

Umfragen im Umfeld der Bundestagswahl 2002: Offline und Online im Vergleich

Faas, Thorsten

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Faas, T. (2003). Umfragen im Umfeld der Bundestagswahl 2002: Offline und Online im Vergleich. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 52, 120-135. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-198977>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Umfragen im Umfeld der Bundestagswahl 2002: Offline und Online im Vergleich

von Thorsten Faas¹

Zusammenfassung

Der Beitrag vergleicht Ergebnisse dreier Umfragen, die anlässlich der Bundestagswahl 2002 auf methodisch sehr unterschiedliche Weise durchgeführt wurden. Es handelt sich erstens um eine repräsentative mündliche Bevölkerungsumfrage, zweitens um eine repräsentative Online-Erhebung unter Internet-Nutzern sowie drittens um eine Online-Erhebung mit selbstrekrutierten Teilnehmern. Der Vergleich der drei Umfragen zeigt, dass sich die Umfragen sowohl hinsichtlich sozialstruktureller Variablen (Alter, Bildung und Geschlecht) als auch hinsichtlich substantieller Fragen (Wahlverhalten, politisches Interesse) deutlich voneinander unterscheiden. Zudem wird gezeigt, dass eine sozialstrukturelle Gewichtung nach Alter und Geschlecht diese substantiellen Unterschiede nicht beseitigt.

Abstract

The article compares the results of three surveys that were conducted in methodologically very different ways in the run-up to the last German federal election. The first survey is a representative sample of the German population, the second one is a representative web survey of internet users, the third one is an unsolicited web survey with self-selected participants. The comparison yields considerable differences among the three surveys concerning demographic (age, sex, education) as well as substantial variables (voting behaviour, interest in politics). It is also shown that these differences continue to exist after weighting the samples by sex and age.

1 **Thorsten Faas**, M.Sc. (LSE), ist wissenschaftlicher Mitarbeiter im DFG-Projekt „Politische Einstellungen, politische Partizipation und Wählerverhalten im vereinigten Deutschland“ am Lehrstuhl für Politikwissenschaft II der Otto-Friedrich-Universität Bamberg, Feldkirchenstraße 21, 96045 Bamberg, Email: thorsten.faas@sowi.uni-bamberg.de, URL: www.thorsten-faas.de

1 Einleitung²

Die Vergleichbarkeit von online und offline erhobenen Daten ist eine der methodischen Fragen, die derzeit sowohl kommerzielle Markt- und Meinungsforscher als auch mit Umfragen arbeitende Wissenschaftler umtreibt.³ Das Interesse an Online-Umfragen ist verständlich, liegen doch ihre Vorteile auf der Hand, zumindest auf den ersten Blick: Online-Umfragen – d.h. Umfragen, die via Internet oder per Email durchgeführt werden – sparen vor allem Zeit und Geld. Die Idee, repräsentative Umfragen online durchzuführen, erscheint daher angesichts weit verbreiteter Zeit- und Mittelknappheit sehr reizvoll. Der Schein könnte aber trügerisch sein, wenn nämlich eine geringere Datenqualität die Folge wäre.

Diesbezügliche Zweifel beziehen sich vor allem, aber keineswegs ausschließlich, auf Internet-Umfragen mit selbstrekrutierten Teilnehmern, in deren Rahmen man eine Umfrage ins Netz stellt, sie möglicherweise intensiv bewirbt⁴ – und auf Teilnehmer wartet (vgl. u.a. *Bandilla* und *Bosnjak* 2000, *Hauptmanns* 1999). Dieser Ansatz ist durchaus mit jenem vergleichbar, der der (falschen) Prognose des „Literary Digest“ anlässlich der amerikanischen Präsidentschaftswahl aus dem Jahre 1936⁵ zugrunde lag und seitdem als Inbegriff falscher Umfragemethodik dient. Entsprechend sehen die „Standards zur Qualitätssicherung für Online-Befragungen“ (vgl. *ADM* et al. 2001) auch explizit vor, dass „Teilnehmer von Online-Umfragen ‚aktiv‘ ... rekrutiert werden“ müssen (*ADM* et al. 2001: 2).

Allerdings führt dies zu der Frage, welche Teilnehmer überhaupt gezielt angesprochen werden können. Eine vollständige Liste aller deutschen Internet-Nutzer existiert bekanntlich ebenso wenig wie ein dem Random-Route-Verfahren vergleichbarer

2 Besonderer Dank gilt *Harald Schoen* für hilfreiche Kommentare und Anregungen zu einer Vorgänger-Version dieses Textes.

3 Eine detaillierte Darstellung zu Formen, Potenzialen und Problemen von Online-Umfragen findet sich u.a. in *Couper* (2000) und *Schonlau, Fricker* und *Elliot* (2002).

4 Eine Analyse verschiedener Werbemethoden hinsichtlich ihrer Effekte auf die Teilnehmerstruktur und -zahl einer Online-Umfrage findet sich in *Alvarez, Sherman* und *van Beselaere* (2003). Die Ergebnisse zeigen, dass der Erfolg von Bannerwerbung eher bescheiden ist, wohingegen „subscription campaigns“ erfolgreicher sind. Bei letzteren werden Internet-Nutzer, die sich für bestimmte (andere) Services registrieren, gleichzeitig zur Teilnahme an WWW-Umfragen eingeladen. Wichtig im hiesigen Zusammenhang ist aber das Fazit der Autoren, demzufolge beide Methoden zu strukturell unterschiedlichen Samples führten, wobei keine der beiden Samples wiederum der Struktur der Internet-Nutzer insgesamt entsprach.

5 Die Zeitschrift „Literary Digest“ verteilte damals rund 10 Millionen Fragebögen zur bevorstehenden Präsidentschaftswahl; 2,3 Millionen dieser Fragebögen wurden von den Lesern zurückgesandt. Die Ergebnisse dieser Umfrage deuteten einen klaren Sieg des republikanischen Kandidaten Landon an. Gleichzeitig prognostizierte aber George Gallup auf der Basis einer weitaus kleineren Zufallsstichprobe einen Sieg des Demokraten Roosevelt und behielt damit Recht.

