

VOTO

Methodenexperiment

Evaluation schriftlicher Befragungsmethoden
für Abstimmungsnachbefragungen



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

sotomo
gesellschaft, politik & raum ■ ■ ■

Auftraggeber

Schweizerische Bundeskanzlei
Bundeshaus West
3003 Bern

Auftragnehmer

Forschungsstelle sotomo
Dolderstrasse 24
8032 Zürich

Autoren:

Michael Hermann
David Krähenbühl

Zürich, August 2019



Inhaltsverzeichnis

1	Zu dieser Studie	3
1.1	Erhebungsdesign	3
1.2	Inhalte der Analyse	4
1.3	Methodisches	5
2	Rücklaufquote	6
2.1	Rücklauf: Erhebungsmethoden im Vergleich	6
2.2	Incentivierung und Erinnerungsschreiben	7
2.3	Fragebogenlänge und Versandprozedur	9
2.4	Zeitlicher Verlauf des Rücklaufs	11
3	Repräsentativität	13
3.1	Demographie und Wohnort	13
3.2	Problematik kleiner Subgruppen	20
3.3	Soziokulturelles Profil	23
3.4	Politisches Verhalten	26
3.5	Struktur der Onlineteilnahme	30
3.6	Einfluss der Variationen auf Repräsentativität	33
4	Antwortqualität	37
4.1	Vollständigkeit der Angaben	37
4.2	Qualität offener Fragen	39
4.3	Antwortkategorie «Weiss nicht»	43
4.4	Unterschiede im Antwortverhalten	44
5	Fazit	50
6	Quellenverzeichnis	i

1 Zu dieser Studie

Im Hinblick auf die nächste WTO-Ausschreibung für die Nachbefragung zu den eidgenössischen Volksabstimmungen (VOTO) zieht die Bundeskanzlei als Auftraggeberin den Wechsel der Erhebungsmethode in Betracht. Bei der aktuellen Erhebungsmethode (Telefon, CATI) besteht die Herausforderung generell sinkender Ausschöpfungsquoten sowie einer spezifischen Unterrepräsentation junger Erwachsener. Im Rahmen eines Methodenexperiments hat die Forschungsstelle sotomo als mögliche Alternative zum aktuellen Ansatz eine schriftliche Erhebungsmethode getestet und einer vergleichenden Analyse unterzogen. Hierzu wurden im Nachgang zu den Abstimmungen vom 10. Februar 2019 und vom 19. Mai 2019 parallel zur telefonischen VOTO-Befragung eine schriftliche Erhebung via Print- und Onlinefragebogen durchgeführt. Mit dieser Versuchsanlage können die Stärken und Schwächen der beiden Erhebungsmethoden in einer idealtypischen experimentellen Situation direkt verglichen werden. Zugleich werden die Effekte verschiedener Variationen der schriftlichen Befragung untersucht.

1.1 Erhebungsdesign

Am 10. Februar 2019 wurde über nur eine Vorlage abgestimmt: die Zersiedelungsinitiative. Am 19. Mai wurde über zwei Vorlagen abgestimmt, nämlich die STAF (Steuerreform und AHV-Finanzierung) und das EU-Waffenrecht. Somit beinhaltet die Evaluation auch den Vergleich einer zahlenmässig und inhaltlich eher einfachen Konstellation (Februar) mit einer Konstellation (Mai) mit mehreren Abstimmungen wovon eine davon (die STAF) eher komplex ist. Ziel des Methodenexperiments ist die Verbesserung von Ausschöpfung und Repräsentativität der VOTO-Studie. Die Analyse einer reinen Online-Befragung wurde deshalb ausgeschlossen. Damit liesse zwar die Teilnahme junger Erwachsener verbessern, zugleich würde dies neue Beteiligungshürden für andere Personengruppen (Ältere, weniger Gebildete) schaffen. Als Alternative zur aktuellen telefonischen Befragung wurde deshalb eine schriftliche Mixed-Mode-Befragung gewählt. Die angeschriebenen Personen erhielten die Möglichkeit entweder online oder per gedrucktem Fragebogen teilzunehmen.

Um eine direkte Vergleichbarkeit zu ermöglichen, wurde der für die telefonische Befragung konzipierte Fragebogen verschriftlicht und durch das Bundesamt für Bauten und Logistik (BBL) gleichzeitig mit dem Einladungsschreiben für die telefonische VOTO verschickt. Analog zu VOTO erfolgte die Befragung auf Basis einer Stichprobe aus dem Stichprobenrahmen des Bundesamts für Statistik (BFS). Die Bruttostichprobe von 5000 Stimmberechtigten pro Abstimmungstermin wurde analog nach Sprachregion stratifiziert mit einem überproportionalen Anteil an Personen aus der französischen und italienischen Schweiz.

Die versandten Einladungsbriefe enthielten einen gedruckten Fragebogen mit Couvert für eine Geschäftsantwortsendung. Der gedruckte Fragebogen wurde jeweils nur einmal verschickt. Im Einladungsbrief sowie in allfälligen Erinnerungsschreiben fand sich zudem der personalisierte Zugang zur Online-Befragung. Dabei wurden den Teilnehmenden zwei Zugangswege zur Auswahl gestellt:

- URL + Zugangscode
- QR-Code mit Zugangskoordinaten

Um die Effekte der Ausgestaltung des Erhebungsdesigns zu testen, wurde mit verschiedenen Variationen der Befragung gearbeitet. Die Stichprobe wurde dabei in verschiedene Subgruppen geteilt. So erhielt zum Beispiel eine Hälfte der angeschriebenen Personen bei der Februar-Befragung maximal ein und die andere Hälfte maximal zwei Erinnerungsschreiben. Analoge Variationen wurden in Bezug auf Incentives, die Fragebogenlänge sowie den Ablauf des Versands vorgenommen.

1.2 Inhalte der Analyse

Der Vergleich der Erhebungsmethoden in dieser Studie ist entlang von drei thematischen Dimensionen gegliedert: Rücklaufquote (Kapitel 2), Grad der Repräsentativität der realisierten Stichprobe in Bezug auf die Grundgesamtheit (Kapitel 3) sowie Qualität der Antworten (Kapitel 4). Im Zentrum der Analyse steht dabei immer der direkte Vergleich zwischen schriftlicher und telefonischer Befragung. Dabei werden zum einen die mit den beiden Erhebungsmethoden tatsächlich realisierten Interviews (Nettostichprobe) mit den vom BFS zur Verfügung gestellten Personensample (Bruttostichprobe) verglichen. Zum anderen werden die Antwortverteilungen in den Nettostichproben von schriftlicher und telefonischer Befragungen verglichen. Zusätzlich werden die Effekte verschiedener Variation der Befragung getestet. Dabei steht insbesondere die Frage im Raum, mit welchen Ansätzen die Rücklaufquote sowie die Quote der Online-Teilnahme erhöht werden kann. Berücksichtigt werden dabei immer auch allfällige Effekte auf die Repräsentation und die Antwortqualität.

Während sich Rücklauf und Repräsentativität direkt mittels statistischer Kennzahlen ermitteln lassen, muss die Qualität der Antworten indirekt erschlossen werden. Dabei wird zunächst die Häufigkeit und Verteilung der Antwortkategorien «keine Angabe» (Item-Nonsresponse) und «weiss nicht» untersucht. Ein spezielles Augenmerk wird auf die Beantwortung der offenen Frage zum Grund des Stimmentscheids gelegt. Schliesslich wird untersucht, ob es bei der Beantwortung der geschlossenen Einschätzungsfragen zu systematischen Abweichungen kommt, die aufgrund der Erhebungsmethode zustande kommen.

1.3 Methodisches

Der Vergleich der Häufigkeitsverteilungen der Ausprägungen zwischen Netto- und Bruttostichprobe aber auch zwischen den Nettostichproben der telefonischen und der schriftlichen Befragungen erfolgt mittels Chi-Quadrat-Homogenitätstests inklusive Monte-Carlo-Simulation der p-Werte. Damit kann untersucht werden, ob sich die Verteilungen bezogen auf eine gemeinsame Grundgesamtheit signifikant unterscheiden. Mit Logit-Regressionen lassen sich gegenseitige Abhängigkeiten überprüfen. Für die Analyse des Effekts der verschiedenen Variationen des Experiments, die teilweise in derselben Befragung vorgenommen wurden, wird mit einer Conjoint-Analyse gearbeitet, die den Einfluss der einzelnen Ausprägungen identifizieren kann (Hainmueller et al., 2014; Häusermann et al., 2018).

2 Rücklaufquote

Im ersten Teil dieser Studie steht die Rücklaufquote beziehungsweise die Ausschöpfung im Fokus. Es geht dabei zunächst um den Vergleich zwischen der klassischen telefonischen VOTO-Befragung und der experimentell eingesetzten schriftlichen Befragung. Im Zentrum steht die Frage, ob mit einer schriftlichen Befragung eine bessere Ausschöpfung der Bruttostichprobe erzielt werden kann als mit der aktuellen Befragungsmethode von VOTO. In einem weiteren Schritt werden die Auswirkungen verschiedener Befragungsvarianten auf den Rücklauf der schriftlichen Befragung untersucht.

2.1 Rücklauf: Erhebungsmethoden im Vergleich

Grundlage für die heutige VOTO-Befragung bildet der Stichprobenrahmen des BFS. Die aus dem Stichprobenrahmen gezogene Bruttostichprobe setzt sich aus einer Hauptstichprobe und einer Reserve zusammen. Die Hauptstichprobe umfasst 5000 Stimmberechtigte, die zu 30 Prozent ausgeschöpft werden soll. Das heisst, es sollen insgesamt 1500 telefonische Interviews durchgeführt werden. Falls die Hauptstichprobe nicht ausreicht, um im vorgegebenen Zeitrahmen genügend Interviews durchzuführen, wird die Bruttostichprobe durch die Stimmberechtigten aus der Reserve aufgestockt. Während in der Befragung zur Abstimmung vom Februar 2019 ausschliesslich die Hauptstichprobe verwendet werden konnte, musste bei der Erhebung zur Abstimmung vom Mai 2019 auf die Reserve zurückgegriffen werden. Die Bruttostichprobe wurde damit auf 5650 Stimmberechtigte angehoben.

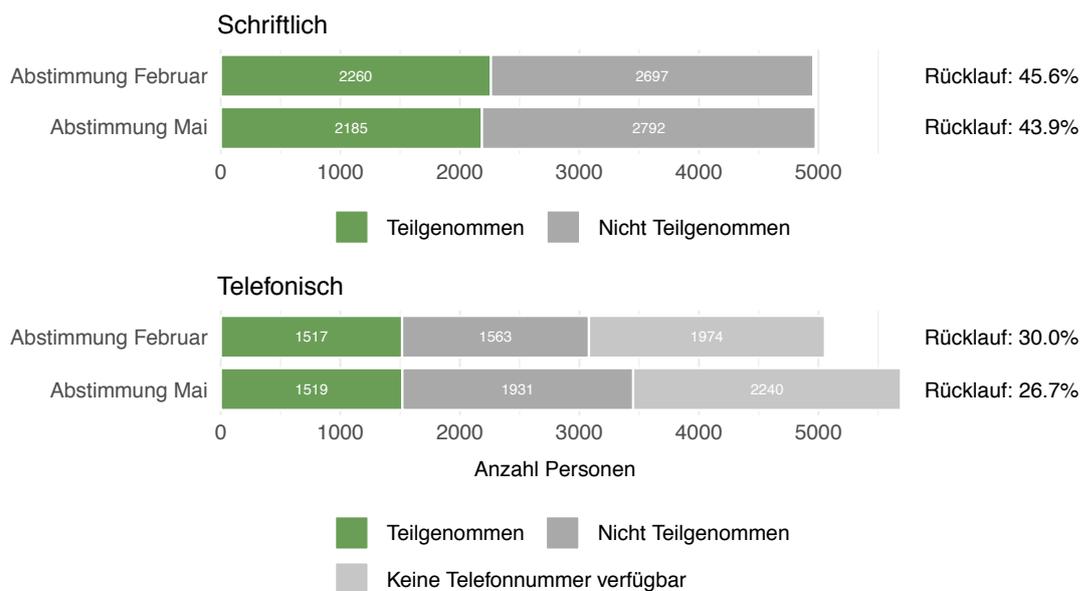


Abbildung 1: Ausschöpfung der Bruttostichprobe bei telefonischen und schriftlichen Befragungen.

Abbildung zeigt die Ausschöpfung im Vergleich. Dabei wird sichtbar, dass sich mit einer schriftlichen Befragung – unter vergleichbaren Bedingungen – eine markant grössere Ausschöpfung der Stichprobe erzielen lässt. Bei der Befragung vom Februar betrug die Rücklaufquote 46 Prozent, bei jener vom Mai 44 Prozent. Die telefonische Befragung erreichte im Februar die angestrebte Ausschöpfung von 30 Prozent. Bei der Mai-Abstimmung musste die Bruttostichprobe erweitert werden, um die angestrebte Zahl an Interviews zu erreichen. Dies senkte die Ausschöpfung auf 27 Prozent.

Die unterschiedliche Ausschöpfung lässt im Wesentlichen auf das Fehlen von telefonischen Verbindungsdaten bei rund 39 Prozent der Personen in der Bruttostichprobe zurückzuführen. Die Postadresse bildet der einzige Verbindungsweg, der für die gesamte Wohnbevölkerung erfasst und mit dem offiziellen Personenregister verknüpft ist. An einer schriftlichen Befragung können heute alle Stimmberechtigten direkt teilnehmen (sofern sie des Lesens mächtig sind). An einer telefonischen Befragung können dies nur etwas mehr als 60 Prozent. Nimmt man nur die Personen mit einer bekannten Telefonverbindung, entspricht die Ausschöpfung der telefonischen jener der (optimal ausgestalteten) schriftlichen Befragung. Das heisst mit anderen Worten: Die tiefere Ausschöpfung der telefonischen Befragung liegt nicht in einer höheren Verweigerungsquote oder in nicht-entgegengenommenen Anrufen, sondern alleine an fehlenden telefonischen Verbindungsdaten.

2.2 Incentivierung und Erinnerungsschreiben

In beiden Durchgängen des Methodenexperiments wurden verschiedene Variationen am Versand und dem Fragebogen vorgenommen und getestet, unter anderem mit dem Ziel, den Rücklauf zu erhöhen, aber auch um Trade-offs zwischen Rücklauf und anderen Anforderungen zu erkennen. Im ersten Durchgang im Februar 2019 wurden zwei Variationen im Befragungsdesign mit je zwei Ausprägungen vorgenommen.

Incentivierung: Mit der Möglichkeit zur Teilnahme an einem Wettbewerb um 10x100 Franken wurde der Einfluss eines finanziellen Anreizes auf die Rücklaufquote untersucht. Dabei wurde eine Hälfte der Stichprobe mit dem Wettbewerb incentiviert, die andere Hälfte nicht.

Erinnerungsschreiben: Mit dem Versand eines zweiten Erinnerungsschreibens bei einer Hälfte der Stichprobe wurde dessen Effekt auf den Rücklauf untersucht. Die andere Hälfte erhielt lediglich ein Erinnerungsschreiben. In beiden Erinnerungsschreiben wurde der Papierfragebogen nicht mehr beigelegt.

Die Einteilung in die jeweiligen Gruppen erfolgte zufällig, so dass alle Variationen gleichermassen in der Stichprobe abgebildet waren. Abbildung 2 zeigt, wie sich der Rücklauf in den verschiedenen Varianten unterscheidet. Dabei ist auch die Art der

Teilnahme (online oder print) dargestellt. Um Störfaktoren auszuschliessen und den Effekt der einzelnen Variationen zu isolieren, wurde zusätzlich eine Conjoint-Analyse durchgeführt. Mit einem entsprechenden Logit-Regressionsmodell lässt sich der Einfluss der interessierenden Faktoren (hier die Versand- und Fragebogenvarianten) auf die Eintretenswahrscheinlichkeit einer abhängigen Variablen untersuchen (hier die Umfrageteilnahme) (Hainmueller et al., 2014; Häusermann et al., 2018). Die Conjoint-Analyse bestätigt dabei die Zusammenhänge, die in Abbildung 2 sichtbar werden.

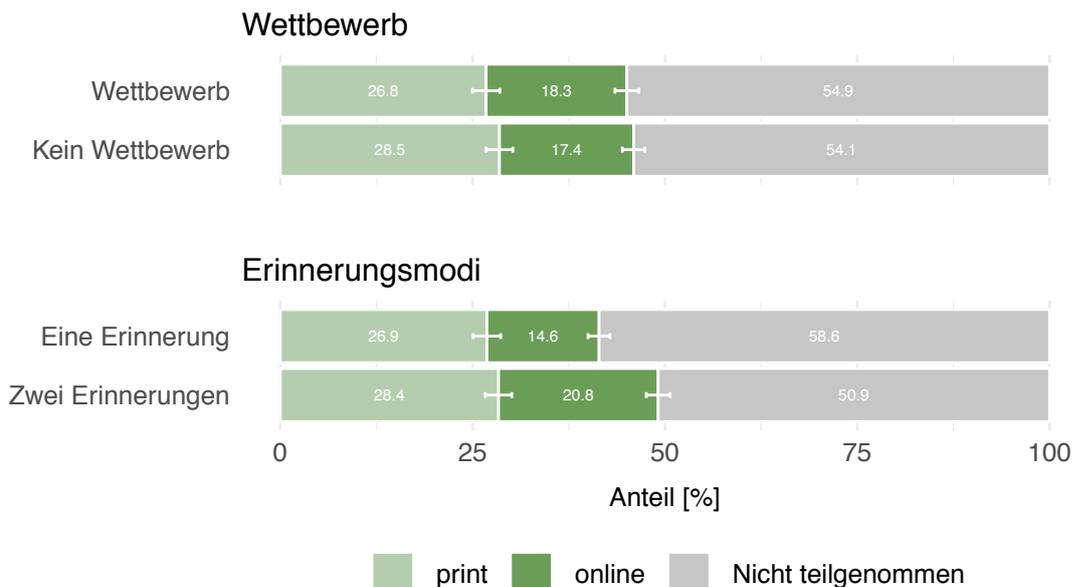


Abbildung 2: Effekt von Incentivierung und Erinnerungsschreiben auf den Rücklauf. (Die horizontalen Linien zeigen den Unsicherheitsbereich.)

Incentivierung: Der Wettbewerb in der Form einer Verlosung von 10 mal 100 Franken hat keinerlei positiven Effekt auf die Teilnahmewahrscheinlichkeit und auch keinen signifikanten Effekt auf das Verhältnis zwischen Online- und Printteilnahme. Damit ist nicht ausgeschlossen, dass andere Incentivierungsformen und insbesondere grössere Wettbewerbssummen einen signifikanten Effekt haben könnten. Der gänzlich ausbleibende Effekt der 1000 Franken, die eingesetzt wurden, legt jedoch zumindest nahe, dass eine monetäre Incentivierung mittels eines Wettbewerbs (in realistischem Rahmen) zu keiner substantiellen Verbesserung des Rücklaufs beitragen kann.

Erinnerungsschreiben: Anders als der Wettbewerb hat das Versenden einer zweiten Erinnerung einen starken positiven Effekt auf den Rücklauf. Gemäss den Resultaten des Logit-Regressionsmodells erhöhen zwei Erinnerungen im Vergleich zu nur einer Erinnerung die Teilnahmewahrscheinlichkeit bei sonst gleichen Bedingungen um 7,5 Prozentpunkte. Der zusätzliche Rücklauf, der mit einem zweiten Erinnerungsschreiben generiert wird, fällt zudem fast ausschliesslich online an.

Dies ist naheliegend, denn der gedruckte Fragebogen wurde bei den Erinnerungen nicht mehr beigelegt und es waren vermehrt junge Befragte, die auf das zweite Erinnerungsschreiben reagierten.

Im Unterscheid zu monetären Incentives sind zwei Erinnerungsschreiben zielführend für die Verbesserung des Rücklaufs. Da die zusätzlichen Teilnahmen vorwiegend online erfolgen, halten sich dabei auch die Kosten in Grenzen. Zu beachten gilt, dass der offizielle Charakter der Anschreiben mit dem Absender Bundeskanzlei bei den Adressaten den Eindruck erwecken kann, dass die Teilnahme an der Befragung obligatorisch sei. Es kann zudem davon ausgegangen werden, dass dieser Eindruck durch den Versand von zwei Erinnerungen in einer kurzen zeitlichen Abfolge noch weiter verstärkt wird. Dies erhöht den Rücklauf, kann aber eventuell auch zu Verwirrung bei den Angeschriebenen führen.

2.3 Fragebogenlänge und Versandprozedur

Im zweiten Teil des Experiments wurde eine etwas komplexere Kombination von Befragungsvariationen getestet. Dabei ging es zum einen um die Länge des Fragebogens, zum anderen um die Versandprozedur für die Anschreiben. Auf eine Incentivierung wurde hingegen verzichtet.

Fragebogenlänge: Während im Februar 2019 nur über eine Vorlage abgestimmt wurde, kamen im Mai 2019 zwei Vorlagen zur Abstimmung. Parallel zur Zahl der Vorlagen nimmt die Zahl der Fragen zu. Trotz kompakterer Umsetzung des gedruckten Fragenbogens umfasste der Mai-Fragebogen sieben gedruckte Seiten. Anders als bei einer telefonischen Befragung ist die Länge bzw. der Umfang eines gedruckten Fragebogens sofort sichtbar. Wie sich die Länge des Fragebogens auf den Rücklauf auswirkt, wurde durch den Versand von zwei unterschiedlichen Fragebogen untersucht. Einer Hälfte der Befragten wurde der vollständige, von VOTO vorgegebene Fragebogen zugeschickt. Der anderen Hälfte wurde eine gekürzte Form desselben Fragebogens vorgelegt, bei dem insbesondere die eher spezifischen soziodemographischen Detailfragen weggelassen wurden. Der kurze Fragebogen umfasste in der ausgedruckten Form vier Seiten, das heisst drei Seiten weniger als der vollständige VOTO-Fragebogen.

