

# Von Telefon- zu Online-Erhebungen in bevölkerungsrepräsentativen Umfragen

## Chancen und Herausforderungen des Methodenwechsels im Wissenschaftsbarometer

Bevölkerungsrepräsentative Umfragen sind ein wesentlicher Bestandteil der quantitativen Sozialforschung und helfen dabei, öffentliche Ansichten, Meinungen oder Einstellungen zu erfassen und zu quantifizieren. Auf diese Weise liefern sie wertvolle Daten über gesellschaftliche Entwicklungen und Grundstimmungen in Bezug auf verschiedene Themen wie beispielsweise Politik und Parteipräferenz vor Wahlen, mentale Gesundheit in der Bevölkerung oder Einstellungen zur Wissenschaft.

Für die Erhebung bevölkerungsrepräsentativer Daten wurde noch bis vor wenigen Jahren in der Regel auf Telefonumfragen, sogenannte „Computer Assisted Telephone Interviews“ (CATI) zurückgegriffen (Mohorko et al., 2013). Lösten diese CATI-Umfragen einst die verhältnismäßig kostenintensiven ‚face-to-face‘-Umfragen ab (Nathan, 2001), haben solche Umfragen heutzutage mit der sinkenden Popularität der Festnetztelefonie, sinkenden Antwortraten und einem generell veränderten Kommunikationsverhalten der Bevölkerung zu kämpfen (Berinsky, 2017; Mohorko et al., 2013a; Mohorko et al., 2013b). Mit der Entwicklung des Internets jedoch stellen inzwischen „Computer Assisted Web Interviews“ (CAWI) in Form von Online-Umfragen oder Online Panels wiederum kosteneffektive Alternativen für Telefonumfragen dar (Blom et al., 2017; Mohorko et al., 2013).

Diese Entwicklungen der Datenerhebung betreffen auch das *Wissenschaftsbarometer*, mit dem *Wissenschaft im Dialog* seit 2014 jährlich bevölkerungsrepräsentative Daten zu gesellschaftlichen Einstellungen gegenüber Wissenschaft und Forschung in

Deutschland erhebt. Bislang wurden diese Daten via Telefonumfrage erhoben, wobei auch das *Wissenschaftsbarometer* trotz eines Mobilfunkanteils von 20% in den letzten Jahren zunehmend mit der schlechten telefonischen Erreichbarkeit von Bürger:innen zu kämpfen hatte. Dies hatte eine deutliche Steigerung der Kosten für die Datenerhebung zur Folge. Aus diesem Grund wird das *Wissenschaftsbarometer* ab 2025 via Online-Panel erhoben, um eine bessere Repräsentativität der Befragung für die deutsche Gesamtbevölkerung zu gewährleisten und deren Erhebung gleichzeitig kosteneffizienter zu gestalten.

### Methodik im Wissenschaftsbarometer – damals und heute

Seit der ersten Erhebung des *Wissenschaftsbarometer* 2014 wurden die bevölkerungsrepräsentativen Daten via CATI als Zufallsstichprobe mit einer Stichprobengröße von über 1.000 Fällen nach ADM gezogen. Dabei werden auf Basis einer auf Initiative des Arbeitskreises Deutscher Marktforschungsinstitute (ADM) unter Zugrundelegung des Gabler-Häder-Verfahrens erstellten nationalen Telefonstichprobe zufällig Haushalte gezogen, innerhalb derer wiederum eine Zufallsauswahl der Zielperson stattfindet. Für den 20%igen Mobilfunkanteil, der seit 2017 zusätzlich erhoben wurde, entfiel diese systematische Zielpersonenauswahl. Die Umfrage war zudem in eine Mehrthemenumfrage eingebettet, um auch Personen zu befragen, die sich sonst nicht zu Themen rund um Wissenschaft und Forschung äußern würden. Im Anschluss an die

Erhebung wurde die Stichprobe hinsichtlich der Merkmale Bundesland, Ortsgröße, Geschlechter, Alter, Berufstätigkeit, Schulbildung und Haushaltsgröße gewichtet. Dies gewährleistet, dass die gezogene Stichprobe in ihrer Zusammensetzung der Struktur der Grundgesamtheit – der deutschsprachigen Wohnbevölkerung in deutschen Privathaushalten ab 14 Jahren – entspricht und die Ergebnisse auf diese Grundgesamtheit verallgemeinert werden können, also bevölkerungsrepräsentativ sind.