Ansatz: Wie also dann eine Online-Zufallsstichprobe ziehen, wie es die reine Lehre vorsieht? Auch hier geben die ADM-Empfehlungen Ratschläge. Dort heißt es nämlich, „dass Internet-Umfragen ... auf der Grundlage einer vorherigen Offline-Auswahl bzw. Offline-Rekrutierung mittels geeigneter Screening-Techniken durchgeführt werden“ sollten (ADM et al. 2001: 2). Konkret bedeutet dies also, dass in persönlichen oder telefonischen bevölkerungsrepräsentativen Umfragen erhoben wird, ob a) ein Computer mit Internet-Anschluss und b) die Bereitschaft der Befragungsperson, zukünftig an Online-Umfragen teilzunehmen, vorhanden ist. Ist beides der Fall, so wird die Person Mitglied eines sogenannten Access-Panels. Für eine konkrete Online-Umfrage werden dann aus diesem Pool der Mitglieder des Access-Panels zufällig Personen ausgewählt, aktiv angeschrieben und zur Teilnahme an der Umfrage eingeladen. Auf diesem Weg, so die Idealvorstellung, können dann repräsentative Aussagen für die Internet-Nutzer getroffen werden, da es sich um eine mehrstufige Zufallsauswahl handelt: Zunächst werden zufällig mit den etablierten Verfahren Zielpersonen für persönliche oder telefonische bevölkerungsrepräsentative Umfragen ausgewählt. Darunter sollte sich ebenfalls ein Zufallsample der Internet-Nutzer befinden, die wiederum eingeladen werden, Mitglied des Access-Panels zu werden. Schließlich werden aus diesem Access-Panel zufällig Personen zur Teilnahme an einer konkreten Umfrage ausgewählt.

Inwieweit sich dieses idealtypische Verfahren in der Praxis bewährt, ist allerdings bisher kaum bekannt. Eine gewisse Skepsis jedenfalls scheint angebracht: Denn an jeder Stufe der zufälligen Auswahl können auch systematische Fehler wirken, zu deren Quellen beispielsweise niedrige Ausschöpfungsquoten bei bevölkerungsrepräsentativen Umfragen⁶, aber auch systematische Unterschiede in der Bereitschaft, sich für Online-Umfragen zur Verfügung zu stellen (und dann auch tatsächlich teilzunehmen), zählen. Schließlich kommt das für Panels typische Problem des Conditioning hinzu: Da die Mitglieder des Panels mehrfach an Befragungen teilnehmen, können sich leicht Lerneffekte oder bestimmte Antwortstrategien entwickeln.

Was also idealtypisch einfach klingt, erweist sich in der Umsetzung als deutlich schwieriger. Insofern ist es kaum verwunderlich, dass sich Online-Umfragen aufbauend auf Selbstselektion trotz aller Nähe zum Literary Digest weiter Beliebtheit erfreuen, gemäß dem Motto: Wenn auch bei mit hohem Aufwand durchgeführten Umfragen, die der reinen Lehre entsprechen, Ausschöpfungs- und Verzerrungsprobleme unvermeidlich sind, warum nicht gleich den einfachen Weg gehen?

6 **Koch** (2002) etwa berichtet von Ausschöpfungsquoten von 49,1 Prozent und 54,2 Prozent für die ALLBUS-Erhebungen der Jahre 2000 bzw. 1996.

Letztlich können nur systematische Vergleiche von Daten, die parallel auf Offline- und Online-Erhebung sowie auf aktiver und selbstselektiver Rekrutierung beruhen, gesicherte Erkenntnisse darüber bringen, welche der verschiedenen Verfahren zu validen und reliablen Ergebnissen führen. Zudem kann auch nur auf diesem Wege geprüft werden, ob es möglich ist, Gewichtungsmaßnahmen zu entwickeln, die die auftretenden Unterschiede zwischen Umfragen korrigieren können.⁷ Einen solchen systematischen Vergleich soll der vorliegende Beitrag liefern, wobei auf drei Umfragen, die anlässlich der Bundestagswahl 2002 durchgeführt wurden, zurückgegriffen wird.^{8,9}

2 Datenbasis

Die drei zugrunde liegenden, anlässlich der Bundestagswahl 2002 durchgeführten Umfragen wurden auf sehr verschiedene Weise durchgeführt; einen entsprechenden Überblick gibt Tabelle 1. Bei der ersten Umfrage handelt es sich um eine repräsentative Bevölkerungsumfrage, in deren Rahmen im Zeitraum vom 12. August bis zum 21. September 2002 1.665 Personen persönlich befragt wurden. Diese Personen wurden zufällig – über Sample Points, Random Route und Schwedenschlüssel – ausgewählt. Die Ausschöpfungsquote lag bei 63,8 Prozent. Die zweite Erhebung ist eine repräsentative, Internet-basierte Online-Umfrage unter 1.165 deutschen Internet-Nutzern. Diese wurden zufällig aus dem Access-Panel des Meinungsforschungsinstituts INRA ausgewählt und im Zeitraum vom 13. September bis zum 1. Oktober 2002 befragt. Die Ausschöpfungsquote lag hier bei 74,2 Prozent. Schließlich wurde drittens noch eine weitere Internet-basierte Online-Umfrage durchgeführt, die unter www.wahlumfrage2002.de erreichbar war. Hier konnten sich im Zeitraum vom 20. August bis zum 22. September 2002 Internet-Nutzer selbst rekrutieren – insgesamt machten davon 34.098 Gebrauch. Zieht man von diesen nochmals Eingänge ab, bei denen weniger als fünf gültige Antworten zu verzeichnen

7 Dabei ist natürlich angesichts der weiterhin bestehenden Unterschiede zwischen Internet- und Gesamtbevölkerung zu fragen, an welche Grundgesamtheit die Gewichtung anpassen soll. Im Rahmen dieses Beitrags wird – vor allem im Sinne des Methodenexperiments – die Gesamtbevölkerung gewählt.

