken) in der Analyse erhalten bleiben.

---

34 Bei der Berücksichtigung von Alter (5 mögliche Merkmalsausprägungen, ohne fehlende Angaben) und Landesverband (16 mögliche Merkmalsausprägungen) entsteht eine Tabelle mit 80 Zellen (5 x 16). Wird darüber hinaus das Geschlecht zusätzlich mit einbezogen, verdoppelt sich die Anzahl der Zellen auf 160 (5 x 16 x 2).

35 Vgl. Arzheimer 2009, S. 363f.