Genau diese Nachgewichtung ist jedoch in umso höherem Maße erforderlich, je schlechter bestimmte Bevölkerungsgruppen wie etwa Jüngere oder Personen mit formal niedrigerer Bildung via CATI erreicht werden können. Die Folge ist eine immer schlechtere Repräsentativität der Daten für die Grundgesamtheit. Aus diesem Grund wurde ein **Wechsel der Erhebungsmethode von CATI auf CAWI für das Wissenschaftsbarometer** geplant, der 2025 umgesetzt wurde. Hierfür wurde eines der größten Online-Panels in Deutschland ausgewählt: das [Payback Online Access Panel](#).

Dieses vom Erhebungsinstitut *Verian* bereitgestellte Online Panel wurde auf Basis der mehr als 34 Mio. deutschen Payback-Kund\*innen rekrutiert und wird laufend erweitert. Mit einer sorgfältigen Vor-Selektion werden aus diesem Kundenstamm Menschen ins Online-Panel eingeladen, momentan beläuft sich die Zahl der aktiven Panelist\*innen in Deutschland auf über 158.000 (Stand Dezember 2024). Diese können dann zu einer maximal begrenzten Anzahl an Studien im Jahr per E-Mail oder postalisch eingeladen werden, wobei es eine sehr hohe Response-Rate (>50%) gibt. Für die Teilnahme erhalten die Befragten eine Aufwandsentschädigung, welche in Form von Payback-Punkten ausgezahlt wird. In Bezug auf die Soziodemographie unterscheidet sich die Grundgesamtheit der Payback-Kund\*innen nicht von der Grundgesamtheit Deutschlands. Leichte Abweichungen sind lediglich bei Einstellungen zum Thema Datenschutz zu beobachten, welcher den Panelist\*innen etwas weniger wichtig ist.

Bei der CAWI-Umfrage liegt, anders als bei der CATI-Umfrage, keine repräsentative Zufallsstichprobe zugrunde. Für das *Wissenschaftsbarometer* basiert die Stichprobenziehung auf einer quotierten Auswahl: einer zweidimensionalen Quotierung anhand von Alter (in vier Kategorien) und Geschlecht, Schulbildung und Bundesland. Diese soziodemographischen Variablen sind gleichzeitig auch die Grundlage der Gewichtungvariable, die anschließende Gewichtung ist identisch mit der bisherigen bevölkerungsrepräsentativen Gewichtung der CATI-Daten.

Dieser Wechsel der Erhebungsmethode bedeutet methodische und interpretatorische Herausforderungen. Insbesondere können die unterschiedlichen Methoden zur Stichprobenziehung und Datenerhebung bei CATI- und CAWI-Umfragen die resultierenden Daten beeinflussen (Berinsky, 2017; Sarracino et al., 2017). Dies führt zu einer erschwerten Vergleichbarkeit von Daten, vor

allem über längere Zeiträume, wenn diese erst via CATI- und anschließend via CAWI-Umfrage erhoben wurden. So ist es kaum möglich, nach einer Änderung der Erhebungsmethode Aussagen darüber zu treffen, ob Unterschiede in den resultierenden Daten auf die veränderte Methode, oder aber auf tatsächliche Einstellungsänderungen zurückgehen.

Um diesem Problem vorzubeugen, hat das *Wissenschaftsbarometer* in Voraussicht auf den anstehenden Methodenwechsel von 2021 bis 2025 zusätzlich zu der standardmäßigen CATI-Umfrage Referenzdaten via CAWI-Umfragen über das Payback Online Panel erhoben. Die Referenzdaten der Parallel-Erhebungen ermöglichen es, etwaige methodenbedingte Unterschiede zwischen den beiden Verfahren festzustellen, welche beim zukünftigen Vergleich früherer Daten aus den Telefonumfragen mit aktuellen Daten aus den Online-Umfragen berücksichtigt werden können.