Versandprozedur: Der zweite Test betraf die Versandprozedur. Einem zufällig ausgewählten Teil der Befragten wurde das Einladungsschreiben bereits am Donnerstag vor dem Abstimmungswochenende zugestellt. Dieses Schreiben enthielt den Zugang zur Online-Abstimmung, aber noch keinen gedruckten Fragebogen. Diesen erhielten die Befragten erst am Mittwoch nach der Abstimmung als eine Form der ersten Erinnerung. 1,5 Wochen später folgte eine zweite Erinnerung (ohne gedruckten Fragebogen). Diesem Test lag die Hypothese zugrunde, dass sich mit der geänderten Prozedur primär die Online-Teilnahme und sekundär auch der Rücklauf insgesamt erhöhen liesse. Dies einerseits aus dem Grund, dass die

Angeschriebenen zunächst nur die Möglichkeit erhielten, online an der Befragung teilzunehmen. Andererseits weil damit der Zeitraum der potenziellen Teilnahme an der Befragung verlängert wurde und die durchschnittliche zeitliche Distanz zum Abstimmungsentscheid und der Auseinandersetzung mit den Sachverhalten in der Befragung verkürzt wurde. Bei zwei Dritteln der Befragten wurde dieses Prozedere gewählt. Beim übrigen Drittel wurde das bereits in der Februarabstimmung getestete Verfahren angewandt. Das Einladungsschreiben mit Printfragebogen wurde am Montag nach der Abstimmung versendet. Jeweils im Abstand einer Woche erhielten die Befragten, die noch nicht teilgenommen hatten, ein erstes und dann ein zweites Erinnerungsschreiben.

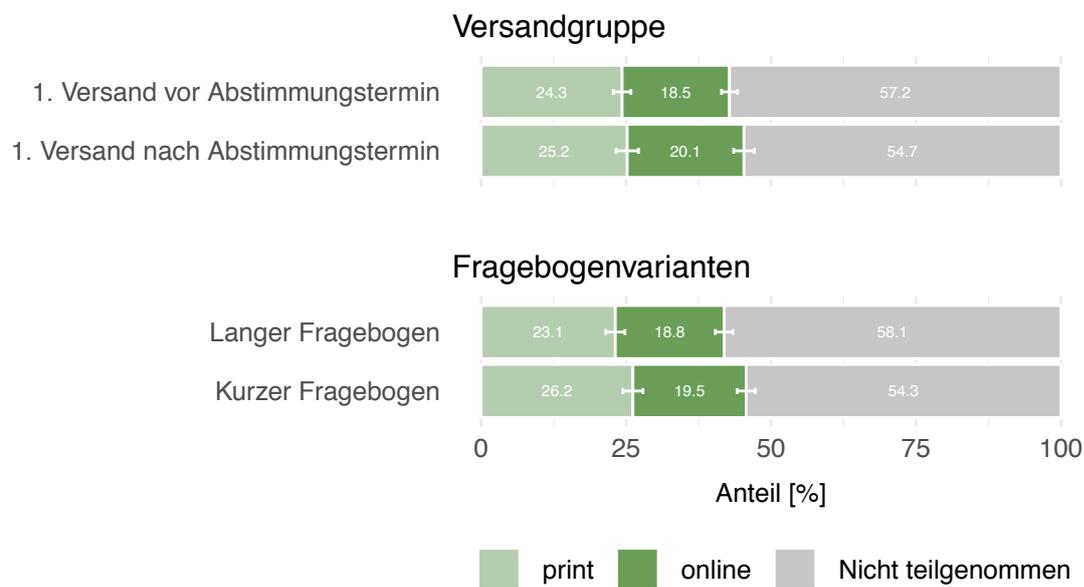


Abbildung 3: Effekt von Fragebogenlänge und Versandprozedur auf den Rücklauf.

Analog zur Februar-Abstimmung wurden die Effekte auf den Rücklauf ebenfalls mit einem Conjoint-Ansatz und einem anschliessenden Logit-Regressionsmodell getestet. Abbildung 3 zeigt zudem die jeweiligen Quoten. Folgende Schlüsse lassen sich daraus ableiten:

Fragebogenlänge: Die Länge des Fragebogens hat einen gewissen Einfluss auf den Rücklauf. Beim längeren Fragebogen ist die vorhergesagte Teilnahmewahrscheinlichkeit an der Umfrage um rund 4 Prozentpunkte tiefer als beim kürzeren. Die Resultate des Logit-Regressionsmodells zeigen allerdings, dass dieser Wert mit einer relativ grossen Unsicherheit behaftet ist. Das Konfidenzintervall auf einem Signifikanzniveau von 95% reicht von 0 bis -7. Die Ergebnisse bestätigen, dass die Länge des Fragebogens ein Thema ist, das bei einer allfälligen schriftlichen Durchführung von VOTO beachtet werden muss. Mit dem heutigen Fragekatalog zeigen sich bereits bei zwei Vorlagen Anzeichen für eine negative Wirkung auf

den Rücklauf. Oftmals kommen jedoch auch drei oder vier Vorlagen pro Termin zur Abstimmung.

Versandprozedur: Entgegen der Erwartung hat die Variation der Versandprozedur keine positive Auswirkung sowohl auf die Rücklaufquote wie auch auf den Anteil der Online- Teilnahmen. Die Möglichkeit zur Online-Teilnahme an der Umfrage bereits vor dem Abstimmungswochenende ist damit offensichtlich nicht zielführend. Bemerkenswert ist dabei insbesondere, dass mit dem angepassten Timing der Anteil der Online-Teilnahmen nicht gesteigert werden kann. Eine plausible Erklärung kann jedoch darin liegen, dass beide getesteten Prozeduren aus drei Versänden bestehen und nur in einem davon der gedruckte Fragebogen mitgeschickt wird. Bei der Grundvariante mit dem Printfragebogen im ersten Schreiben führt insbesondere die zweite Erinnerung zu einer Erhöhung der Online-Teilnahmen. Damit wird offenbar der Effekt der frühzeitig möglichen Online-Teilnahme in der abgewandelten Prozedur kompensiert.

2.4 Zeitlicher Verlauf des Rücklaufs

In den nächsten Abschnitten wird der zeitliche Rücklauf für die beiden Durchgänge des Methodenexperiments aufgezeigt. Anhand des zeitlichen Verlaufs des Rücklaufs kann analysiert werden, wie der direkte Einfluss der Erinnerungen auf die Rücklaufquote ist. Weil für die Abstimmung im Februar nicht erhoben wurde, wann die Printfragebögen beim Scanzentrum angekommen sind, ist die folgende Auswertung auf die Abstimmung vom Mai fokussiert.

2.4.1 Zeitlicher Verlauf Print

Abbildung 4 zeigt den Zeitverlauf des Rücklaufs für die ausgedruckten Fragebögen. Ein erster Anstieg ist zwischen dem 22. und dem 24. Mai zu verzeichnen. Dies sind vermutlich insbesondere Teilnehmende, die sich in Versandgruppe A befanden, der Gruppe, die den ausgedruckten Fragebogen schon am Montag nach dem Abstimmungstag (20. Mai) zugestellt bekam. Der Anstieg beim 28. Mai ist wohl nicht primär mit dem Versand der Erinnerungen zu diesem Zeitpunkt zu erklären, sondern Damit, dass der 28. Mai der Dienstag nach dem ersten Wochenende nach den Einladungen war. Es kann angenommen werden, dass wohl viele Teilnehmende erst am Wochenende dazu gekommen sind, die Umfrage auszufüllen und die Fragebögen mit der Post am Dienstag nachher bei publitest angekommen sind. Dies gilt vermutlich insbesondere auch für die Befragten, die in der Versandgruppe B waren, also der Gruppe, die den ausgedruckten Fragebogen erst am Mittwoch nach der Abstimmung (22. Mai) erhielten. Am 4. Juni wurde für alle Versandgruppen die zweite Erinnerung verschickt. Danach nahm der Rücklauf zwar stetig ab, es ist aber doch ersichtlich, dass diese Erinnerung einen Einfluss hatte.

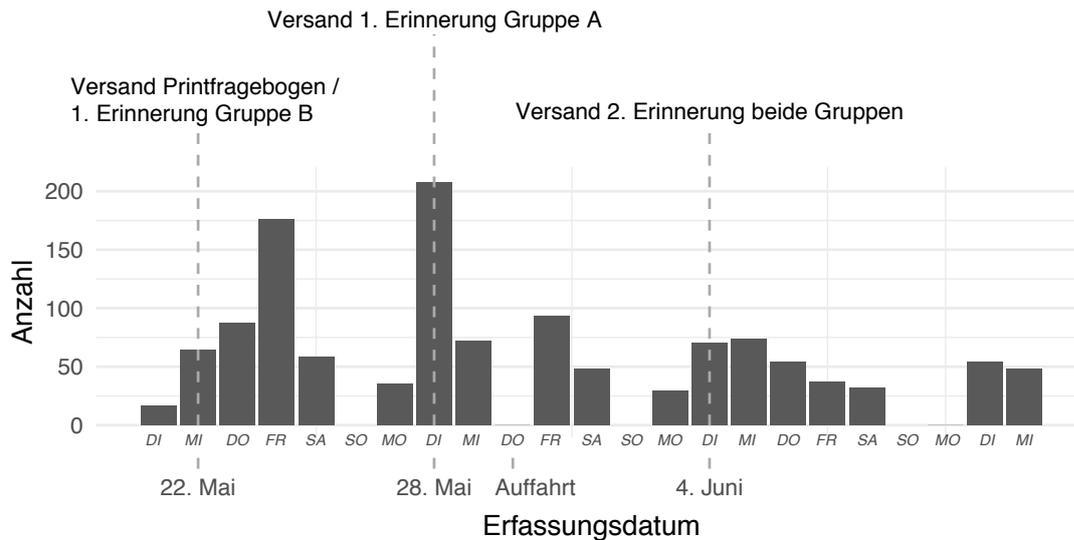


Abbildung 4: Zeitlicher Verlauf des Rücklaufs, Printbefragung Abstimmung Mai (gestrichelte Linien: Versand Einladung/Erinnerungen)

2.4.2 Zeitlicher Verlauf Online

Aus Abbildung 5, die den Zeitverlauf des Rücklaufs der Onlinebefragung zeigt, wird ersichtlich, dass der Rücklauf der Onlinebefragung sehr direkt auf den Versand der Einladungen und Erinnerungen reagiert. Versandgruppe B hat die Einladung bereits drei Tage vor dem Abstimmungstermin, am 16. Mai, erhalten. In diesen drei Tagen haben immerhin 143 Personen an der Umfrage teilgenommen. Danach sieht man bei den Erinnerungen für Gruppe B (22. Mai) und A (28. Mai) jeweils einen Anstieg der Teilnehmenden. Insbesondere am 4. Juni, wo beiden Versandgruppen gleichzeitig eine Erinnerung zugesandt wurde, stieg der Rücklauf nochmals sehr stark an.

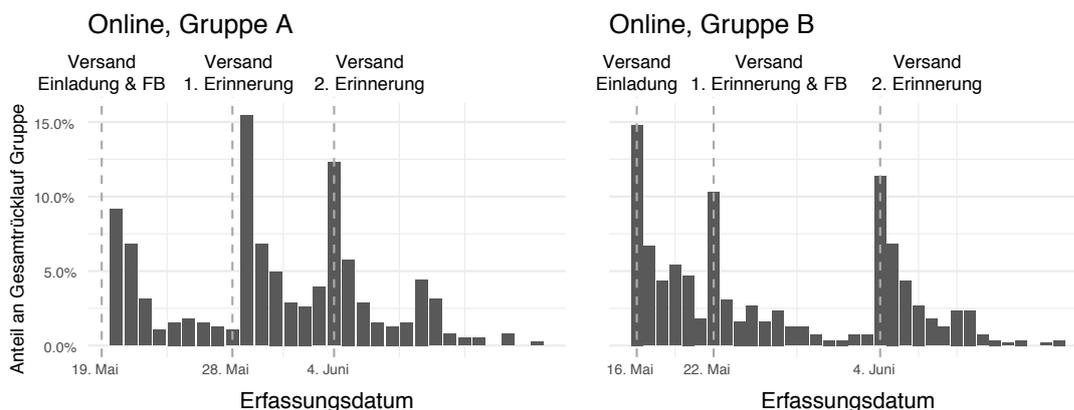


Abbildung 5: Zeitlicher Verlauf des Rücklaufs, Onlinebefragung Abstimmung Mai (gestrichelte Linien: Versand Einladung/Erinnerungen)

3 Repräsentativität

Im zweiten Teil dieser Studie steht die Repräsentativität der Stichprobe im Fokus. Grundsätzlich basieren beide Erhebungsmethoden auf einer äusserst repräsentativen Stichprobe aus dem Stichproberahmen SRPH des BFS. Mit beiden Erhebungsmethoden können jedoch nicht alle Personen der BFS-Stichprobe befragt werden. Es muss deshalb zwischen Brutto- und Nettostichprobe unterschieden werden. Die *Bruttostichprobe* entspricht den Personen, die zufällig aus dem Stichprobenrahmen gezogen werden und damit potenziell an der Befragung teilnehmen können. Die *Nettostichprobe* entspricht den Personen, die tatsächlich an der Befragung teilnehmen. Die Repräsentativität der Erhebungsmethode hängt davon ab, wie gut die Struktur der Nettostichprobe mit jener der Bruttostichprobe übereinstimmt. Abweichungen entstehen, wenn die Erhebungsmethode dazu führt, dass bestimmte Gruppen vermehrt andere dagegen weniger oft an der Befragung teilnehmen. Die Repräsentativität der Netto- im Vergleich zur Bruttostichprobe ist ein wichtiges Qualitätsmerkmal der Erhebungsmethode.

Ein direkter Vergleich der Struktur der Brutto- und der Nettostichprobe ist nur für ausgewählte demographische Merkmale möglich. Die Repräsentativität anderer wichtiger Merkmale lässt sich allerdings entweder anhand der bekannten Verteilung in der Grundgesamtheit ermitteln (z.B. für die Abstimmungsteilnahme), oder es können Ergebnisse amtlicher Statistiken beigezogen werden (z.B. für die Anteile der Bildungsabschlüsse).

3.1 Demographie und Wohnort

Den Effekt der Erhebungsmethode auf die Repräsentativität der Nettostichprobe lässt sich direkt für alle Merkmale bestimmen, deren Ausprägung auch in der Bruttostichprobe bekannt ist. Das betrifft die Merkmale, die in den Registerdaten des BFS erfasst sind und über den Stichprobenrahmen den Nutzenden zugänglich gemacht werden. Es handelt sich dabei namentlich um die demographischen Hauptmerkmale Geschlecht, Alter und Wohnort. Bei diesen Merkmalen können Abweichungen in der Verteilung zwischen Brutto- und Nettostichprobe unmittelbar untersucht und dargestellt werden.

3.1.1 Geschlecht

Aus Abbildung 6 wird ersichtlich, dass die Geschlechteranteile der Nettostichprobe sowohl bei der schriftlichen als auch der telefonischen Befragung an beiden Abstimmungsterminen sehr genau mit den Anteilen in der Bruttostichprobe übereinstimmen. Die Repräsentativität in Bezug auf das Geschlecht stellt damit unabhängig von der Erhebungsmethode keine besondere Herausforderung dar.

Der Chi-Quadrat-Test auf Homogenität der Verteilung zeigt, dass es für die schriftlichen Befragung keine signifikanten Unterschiede zwischen Brutto- und Nettostichprobe gibt. Der p-Wert liegt weit über dem typischen Kriterium von 0.05. (Abstimmung Februar: $\tilde{\chi}^2 = 0.04$, $p = 0.84$ / Abstimmung Mai: $\tilde{\chi}^2 = 0.43$, $p = 0.52$). Ähnlich klar sind die Verhältnisse bei der telefonischen Befragung. (Abstimmung Februar: $\tilde{\chi}^2 = 0.4$, $p = 0.53$ / Abstimmung Mai: $\tilde{\chi}^2 = 0.06$, $p = 0.82$). Dennoch fällt auf, dass die Männer in allen vier Nettostichproben leicht – aber eben nicht signifikant – übervertreten sind.

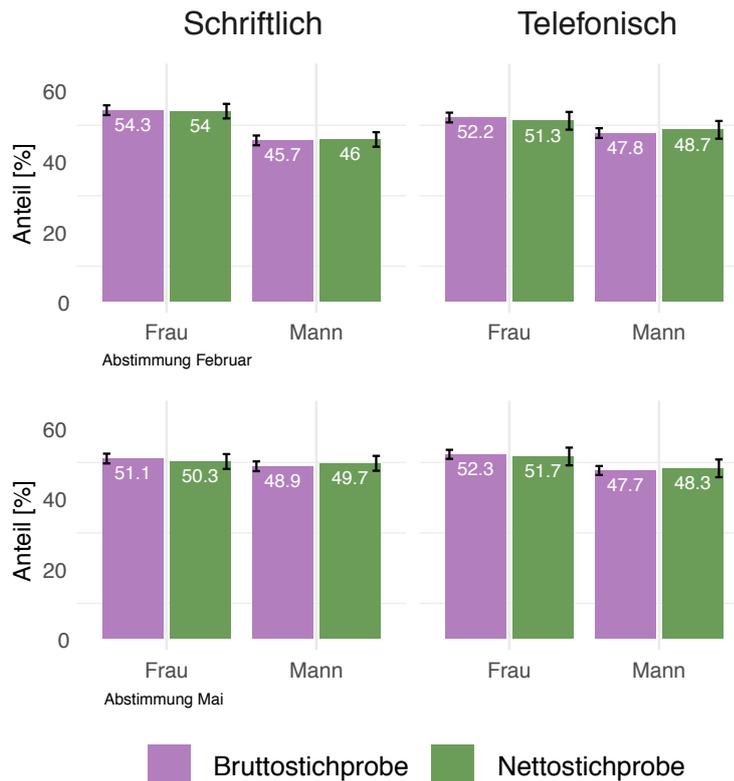


Abbildung 6: Geschlecht: Vergleich Netto- und Bruttostichprobe der beiden Befragungsmodi

3.1.2 Alter

Eine grössere Herausforderung als das Geschlecht stellen die Altersgruppen dar. Der Zugang zu den verschiedenen Befragungsmethoden hängt stark vom Alter der Befragten ab. Ältere sind schwieriger online zu erreichen, Jüngere haben oft keinen Festnetzanschluss. Entsprechend zeigen sich hier grössere Abweichungen in der Zusammensetzung der Stichproben. Abbildung 7 zeigt die Altersverteilung für beide Erhebungsmethoden und beide Abstimmungstermine. Dabei zeigen sich insbesondere bei der telefonischen Befragung markante Abweichungen in der Alterszusammensetzung der Netto- im Vergleich zur Bruttostichprobe. In der Nettostichprobe ist die jüngere Generation stark unter- und die ältere stark über-

vertreten. Auch bei der schriftlichen Befragung sind die Jüngeren untervertreten, jedoch in weit geringerem Ausmass. Leicht übervertreten sind hier die 40- bis 70-Jährigen.



Abbildung 7: Alter: Vergleich der Netto- und Bruttostichprobe der beiden Befragungsmodi

Der Chi-Quadrat-Test auf Basis einer fünfstufig gruppierten Altersvariable bestätigt was die Alterskurven in Abbildung 7 zeigen. Die Altersverteilung innerhalb der tatsächlich befragten Personen weicht sowohl bei der schriftlichen als auch bei der telefonischen Befragung signifikant von der Bruttostichprobe ab. Bei der schriftlichen Befragung fällt die Abweichung aber geringer aus (Abstimmung Februar: $\tilde{\chi}^2 = 14.99$, $p = 0.004$ / Abstimmung Mai: $\tilde{\chi}^2 = 17.37$, $p = 0.001$) als bei der telefonischen (Abstimmung Februar: $\tilde{\chi}^2 = 78.36$, $p = 0.000$ / Abstimmung Mai: $\tilde{\chi}^2 = 123.52$, $p = 0.000$).

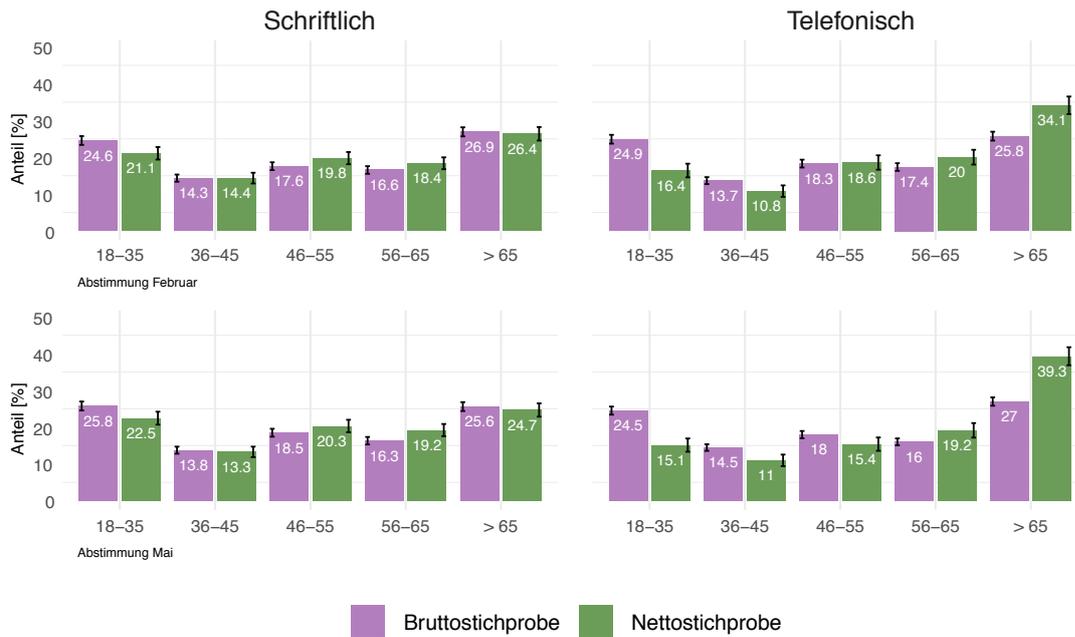


Abbildung 8: Altersgruppen: Vergleich der Netto- und Bruttostichprobe der beiden Befragungsmodi

3.1.3 Siedlungstyp

Die Repräsentativität in Bezug auf den Wohnort wird nicht auf Stufe der Wohn-gemeinde untersucht, sondern auf Basis relevanter Regionalisierungen. Eine für politische Fragen wichtige Regionalisierung bildet jene nach dem Siedlungstyp, konkret ist vor allem der Gegensatz zwischen Stadt und Land politisch bedeutsam. Abbildung 9 zeigt, dass die Verteilung der Siedlungstypen in den Nettostichproben beider Abstimmungstermine in der schriftlichen Befragung relativ nahe an derjenigen in der jeweiligen Bruttostichprobe ist. Einzig die leichte Übervertretung des Siedlungstyps «Grosse Agglomeration» in beiden Befragungen fällt ins Auge. Dennoch zeigt der Chi-Quadrat-Test keine signifikanten Abweichungen für die beiden schriftlichen Befragungen. (Abstimmung Februar: $\tilde{\chi}^2 = 4.33$, $p = 0.37$ / Abstimmung Mai: $\tilde{\chi}^2 = 1.45$, $p = 0.84$).