8 Ähnliche Vergleiche zwischen offline- und online-erhobenen Daten finden sich u.a. auch in *Bandilla, Bosnjak* und *Altdorfer* (2001) sowie *Berrens* et al. (2003).

9 Auch in anderen Ländern sind zu Zwecken der Wahlforschung virtuelle Elemente (ergänzend) eingesetzt worden, vgl. etwa *Gibson* und *McAllister* (2002) für Australien, *Sanders* et al. (2002) für Großbritannien und *Krosnick* und *Chang* (2002) für die USA.

Tabelle 1 Details der drei Umfragen

| | Repräsentative Bevölkerungsumfrage | Repräsentative Online-Umfrage | Wahlumfrage2002 |
|----------------------|---|--|------------------------------------|
| Feldzeit | 12. August – 21. September 2002 | 13. September – 1. Oktober 2002 | 20. August – 22. September 2002 |
| Teilnehmer | 1.665 | 1.165 | 34.098/29.583 |
| Rekrutierung | Zufällige Auswahl über Sample Points, Random Route und Schwedenschlüssel | Zufällige Auswahl aus einem Access Panel | Selbstrekrutierung |
| Befragungs- modus | PAPI | CASI ¹⁰ | CASI |
| Ausschöpfung | 63,8% | 74,2% | — |

waren, so verbleiben 29.583 Antworteingänge, die die Datenbasis für die Ergebnisse dieser Umfrage bilden.¹¹

Zusätzlich wurden die drei Umfragen gewichtet, indem die gemeinsame Verteilung von Alter und Geschlecht in jeder der drei Umfragen an die aus der Amtlichen Statistik bekannte (gemeinsame) Verteilung dieser Variablen in der Bevölkerung angepasst wurde.¹² Somit liegen drei ungewichtete Stichproben und drei gewichtete Stichproben vor, wobei letztere die Struktur der Gesamtbevölkerung hinsichtlich Alter und Geschlecht widerspiegeln. Damit lässt sich prüfen, ob auftretende (ungewichtete) Unterschiede zwischen den Stichproben primär auf sozialstrukturelle Abweichungen zurückzuführen sind oder ob die Ursachen der Abweichungen darüber hinausgehen. Diese Stichproben sollen nun sukzessive verglichen werden, wobei der Vergleich sowohl sozialstrukturelle Variablen (Alter, Geschlecht, Bildung) als auch substantielle Fragen (beabsichtigtes Wahlverhalten, politisches Interesse) umfasst.

10 Computer Assisted Self-Administered Interview.

11 Weitere Details zur Wahlumfrage2002 finden sich in *Faas* (2003).

12 Folgende Altersklassen wurden der Gewichtung zugrunde gelegt: 16 bis 17, 18 bis 24, 25 bis 34, 35 bis 44, 45 bis 59, über 60 Jahre. Zusätzlich wurde die bevölkerungsrepräsentative Umfrage an die gemeinsame Verteilung von Ortsgröße und Bundesland angepasst, die beiden Internet-basierten Umfragen mangels Informationen über die Ortsgröße zusätzlich an die Verteilung der Bundesländer.

3 Vergleich der ungewichteten Stichproben

Vergleicht man die drei Umfragen zunächst hinsichtlich ihrer sozialstrukturellen Zusammensetzung, so treten – wie ein Blick auf Tabelle 2 zeigt – deutliche Unterschiede zu Tage. Diese umfassen Geschlecht, Alter und auch Bildung.

Tabelle 2 Sozialstrukturelle Zusammensetzung der Teilnehmer der drei Umfragen, in Prozent (ungewichtete Daten)

| | Repräsentative Bevölkerungsumfrage | Repräsentative Online-Umfrage | Wahlumfrage2002 |
|-------------------------------------|---------------------------------------|----------------------------------|---------------------|
| Geschlecht | | | |
| Männer | 51,5 | 58,8 | 77,9 |
| Frauen | 48,5 | 41,2 | 22,1 |
| Summe | 100,0 (n=1.663) | 100,0 (n=1.165) | 100,0 (n=24.928) |
| Alter | | | |
| 16-24 | 9,3 | 21,6 | 30,3 |
| 25-34 | 13,6 | 23,3 | 33,7 |
| 35-44 | 19,1 | 28,8 | 18,6 |
| 44-59 | 22,2 | 20,3 | 13,2 |
| über 60 | 35,8 | 5,9 | 4,2 |
| Summe | 100,0 (n=1.661) | 100,0 (n=1.165) | 100,0 (n=24.727) |
| Mittleres Alter (in Jahren) | 50,0 | 36,7 | 32,8 |
| Bildung | | | |
| Noch Schüler | 2,0 | 6,9 | 7,2 |
| Volks-, Hauptschul- abschluss | 42,8 | 13,0 | 5,1 |
| Mittlere Reife | 28,7 | 36,1 | 16,9 |
| Abitur | 26,6 | 44,0 | 70,8 |
| Summe | 100,0 (n=1.634) | 100,0 (n=1.165) | 100,0 (n=23.828) |

Bei der Stichprobenziehung der bevölkerungsrepräsentativen Umfrage wurde Ostdeutschland überrepräsentiert. Dies wurde hier durch ein reines Ost-West-Gewicht korrigiert.