## Von CATI zu CAWI – ein notwendiger, aber herausfordernder Schritt

Sogenannte „Computer Assisted Telephone Interviews“ (CATI) wurden in Deutschland in den 1980er Jahren eingeführt (Porst & Schneidel, 1991). Funktionierte diese Methode unter der Verwendung trainierter Interviewer\*innen per Telefon einige Zeit lang sehr gut, sind die Kontakt- und Antwortraten in den letzten Jahren stark gesunken (Berinsky, 2017). So reduzierte sich die Antwortquote bei Umfragen des Pew Research Centers von 36% im Jahr 1997 auf gerade einmal 6% im Jahr 2018 (Kennedy & Hartig, 2019). Neben steigenden finanziellen Aufwand können derart niedrige Antwortquoten die Aussagekraft der Umfrageergebnisse einschränken, wenn die geringe Zahl der Umfrageteilnehmenden sich in ihren Eigenschaften systematisch von Nicht-Teilnehmenden unterscheidet (Berinsky, 2017).

Dies stellt infrage, inwieweit reine Telefonumfragen zur Erhebung bevölkerungsrepräsentativer Daten in Zukunft sinnvoll und gerechtfertigt sind. Eine mögliche Alternative ist die Datenerhebung mittels „Computer Assisted Web Interviews“ (CAWI), welche die Möglichkeit bietet, effektiv und kostengünstig viele Menschen zu erreichen und den Befragten die einfache und selbstständige Durchführung der Umfrage in browserbasierten Fragebögen vorzunehmen. Dies ist Kosten und Ressourcen sparend und hat gleichzeitig eine positive Auswirkung auf die Response Rate (Brick, 2011; Grewenig et al., 2023), denn es bedarf keiner gesonderten Terminfindung. Außerdem sinken organisatorische Fehlerwahrscheinlichkeiten, da keine Datenumwandlung über verschiedene Datenformate mehr nötig ist. Die Antworten verbleiben im computergestützten System und werden in diesem auch ausgewertet – ohne Schwierigkeiten der Übertragung von Papierfragebögen. Darüber hinaus können Gruppenauswahlen bereits vorher getroffen werden und es muss nicht erst nach der Zu-

stimmung durch die potentiell motivierten Befragten festgestellt werden, dass diese nicht in die Studie passen (Grewenig et al., 2023). Gleichzeitig gibt es hier Zweifel an der Repräsentativität der gezogenen Stichproben. Das Fehlen eines Gesamtverzeichnis aller Internetanschlüsse oder sonstiger Identifikationsmöglichkeiten verunmöglicht Stand jetzt eine zufällige Stichprobenziehung, welche alle Deutschen abbildet.

Folglich ergeben sich Unterschiede zu den bisher etablierten Erhebungsmethoden, deren Auswirkungen berücksichtigt werden müssen. Um verschiedene Erhebungsmethoden – in diesem Fall CATI und CAWI – miteinander zu vergleichen, sind vor allem zwei mögliche Effektarten relevant: **Sampling- und Modus-Effekte** (Sarracino et al., 2017). Erstere entstehen aus den Besonderheiten der Stichprobenziehung, also der Auswahl der Befragten, letztere aus den unterschiedlichen Durchführungen von Erhebungen. Da bei dem Methodenwechsel des *Wissenschaftsbarometer* von CATI auf CAWI beide Effektkategorien eine Rolle spielen, werden im Folgenden mögliche Implikationen näher beleuchtet.

### Sampling-Effekte

Sampling-Effekte beschreiben die Auswirkungen der Datengrundlage und der darauf folgenden Stichprobenziehung auf das Forschungsvorhaben. Dabei ist vor allem relevant, wer die gewünschte Zielgruppe darstellt und auf welche Art sich diese Gruppe statistisch akkurat abbilden lässt.

- **Zufallsbasierte vs. gezielte Stichproben:** Die nicht rein zufällige Ziehung bei CAWI-Erhebung ist vermutlich der größte Nachteil im Vergleich zu CATI-Befragungen. Während für CATI-Umfragen in der Vergangenheit eine randomisierte Ziehung aus der deutschen Grundgesamtheit zumindest theoretisch zugrunde lag, ist dies bei CAWI-Umfragen nicht mehr ohne Weiteres möglich. Im Gegensatz zu den bisherigen Interview-Formaten gibt es keine Liste mit Internetanschlüssen oder gar E-Mail-Verzeichnisse aller Deutschen, aus denen gezogen werden kann. Während es Zweifel daran gibt, dass online rekrutierte Samples sich Wahrscheinlichkeitsstichproben annähern können (Brick, 2011) und die Größe der Stichprobenfehler durchaus kritisch gesehen wird (Liedl & Steiber, 2023), konnten Studien unter vergleichbaren Gegebenheiten nachweisen, dass die onlinebasierte Befragung zwar modusbasierte Unterschiede aufwies, diese aber für die Auswertung nahezu keinen Einfluss auf die statistische Nachweisbarkeit der Effekte oder deren Ausrichtung hatten (Sarracino et al., 2017). Auch Grewenig et al. (2023), die wie *Wissenschaft im Dialog* das Payback Online Panel von Verian nutzen, erreichen durch eine Nachgewichtung auf Basis der soziodemographischen Variablen ein Antwortverhalten der Telefon- und der Online-Stichprobe, das – abgesehen von dem gesteuerten