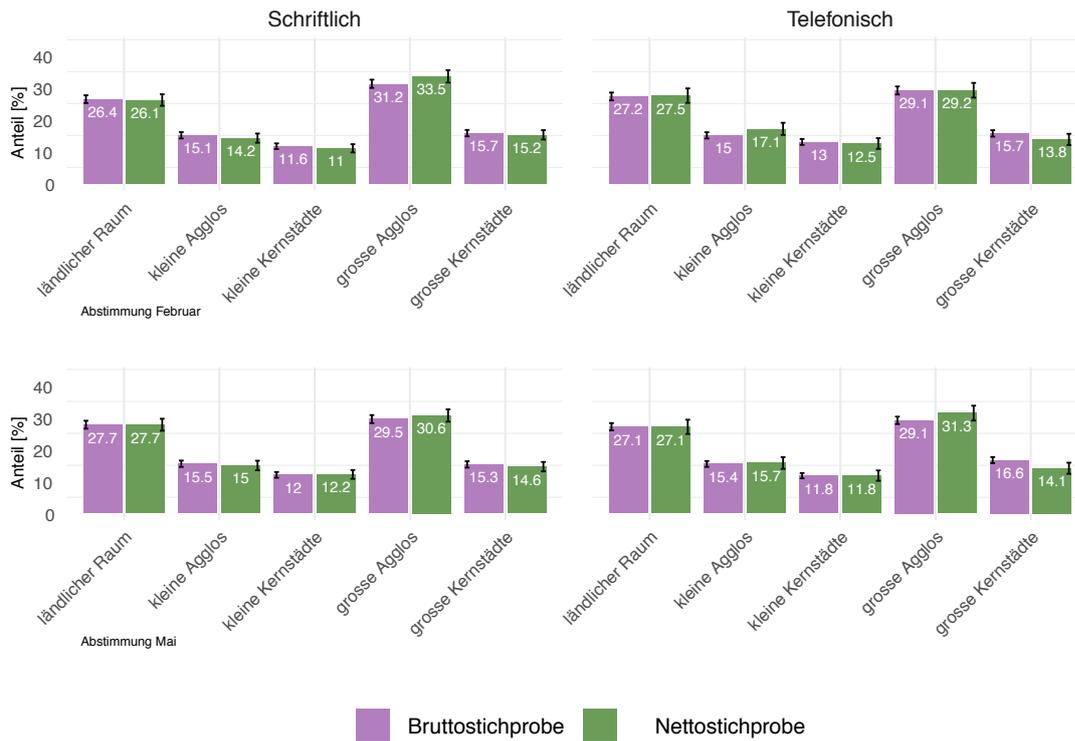


Abbildung 9: sotomo Raumtypen: Vergleich der Netto- und Bruttostichprobe der beiden Befragungsmodi

Auch in der telefonischen Befragung gibt es an beiden Abstimmungsterminen keine signifikanten Unterschiede in der Rücklaufquote zwischen den verschiedenen Siedlungstypen (Abstimmung Februar: $\tilde{\chi}^2 = 5.96$, $p = 0.2$ / Abstimmung Mai: $\tilde{\chi}^2 = 7.65$, $p = 0.1$). Abbildung 9 zeigt jedoch, dass Personen aus grossen Kernstädten eher etwas unter- und solche aus grossen Agglomerationen etwas übervertreten sind. Wie für die schriftliche Befragung liegen diese Befunde beim Chi-Quadrat-Test in einem nicht signifikanten Bereich.

3.1.4 Sprachregionen

In der telefonischen Befragung liegen die Anteile nach Sprachregionen sehr nahe an den Verhältnissen der Bruttostichprobe. Entsprechend zeigen die Chi-Quadrat-Tests keine signifikante Abweichungen in der Verteilung: (Abstimmung Februar: $\tilde{\chi}^2 = 0.96$, $p = 0.62$ / Abstimmung Mai: $\tilde{\chi}^2 = 1.74$, $p = 0.42$). Es ist wichtig, zu erwähnen, dass die jeweiligen Anteile dabei nicht mit den Anteilen in der Grundgesamtheit übereinstimmen, weil Personen aus der französisch- und der italienischsprachigen Schweiz bereits in der Bruttostichprobe übervertreten sind. Für die schriftliche Befragung fällt auf, dass hier offenbar der Rücklauf aus der französischen Schweiz grösser ist als aus den anderen beiden Sprachregionen. Zumindest bei der Mai-Abstimmung zeigt der Chi-Quadrat-Test einen signifikanten

ten Unterschied in der Verteilung (Abstimmung Februar: $\tilde{\chi}^2 = 3.00$, $p = 0.23$ / Abstimmung Mai: $\tilde{\chi}^2 = 17.26$, $p = 0.000$).

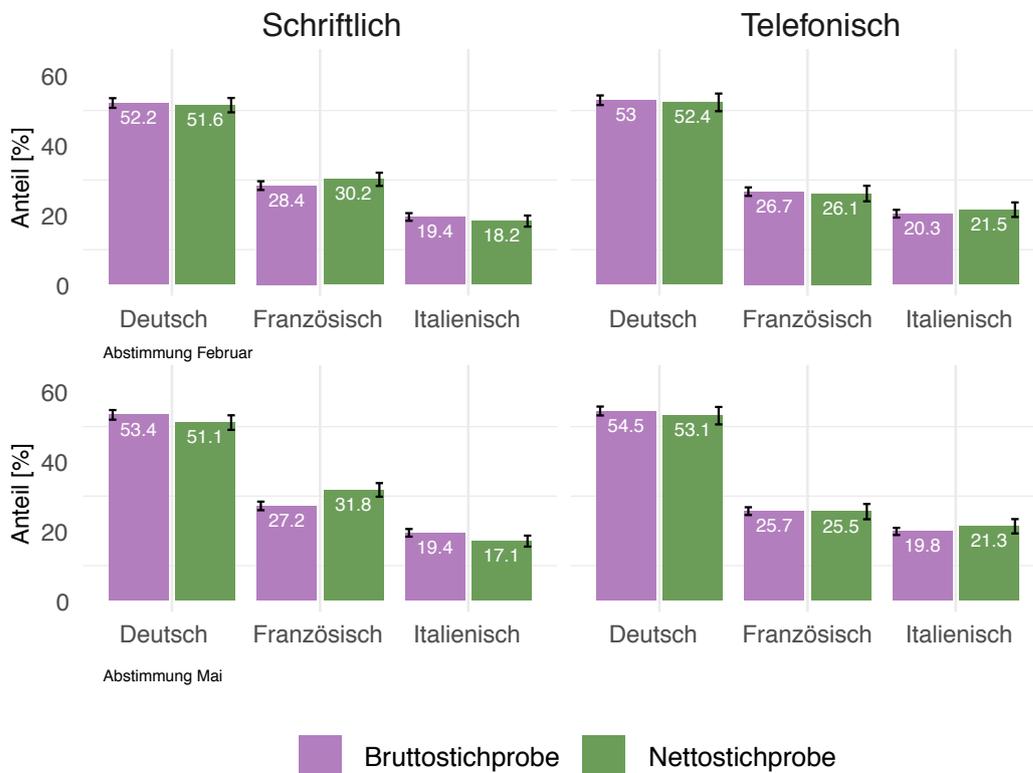


Abbildung 10: Sprachregionen: Vergleich der Netto- und Bruttostichprobe der beiden Befragungsmodi

3.1.5 Integrale Betrachtung

Geschlecht, Alter und Wohnort sind die bekannten Merkmale der aus dem BFS-Stichprobenrahmen gezogenen Personen. In den vorangegangenen Abschnitten wurde die Verteilung dieser Merkmale bei den realisierten Interviews jeweils einzeln untersucht. Um gegenseitige Abhängigkeiten zu berücksichtigen und eine Gesamteinschätzung zu erhalten, werden diese Merkmale im Folgenden in ein Logit-Regressionsmodell integriert. Dieses schätzt den Effekt der verschiedenen Merkmalsausprägungen auf die Wahrscheinlichkeit für eine Teilnahme an der Umfrage.

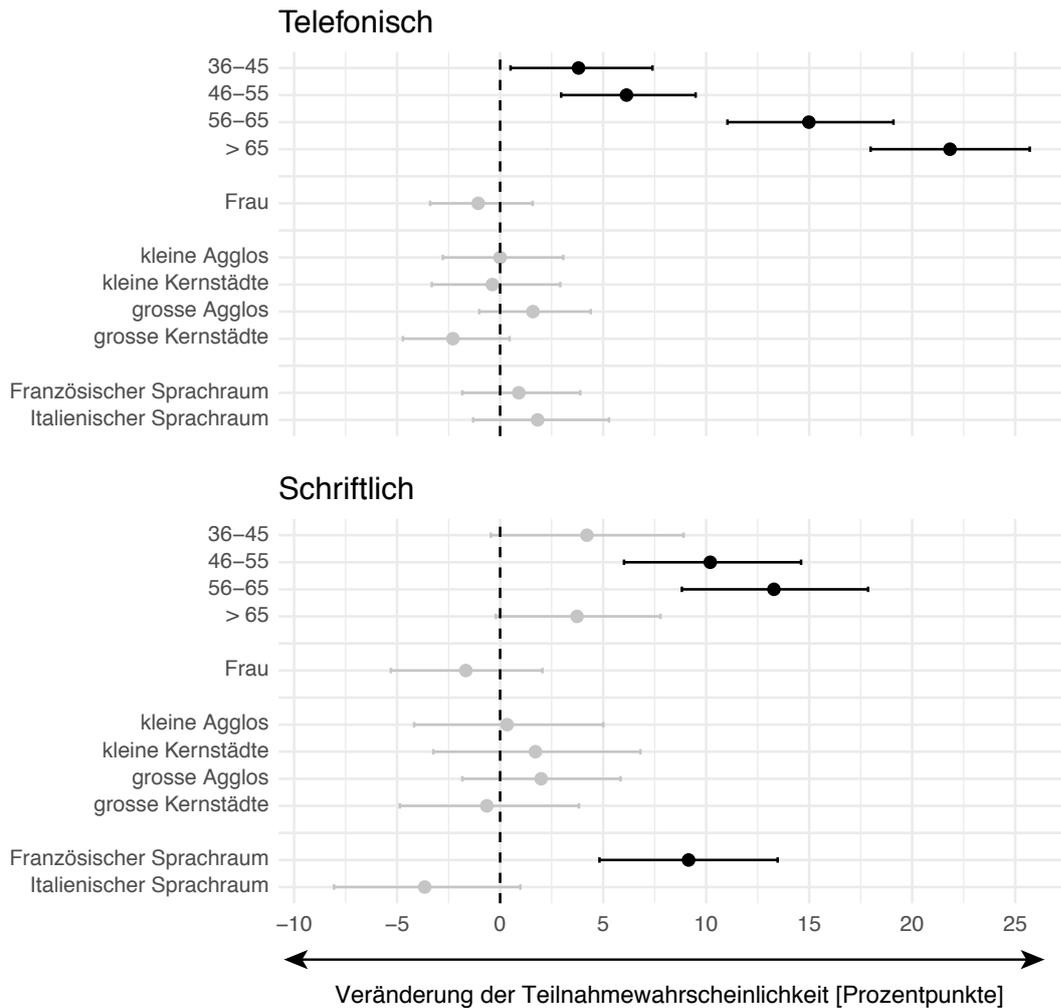


Abbildung 11: Einflussfaktoren für Wahrscheinlichkeit der Teilnahme an der Umfrage: Schriftliche und der telefonische Befragung im Vergleich.

Abbildung 11 zeigt je ein Logit-Regressionsmodell zur Schätzung der Teilnahmewahrscheinlichkeit an der telefonischen bzw. der schriftlichen Befragung im Nachgang zur Abstimmung vom Mai 2019. Dargestellt sind, wie in solchen Logit-Regressionsmodellen üblich, die Abweichungen zur Teilnahmewahrscheinlichkeit der jeweiligen Referenzkategorien in Prozentpunkten. Die einzelnen Kategorien fügen sich dabei zu einer Referenzgruppe aus Deutschschweizer Männern zwischen 18 und 35 Jahren aus dem ländlichen Raum zusammen. So lässt sich aus Abbildung 11 beispielweise herauslesen, dass bei den über 65-Jährigen die Wahrscheinlichkeit an der telefonischen Befragung teilzunehmen um 22 Prozentpunkte höher liegt als bei der Referenzkategorie der 18 bis 35-Jährigen, wenn alle anderen Faktoren konstant gehalten werden. Die Schätzer lassen sich auch kombinieren: So liegt die Teilnahmewahrscheinlichkeit der über 65-Jährigen (+22) aus einer Kernstadt (-2,5) insgesamt 19,5 Prozentpunkte über der Referenzgruppe, wenn die anderen Faktoren konstant gehalten werden.

Eine repräsentative Stichprobe wird dann realisiert, wenn sich die Teilnahme-wahrscheinlichkeit zwischen den verschiedenen Merkmalsausprägungen nicht unterscheidet und die Schätzer für die Kategorien entsprechend nahe beim Nullpunkt liegen. Abbildung 11 zeigt, dass bei der telefonischen Befragung das Alter stark verzerrend wirkt, alle anderen Merkmalsausprägungen jedoch kaum von Null unterscheidbar zu liegen kommen. Bei der schriftlichen Befragung wirkt das Alter weniger stark verzerrend, dafür sind zusätzlich die Teilnehmenden aus der französischen Schweiz signifikant überrepräsentiert (der Unsicherheitsbereich schneidet den Nullpunkt nicht).

Generell fällt auf, dass die Streuung der Schätzer bei der schriftlichen Befragung leicht grösser ist und etwas weniger Schätzer ganz nahe bei null liegen. Diese leicht kleinere Streuung bringt wahrscheinlich zum Ausdruck, dass sich die Zusammensetzung der Nettostichprobe bei einer schriftlichen Befragung weniger gut steuern lässt als bei einer telefonischen. Wenn bei einer telefonischen Befragung die benötigte Zahl der Teilnehmenden einer bestimmten Gruppe erreicht ist, können mittels Quoteneinstellungen gezielt Personen aus anderen Gruppen rekrutiert werden. Wenn zum Beispiel genügend Interviews mit Personen aus dem französischen Sprachraum realisiert sind, dann lässt sich die restliche Erhebung auf die anderen Sprachräume beschränken. Zu einer Verzerrung der Repräsentativität in Bezug auf die demographischen Hauptmerkmale kommt es nur, wenn sich in der vorgegebenen Zeit in einer Kategorie nicht genügend Interviews realisieren lassen. Im Fall der zeitkritischen VOTO-Befragung gilt dies namentlich für junge Erwachsene.

Bei einer schriftlichen Befragung ist eine entsprechende Feinsteuerung weniger notwendig. Alle Personen der Bruttostrichprobe werden gleich oft kontaktiert, weshalb es umso auffälliger ist, dass sich dennoch eine grosse strukturelle Übereinstimmung zwischen der Brutto- und der Nettostrichprobe zeigt.

3.2 Problematik kleiner Subgruppen

Kleinere Verzerrungen in der Zusammensetzung der realisierten Stichprobe lassen sich durch statistische Hochrechnungen und Gewichtungen problemlos korrigieren. Eine grössere Herausforderung stellen spezifische Subgruppen dar, deren Umfang in der Stichprobe so gering sind, dass keine verlässlichen Aussagen möglich sind. Eine zu dünne Datenbasis lässt sich nicht mittels statistischer Hochrechnungs- und Gewichtungsverfahren ausgleichen. Dieser Abschnitt widmet sich der Problematik der zu kleinen Subgruppen.

In Abbildung 13 ist dargestellt, welche dieser Subgruppen in den beiden Stichproben der schriftlichen und der telefonischen Befragungen am stärksten überbeziehungsweise unterrepräsentiert sind. Die Stichproben wurden dazu über beide Abstimmungstermine gepoolt, das heisst, in den Auswertungen sind jeweils die

Daten von beiden Abstimmungsterminen zu einem Datensatz zusammengefasst worden. In der schriftlichen Befragung sind die über 65-jährigen Frauen aus dem italienischen Sprachraum am stärksten unterrepräsentiert: Die Grösse dieser Subgruppe beträgt in der Nettostichprobe nur etwa 70 Prozent im Vergleich zur Bruttostichprobe.

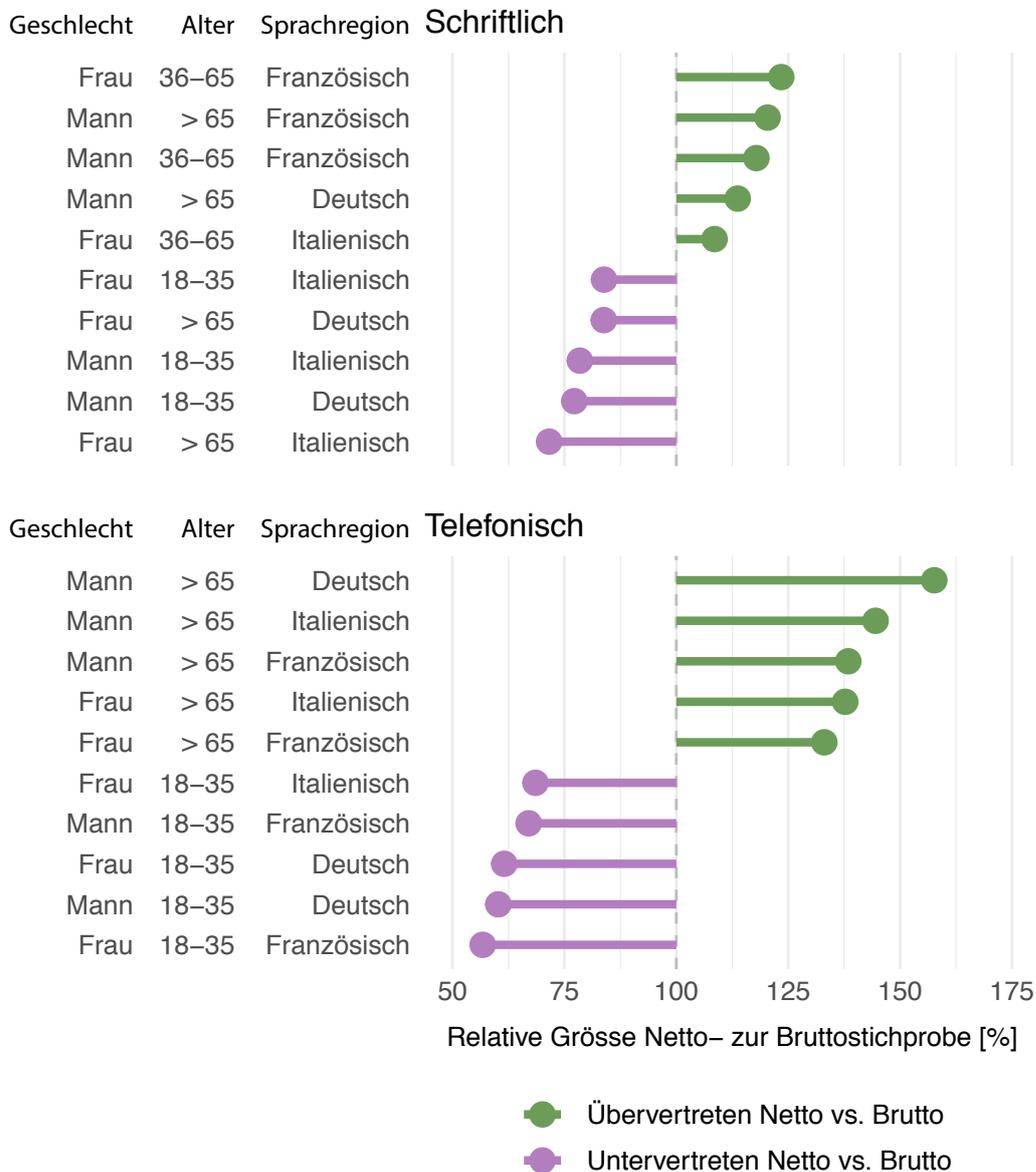


Abbildung 12: Vergleich der Netto- und Bruttostichprobe: Kombination aus Altersgruppen, Geschlecht und Sprachregion. Jeweils fünf über- und unterrepräsentierteste pro Befragungsmodus

Bei den telefonischen Befragungen zeigen sich insgesamt deutlich grössere Abweichungen in der Repräsentation der verschiedenen Subgruppen. Abbildung 7 zeigt, dass junge Frauen aus dem französischen Sprachgebiet am stärksten unterrepräsentiert sind. Im Vergleich zur Bruttostichprobe ist ihr Anteil in der

Nettostichprobe nur unwesentlich mehr als halb so gross (55%). Am stärksten überrepräsentiert sind Deutschschweizer Männer über 65 Jahre (160%)

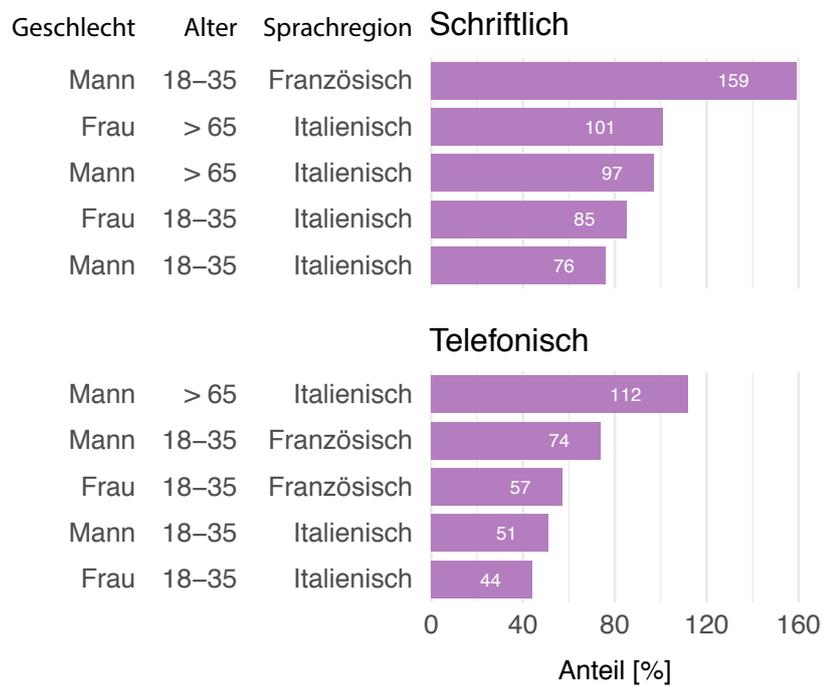


Abbildung 13: Vergleich der Netto- und Bruttostichprobe: Kombination aus Altersgruppen, Geschlecht und Sprachregion. Welche Subgruppen sind am wenigsten Vertreten in den Stichproben (in absoluten Zahlen)?