Dabei gilt, dass sich die bekannten Verzerrungen der Online-Bevölkerung im Vergleich zur Gesamtbevölkerung (die hier erneut bestätigt werden können) bei den selbstrekrutierten Teilnehmern der Wahlumfrage2002 nochmals verschärfen. Konkret schlägt sich dies darin nieder, dass 58,8 Prozent der Teilnehmer der repräsentativen Online-Erhebung Männer sind, verglichen mit nur 51,5 Prozent der Teilnehmer der bevölkerungsrepräsentativen Umfrage. Von den selbstrekrutierten Teilnehmern der Wahlumfrage2002 waren aber sogar 77,9 Prozent männlichen Geschlechts. Gleiches gilt auch für die Altersverteilung: Die Teilnehmer der Wahlumfrage2002 sind mit einem mittleren Alter von 32,8 Jahren deutlich jünger als sowohl die Teilnehmer der repräsentativen Internet-Erhebung (36,7 Jahre) als auch die Teilnehmer der bevölkerungsrepräsentativen Umfrage (50,0 Jahre). Besonders die „Extremgruppen“ sind sehr unterschiedlich besetzt: Während 30,3 Prozent der Teilnehmer der Wahlumfrage2002 zwischen 16 und 24 Jahren alt waren (und nur 4,2 Prozent über 60 Jahre), ist das entsprechende Verhältnis in der bevölkerungsrepräsentativen Umfrage fast spiegelbildlich: 9,3 Prozent gegenüber 35,8 Prozent. Überaus deutlich ist auch der Unterschied bei der Verteilung der Bildungsabschlüsse: 70,8 Prozent aller Teilnehmer an der Wahlumfrage2002 haben Abitur verglichen mit nur 44,0 Prozent der Teilnehmer der Internet-Erhebung bzw. nur 26,6 Prozent der Teilnehmer der bevölkerungsrepräsentativen Umfrage. Der typische Teilnehmer der Wahlumfrage2002 war also männlich, jung und hoch gebildet – und dies in einem noch viel stärkeren Maße, als es der „normale“ Internet-Nutzer im Vergleich zum „Normalbürger“ ohnehin schon ist.

Angesichts der Größenordnung der zu Tage getretenen Unterschiede ist es wenig verwunderlich, dass die Unterschiede durchweg von signifikantem Ausmaß sind, wie Tabelle 3 zeigt. Zudem fällt bei der Betrachtung der Zusammenhangsmaße auf, dass sich die beiden Internet-basierten Umfragen bezüglich der sozialstrukturellen Zusammensetzung „am ähnlichsten“ sind, während – mit Ausnahme des Alters – die Unterschiede zwischen Wahlumfrage2002 und bevölkerungsrepräsentativer Umfrage am größten sind.

Verlässt man nun den sozialstrukturellen Bereich und betrachtet stattdessen Fragen von substantiellem Interesse, so treten auch hier – siehe Tabelle 4 – deutliche Unterschiede hervor. Was das Wahlverhalten betrifft, hier operationalisiert durch die Frage, welche Partei man mit der Zweitstimme zu wählen beabsichtigt¹³, zeigen sich beim Übergang in die virtuelle Welt – d.h. also in beiden Internet-basierten

13 Etwa die Hälfte der Interviews der repräsentativen Internet-Erhebung wurde erst nach der Bundestagswahl durchgeführt. Für diese Personen wurde hier die retrospektive Frage nach der mit der Zweitstimme gewählten Partei herangezogen.

Umfragen – Verschiebungen zu Gunsten der beiden kleinen Parteien. Zudem zeigt sich, dass die Parteiateile der bevölkerungsrepräsentativen Umfrage relativ dicht an den tatsächlichen Parteiateilen gemäß dem Wahlausgang vom 22. September 2002 liegen.¹⁴ Offensichtlich handelt es sich also tatsächlich um eine bevölkerungsrepräsentative Umfrage.

Tabelle 3 Größenordnung und Signifikanz der sozialstrukturellen Unterschiede zwischen den drei Umfragen, χ^2 und Cramers V (ungewichtete Daten)

| | Geschlecht | | Alter | | Bildung | |
|------------------------------------|--|---|---|--|---|--|
| | Wahlumfrage 2002 | Repräsentative Online-Umfrage | Wahlumfrage 2002 | Repräsentative Online-Umfrage | Wahlumfrage 2002 | Repräsentative Online-Umfrage |
| Repräsentative Bevölkerungsumfrage | $\chi^2_{df=1} = 600$ p<0,001 V=0,15 | $\chi^2_{df=1} = 15$ p<0,001 V=0,07 | $\chi^2_{df=4} = 3033$ p<0,001 V=0,34 | $\chi^2_{df=4} = 399$ p<0,001 V=0,38 | $\chi^2_{df=3} = 3508$ p<0,001 V=0,37 | $\chi^2_{df=3} = 311$ p<0,001 V=0,33 |
| Repräsentative Online-Umfrage | $\chi^2_{df=1} = 318$ p<0,001 V=0,09 | | $\chi^2_{df=4} = 174$ p<0,001 V=0,08 | | $\chi^2_{df=3} = 472$ p<0,001 V=0,14 | |

Wie weicht aber die Stimmenverteilung der Internet-Bevölkerung von dieser allgemeinen Verteilung im Einzelnen ab? Beiden Internet-basierten Umfragen – Wahlumfrage2002 wie auch repräsentativer Internet-Erhebung – ist gemein, dass der Anteil der Union hier sehr niedrig liegt: Weniger als ein Viertel der jeweiligen Teilnehmer nennt bei der Frage nach der Zweitstimme die Union.

Bei den übrigen Parteien muss noch einmal zwischen den beiden Internet-basierten Umfragen unterschieden werden. Die Anteile der Grünen und der FDP werden zwar in beiden Umfragen überschätzt, allerdings in der Wahlumfrage2002 nochmals deutlich stärker als in der repräsentativen Online-Umfrage. Die Grünen erhalten von den Teilnehmern der Wahlumfrage2002 21,4 Prozent der Zweitstimmen, die FDP 18,1 Prozent, verglichen mit „nur“ 14,1 Prozent bzw. 11,4 Prozent in der repräsentativen Online-Erhebung. Was schließlich die SPD betrifft, so wird ihr Stimmenanteil

¹⁴ Nach dem amtlichen Ergebnis erreichte die SPD 38,5 Prozent, die Union ebenfalls 38,5 Prozent, die Grünen 8,6 Prozent, die FDP 7,4 Prozent, die PDS 4,0 Prozent und sonstige Parteien 3,0 Prozent. Die Unterschiede zwischen diesen Anteilen und den Anteilen gemäß der bevölkerungsrepräsentativen Umfrage sind nicht signifikant.

in der repräsentativen Online-Umfrage mit 41,1 Prozent überschätzt. In der Wahlumfrage2002 ergeht es ihr allerdings kaum besser als der Union – sie erhält hier nur 29,3 Prozent der Zweitstimmen.