Onlineverhalten der CAWI-Befragten – statistisch nicht mehr voneinander unterschieden werden kann. Das mag auch dadurch bedingt sein, dass durch die push-to-web Rekrutierung, also die postalische Einladung zu der Teilnahme am Payback Online Panel, der mögliche Nachteil von CAWI, ausschließlich besonders aktive Internetnutzende zu erreichen, zusätzlich verkleinert wird. Darüber hinaus wurde bereits bei der Rekrutierung für das Payback Online Panel sowie bei der Einladung für einzelne Umfragen sichergestellt, dass die Payback-Nutzungsdichte sehr breit streut.

- **Unzureichende Haushaltsdurchdringung:** Es haben nicht alle Deutschen einen Zugang zum Internet oder manche verweigern sogar die Nutzung. Diese können daher in der CAWI-Stichprobe nicht berücksichtigt werden (Di Franco, 2024; Grewenig et al., 2023). Während in der Studie von Di Franco (2024a) noch die Rede von 17% Offlinern in Deutschland ist (Stand 2017), zeigen die aktuellen Daten des D21-Digitalindex, dass sich dieser Anteil in 2024 auf 6% verringert hat. Über die Hälfte der Offliner wurde dabei vor 1945 geboren (*D21-Digital-Index 2023/24 Jährliches Lagebild zur Digital Gesellschaft*, 2024).
- **Über- oder Unterrepräsentation bestimmter Bevölkerungsgruppen:** Für die Auswirkungen des Samplings ist die Methodik der Stichprobenziehung wichtig. Dabei können Gruppen über- oder unterrepräsentiert werden. So werden in CATI-Befragungen deutlich häufiger ältere Menschen erreicht. Bezüglich der Bildung gibt es gemischte Evidenz über die Zusammensetzung der Gruppen, welche mit CATI oder CAWI erreicht werden (Sarracino et al., 2017). Bei den ungewichteten CATI-Stichproben der bisherigen *Wissenschaftsbarometer* zeigte sich aber eine deutliche Überrepräsentation formal höher gebildeter sowie auch älterer Personen. Nichtprobabilistische Web-Samples wiederum können Menschen überrepräsentieren, die sich mit dem Thema der Umfrage besonders gut auskennen (Chang & Kroshnick, 2009) bzw. Menschen, die das Internet häufiger nutzen als andere (Grewenig et al., 2023; Yun & Trumbo, 2006). Während ersteres nicht auf das *Wissenschaftsbarometer* zutrifft, da keine gezielte, themenspezifische Rekrutierung erfolgt, sind die Personen des Payback Online Panels potentiell internetaffiner und weniger kritisch mit Blick auf Datenschutzbelange als der Bevölkerungsdurchschnitt.

### Modus-Effekte

Modus-Effekte beschreiben Unterschiede in den Antworten aufgrund verschiedener Interaktionen mit dem Interviewmodus, also der Art und Weise, wie die Studie erhoben wird. Dazu zählen die An- oder Abwesenheit einer interviewenden Person, sowie die Präsentation der Fragen, ob diese also übers Telefon vorgelesen

werden oder aber auf dem eigenen Computer oder Handy beantwortet werden können.