Wie eingangs dieses Abschnitts erwähnt, sind nicht repräsentativ abgebildete Subgruppen insbesondere dann ein Problem, wenn sie zu sehr kleinen Fallzahlen führen, die keine verlässlichen Aussagen zulassen. Abbildung 13 zeigt deshalb den Umfang der kleinsten Subgruppen in absoluten Zahlen bei der schriftlichen und der telefonischen Befragung. Sehr kleine Subgruppen zeigen sich typischerweise im italienischen Sprachraum. Um überhaupt Aussagen über den drittgrössten Sprachraum machen zu können, wurde dessen Anteil bereits in der Bruttostichprobe angehoben. Dennoch wurden in der telefonischen Befragung nur 44 Frauen und 51 Männer bis 35 Jahre erfasst. Auch in der schriftlichen Befragung haben diese beiden demographischen Gruppen die kleinsten Fallzahlen. Ihre Fallzahl liegt jedoch bei 85 Frauen und 76 Männer. Der Grund für die grösseren Fallzahlen liegt weniger einer besseren Repräsentation als an der grösseren Rücklaufquote der schriftlichen Befragung. Würden bei der schriftlichen Befragung ebenfalls nur 1500 Personen interviewt, läge die Fallzahl bei umgerechnet 52 Männern, bzw. 58 Frauen bis 35 Jahre aus dem italienischen Sprachraum. Eine Reduktion des Umfangs der Bruttostichprobe führt auch bei einer schriftlichen Befragung zu Fallzahlen der kleinsten Subgruppen, die statistisch verlässliche Aussagen in Frage stellen.

3.3 Soziokulturelles Profil

Verzerrungen in der Repräsentativität aufgrund der Befragungsmethode lassen sich sehr präzise für alle Variablen ermitteln, deren Ausprägung für alle Personen des Stichprobenrahmens bekannt sind. Bei anderen wichtigen Merkmalen sind die Ausprägungen nur bei jenen bekannt, die tatsächlich an der Befragung teilgenommen haben. Dies gilt namentlich für das soziokulturelle Profil der Befragten. Hier ist nur eine indirekte Einschätzung der Repräsentativität möglich. Als Referenz wird die BFS-Strukturerhebung herangezogen. Dabei handelt es sich zwar ebenfalls um eine Stichprobe. Da die Teilnahme an der Strukturerhebung für die angeschriebenen Personen obligatorisch ist und jedes Jahr mindestens 200 000 Personen befragt werden, spielen hier Verzerrungen durch die Erhebungsmethode eine vergleichsweise marginale Rolle. Die Vergleichsgrösse für die folgenden Auswertungen bilden die in der Strukturerhebung von 2016 erfassten, in der Schweiz stimmberechtigten Personen. Die disproportional geschichtete Bruttostrichprobe mit einem grösseren Anteil Befragter aus der französisch- und italienischsprachigen Schweiz wurde hierfür mit einer Designgewichtung ausgeglichen.

3.3.1 Bildungsabschluss

Abbildung 15 zeigt sowohl bei den telefonischen als auch bei den schriftlichen Befragungen eine leichte, aber eindeutige Verzerrung des Bildungsprofils der Befragten im Vergleich zur Strukturerhebung. Kaum vertreten in den Nettostichproben sind Personen ohne abgeschlossene Schulbildung. Auch bei der nächst höheren Bildungskategorie zeigt sich eine deutliche Untervertretung. Die Untervertretung ist bei den schriftlichen Befragungen jeweils etwas ausgeprägter als bei der telefonischen Erhebung. Gut repräsentiert ist der Anteil der Personen mit einer Berufslehre als höchstem Bildungsabschluss. Klar überrepräsentiert sind einzig die Hochschulabgänger und -abgängerinnen. Hier ist die Überrepräsentation bei den telefonischen Befragungen stärker ausgeprägt. Besonders markant tritt dies beim Mai-Termin zu Tage. Die Abweichungen zu den Werten der Strukturerhebung sind bei beiden Methoden und beiden Terminen hochsignifikant. Allerdings zeigen sich relativ stabile Verteilungsmuster. Grössere Verzerrungen zeigen sich nur am obersten und am untersten Rand der Bildungsskala. Es sind höchst wahrscheinlich Verzerrungen, die bei Befragungen mit freiwilliger Teilnahme zu Themen ohne unmittelbaren Alltagsbezug nicht zu vermeiden sind.

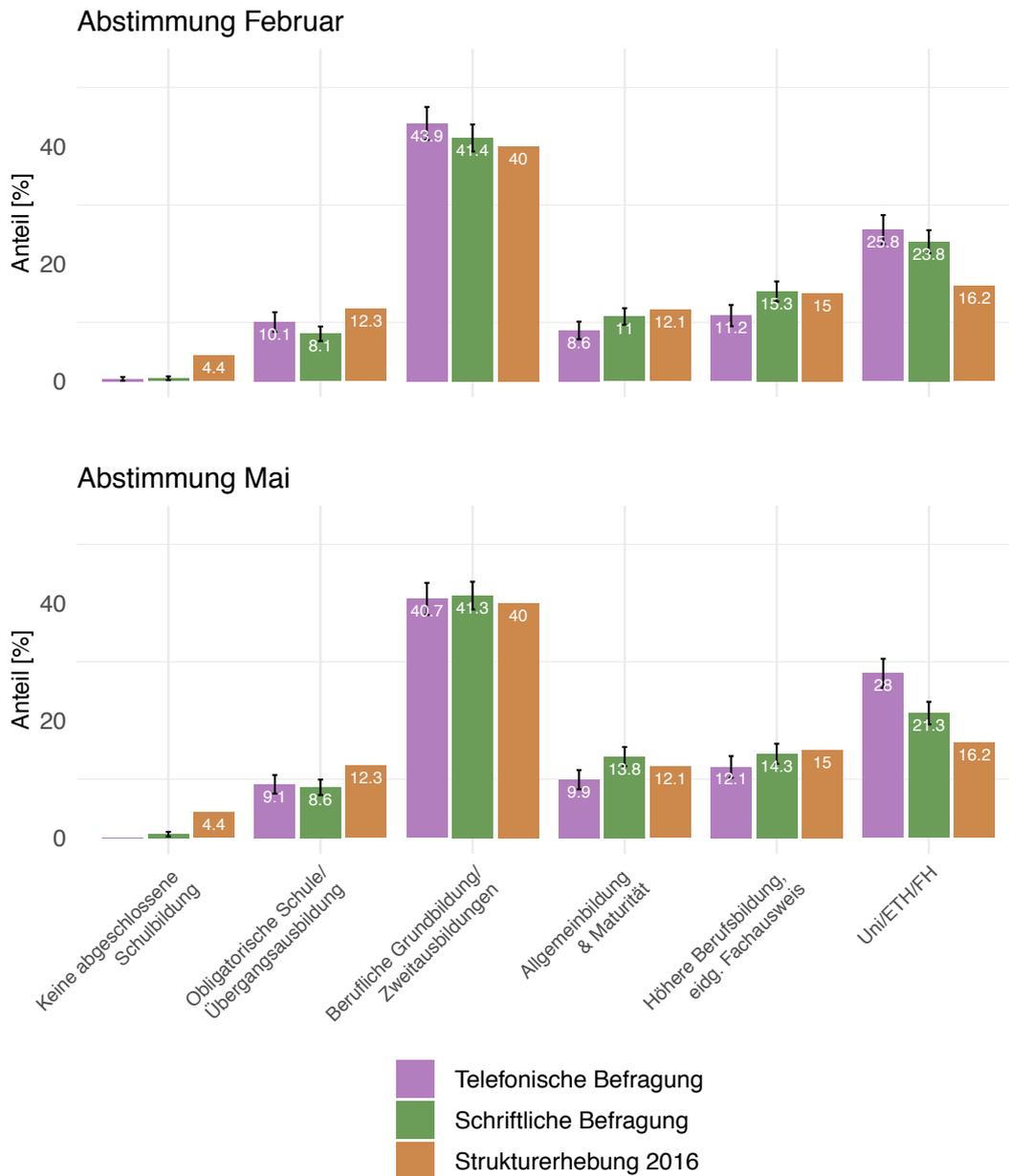


Abbildung 14: Bildungsabschlüsse: Anteile nach Erhebungsmethode im Vergleich zur Verteilung gemäss Strukturerhebung (2016); mit Designgewichtung

3.3.2 Konfession

Gut repräsentiert durch beide Erhebungsmethoden ist die Konfession. Auffällig ist, dass die Zahl der Konfessionslosen im Vergleich zur Strukturerhebung etwas übervertreten ist. Allerdings stammen die Daten der Strukturerhebung von 2016 und der Anteil der Konfessionslosen hat seither bekanntlich weiter zugenommen. Etwas unterrepräsentiert sind neben den muslimischen ebenfalls die katholischen Stimmberechtigten.

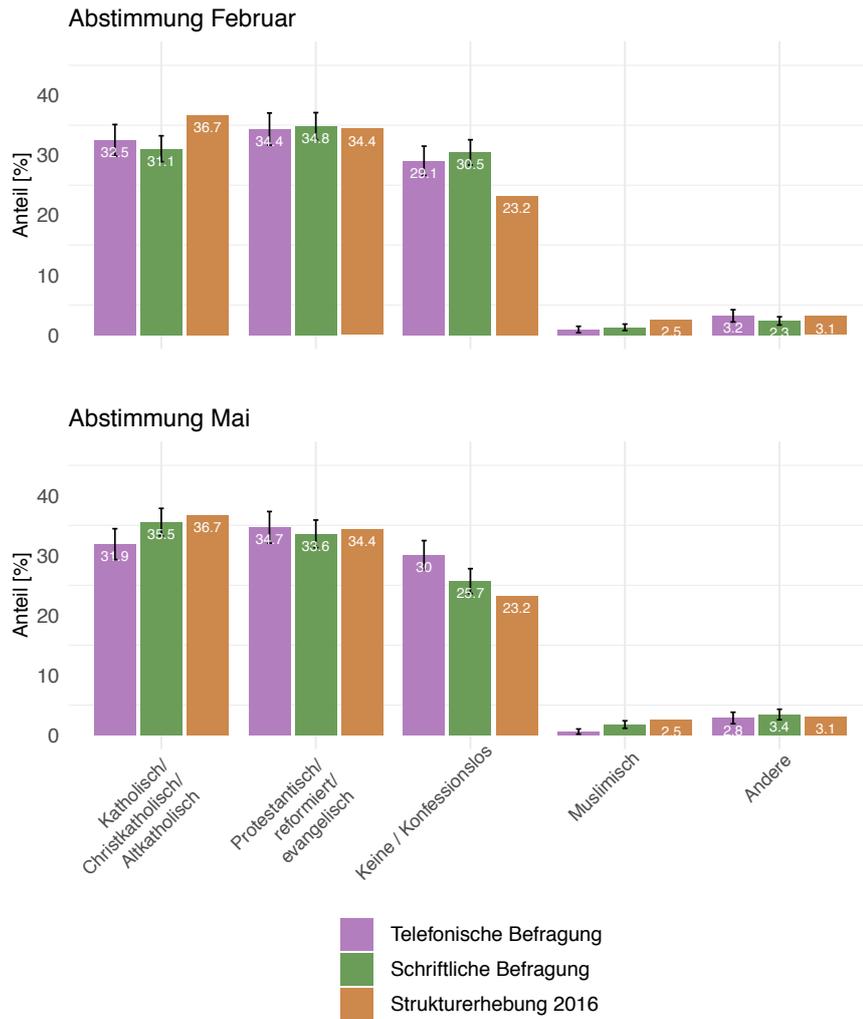


Abbildung 15: Konfession: Anteile nach Erhebungsmethode im Vergleich zur Verteilung gemäss Strukturerhebung (2016); mit Designgewichtung

3.3.3 Haushaltseinkommen

In der Strukturerhebung werden keine Angaben zum Haushaltseinkommen erfasst. Die Strukturerhebung kann deshalb bei dieser wichtigen sozialen Dimension nicht als Referenz genommen werden. Sichtbar wird jedoch, dass die Einkommensverteilung bei beiden Erhebungsmethoden sehr ähnlich ist. Dies gilt insbesondere bei der Befragung zum Februar-Termin. Beim Mai-Termin ist der Anteil der tiefen Einkommen bis 6000 Franken in der telefonischen Befragung allerdings etwas grösser.

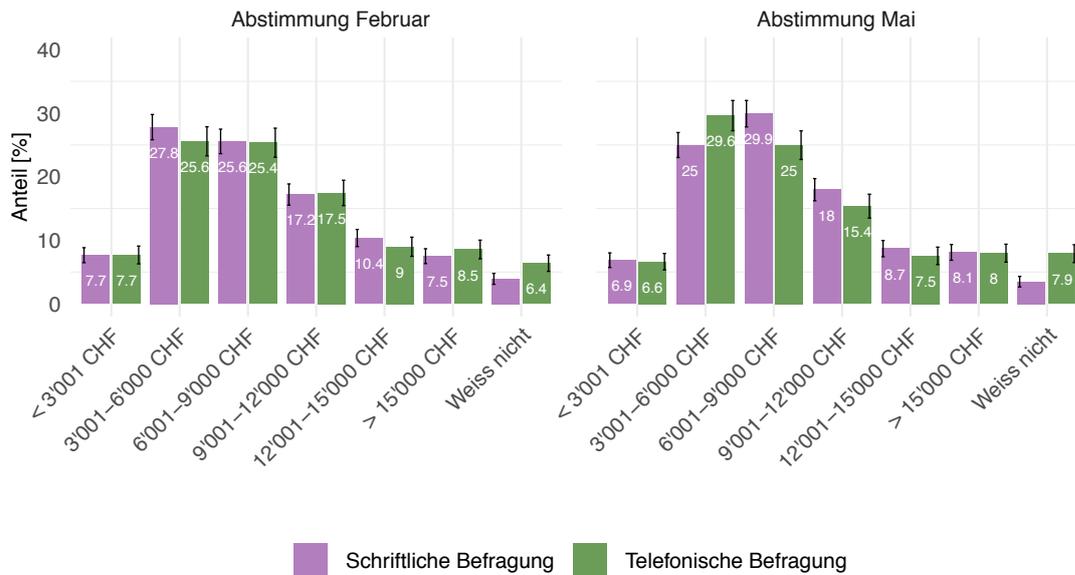


Abbildung 16: Haushaltseinkommen gruppiert

3.4 Politisches Verhalten

Im Kontext der VOTO-Befragung besonders bedeutsam ist die Repräsentativität der Stichprobe in Bezug auf politische Indikatoren. Zumindest für die Stimmbeteiligung und den Abstimmungsentscheid ist dabei zwar kein Vergleich mit der Bruttostichprobe möglich, dafür jedoch ein Vergleich mit der Grundgesamtheit, nämlich der stimmberechtigten Bevölkerung der Schweiz.

3.4.1 Abstimmungsteilnahme

Die Stimmbeteiligung wird, wie aus Abbildung 17 ersichtlich wird, von beiden Methoden nicht repräsentativ abgebildet. Sowohl die schriftliche als auch die telefonische Befragung überschätzen die Stimmbeteiligung bei beiden Abstimmungsterminen sehr deutlich. Beim Abstimmungstermin vom Februar 2019 lag die tatsächlich ausgewiesene Stimmbeteiligung bei 37,9 Prozent. Von den schriftlich Befragten gaben jedoch 58,7 Prozent an, an der Abstimmung teilgenommen zu haben. Von den telefonisch Befragten waren es 60,6 Prozent. Der Unterschied zwischen den Methoden ist nicht signifikant, und beide führen zu einer starken Überrepräsentation der aktiven Stimmbewölkerung.

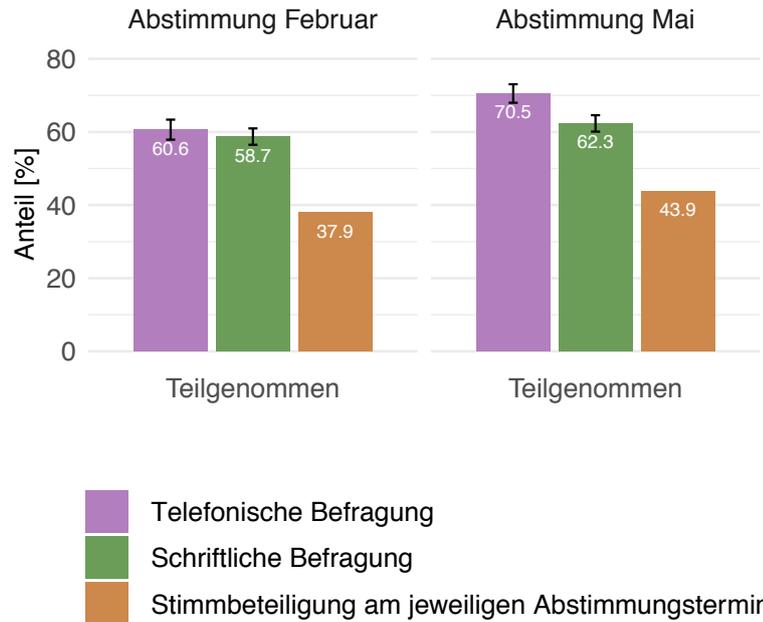


Abbildung 17: Abstimmungsteilnahme nach Erhebungsmethode im Vergleich zur tatsächlichen Stimmbeteiligung; mit Designgewichtung.

Beim Mai-Termin zeigt sich ein analoger Sachverhalt. Die tatsächliche Stimmbeteiligung lag bei 43,9 Prozent. Unter den schriftlichen Befragten waren es jedoch 62,3 Prozent, die angegeben haben, sie hätten an der Abstimmung teilgenommen. Unter den telefonisch Befragten lag die Stimmbeteiligung sogar bei 70,5 Prozent. Anders als beim Februartermin zeigt sich hier ein signifikanter Teilnahmeunterschied ($\tilde{\chi}^2 = 26.65$, $p = 0.000$) zwischen den Erhebungsmethoden. Diese Diskrepanz lässt sich allerdings mit der unterschiedlichen Zusammensetzung der beiden Stichproben erklären. Abbildung 18 zeigt, dass sich sowohl das Alter als auch die formelle Bildung deutlich auf die Wahrscheinlichkeit der Teilnahme an der Abstimmung auswirkt. Nach der Kontrolle dieser Indikatoren hat die Erhebungsmethode weder bei der dargestellten Abstimmung vom Mai, noch bei jener vom Februar einen signifikanten Einfluss auf die angegebene Teilnahme an der Abstimmung. Die höhere Stimmbeteiligung der telefonisch Befragten beim Mai-Termin ist deshalb eine Folge der verstärkten Überrepräsentation von Älteren und Personen mit Hochschulabschluss in der realisierten Stichprobe.