Tabelle 4 (Beabsichtigtes) Wahlverhalten und politisches Interesse der Teilnehmer der drei Umfragen, in Prozent (ungewichtete Daten)

| | Repräsentative Bevölkerungsumfrage | Repräsentative Online-Umfrage | Wahlumfrage2002 |
|------------------------------|---------------------------------------|----------------------------------|---------------------|
| Wahlverhalten | | | |
| SPD | 38,2 | 41,1 | 29,3 |
| CDU/CSU | 35,9 | 24,0 | 23,6 |
| Bündnis 90/Die Grünen | 8,9 | 14,1 | 21,4 |
| FDP | 8,6 | 11,4 | 18,1 |
| PDS | 5,8 | 4,7 | 4,2 |
| Sonstige | 2,6 | 4,7 | 3,4 |
| Summe | 100,0 (n=1.280) | 100,0 (n=1084) | 100,0 (n=27.163) |
| Politisches Interesse | | | |
| Sehr stark (+2) | 10,9 | 8,2 | 35,2 |
| Stark (+1) | 23,2 | 30,5 | 40,0 |
| Mittelmäßig (0) | 41,8 | 46,9 | 21,2 |
| Weniger stark (-1) | 17,9 | 12,7 | 3,1 |
| Überhaupt nicht (-2) | 6,1 | 1,8 | 0,5 |
| Summe | 100,0 (n=1.648) | 100,0 (n=1.162) | 100,0 (n=27.803) |
| Mittleres Interesse | 0,2 | 0,3 | 1,1 |

Bei der Stichprobenziehung der bevölkerungsrepräsentativen Umfrage wurde Ostdeutschland überrepräsentiert. Dies wurde hier durch ein reines Ost-West-Gewicht korrigiert.

Insgesamt ergibt dies – betrachtet man politische Lager – eine Verzerrung der Internet-Stichproben zugunsten von Rot-Grün, vor allem zu Lasten der Christdemokraten.

Neben dem Wahlverhalten soll hier eine weitere Größe betrachtet werden: das politische Interesse. Die erneut erkennbaren Unterschiede zwischen den drei Umfragen – aber insbesondere zwischen der Wahlumfrage2002 auf der einen Seite und den beiden repräsentativen Umfragen auf der anderen Seite – sind beachtlich. Bei den (selbstrekrutierten) Teilnehmern der Wahlumfrage2002 handelt es sich offensichtlich um eine politisch höchst interessierte Gruppe, die sich diesbezüglich deutlich vom Rest der Bevölkerung unterscheidet: Drei Viertel der Teilnehmer geben an, politisch sehr stark oder stark interessiert zu sein, im Vergleich dazu sind es nur rund 40 Prozent bzw. rund 33 Prozent in der repräsentativen Online- bzw. Bevölkerungsumfrage. Somit gilt erneut: In den Ergebnissen der Wahlumfrage2002 potenzieren sich Verzerrungen, die sich ohnehin schon beim Übergang von der realen in die virtuelle Welt ergeben.

Tabelle 5 Größenordnung und Signifikanz der substanziellen Unterschiede zwischen den drei Umfragen, χ^2 und Cramers V (ungewichtete Daten)

| | Wahlverhalten | | Politisches Interesse | |
|------------------------------------|--|---|---|---|
| | Wahlumfrage 2002 | Repräsentative Online-Umfrage | Wahlumfrage 2002 | Repräsentative Online-Umfrage |
| Repräsentative Bevölkerungsumfrage | $\chi^2_{df=5} = 273$ p<0,001 V=0,10 | $\chi^2_{df=5} = 56$ p<0,001 V=0,15 | $\chi^2_{df=4} = 2149$ p<0,001 V=0,27 | $\chi^2_{df=4} = 64$ p<0,001 V=0,15 |
| Repräsentative Online-Umfrage | $\chi^2_{df=5} = 107$ p<0,001 V=0,06 | | $\chi^2_{df=4} = 923$ p<0,001 V=0,18 | |

Betrachtet man abschließend die Signifikanz der berichteten Unterschiede, so erweisen sich alle Unterschiede als signifikant (siehe Tabelle 5). Zudem zeigt die Betrachtung der Zusammenhangsmaße, dass sich bei der Betrachtung des Wahlverhaltens die beiden Internet-basierten Umfragen weniger stark unterscheiden, die bevölkerungsrepräsentative Umfrage also der deutlichste „Ausreißer“ ist, während dies beim politischen Interesse eher die Wahlumfrage2002 ist.

Ein Zwischenfazit an dieser Stelle muss zwei Punkte umfassen. Erstens sind die Gefahren der Selbstselektion deutlich hervorgetreten. An der Wahlumfrage2002 haben sich primär politisch höchst interessierte, hoch gebildete Internet-Nutzer

beteiligt. Das Ausmaß der Verzerrungen, das diese Selbstselektion produziert hat, ist immens. Zweitens liegen – trotz einer Internet-Penetration von rund 50 Prozent – auch zwischen repräsentativer Internet- und repräsentativer Bevölkerungsumfrage deutliche Unterschiede. Rückschlüsse auf die Bevölkerung insgesamt auf der Basis von Online-Umfragen sind also weiterhin mit Vorsicht zu genießen. Allerdings ist bisher noch keinerlei Gewichtung in die Analysen eingeflossen, die die beobachteten Verzerrungen korrigieren könnten. Dies soll nun im nächsten Schritt geschehen.