- **Soziale Erwünschtheit:** In Interviewformaten mit Interviewenden (z.B. CATI) führt die Anwesenheit einer befragenden Person oft zu einer höheren sozialen Erwünschtheit. Damit ist die Neigung gemeint, sich selbst in einem positiveren Licht darzustellen und sozial unerwünschte Eigenschaften nur abgeschwächt wiederzugeben. Liedl & Steiber (2023) zeigen, dass vor allem die finanzielle, soziale und gesundheitliche Selbsteinschätzung in Anwesenheit eines Interviewenden nach oben „korrigiert“ wird. Auch beim *Wissenschaftsbarometer* gibt es eine Reihe von Fragen – etwa zum Interesse an Wissenschaft oder zum Grad der Informiertheit über wissenschaftliche Themen – die am Telefon eine soziale Erwünschtheit suggerieren können.
- **Präsentation der Fragen und Antwortmöglichkeiten:** Verglichen mit CATI-Befragungen scheint es bei CAWI-Befragungen eine Tendenz zur häufigeren Auswahl negativer Kategorien zu geben (Sarracino et al., 2017). Eine Meta-Analyse von Ye et al. (2011) deutet hingegen auf eine zusätzliche, modus-relevante Nutzung der positiven Extremkategorien für CATI-Befragungen hin, da telefonisch Befragte tendenziell einen positiven Eindruck bei den Interviewenden hinterlassen wollen, was durch soziale Erwünschtheit zusätzlich verstärkt werden kann. Zusätzlich ist die Unterscheidung zwischen Primacy- und Recency-Tendenzen wichtig. Dabei werden bevorzugt die zuerst genannten Antwortmöglichkeiten (Primacy; potenziell höher in CAWI) oder die zuletzt genannten Antwortalternative (Recency; potenziell höher in CATI) ausgewählt (Sarracino et al., 2017). Besonders bei CAWI-Umfragen kann es außerdem zu Satisficing kommen. Dabei antworten Befragte tendenziell homogen auf Itembatterien, um kognitive Ressourcen zu schonen oder den Zeitaufwand einzuschränken (Chang & Krosnick, 2009; Fricker, 2005; Tourangeau, 2004). Da die meisten Fragen des *Wissenschaftsbarometer* via fünfstufiger Likert-Skala erhoben werden, ist insbesondere die potentiell vom Modus beeinflusste stärkere Nutzung von Extrem- oder Mittelkategorien ein relevanter Effekt. Auch die häufig genutzten Itembatterien, die mehrere themenbezogene Fragen clustern, können potenziell Satisficing begünstigen.
- Um künftige Vergleiche von Einstellungen gegenüber Wissenschaft über die Zeit trotz des Methodenwechsels von Telefon zu Online anstellen zu können, ist es wichtig, abschätzen zu können, inwieweit derartige Modus- und Sampling-Effekte eine Rolle bei der Datenerhebung des *Wissenschaftsbarometer* spielen. Hierfür werden im Folgenden einige Fragen analysiert, die seit 2021 sowohl

via CATI als auch via CAWI erhoben wurden, um potentielle Sampling- und Modus-Effekte in den jeweiligen Erhebungsarten festzustellen. Sollten derartige Effekte feststellbar sein, sind diese bei den Längsschnittbetrachtungen der betreffenden Einstellungen gegenüber Wissenschaft unbedingt zu berücksichtigen.

## Unterschiede zwischen CATI und CAWI beim Wissenschaftsbarometer

Nach der ausschließlich telefonischen Erhebung des *Wissenschaftsbarometer* seit 2014 wurden in Anbetracht der Entwicklungen der Erhebungsmethoden von 2021 bis 2025 Parallelmessungen via CATI und CAWI durchgeführt. Im Rahmen der parallelen Erhebung wurden soziodemographische Variablen sowie Fragen zur Wissenschaftsnähe erhoben, aber auch ausgewählte Fragen des CATI-Fragebogens aus dem jeweiligen Jahr. Dazu zählten vor allem das Interesse an sowie das Vertrauen in Wissenschaft und Forschung, dahingehende Nutzen- und Risiko-Wahrnehmungen als auch Bewertungen des Verhältnisses von Wissenschaft und Politik.

### Sampling-Effekt: Über- bzw. Unterrepräsentation von Bevölkerungsgruppen

Ungewichtet unterscheiden sich die Stichproben der jeweiligen Erhebungsmethoden gerade mit Blick auf soziodemographische Aspekte stark. Während der durchschnittliche CATI-Teilnehmende 59 Jahre alt ist, ist der durchschnittliche CAWI-Befragte mit knapp 50 Jahren deutlich jünger. Auch inferenzstatistisch gibt der Mann-Whitney-U-Test, der bei nicht normalverteilten Daten dieser Art die Effektstärke von Unterschieden aufzeigt, einen mittelschwachen Effekt mit einem  $r$  von 0,25 aus, was auf deutliche Unterschiede in der Verteilung hinweist. Diese Überrepräsentation älterer Menschen in der CATI-Befragung deckt sich mit den oben vorgestellten aktuellen Befunden über die Telefonnutzung in Deutschland.