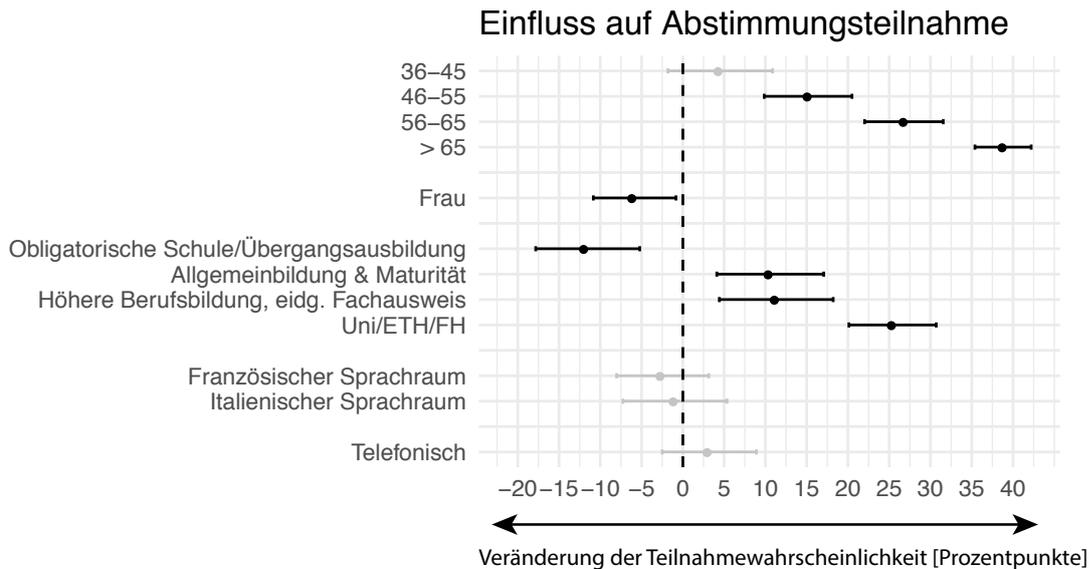


Abbildung 18: Einflussfaktoren für Wahrscheinlichkeit der Teilnahme an der Abstimmung vom 19. Mai 2019. Vergleich zur Referenzgruppe.

Die Analyse zeigt, dass es sich hier um einen Befragungs- und nicht um einen Methodeneffekt handelt. Es besteht eine starke, methodenübergreifende Überrepräsentation der aktiven Stimmberechtigten. Es ist dies ein bekanntes Resultat politischer Befragungen: Menschen, die eher politisch engagiert und interessiert sind, nehmen auch eher an politischen Befragungen teil (Clausen, 1968; Karp & Brockington, 2005, s. 827). Der Zusammenhang relativiert sich dabei insofern, als dass bei der VOTO-Befragung jene Personen im Fokus stehen, die tatsächlich an der Abstimmung teilgenommen haben. Potenziell problematischer ist, wenn Befragte aus Gründen der sozialen Erwünschtheit trotz Stimmbeteiligung angeben, teilgenommen zu haben. Aufschlussreich ist hier allerdings, dass kein Methodeneffekt festgestellt werden kann. Personen, die von einem persönlichen Gegenüber am Telefon befragt werden, geben nicht häufiger an, teilgenommen zu haben, als Personen, welche die Befragung schriftlich und ohne Gegenüber durchgeführt haben. Dies ist zudem ein Indiz dafür, dass es hier keinen massgeblichen Effekt der sozialen Erwünschtheit gibt.

3.4.2 Abstimmungsentscheide

Der Stimmentscheid wird, im Vergleich zur Stimmbeteiligung, von beiden Befragungsmethoden einiges genauer repräsentiert (Abb. 19). Dies gilt insbesondere für die Abstimmung zur Zersiedelungsinitiative, über die am 1. Februar 2019 abgestimmt wurde. Sowohl das Resultat der telefonischen Befragung als auch dasjenige der schriftlichen Befragung liegen nahe am tatsächlichen Abstimmungsergebnis. Sowohl der Ja- als auch der Nein-Anteil liegen dabei leicht zu tief. Dies liegt offensichtlich vor allem daran, dass bei den Befragungen zusätzlich die Kategorie «weiss nicht» ausgewählt werden kann.

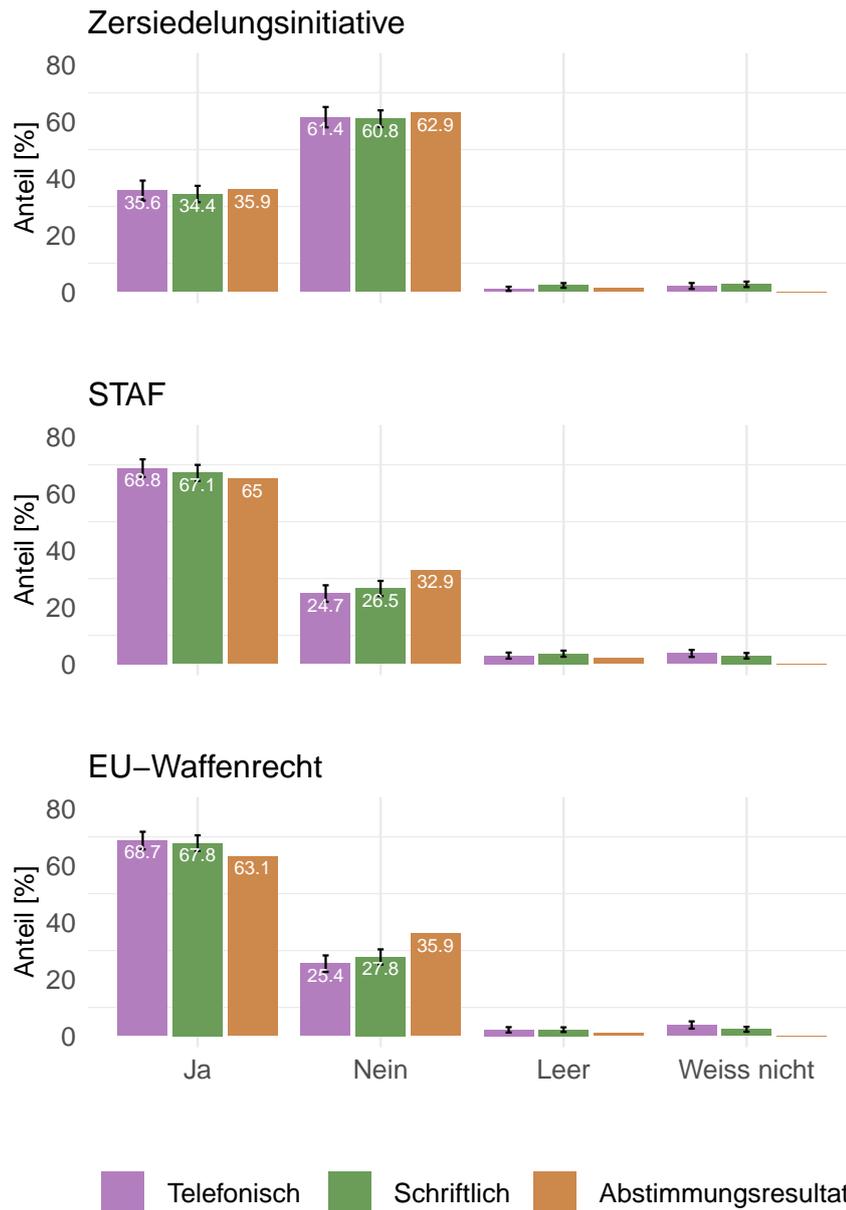


Abbildung 19: Vergleich telefonisch vs. schriftlich: Zustimmung zu den Abstimmungsvorlagen aus beiden Durchgängen des Methodenexperiments, mit Designgewichtung.

Etwas weniger gut werden die Resultate beim zweiten untersuchten Abstimmungstermin vom 19. Mai 2019 durch die beiden Nachbefragungen repräsentiert. Bei der Abstimmung über das EU-Waffenrecht führen beide Erhebungsmethoden zu einer signifikanten Überrepräsentation der Ja-Stimmenden im Vergleich zu den Nein-Stimmenden. Ein ähnliches Bild zeigt sich auch bei der Abstimmung zur Steuerreform STAF. Auch hier unterschätzen beide Befragungsmethoden die Ablehnung im Vergleich zur Zustimmung. Die Abweichung vom tatsächlichen Resultat ist allerdings etwas weniger deutlich als bei der Waffenrechtsvorlage.

Zumindest bei den drei untersuchten Vorlagen zeigt sich jeweils eine grössere Differenz zwischen dem tatsächlichen Resultat und den Befragungen als zwischen den beiden Erhebungsmethoden. Während die telefonische Befragung bei der Abstimmung zur Zersiedlungsinitiative vom Februar etwas näher am tatsächlichen Resultat liegt, repräsentiert die schriftliche Befragung die Abstimmungen vom Mai etwas genauer. Diese Unterschiede in Bezug auf die Ja- und Nein-Anteile sind jedoch nicht signifikant. Die Resultate beider Termine deuten darauf hin, dass die Repräsentationsverzerrungen in Bezug auf Abstimmungsentscheide unabhängig von den hier untersuchten Erhebungsmethoden bestehen. Beide Methoden zeigen die grösste Repräsentationsverzerrung bei der Abstimmung zum Waffenrecht. Es handelt sich dabei um eine Vorlage, die politisch im Spannungsfeld zwischen konservativ und progressiv-liberal verläuft. Personen mit einer konservativen Orientierung sind sowohl bei der telefonischen als auch bei der schriftlichen Befragung untervertreten. Diese Verzerrung hat allenfalls zu einem kleineren – statistisch nicht signifikanten – Teil mit der gewählten Erhebungsmethode zu tun und kann mit einem allfälligen Wechsel von einer telefonischen zu einer schriftlichen Befragung nicht grundsätzlich ausgeglichen werden.

Alles in allem zeigt die Analyse der Repräsentativität der beiden Erhebungsmethoden vergleichsweise geringe Verzerrung in Bezug auf die soziodemographischen Variablen, mit Ausnahme des Alters bei der telefonischen Befragung. Eine markante Verzerrung besteht jedoch in Bezug auf die Stimmbeteiligung und in geringerem Ausmass sowie themenabhängig auch beim Stimmentscheid. Ersteres ist primär eine Folge davon, dass politische Interessierte vermehrt an Befragungen zu politischen teilnehmen (Groves et al., 2004). Letzteres ist ein Hinweis dafür, dass Personen mit nationalkonservativer Orientierung seltener an sozialwissenschaftlichen Befragungen teilnehmen als an Volksabstimmungen. Beides zeigt, dass eine soziodemographische Repräsentativität alleine noch nicht zu politischer Repräsentativität der realisierten Stichprobe führt.

3.5 Struktur der Onlineteilnahme

Das Ziel des vorliegenden Methodenexperiments ist der Vergleich der Repräsentativität der schriftlichen und der telefonischen Erhebungsmethode. Das Forschungsdesign ermöglicht keinen Vergleich der Repräsentativität einer reinen Online-Befragung im Vergleich zu einer telefonischen oder einer Befragung nur mit Papierfragebogen mit anderen Erhebungsmethoden. Alle Teilnehmenden der schriftlichen Befragung hatten die Möglichkeit entweder online oder per Papierfragebogen teilzunehmen, so dass der Effekt einer Beschränkung auf bloss einen der beiden Kanäle nicht untersucht werden konnte. Aussagen zur Struktur der Befragten nach Erhebungskanal sind deshalb keine Aussagen über die Repräsentativität einer Befragung, die auf einen der beiden Kanäle verzichtet. Sie zeigen aber die Präferenzen der verschiedenen untersuchten Bevölkerungsgruppen.

Der Anteil der Online-teilnahme betrug in der Befragung vom Februar 39,2 Prozent und bei jener vom Mai 43,6 Prozent. Der tiefere Anteil bei der Februar-Befragung ist vor allem eine Folge davon, dass hier nur bei der Hälfte der Befragten bei Bedarf ein zweites Erinnerungsschreiben verschickt wurde. Entsprechend ist der Anteil der Online-Teilnahme bei der zweiten Erinnerung ohne erneute Beilage des Papierfragebogens besonders hoch. Der Vergleich konzentriert sich deshalb vor allem auf die Befragung zum Abstimmungstermin vom Mai.

3.5.1 Soziodemographie

Wie zu erwarten, hängt der gewählte Kanal zur Teilnahme an der Befragung von soziodemographischen Faktoren ab (Abb. 20). Jüngere Männer mit einem höheren Bildungsabschluss nehmen am ehesten online an der Befragung teil. Bei der Bildung zeigen sich dabei nur geringe Unterschiede zwischen Hochschulabschluss und anderen tertiären Ausbildungen, jedoch grosse Unterschiede zwischen einem schul- und einem berufsbezogenen Abschluss der Sekundarstufe II (Mittelschule vs. Berufslehre).

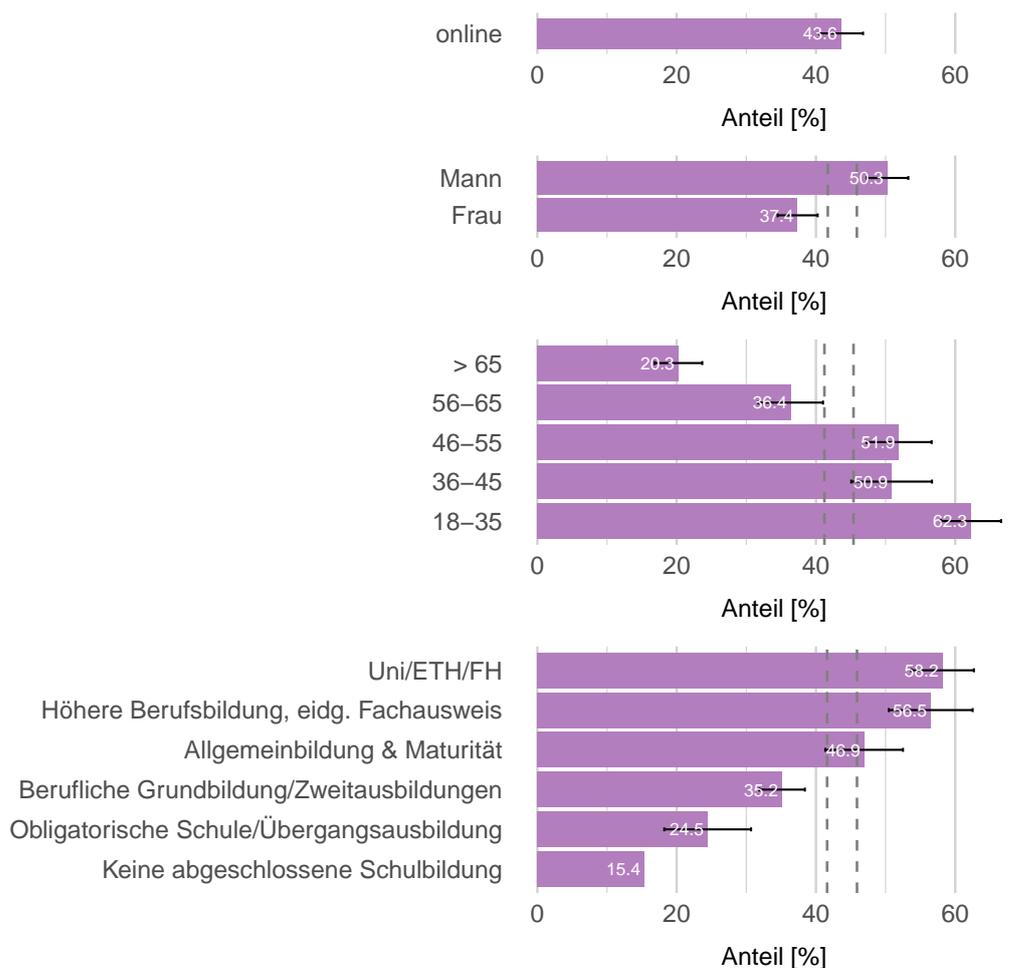


Abbildung 20: Onlinequote soziodemografischer Subgruppen (Mai-Termin).

3.5.2 Politisches Verhalten

Abbildung 21 zeigt, dass im ersten Teil des Methodenexperiments die Stimmbeteiligung der Online-Teilnehmenden signifikant unter der Stimmbeteiligung jener lag, die per Brief mitgemacht haben ($\tilde{\chi}^2 = 59.74$, $p = 0.000$). Im zweiten Teil der Befragung zeigt sich diese Differenz allerdings nicht. Hier ist die Stimmbeteiligung Online und Print sehr ähnlich (fehl-)repräsentiert ($\tilde{\chi}^2 = 1.01$, $p = 0.30$). Diese Variation, für die es keine offensichtliche Erklärung gibt, zeigt, dass es keine belastbaren Zusammenhänge zwischen der Beteiligungswahrscheinlichkeit an der Abstimmung und dem bevorzugten Teilnahme Kanal gibt.

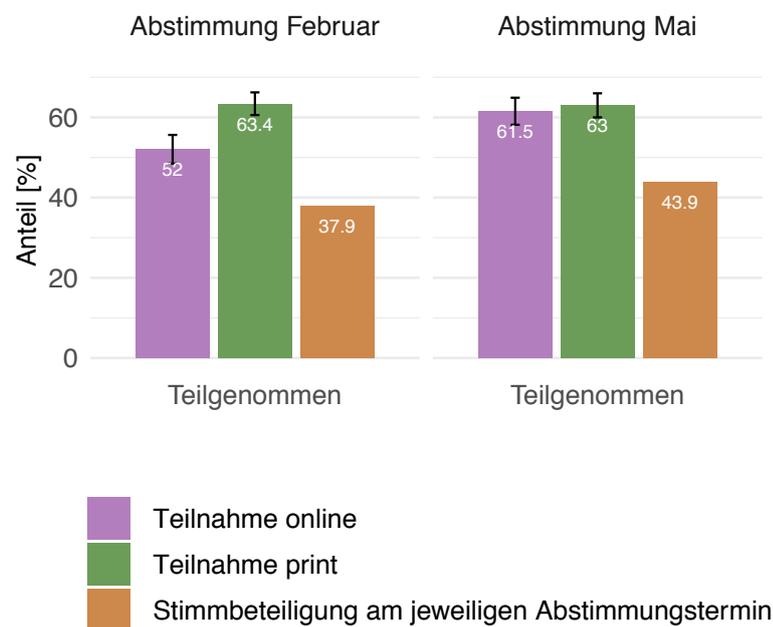


Abbildung 21: Vergleich online vs. print: Teilnahme an jeweiliger Abstimmung, mit Designgewichtung

Bei den Abstimmungsentscheiden zeigt sich eine relativ grosse Übereinstimmung zwischen den Teilnehmenden der beiden schriftlichen Teilnahme Kanälen. Wie Abbildung 22 zeigt, gilt dies insbesondere für den Abstimmungsentscheid zur Zersiedelungsinitiative. Auffällig ist, dass mehr Print-Teilnehmende angeben, sich nicht an ihren Entscheid erinnern zu können.

Auch bei der STAF ist in der Printversion der Umfrage der Anteil derer, die angaben, nicht mehr zu wissen, was sie gestimmt haben, etwas höher als in der Onlineversion. Daneben gibt es aber auch hier keine signifikanten Unterschiede in der Zustimmung zur Vorlage zwischen Online- und Print-Teilnehmenden.

Signifikante Unterschiede zeigen sich dagegen bei der Abstimmung über das EU-Waffenrecht. Die Unterrepräsentation der Nein-Stimmenden ist bei den Online-

Teilnehmenden noch einmal deutlicher ausgeprägt als bei den Print-Teilnehmenden. Die Erklärung dafür liegt hauptsächlich in der starken Bildungs- und Altersverzerrung bei der Online-Teilnahme, die dazu führt, dass konservativ Orientierte eher den gedruckten Fragebogen ausfüllen.

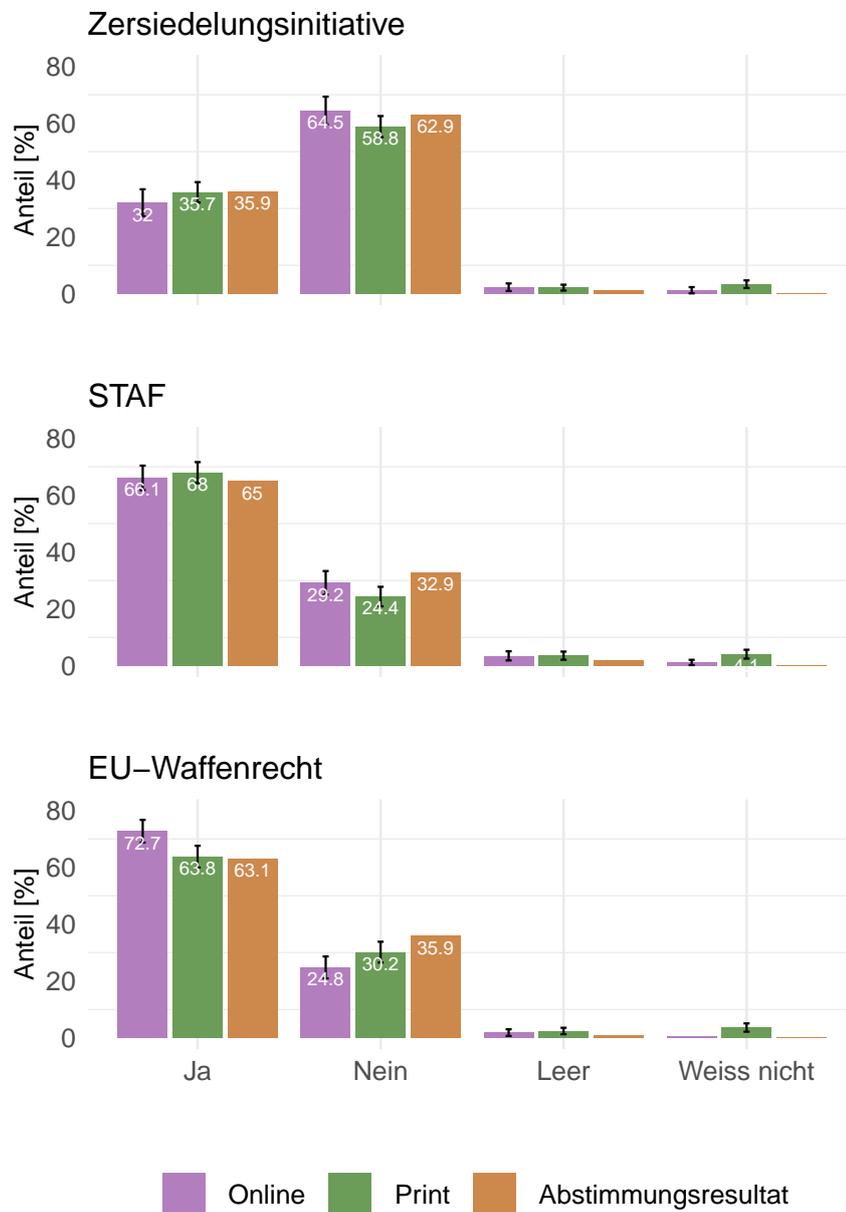


Abbildung 22: Vergleich online vs. print: Abstimmungsentscheide zu den jeweiligen Vorlagen, mit Designgewichtung

3.6 Einfluss der Variationen auf Repräsentativität

Im ersten Teil dieser Studie wurde der Einfluss der experimentell eingeführten Variationen auf den Rücklauf untersucht. Es hat sich gezeigt, dass insbesondere die Zahl der Erinnerungsschreiben sowie die Länge des Fragebogens Einfluss

auf die Rücklaufquote der Befragung haben. Bei diesen beiden Variationen ist somit auch der Einfluss auf die Repräsentativität von Belang. Problematisch wäre dabei vor allem, wenn sich ein grösserer Rücklauf nur auf Kosten einer geringeren Repräsentativität einstellen würde.