4 Vergleich der gewichteten Stichproben

Beim Vergleich der gewichteten Ergebnisse ist natürlich die Betrachtung der sozialstrukturellen Verteilungen redundant, da – mit Ausnahme der Bildung – diese Variablen zur Gewichtung herangezogen wurden. Daher können wir uns hier auf die Darstellung der beiden substanziellen Variablen beschränken. Zuvor aber sind noch einige Angaben zu den Größenordnungen der einzelnen Gewichte zu machen. Für die bevölkerungsrepräsentative Umfrage beträgt das minimale (individuelle) Gewicht 0,053, das maximale 7,061; das untere Quartil der Verteilung der Gewichte liegt bei 0,531, das obere bei 1,299. Für die repräsentative Online-Erhebung liegt das minimale individuelle Gewicht bei 0,256, das maximale bei 20,321, das untere Quartil liegt bei 0,568, das obere bei 0,894. Schließlich betragen die entsprechenden Werte der Wahlumfrage2002 0,160 bzw. 44,670 sowie 0,317 bzw. 0,994. Es wird damit deutlich, dass für die beiden Internet-basierten Erhebungen deutlich stärkere Anpassungen nötig sind als in der bevölkerungsrepräsentativen Umfrage, wobei insbesondere einige wenige Individuen (dies zeigt jeweils der Wert des oberen Quartils) deutlich „hochgewichtet“ werden müssen. Dabei handelt es sich vor allem um ältere Personen, insbesondere ältere Frauen.

Betrachtet man nun die gewichteten Ergebnisse (Tabelle 6), so zeigt sich, dass die zwischen den Umfragen beobachteten Unterschiede durch die sozialstrukturelle Gewichtung nicht verschwinden, vielmehr weitgehend unverändert und signifikant (siehe Tabelle 8) bestehen bleiben. Die auffälligste durch die Gewichtung hervorgerufene Veränderung in der bevölkerungsrepräsentativen Umfrage ist, was das Wahlverhalten betrifft, die Annäherung der beiden großen Parteien: Die Union gewinnt durch die Gewichtung 0,8 Prozentpunkte hinzu, während die SPD 0,6 Prozentpunkte verliert. Für die übrigen Parteien ergeben sich kaum Unterschiede.

Tabelle 6 (Beabsichtigtes) Wahlverhalten und politisches Interesse der Teilnehmer der drei Umfragen, in Prozent sowie - in Klammern - Veränderungen gegenüber ungewichteten Daten (sozialstrukturell gewichtete Daten)

| | Repräsentative Bevölkerungsumfrage | Repräsentative Online-Umfrage | Wahlumfrage2002 |
|------------------------------|---------------------------------------|----------------------------------|---------------------|
| Wahlverhalten | | | |
| SPD | 37,6 (-0,6) | 38,5 (-2,6) | 31,5 (+2,2) |
| CDU/CSU | 36,7 (+0,8) | 23,8 (-0,2) | 23,1 (-0,5) |
| Bündnis 90/Die Grünen | 9,0 (+0,1) | 14,3 (+0,2) | 21,4 (+/-0,0) |
| FDP | 8,3 (-0,3) | 13,1 (+1,7) | 15,4 (-2,7) |
| PDS | 5,6 (-0,2) | 6,6 (+1,9) | 5,3 (+1,1) |
| Sonstige | 2,8 (+0,2) | 3,7 (-1,0) | 3,3 (-0,1) |
| Summe | 100,0 (n=1.278) | 100,0 (n=1.110) | 100,0 (n=27.062) |
| Politisches Interesse | | | |
| Sehr stark (+2) | 9,8 (-1,1) | 8,7 (+0,5) | 35,0 (-0,2) |
| Stark (+1) | 22,2 (-1,0) | 37,4 (+6,9) | 39,9 (-0,1) |
| Mittelmäßig (0) | 43,2 (+1,4) | 43,3 (-3,6) | 21,4 (+0,2) |
| Weniger stark (-1) | 18,2 (+0,3) | 9,2 (-3,5) | 2,9 (-0,2) |
| Überhaupt nicht (-2) | 6,5 (+0,4) | 1,4 (-0,4) | 0,7 (+0,2) |
| Summe | 100,0 (n=1.649) | 100,0 (n=1.163) | 100,0 (n=27.602) |
| Mittleres Interesse | 0,1 (-0,1) | 0,4 (+0,1) | 1,1 (+/-0,0) |

Würde man auf Basis der Daten eine Wahlprognose machen wollen, so fiel diese auf Basis der gewichteten Daten etwas besser aus als zuvor auf Basis der ungewichteten. Im Mittel würden die Parteianteile im Vergleich zum tatsächlichen Wahlausgang vom 22. September 2002 um 1,0 Prozentpunkte verfehlt, die maximale Abweichung würde 1,8 Prozentpunkte betragen; dies im Vergleich zu 1,1 bzw. 2,6 Prozentpunkten auf der Basis ungewichteter Daten (vgl. Tabelle 7).

Was das politische Interesse, die zweite Variable von substantiellem Interesse in diesem Zusammenhang, betrifft, so kommt es zu einer Verschiebung hin zu weniger starkem Interesse. Entsprechend sinkt der Mittelwert um einen Zehntelskalenpunkt.

Für die repräsentative Online-Erhebung ergeben sich insgesamt stärkere Verschiebungen bei der Betrachtung des Wahlverhaltens: Die Anteile der SPD sowie der sonstigen Parteien sinken um 2,6 bzw. 1,0 Prozentpunkte, während die Anteile der FDP bzw. der PDS um 1,7 bzw. 1,9 Prozentpunkte steigen. Allerdings führen diese Verschiebungen nicht zu einem einheitlichen Trend, demzufolge die Anteile der einzelnen Parteien nun näher am tatsächlichen Wahlergebnis lägen. Zwar wird der SPD-Anteil nach der Gewichtung exakt getroffen, die (deutlichen) Verschiebungen bei FDP und PDS dagegen führen dazu, dass deren Anteile noch deutlicher verfehlt werden. Auch die Union, für die die Gewichtung nur minimale Veränderungen ergibt (-0,2 Prozentpunkte), bleibt auf dem viel zu niedrigen Niveau von 23,8 Prozent. Betrachtet man hier eine hypothetische Wahlprognose auf Basis dieser Daten, so hat die Gewichtung kontraproduktive Auswirkungen: Ohne Gewichtung wurden die Anteile der Parteien im Mittel um 4,8 Prozentpunkte verfehlt, die größte Abweichung lag (für die Union) bei 14,5 Prozentpunkten. Mit gewichteten Daten betragen die entsprechenden Werte aber sogar 4,9 bzw. 14,7 Prozentpunkte.