Besonders stark ist der Unterschied auch bei der formalen Bildung der Befragten, die bei den CATI-Befragungen im Mittel deutlich höher ist. Der Mann-Whitney-U-Test weist hier eine Effektstärke von 0,27 auf. Zwar muss die Verteilung der Werte für Personen mit mittlerer Bildung in beiden Stichproben nur geringfügig durch die Gewichtung angepasst werden, die Werte für CATI-Befragte mit niedriger und insbesondere hoher Bildung bedürfen aber erheblichen Korrekturen. So wird zum Beispiel der Anteil formal höher gebildeter CATI-Befragter im Jahr 2025 von 60% auf 30,3% reduziert, während für die CAWI-Befragten der Anteil von 36,7% auf 37% nur minimal erhöht wird.

Über die Wellen hinweg hat sich zudem ein kleiner, aber doch merklicher Unterschied bzgl. des Geschlechts der Befragten ge-





zeigt: An CATI-Befragungen nehmen grundsätzlich eher Männer als Frauen teil. Dieser aber auch die oben genannten Unterschiede in den soziodemographischen Variablen können durch die Gewichtung weitestgehend ausgeglichen werden (siehe Tabelle 1 im Anhang).

### Sampling-Effekt: Unzureichende Haushaltsdurchdringung

Mit Blick auf die unterschiedliche Erreichbarkeit von On- und Offlinern via CATI bzw. CAWI zeigen die Parallelerhebungen Unterschiede in der ungewichteten Verteilung: So geben CATI-Befragte einerseits häufiger an, sich nie im Internet über Wissenschaft und Forschung zu informieren, gleichzeitig sagen CATI-Befragte auch öfter als CAWI-Befragte, dass sie sich sehr häufig im Internet informieren. Durch die Gewichtung wird dieser Unterschied jedoch reduziert, sodass der Mann-Whitney-U-Test keinen signifikanten Effekt mehr zeigt. Durchschnittlich 40,8% der CATI-Befragten geben an, sich häufig oder sehr häufig im Internet zu informieren, bei den CAWI-Befragten sind es 33,7%. Die über das Payback Online Panel rekrutierten Befragten scheinen also online nicht deutlich aktiver zu sein.

Darüber hinaus ist jedoch auffällig, dass sich die CATI- und CAWI-Stichproben in Bezug auf die Nähe zur Wissenschaft nachhaltig unterscheiden – selbst nach der Gewichtung. Während bei der gewichteten CATI-Stichprobe im Schnitt 15% angeben, aktuell in Wissenschaft und Forschung zu arbeiten oder dies in der Vergangenheit getan zu haben, sind es in der CAWI-Stichprobe durchschnittlich 11%. Ein noch deutlicherer Unterschied tritt bei den Befragten zutage, die dies nicht taten, aber angaben, eine\*n Wissenschaftler\*in persönlich zu kennen: Während CATI-Befrag-

te dies im Schnitt zu 35% bejahten, traf dies für CAWI-Befragte nur zu 18% zu.

### Modus-Effekt: Präsentation der Fragen und Antwortmöglichkeiten

Auch bei anderen Fragen werden trotz Gewichtung Unterschiede im Antwortverhalten erkennbar. Diese Unterschiede können auf den verschiedenen, oben genannten Modus-Effekten (Tendenzen zu Extrem- bzw. Mittelkategorie, Primacy- bzw. Recency-Effekten, Satisficing) basieren, teilweise aber auch auf soziale Erwünschtheit zurückzuführen sein. Durch die potentielle Überlagerung modusspezifischer Effekte ist diese Auftrennung schwierig, daher werden Einstellungen, bei denen primär soziale Erwünschtheit auftreten könnte, im Nachfolgenden gesondert betrachtet.

Nach der Gewichtung der Daten gibt es zwischen den meisten Variablen keine signifikanten Unterschiede. Für die Variablen, bei denen es Unterschiede gibt, bewegen diese sich in den meisten Fällen bei einer Effektstärke von unter 0,10, was sehr kleinen Effekten entspricht (siehe Tabelle 2 im Anhang). Trotz dieser schwachen Effekte ist auffällig, dass CATI-Befragten die meisten Fragen deutlich häufiger mit der äußersten Zustimmungskategorie beantworten (z. B. „stimme voll und ganz zu“), während CAWI-Befragte öfter von gemäßigten Zustimmungskategorien Gebrauch machen (z. B. „stimme eher zu“). Diese Tendenz der CATI-Befragten, die äußerste Kategorie zu nennen (siehe Abbildung 1), kann sowohl auf soziale Erwünschtheit (bei Nennung positiver Extremkategorie) als auch durch den Recency-Effekt (bei Nennung negative Kategorie zuletzt genannt) zurückzuführen sein.