Die Variation von Elementen einer Umfrage in derselben Welle lässt den Vergleich zu, ob die unterschiedlichen Varianten zu Differenzen in den Resultaten führen. Solche Differenzen wären bei einer Abstimmungsbefragung wie VOTO natürlich insbesondere bezüglich der politischen Variablen suboptimal, weil dann je nach Version der Umfrage eine unterschiedliche Grundgesamtheit erhoben würde. Um darüber Erkenntnisse zu gewinnen, wurde deshalb im zweiten Durchlauf des Methodenexperiments im Mai 2019 die Länge des Fragebogens sowie der Zeitraum des Versands der Unterlagen variiert.

3.6.1 Zahl der Erinnerungsschreiben

Werden die Befragten mit maximal zwei, statt nur mit einem Erinnerungsschreiben zur Teilnahme aufgefordert, erhöht sich die Wahrscheinlichkeit einer Teilnahme an der Befragung um 7,5 Prozentpunkte. Ein zweites Erinnerungsschreiben führt ausserdem zu einer besseren Repräsentation der Altersklassen. Dies zeigt Abbildung 23. Bei zwei Erinnerungsschreiben wirkt sich das Alter der Personen weniger stark auf die Teilnahmewahrscheinlichkeit aus als bei nur einem. Die beiden Verteilungen unterscheiden sich nicht signifikant. Der Effekt zeigt sich allerdings robust über alle Altersklassen. Ausserdem findet sich praktisch dieselbe Altersverteilung bei der Befragung vom Mai, die durchgehend mit zwei Erinnerungsschreiben durchgeführt wurde. Die Überrepräsentation der Befragten aus grossen Agglomerationen bei zwei Erinnerungsschreiben zeigt sich demgegenüber nur bei der Abstimmung vom Februar, nicht jedoch bei jener vom Mai.

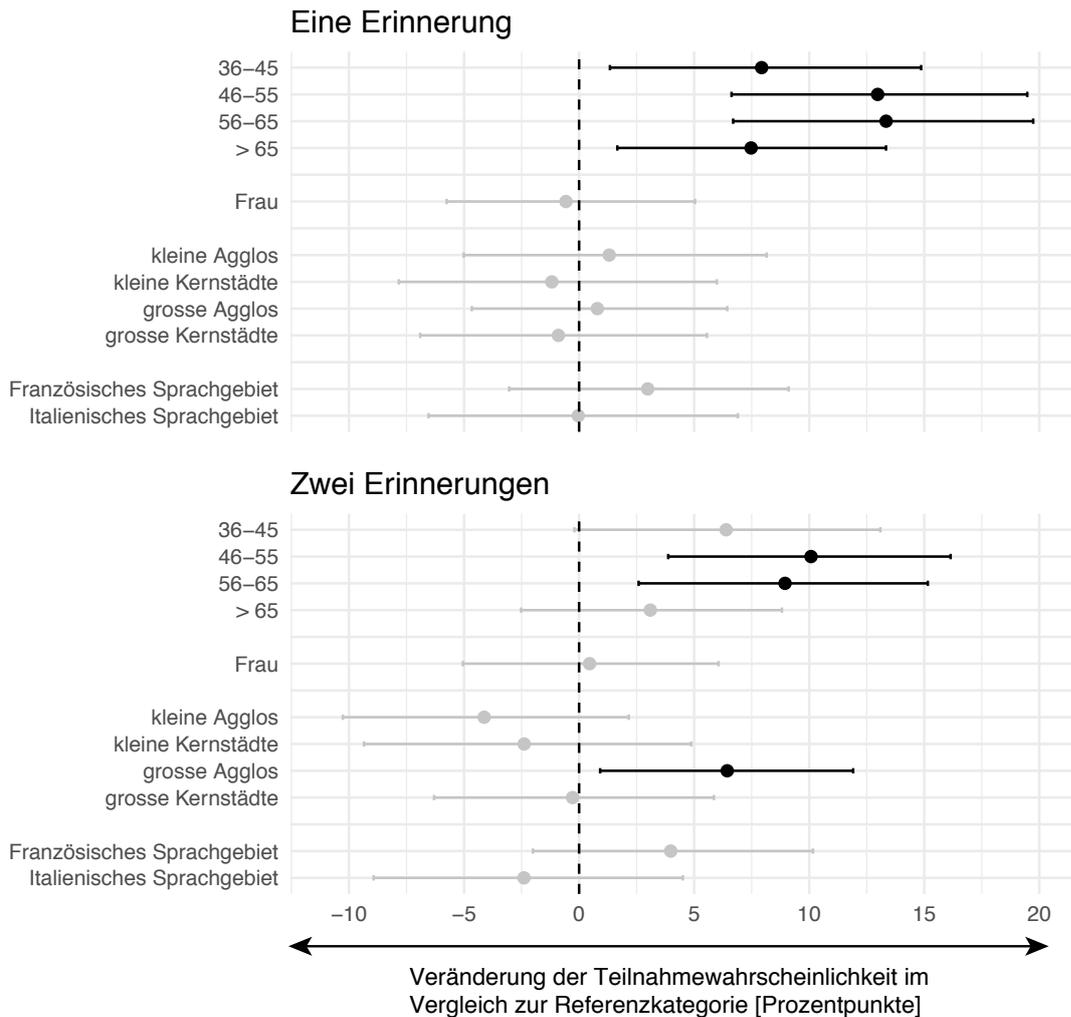


Abbildung 23: Faktoren für eine Teilnahme an der schriftlichen Befragung bei einem und bei zwei Erinnerungsschreiben.

3.6.2 Fragebogenlänge

Auch die Fragebogenlänge hat einen signifikanten, wenn auch weniger starken Einfluss auf die Teilnahmewahrscheinlichkeit. Dabei zeigt sich, dass sich das soziodemographische Profil der Teilnehmenden nicht zwischen dem kurzen und dem langen Fragebogen unterscheiden. Von besonderem Interesse ist dabei der Zusammenhang mit dem Bildungsabschluss. Möglich wäre, dass sich weniger Gebildete von einem längeren Fragebogen eher abschrecken lassen. Abbildung 24 zeigt jedoch, dass entgegen dieser Erwartung eher Personen mit einer tiefen Bildung beim langen Fragebogen übervertreten sind, wenn auch nicht signifikant. Dies zeigt, dass die Fragebogenlänge, zumindest im untersuchten Bereich, zu keiner zusätzlichen Unterrepräsentation tieferer Bildungsschichten führt.

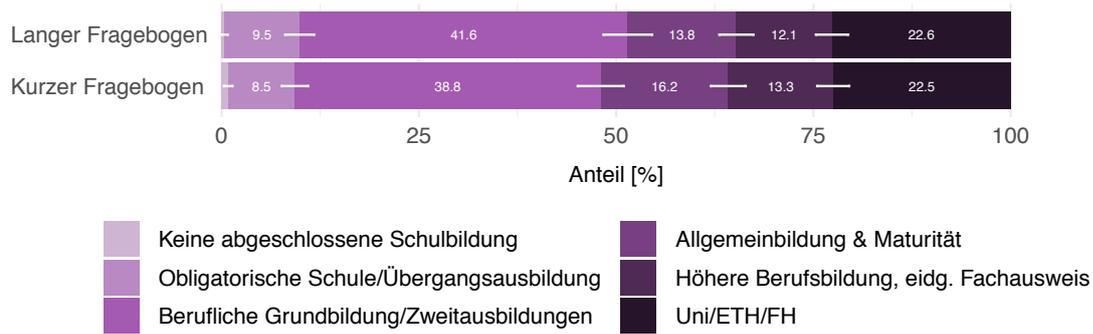


Abbildung 24: Bildungsstruktur der Befragten abhängig von der Fragebogenlänge (ohne Designgewichtung).

3.6.3 Versandprozedur

Die Versandprozedur hat keinen signifikanten Einfluss auf den Rücklauf der Befragung. Sie wirkt sich auch nicht auf die soziodemographische Repräsentativität aus. Der Versand des ersten Einladungsschreibens bereits vor dem Abstimmungstermin könnte sich allerdings auf die politische Beteiligung auswirken. Die Einladung kann als Erinnerung und Aufforderung zum Stimmen verstanden werden. Abbildung 25 zeigt, dass die Stimmbeteiligung der Gruppe mit Einladungsschreiben vor dem Abstimmungstermin tatsächlich leicht höher ist als bei der andern. Dieser Unterschied ist allerdings nicht signifikant.

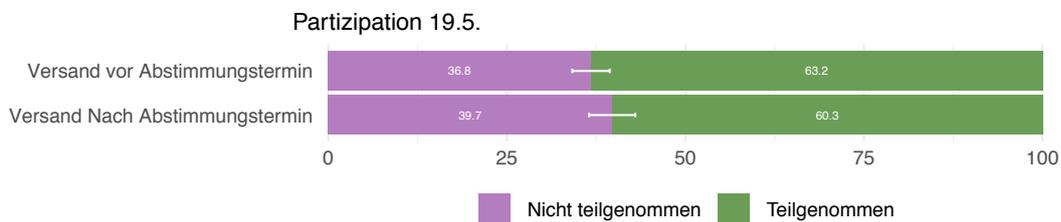


Abbildung 25: Stimmbeteiligung der Befragten am 19.5. abhängig von der Versandprozedur.

4 Antwortqualität

Thema im letzten Teil der Studie ist die Qualität der Antworten der beiden untersuchten Erhebungsmethoden im Vergleich. Anders als die Repräsentativität oder der Rücklauf lässt sich die Qualität der Antworten nicht ohne Weiteres messen und vergleichen. Die folgenden Auswertungen sollen jedoch Hinweise geben über systematische, quantifizierbare Unterschiede in der Art der Beantwortung der Fragen im schriftlichen und im telefonischen Befragungskontext. Die Auswertungen zu fehlenden Angaben und zur Antwortkategorie «weiss nicht» beziehen sich dabei auf die Befragung zur Abstimmung vom Mai 2019.

4.1 Vollständigkeit der Angaben

Einfach messen und vergleichen lässt sich das Qualitätsmerkmal der Item-Nonresponse. Dabei geht es um die Nichtbeantwortung einzelner Fragen bzw. die Auswahl der Antwortkategorie «keine Angabe». Die Auswertung in Abbildung 26 zeigt einen deutlichen Unterschied zwischen der telefonischen und der schriftlichen Erhebung. Während bei der telefonischen Befragung insgesamt nur bei 0,6 Prozent aller Fragen keine Antwort gegeben wurde, liegt der entsprechende Anteil bei der schriftlichen Befragung bei 4,4 Prozent.

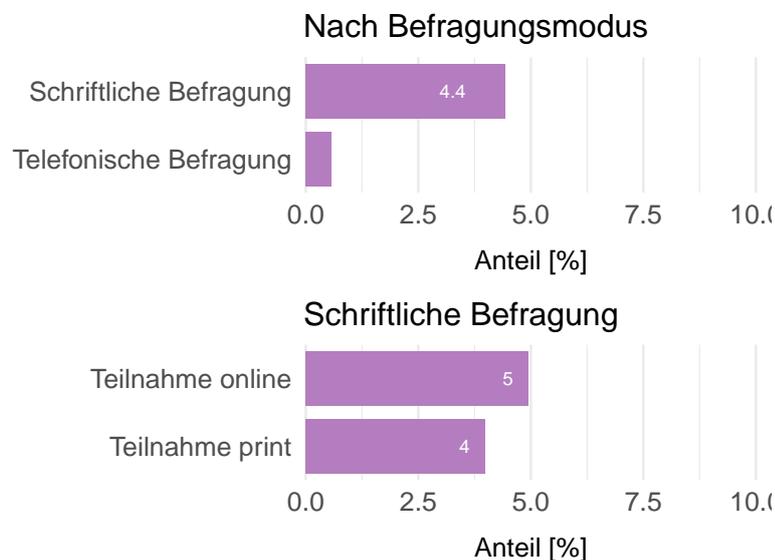


Abbildung 26: Durchschnittlicher Anteil fehlender Werte Total, Vergleich zwischen den Befragungsmodi und den Teilnahmemodi schriftlich (Abstimmung Mai)

Dabei kommen Item-Nonresponses bei der Online-Teilnahme noch etwas häufiger vor als bei der Teilnahme mit dem gedruckten Fragebogen. Bei einem Papierfragebogen ist es immer möglich, Items zu überspringen. Bei einer Online-Teilnahme kann demgegenüber mit Pflichtfeldern gearbeitet werden, die nicht übersprun-

gen werden können. Auf solche Pflichtfelder wurde allerdings in der Befragung verzichtet.

Abbildung 27 zeigt den Anteil der fehlenden Angaben bzw. der Item-Nonresponses für jeden Themenblock der Umfrage. Die dargestellte Reihenfolge der Themenblöcke entspricht ihrer Position im Fragebogen. Dabei wird sichtbar, dass bei der schriftlichen Befragung der Anteil «keine Antwort» bereits nach dem Einstieg deutlich ansteigt, dann jedoch bis zum zweitletzten Frageblock konstant bleibt. Dies deutet darauf hin, dass es keine progressive Zunahme von fehlenden Antworten im Verlauf des Fragebogens gibt. Der höhere Anteil fehlender Antworten im letzten Block ist primär eine Folge der dort gestellten (persönlichen) Fragen.

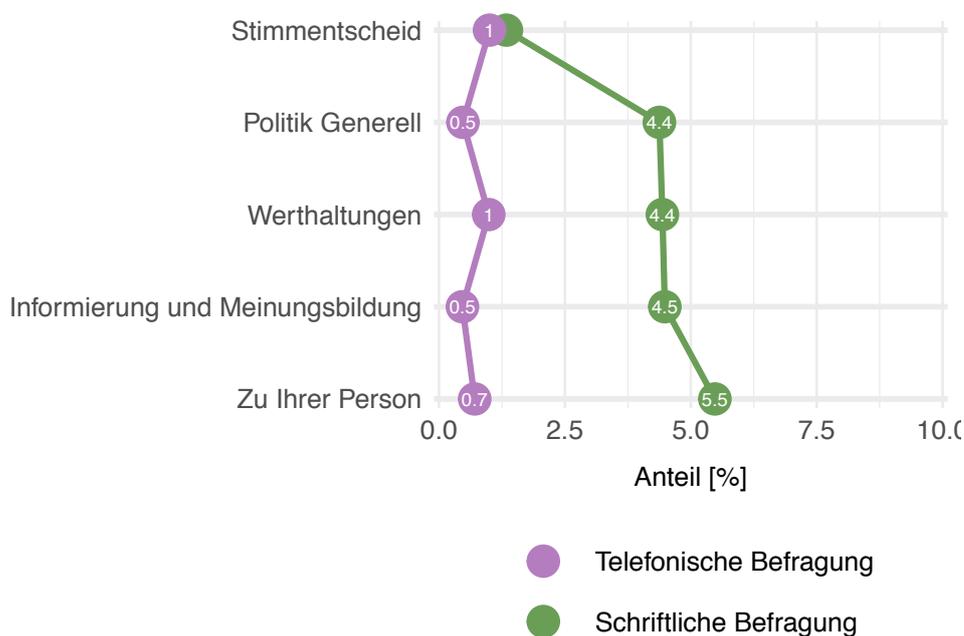


Abbildung 27: Durchschnittlicher Anteil fehlender Werte pro Fragebogenabschnitt (Abstimmung Mai)

Der Grund für die relative hohe Zahl an Non-Responses liegt an den persönlichen Daten, die im letzte Frageblock abgefragt werden. Abbildung 28 zeigt die Items mit den höchsten Anteilen bei der telefonischen und der schriftlichen Befragung. Am häufigsten verweigert wird die Beantwortung der Frage nach dem Haushaltseinkommen. Zu dieser Frage machen bei der telefonischen Befragung 6,6 Prozent keine Angaben – bei der schriftlichen sind es mit 13,9 Prozent mehr als doppelt so viele. Bei der telefonischen Befragung ist das Einkommen das einzige Item mit einem substantiellen Anteil an Non-Responses. Demgegenüber gibt es bei der schriftlichen Befragung auch grössere Anteile bei der Konfession oder der Links-rechts-Selbstpositionierung.

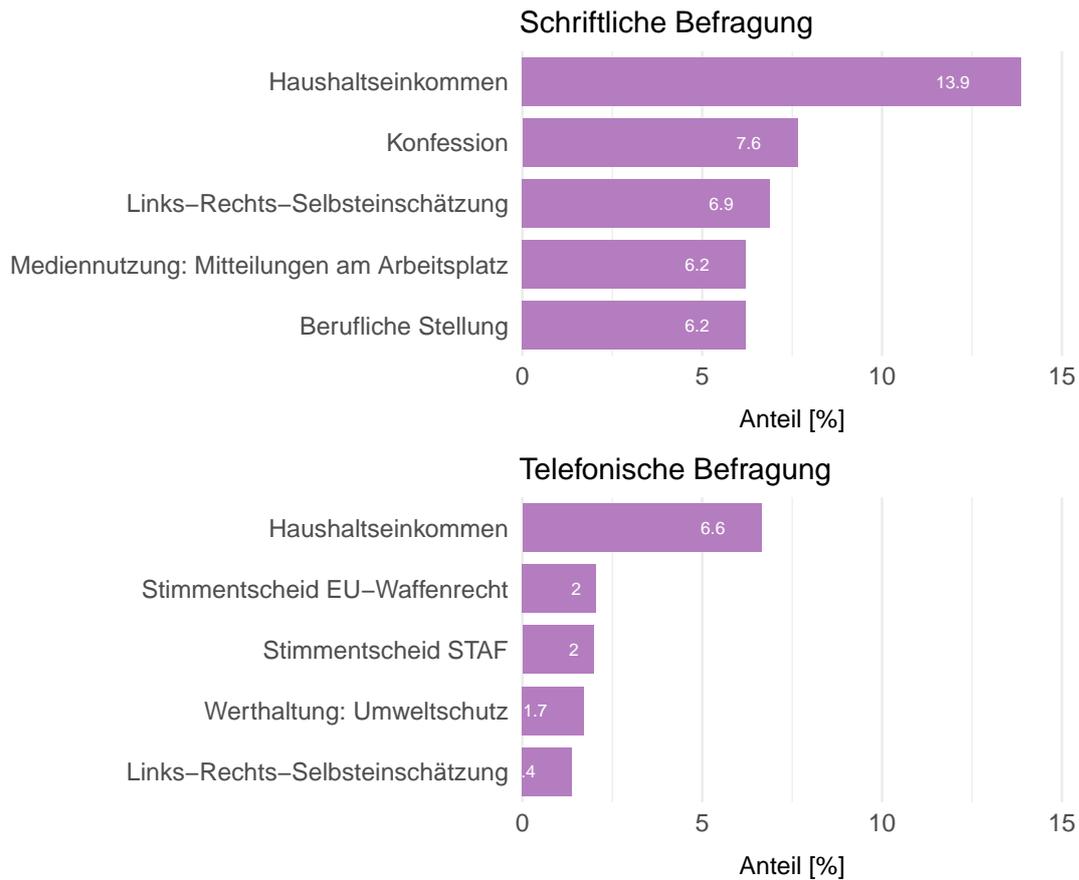


Abbildung 28: Variablen mit dem höchsten Anteil fehlender Werte (Abstimmung Mai)

Die persönliche Befragungssituation am Telefon macht es für die Befragten schwieriger, nicht auf Fragen zu antworten. Eine tiefere Rate von Item-Nonresponses bildet einen spezifischen Vorteil der telefonischen Erhebungsmethode. Diesem Vorteil steht allerdings der Nachteil der Unit-Nonresponses gegenüber. Damit ist die Nicht-Beantwortung des ganzen Fragebogens gemeint. So gab es bei der telefonischen Befragung vom Mai 73,3 Prozent Unit-Nonresponses, bei der schriftlichen dagegen nur 56,1 Prozent. Werden die Item-Nonresponses dazu gezählt, dann gab es bei der telefonischen Befragung insgesamt 73,5 Prozent Nicht-Antworten und bei der schriftlichen 58 Prozent. Dies zeigt, dass die Item-Nonresponse im Vergleich zu den Unit-Nonresponse deutlich weniger stark ins Gewicht fällt.