Tabelle 7 Abweichungen hypothetischer Wahlprognosen auf Basis der ungewichteten und gewichteten Daten vom tatsächlichen Wahlergebnis

| | Repräsentative Bevölkerungsumfrage | | Repräsentative Online-Umfrage | | Wahlumfrage2002 | |
|--|------------------------------------|-----------|-------------------------------|-----------|------------------|-----------|
| | Unge- wichtet | Gewichtet | Unge- wichtet | Gewichtet | Unge- wichtet | Gewichtet |
| Durchschnittliche Abweichung für alle Parteien | 1,1 | 1,0 | 4,8 | 4,9 | 8,0 | 7,5 |
| Maximale Abweichung | 2,6 | 1,8 | 14,5 | 14,7 | 14,9 | 15,4 |

Die Effekte der Gewichtung auf die Verteilung des politischen Interesses sind ebenfalls deutlich. Allerdings ist ihre Richtung zweifelhaft. Schließlich war ohnehin schon in der repräsentativen Online-Erhebung ein höheres politisches Interesse im Vergleich zur bevölkerungsrepräsentativen Erhebung zu beobachten. Während dort die Gewichtung zu höherem politischen *Desinteresse* geführt hat, gilt für die repräsentative Online-Erhebung, dass die Gewichtung zu (noch) höherem Interesse führt, so dass die beiden Stichproben sich durch die Gewichtung diesbezüglich noch weiter voneinander entfernen, anstatt sich anzunähern.

Für die Wahlumfrage2002 schließlich hat die Gewichtung deutliche Effekte auf die Stimmenanteile von SPD (+2,2 Prozentpunkte), FDP (-2,7 Prozentpunkte) und PDS (+1,1 Prozentpunkte). Interessant ist dabei der Vergleich zur repräsentativen Online-Erhebung, wo für SPD und FDP genau gegenläufige Effekte der Gewichtung zu beobachten waren. Dagegen ist für die Union, deren Anteil am weitesten neben dem tatsächlichen Wert liegt, erneut praktisch kein Effekt der Gewichtung zu beobachten. Für eine hypothetische Prognose bedeutet dies, dass der durchschnittliche Fehler im Gegensatz zu den ungewichteten Daten um 0,5 Prozentpunkte auf 7,5 Prozentpunkte fallen würde (da die Anteile von SPD und FDP in die richtige Richtung korrigiert werden), der maximale Fehler allerdings (für die Union) auf 15,4 Prozentpunkte steigt. Insgesamt kann das Fazit bezüglich Prognosen damit nur lauten, dass Wahlprognosen auf Basis von ungewichteten oder nur sozialstrukturell gewichteten Online-Erhebungen (zumindest bisher) keinerlei Wert haben. Was schließlich das politische Interesse betrifft, so sind die durch die Gewichtung verursachten Änderungen eher kosmetischer Natur. Die zugrunde liegende Selbstselektion umfasst also nicht nur sozialstrukturelle Größen. Es gilt auch nach Gewichtung weiterhin, dass 75 Prozent der Teilnehmer der Wahlumfrage2002 politisch sehr stark oder stark interessiert sind.

Tabelle 8 Signifikanz der substantiellen Unterschiede zwischen den drei Umfragen, χ^2 und Cramers V (sozialstrukturell gewichtete Daten)

| | Wahlverhalten | | Politisches Interesse | |
|------------------------------------|--|---|---|--|
| | Wahlumfrage 2002 | Repräsentative Online-Umfrage | Wahlumfrage 2002 | Repräsentative Online-Umfrage |
| Repräsentative Bevölkerungsumfrage | $\chi^2_{df=5} = 241$ p<0,001 V=0,09 | $\chi^2_{df=5} = 62$ p<0,001 V=0,16 | $\chi^2_{df=4} = 2224$ p<0,001 V=0,28 | $\chi^2_{df=4} = 135$ p<0,001 V=0,22 |
| Repräsentative Online-Umfrage | $\chi^2_{df=5} = 50$ p<0,001 V=0,04 | | $\chi^2_{df=4} = 617$ p<0,001 V=0,15 | |

Insgesamt, das bestätigt nochmals Tabelle 8, hat die Gewichtung damit keineswegs dazu geführt, dass sich die Ergebnisse der drei Umfragen angenähert haben. Die Unterschiede sind weiterhin signifikant, einige sind zwar etwas kleiner geworden, andere aber auch größer.

5 Schluss

Die Erwartungen an Online-Umfragen sind hoch, können sie aber auch erfüllt werden? Die Ergebnisse des vorliegenden Vergleichs von drei Umfragen sind eher Wasser auf die Mühlen der Skeptiker. Sie untermauern insbesondere die bestehenden Zweifel an Umfragen, die auf Selbstrekrutierung beruhen. Die selbstrekrutierten Teilnehmer der Wahlumfrage 2002 sind eindeutig systematisch verzerrt in Richtung junger, formal hoch gebildeter und politisch interessierter Personen, auch ihr Wahlverhalten unterscheidet sich signifikant vom Wahlverhalten der Teilnehmer der beiden anderen Umfragen. Aber auch zwischen repräsentativer Bevölkerungs- und repräsentativer Online-Umfrage wurden signifikante Unterschiede festgestellt, die auch nach sozialstruktureller Gewichtung (nach Alter und Geschlecht) bestehen blieben.