4.2 Qualität offener Fragen

Ein Kernthema der VOTO-Befragungen ist es, den Motiven eines Abstimmungsentscheids auf den Grund zu gehen. Dafür gibt es in der Befragung zum einen die geschlossen abgefragten Argumente. Zum anderen existieren offene Fragen, in denen die Befragten den Hauptgrund und weitere Gründe für den Wahlentscheid angeben können. Während die Befragten bei einer telefonischen Befragung bei

offenen Fragen durch die Interviewenden unterstützt werden können, ist dies bei einer schriftlichen Befragung nicht möglich. Wie wirkt sich dies auf die Qualität aus? Ein einfach zu operationalisierendes Merkmal ist hier wiederum die Zahl der Nicht-Antworten.

4.2.1 Anteil und Anzahl der Antworten

Abbildung 29 lässt erkennen, dass beim Hauptgrund in der telefonischen Befragung ein deutlich höherer Anteil der Befragten mit Stimmentscheid eine Angabe macht als bei der schriftlichen. Dies gilt sowohl für den Hauptgrund zur Annahme als auch zur Ablehnung der Zersiedelungsinitiative. Bei den weiteren Gründen gibt es keine signifikanten Unterschiede zwischen den beiden Befragungsmodi. Die Differenz im Anteil der Item-Nonresponses ist beim ersten Grund etwa doppelt so gross wie bei den geschlossenen Fragen. Aufgrund des grösseren Rücklaufs bei der schriftlichen Befragung ist dennoch die Zahl der Antworten grösser als bei der telefonischen Befragung.

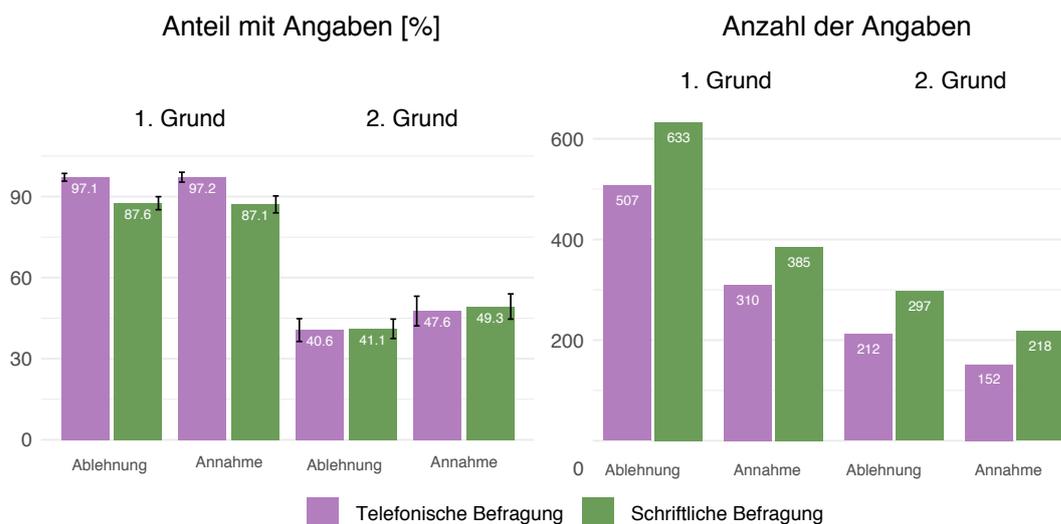


Abbildung 29: Anteil und Anzahl der Stimmenden mit Antwort(en) zur offenen Motivfrage zur Zersiedelungsinitiative (10.2.).

Am Abstimmungstermin vom 19. Mai wurde über zwei Vorlagen abgestimmt. Jede Person, die bei beiden Vorlagen gestimmt hat, wurde deshalb aufgefordert, bis zu vier Gründe für den Stimmentscheid anzugeben. Während bei der telefonischen Befragung annähernd gleich viele erste Gründe angegeben wurden wie bei der Abstimmung vom Februar, lag die Quote der Antworten bei der schriftlichen Befragung zum Mai-Termin nochmals tiefer (Abb. 30). Dies gilt allerdings nur für den ersten Grund. Wie beim Februar-Termin ist der Anteil der zweiten Gründe bei beiden Erhebungsmethoden annähernd gleich gross.

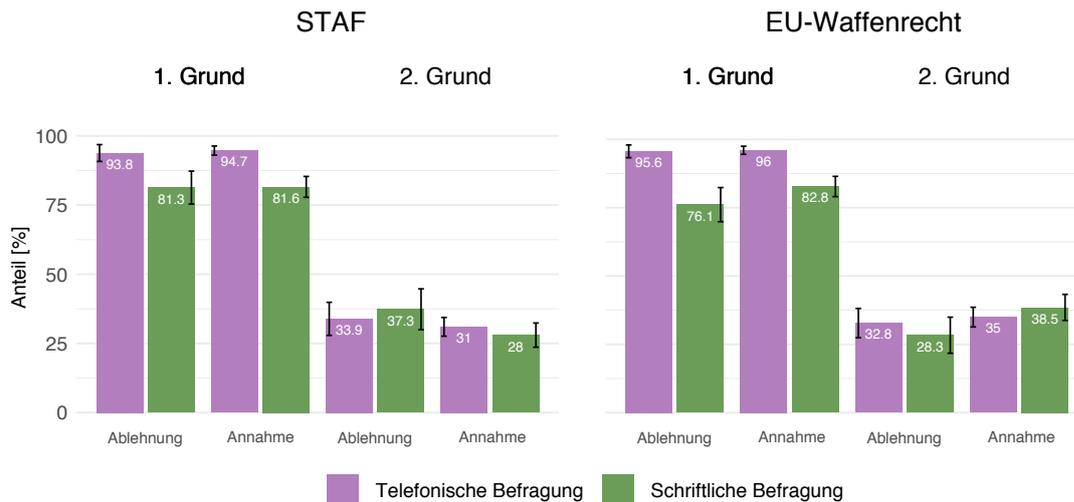


Abbildung 30: Anteil der Stimmenden mit Antwort(en) zur offenen Motivfrage zur STAF und zum EU-Waffenrecht (19.5.)

Auffällig ist, dass bei der telefonischen Befragung die Antwortquote (1. Grund) bei der EU-Waffenrechtsvorlage etwas grösser ist als bei der STAF. Bei der schriftlichen Befragung hat dagegen fast ein Viertel der Nein-Stimmenden keinen Grund für die Ablehnung der EU-Waffenrechtsvorlage angegeben. Insgesamt wurden bei dieser Vorlage weniger erste Gründe genannt, obwohl diese etwas weniger komplex als die STAF erscheint. Ein möglicher Faktor ist die Position im Fragebogen: Die Fragen zur Waffenrechtsvorlage wurden nach jenen zur STAF gestellt. Die insgesamt tiefere Antwortquote beim Mai-Termin mit zwei Vorlagen und die tiefere Antwortquote bei der zweiten abgefragten Vorlage sind Indizien dafür, dass sich die Länge des Fragebogens verstärkt bei den offenen Fragen auswirkt. Ob es mehr als Indizien sind, lässt sich allerdings auf Basis der beiden experimentellen Erhebungen nicht definitiv festlegen.

4.2.2 Vergleich der Ergebnisse

Um die Qualität der offenen Fragen besser abschätzen zu können, wurde im Rahmen des Experiments die Motive zu einer der Abstimmungsvorlagen codiert und analog zur VOTO ausgewertet. Diese Auswertung auf Basis des Code-Buchs von VOTO wurde für die Vorlage zum EU-Waffenrecht vorgenommen. Dabei zeigte sich, dass die Befragten die Beantwortung ernst genommen haben. Im gesamten Sample gab es keine einzige Jux-Antwort. Die Abbildungen 31 und 32 zeigen die Ergebnisse der Auswertung der Gründe für den Stimmentscheid für und gegen das EU-Waffenrecht.

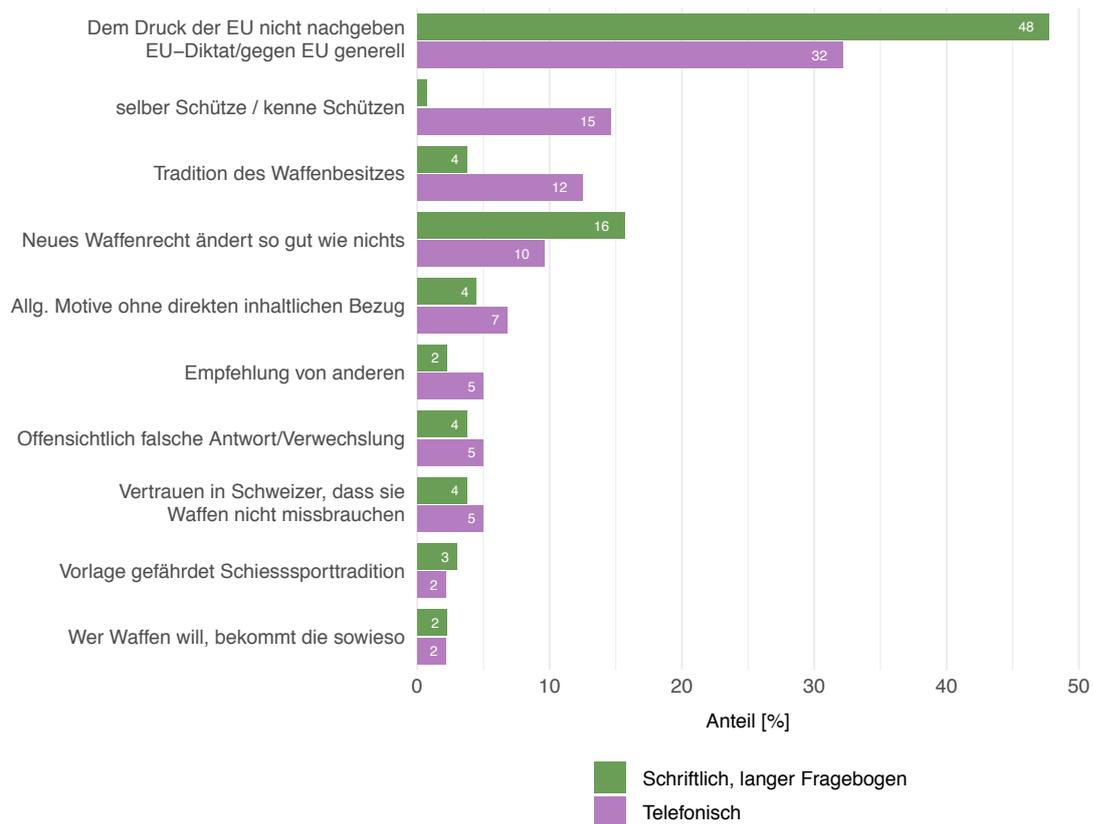


Abbildung 31: Offen abgefragte Motive gegen die Vorlage zum EU-Waffenrecht

Die Auswertungen der schriftlichen und telefonischen Befragungen zeigen grundsätzlich ein ähnliches Motivbild. Bei den konkreten Kategorien weichen die beiden Auswertungen jedoch teilweise erheblich voneinander ab. Da die Codierung von VOTO und die experimentelle Codierung der schriftlichen VOTO nicht von derselben Person vorgenommen wurde, lässt sich abschliessend nicht beantworten, ob die Unterschiede auf die Codierung oder auf tatsächlich abweichende Antworten zurückgehen. Unterschiede in der Beantwortung sind durchaus möglich. Die Auswertung zu Testzwecken zeigt jedoch, dass offene Motivfragen grundsätzlich auch bei einer schriftlichen Befragung funktionieren.

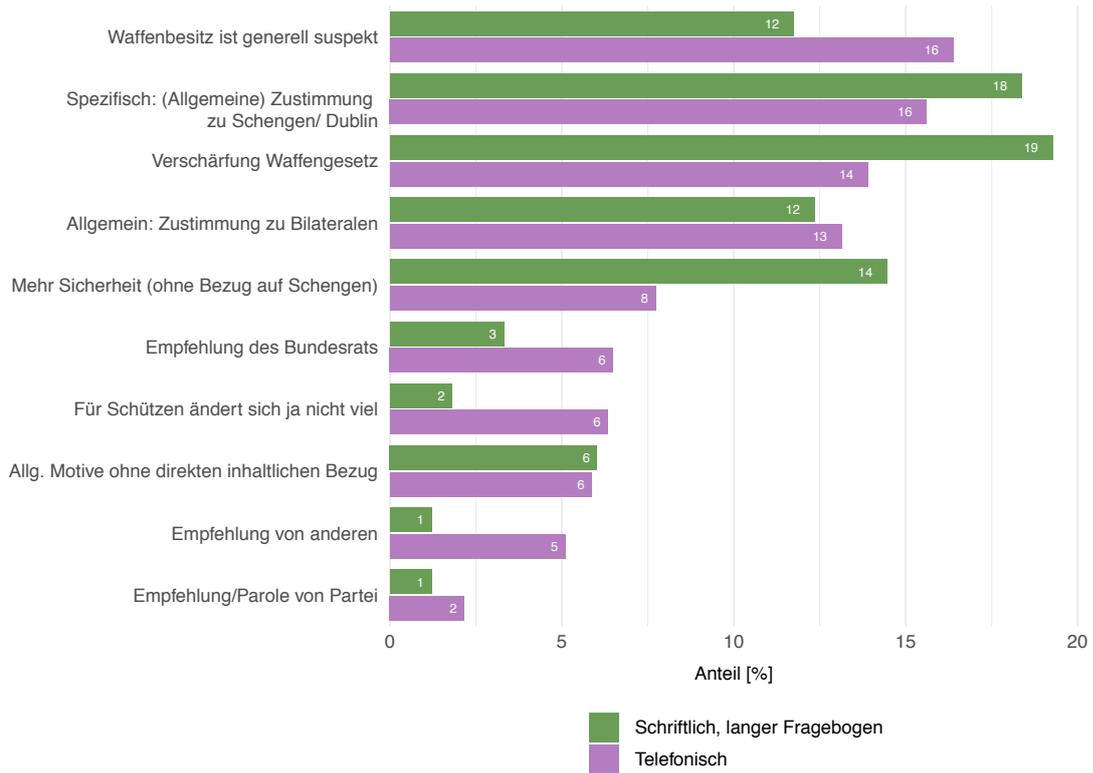


Abbildung 32: Offen abgefragte Motive für die Vorlage zum EU-Waffenrecht

4.3 Antwortkategorie «Weiss nicht»

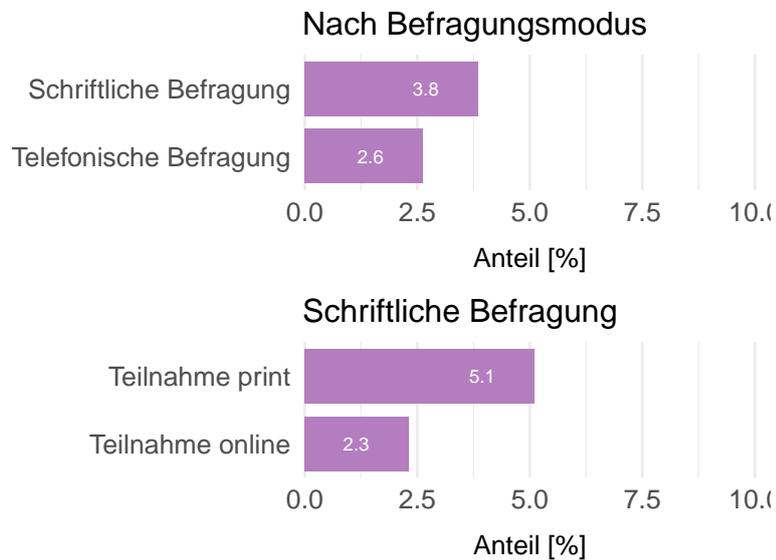


Abbildung 33: Durchschnittlicher Anteil «weiss nicht» Angaben Total, Vergleich zwischen den Befragungsmodi und den Teilnahmemodi schriftlich (Abstimmung Mai)

Ein weiterer Qualitätstest bezieht sich auf die Antwortkategorie «weiss nicht». Auch diese Kategorie wird bei der schriftlichen Befragung häufiger gewählt als bei der telefonischen. Die Unterschiede sind allerdings geringer als bei «keine Antwort». «Weiss nicht» wird dabei insbesondere bei der Teilnahme mittels Printfragebogen häufig gewählt, hier sind über 5 Prozent der Antworten «weiss nicht».

Die Auswertung nach Frageblöcken zeigt, dass in der schriftlichen Befragung die Antwort «weiss nicht» besonders oft bei den Themen «Politik generell» sowie «Werthaltungen» gewählt wurde. Es erscheint als plausibel, dass ein Teil der Befragten gerade bei diesen Themen tatsächlich das eine oder andere nicht weiss. Anders als «keine Antwort» ist die Häufigkeit von «weiss nicht» nicht unmittelbar mit der Qualität der Erhebungsmethode in Verbindung zu bringen. Möglich ist, dass sich die Befragten im anonymen schriftlichen Kontext eher getrauen, «weiss nicht» anzugeben als wenn sie am Telefon ein persönliches Gegenüber haben. Möglich ist aber auch, dass die Befragten sich am Telefon eher anstrengen, eine Antwort zu finden.

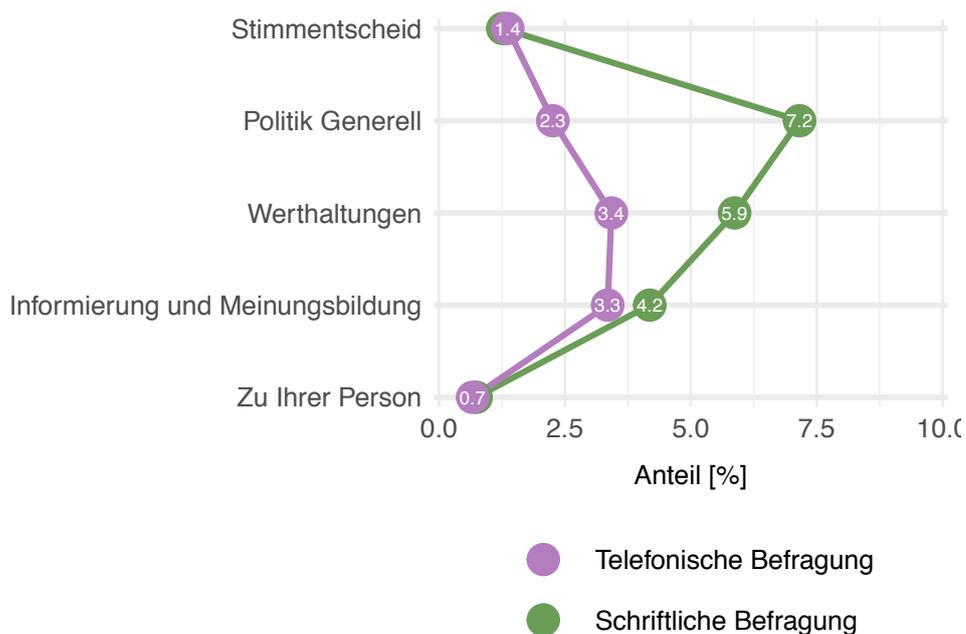


Abbildung 34: Durchschnittlicher Anteil «weiss nicht» Angaben pro Fragebogenabschnitt (Abstimmung Mai)

4.4 Unterschiede im Antwortverhalten

Zum Abschluss der Qualitätsbetrachtung folgt ein Vergleich der Antworten zu ausgewählten inhaltlichen Fragen, deren unterschiedliche Beantwortung sich womöglich auf die Erhebungsmethode zurückführen lässt.

4.4.1 Finanzielle Engpässe

Ein auffälliges Beispiel für ein unterschiedliches Antwortverhalten zeigt sich bei der Frage zu den finanziellen Engpässen. Bei der schriftlichen Befragung geben rund 48 Prozent der Befragten an, dass sie am Ende des Monats nach Abzug der Fixkosten gut durchkommen. Bei der telefonischen Befragung sind es mit 72 Prozent deutlich mehr. Die Unterschiede bleiben dabei zwischen den Befragungen sehr stabil.

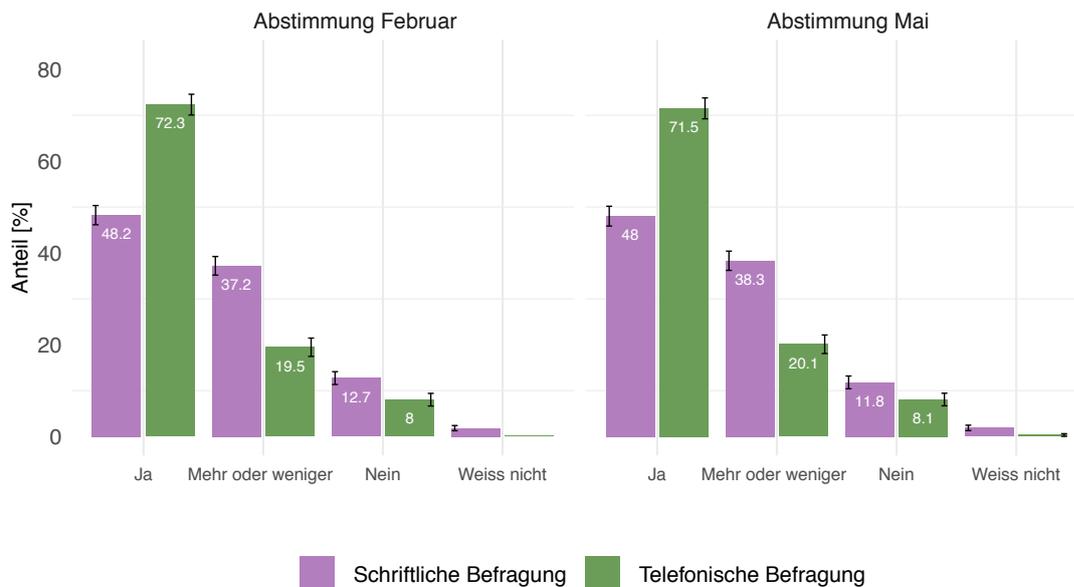


Abbildung 35: Vom Einkommen muss man ja einen Teil gerade wieder ausgeben, z. B. für Miete und Versicherungen. Mit dem, was Ihnen dann noch bleibt: Kommen Sie damit gut aus?