Damit zu enden, wäre aber zu pessimistisch. Zunächst ist darauf hinzuweisen, dass die im Rahmen dieser Analyse vorgenommene Gewichtung nur auf Alter und Geschlecht (sowie Bundesland) aufbaute. Möglicherweise sind für Online-Umfragen umfangreichere Gewichtungsmaßnahmen erforderlich, möglicherweise müssen Variablen wie das politische Interesse selbst in die Gewichtung einbezogen werden. Harris Interactive etwa, ein Unternehmen, das in den USA Online-Umfragen durchführt, verwendet Propensity Weights, die auf sozialstrukturellen Variablen, aber auch auf Einstellungs- und Verhaltensfragen aufbauen, um die in den Online-Stichproben enthaltenen Verzerrungen zu korrigieren. Das Unternehmen reklamiert für sich, auf Basis von Online-Umfragen und dank der verwendeten Gewichtung den Ausgang der US-Wahlen 2000 – im Gegensatz zu anderen Meinungsforschungsinstituten – korrekt vorausgesagt zu haben (*Harris Interactive* 2000).

Zudem sind im Rahmen dieses Beitrags nur Randverteilungen betrachtet worden. Diese sind natürlich – gerade im Umfeld von Wahlen – per se von Interesse. Allerdings gehört zur Aufgabe der Sozialwissenschaften nicht nur die Beschreibung, sondern auch die Erklärung. Offen ist aber, ob auch die Betrachtung von Zusammenhängen innerhalb der einzelnen Umfragen signifikante und relevante Unterschiede zwischen den Umfragen liefern würde. Ergebnisse aus den USA (*Alvarez, Sherman* und *van Beselaere* 2003) und Großbritannien (*Sanders* et al. 2002) deuten an, dass die in offline- vs. online-erhobenen Daten enthaltenen Zusammenhänge

weniger stark verzerrt sind als Randverteilungen. Weitere Forschungen, auch mit den hier zugrunde liegenden Daten, werden dies bestätigen müssen. Zumindest aus sozialwissenschaftlicher Perspektive wäre dieser Befund aber von großer Wichtigkeit und könnte die aufgrund des Vergleichs der Randverteilungen entstandene Skepsis gegenüber Online-Umfragen deutlich relativieren.

Literatur

Alvarez, R. Michael; Sherman, Robert P.; Van Beselaere, Carla 2003: Subject Acquisition for Web-Based Surveys. In: *Political Analysis*, 11: 23-43.

Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute (ADM) et al. 2001: Standards zur Qualitätssicherung für Online-Befragungen, Frankfurt.

Bandilla, Wolfgang; Bosnjak, Michael 2000: Online Surveys als Herausforderung für die Umfrageforschung: Chancen und Probleme. In: **Mohler, Peter; Lüttinger, Paul** (Hrsg.): *Querschnitt: Festschrift für Max Kaase*, Mannheim, S. 9-28.

Bandilla, Wolfgang; Bosnjak, Michael; Altdorfer, Patrick 2001: Effekte des Erhebungsverfahrens? Ein Vergleich zwischen einer Web-basierten und einer schriftlichen Befragung zum ISSP-Modul Umwelt. In: *ZUMA-Nachrichten* 49, 25: 7-28.

Berrens, Robert P.; Bohara, Alok K.; Jenkins-Smith, H.; Silva, Carlo; Weimer, David L. 2003: The Advent of Internet Surveys for Political Research: A Comparison of Telephone and Internet Samples. In: *Political Analysis*, 11: 1-22.

Couper, Mick P. 2000: Web Surveys: A Review of Issues and Approaches. In: *Public Opinion Quarterly*, 64: 464-494.

Faas, Thorsten 2003: www.wahlumfrage2002.de: Ergebnisse und Analysen. In: *Bamberger Beiträge zur Politikwissenschaft: Forschungsschwerpunkt Politische Einstellungen und Verhalten*, Nr. II-11.

Gibson, Rachel; McAllister, Ian 2002: The Future of National Election Surveys? Evaluating Online Election Surveys in Australia. Konferenzpapier anlässlich des 98. Annual Meeting der American Political Science Association (APSA) in Boston, 26. August bis 1. September 2002, abrufbar unter <http://apsaproceedings.cup.org/Site/abstracts/040/040001GibsonRach.htm>.

Harris Interactive 2000: 2000 Election Winners: George W. Bush ... and Online Polling: Harris Interactive scores unprecedented 99% accuracy in predicting 2000 election outcome, Pressemitteilung abrufbar unter <http://www.harrisinteractive.com/news/allnewsbydate.asp?NewsID=208>.

Hauptmanns, Peter 1999: Grenzen und Chancen von quantitativen Befragungen mit Hilfe des Internet. In: **Batinic, Bernad** et al. (Hrsg.): *Online Research*, Göttingen, S. 21-37.

Koch, Achim 2002: 20 Jahre Feldarbeit im ALLBUS: Ein Blick in die Blackbox. In: *ZUMA-Nachrichten* 51, 26: 9-37.

Krosnick, Jon A.; Chang, LinChiat 2002: A Comparison of the Random Digital Dialing Telephone Survey Methodology with Internet Survey Methodology as Implemented by Knowledge Networks and Harris Interactive. Konferenzpapier anlässlich des 98. Annual Meeting der American Political Science Association (APSA) in Boston, 26. August bis 1. September 2002, abrufbar unter <http://apsaproceedings.cup.org/Site/abstracts/040/040001KrosnickJo.htm>.

Sanders, David; Clarke, Harold; Stewart, Marianne; Whiteley, Paul; Twyman, Joe 2002: The 2001 British Election Study Internet Poll: A Methodological Experiment, Konferenzpapier anlässlich des 98. Annual Meeting der American Political Science Association (APSA) in Boston, 26. August bis 1. September 2002, abrufbar unter <http://apsaproceedings.cup.org/Site/abstracts/040/040001TwymanJoe0.htm>.

Schonlau, M.; Fricker, R.D. Jr.; Elliott, M.N. 2002: *Conducting Research Surveys via E-mail and the Web*. Santa Monica, CA.