Grundsätzlich ist es denkbar, dass die Unterschiede zwischen den beiden Erhebungen aufgrund von Unterschieden in der Stichprobe zustande kommen, etwa der unterschiedlichen Altersverteilung der Teilnehmenden. Die Überprüfung der Einflussfaktoren mittels Logit-Regressionsmodell macht jedoch deutlich, dass es sich hier tatsächlich um einen Methodeneffekt handelt (vgl. Abb. 36). Neben der Einkommenshöhe, welche den grössten Einfluss darauf hat, ob jemand mit seinem verfügbaren Geld gut auskommt, spielt das Alter zwar tatsächlich eine Rolle: So kommen bei gleicher Einkommenshöhe die 65-Jährigen mit ihrem Einkommen deutlich besser aus und die 46- bis 55-Jährigen etwas schlechter als die Referenzgruppe der unter 36-Jährigen. Dennoch ist die Alterszusammensetzung nicht entscheidend für die gemessenen Unterschiede zwischen der schriftlichen und der telefonischen Befragung. Unter sonst gleichbleibenden Bedingungen ist der Anteil, der angibt, gut mit dem eigenen Einkommen auszukommen bei den telefonisch Befragten um rund 25 Prozentpunkte grösser als bei den schriftlich Befragten.

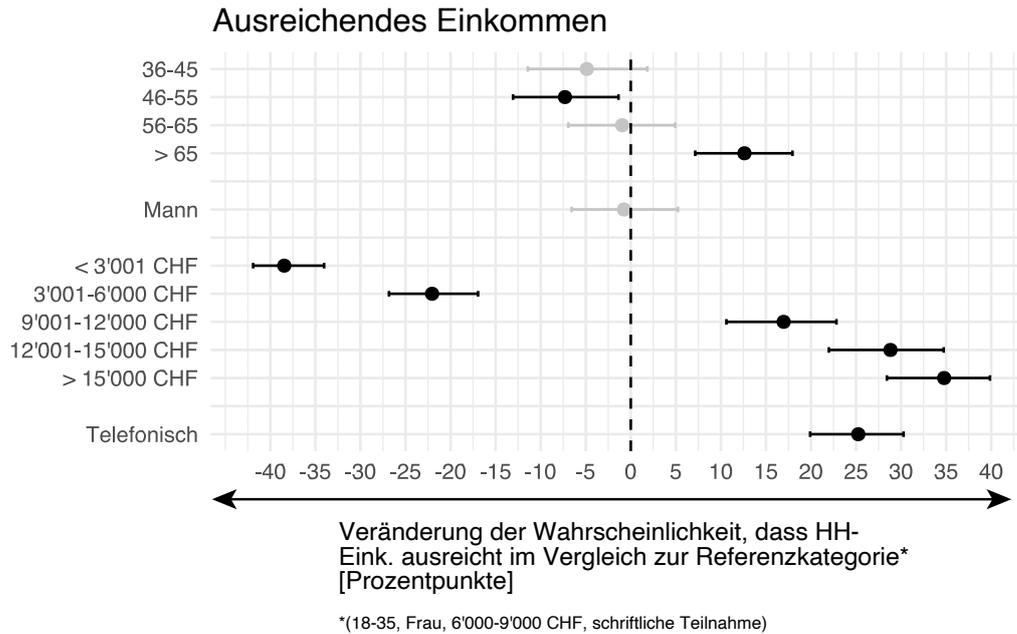


Abbildung 36: Einflussfaktoren auf die Angabe, dass das eigene Einkommen gut ausreicht im Vergleich zur Referenzgruppe

Dass gerade bei dieser Frage ein derart starker Methodeneffekt auftritt, lässt sich wohl nur darauf zurückführen, dass deren Beantwortung stark mit sozialer Erwünschtheit, bzw. mit Scham verbunden ist. Offensichtlich besteht eine Schamswelle einem persönlichen Gegenüber zuzugestehen, dass man finanziell nicht immer gut über die Runden kommt. Dies passt etwa mit der Erkenntnis von Kreuter et al. (2008) zusammen, dass Personen bei schriftlichen Umfragen eher Auskunft über sensitive Positionen geben als bei telefonischen.

4.4.2 Politisches Interesse und Teilnahmehäufigkeit

In der VOTO werden generelle politische Werthaltungen und Einstellungen abgefragt. So wird das politische Interesse mithilfe einer Viererskala, die von «sehr interessiert» bis «überhaupt nicht interessiert» reicht, abgefragt. Abbildung 37 zieht den Vergleich der Verteilung dieser Variable zwischen den Befragungsmethoden. Wie bei den finanziellen Engpässen sind die Unterschiede zwischen schriftlicher und telefonischer Befragung relativ stabil zwischen der ersten und der zweiten Befragung. Und auch hier zeigt sich ein systematischer Unterschied im Antwortverhalten. Bei beiden Umfrageterminen gaben in der telefonischen Befragung signifikant mehr Befragte an, «sehr interessiert» an Politik zu sein. Im Gegensatz dazu gaben auch wieder bei beiden Abstimmungsterminen signifikant mehr Befragte bei der schriftlichen Befragung an, «eher nicht interessiert» zu sein an Politik. Auch dies dürfte ein Beispiel für soziale Erwünschtheit sein.

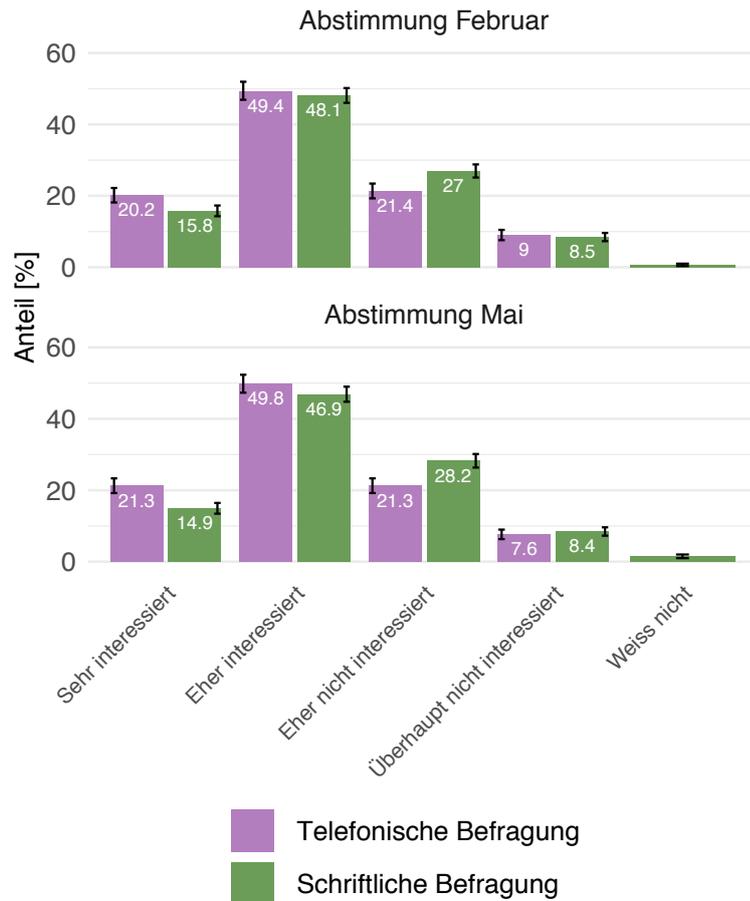


Abbildung 37: Vergleich telefonisch vs. schriftlich: Politisches Interesse

Erwartungsgemäss zeigt sich derselbe Methodeneffekt auch der Einschätzung der Teilnahmehäufigkeit an Abstimmungen. «Angenommen, in einem Jahr gibt es 10 eidgenössische Abstimmungstermine. An wie vielen davon nehmen Sie normalerweise teil?» Auf diese Frage antworten bei der telefonischen Befragung in beiden Abstimmungsterminen signifikant mehr Befragte als in der schriftlichen, dass sie an zehn Abstimmungsterminen teilnehmen würden (siehe Abb. 38).

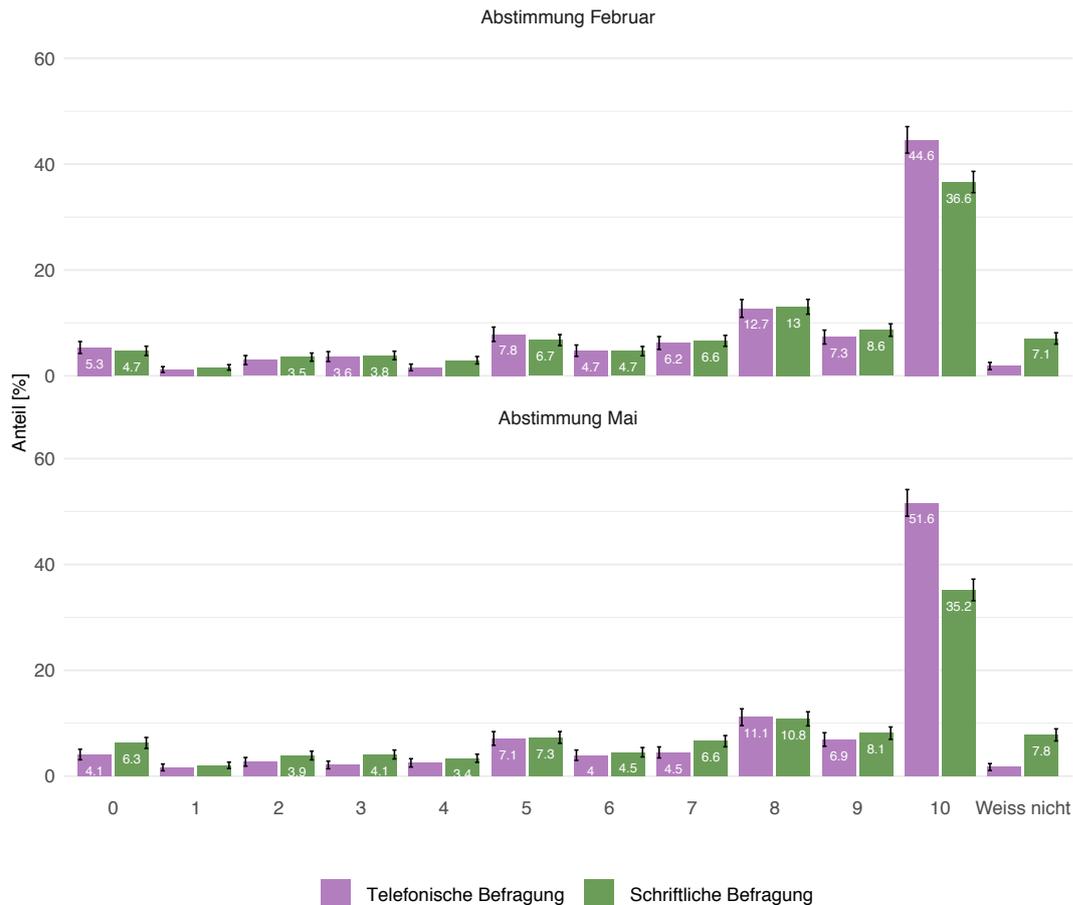


Abbildung 38: Vergleich telefonisch vs. schriftlich: Beteiligungswahrscheinlichkeit bei Volksabstimmungen

4.4.3 Links-rechts-Positionierung

Auch bei der Selbstpositionierung auf der Rechts-Links-Achse (Abb. 39) zeigen sich systematische Unterschiede zwischen den Erhebungsmethoden. Signifikant mehr Befragte geben in der telefonischen Befragung an, sich in der politischen Mitte zu positionieren, während in der schriftlichen Befragung im Vergleich zur telefonischen signifikant mehr Befragte die Kategorie «weiss nicht» wählen. Diese Tendenz zur Mittekategorie ist eine weitere Heuristik, die Umfrageteilnehmende oft im Zusammenhang mit Effekten der sozialen Erwünschtheit anwenden (Bagozzi & Mukherjee, 2012). Statt sich nicht zu positionieren kann die Mittekategorie als einfacher Ausweg gesehen werden, sich zu positionieren ohne wirklich Position beziehen zu müssen. Ein weiterer naheliegender Grund ist, dass die Befragten die Skala nur in einer schriftlichen Befragung visuell vor sich haben. Bei einer telefonischen müssen sie sich diese vorstellen. Die Interviewerin oder der Interviewer benennt dabei nur drei Zahlen: «0 ganz links, 5 die Mitte und 10 ganz rechts». Wer sich nicht ganz extrem positionieren will hört zusätzlich nur den Anker «5».

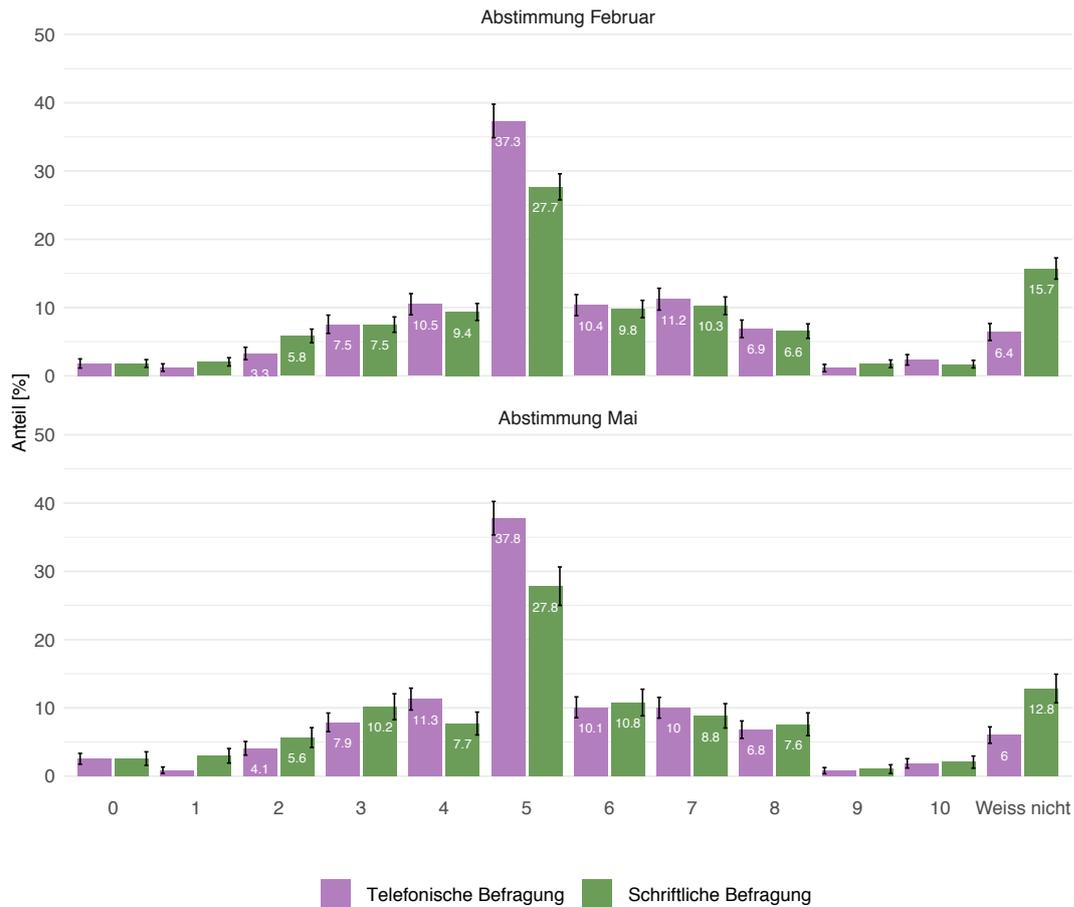


Abbildung 39: Vergleich telefonisch vs. schriftlich: rechts – links Selbstpositionierung

Letztlich ist es nicht entscheidend, ob die Unterschiede im Antwortverhalten zwischen schriftlicher und telefonischer Befragung auf Unterschiede in Bezug auf die soziale Erwünschtheit oder auf andere Methodeneffekte zurückgehen. Bedeutsam ist, dass es systematische Unterschiede im Antwortverhalten gibt. Falls ein Wechsel der VOTO von telefonischer zu schriftlicher Befragung vorgenommen wird, müssen diese Methodeneffekte, insbesondere bei Analysen über die Zeit, berücksichtigt werden.

5 Fazit

Das Methodenexperiment zeigt, dass sich mit der schriftlichen Erhebungsmethode im Rahmen einer VOTO-Befragung eine Ausschöpfung von rund 45 Prozent der Stichprobe erreichen lässt, während bei der telefonischen Befragung die minimal angestrebte Quote von mindestens 30 Prozent realisierter Interviews nicht immer erreicht wird. Diese Diskrepanz ergibt sich daraus, dass für fast zwei von fünf Personen in der Stichprobe telefonische Verbindungsdaten nicht eruiert werden können, während die für die schriftliche Befragung benötigte Postadresse für alle Zielpersonen vorhanden ist. Werden nur Personen mit Kontaktdaten verglichen, ist die erzielte Ausschöpfung beider Erhebungsmethoden vergleichbar.

Die schriftliche Befragung liegt in Bezug auf die Altersverteilung der Teilnehmenden signifikant näher an der Grundgesamtheit als die telefonische. In Bezug auf die anderen untersuchten soziodemographischen Variablen ist die Repräsentativität beider Methoden ähnlich gut, obwohl die Zusammensetzung der realisierten Stichprobe nur bei der telefonischen Befragung mittels Quotenvorgaben gesteuert werden kann. Beide Methoden zeigen eine erhebliche Verzerrung zugunsten von Stimmberechtigten, die an der jeweiligen Abstimmung teilgenommen haben.

Während die Faktoren Ausschöpfung und Repräsentativität klar für eine schriftliche Befragung sprechen, zeigt der Faktor «Qualität der Antworten» ein gemischtes Ergebnis. So ist bei der schriftlichen Befragung der Anteil der nicht beantworteten Fragen (Item-Nonresponse) signifikant höher als bei der telefonischen (4,4 zu 0,6 Prozent). Der höhere Anteil von Item-Nonresponse bei der schriftlichen Befragung fällt allerdings im Vergleich zum höheren Anteil nicht beantworteter Fragebogen (Unit-Nonresponse) bei der telefonischen Befragung kaum ins Gewicht. Werden beide Effekte zusammengezählt, ergibt dies beim Mai-Termin 73,5 Prozent Nicht-Antworten bei der telefonischen Befragung und 58 Prozent bei der schriftlichen. Substanziell mehr Nicht-Antworten bei der schriftlichen als bei der telefonischen Befragung zeigen sich bei den offenen Fragen zum Grund für den Stimmentscheid (bis zu 20 Prozent Nicht-Antworten beim ersten Pro- oder Kontra-Grund). Die Auswertung der Befragung zum Mai-Termin deutet ausserdem darauf hin, dass sich die Antwortmüdigkeit bei offenen Fragen bei einer grösseren Zahl von abgefragten Abstimmungsvorlagen verstärken dürfte. Bei der schriftlichen Befragung hat die Länge des Fragebogens einen grösseren Effekt auf den Rücklauf und wohl auch auf den Anteil der beantworteten Fragen als bei der telefonischen. Die Analyse der Antwortoption «weiss nicht» und der Antwortverteilung zu Einschätzungsfragen deutet demgegenüber darauf hin, dass der Faktor der sozialen Erwünschtheit bei der telefonischen Befragung relevanter ist als bei der schriftlichen.

Unter der Berücksichtigung aller untersuchter Dimensionen lässt sich der Schluss ziehen, dass eine schriftliche Befragung eine taugliche Alternative zur heutigen

telefonischen Befragung darstellt. Mit einem entsprechenden Erhebungsdesign lassen sich die spezifischen Herausforderungen der sinkenden Rücklaufquoten und der Teilnahmeverzerrung nach Altersklassen bewältigen. Die schriftliche Erhebungsmethode ist allerdings sensitiver bezüglich Fragebogenlänge, was insbesondere in Bezug auf die Qualität offener Fragen berücksichtigt werden muss.

Wird eine schriftliche Erhebungsmethode gewählt, so erscheint für die spezifischen Anforderungen von VOTO eine Mixed-Mode-Befragung (online und print) als unerlässlich. Der experimentelle Vergleich zeigt, dass auch bei einer schriftlichen Befragung ein Erstversand auf Montag nach dem Abstimmungswochenende angezeigt ist. Zur Wahrung der Kosteneffizienz wird der Papierfragebogen sinnvollerweise nur beim ersten Versand beigelegt. Während der Versand von bis zu zwei Erinnerungsschreiben angezeigt ist, erscheint eine monetäre Intensivierung nicht als opportun.

6 Quellenverzeichnis

- Bagozzi, B. E., & Mukherjee, B. (2012). A mixture model for middle category inflation in ordered survey responses. *Political Analysis, 20*(3), 369–386.
- Clausen, A. R. (1968). Response validity: Vote report. *The Public Opinion Quarterly, 32*(4), 588–606.
- Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly, 68*(1), 2–31.
- Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis, 22*(1), 1–30.
- Häusermann, S., Kurer, T., Pinggera, M., & Traber, D. (2018). Mehrheitsfähigkeit der altersvorsorge 2020: Die bewertung der reformelemente durch die stimmbürgerinnen und stimmbürger vor der abstimmung. *Swiss Political Science Review, 24*(1), 69–78.
- Karp, J. A., & Brockington, D. (2005). Social desirability and response validity: A comparative analysis of overreporting voter turnout in five countries. *The Journal of Politics, 67*(3), 825–840.
- Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in cati, ivr, and web surveys: The effects of mode and question sensitivity. *Public Opinion Quarterly, 72*(5), 847–865.