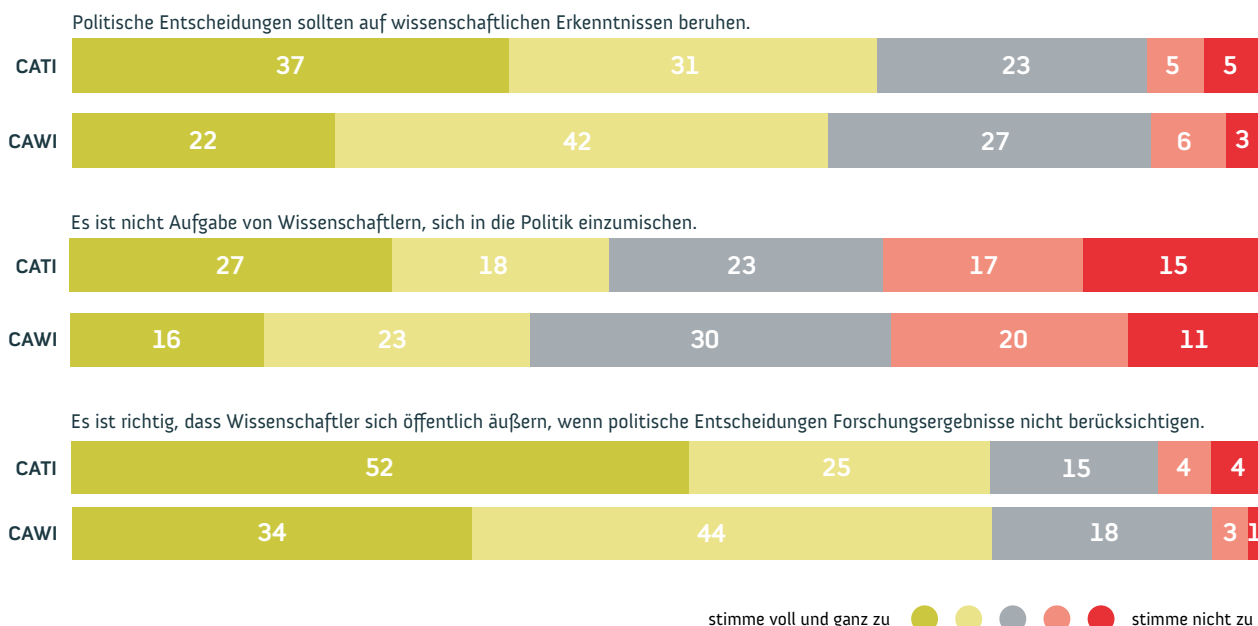


Abbildung 1: Gemittelte Zustimmung zu Fragen bzgl. des Verhältnisses von Wissenschaft von Politik von 2021 bis 2025 (in %)

Bei einer aggregierten Betrachtung der beiden Zustimmungskategorien (Top2) ist dieser Unterschied in den meisten Fällen aber nicht mehr erkennbar. Auch wird erkennbar, dass CAWI-Befragte häufiger die Mittelkategorie auswählen. Dies könnte ein Hinweis auf Satisficing sein.

Eine Ausnahme von diesem Befund stellen die Aussagen zum Verhältnis von Wissenschaft und Politik dar, bei denen über alle Wellen hinweg CATI-Befragte eine leicht stärkere Zustimmung zeigen. Auch hier ist der Effekt mit 0,05 bis 0,12 nur schwach, die Konsistenz dieses Unterschieds ist aber auffällig.

### Modus-Effekt: Soziale Erwünschtheit

Für die Fragen nach dem Interesse an Wissenschaft, Politik und lokalen Nachrichten sowie der Informiertheit über Wissenschaft zeigt sich ebenfalls: CATI-Befragte geben häufiger die extrem positive Kategorie an. Zusätzlich zu den oben genannten Modus-Effekten ist bei dieser Art von Fragen die soziale Erwünschtheit potentiell besonders groß.

Der größte Unterschied im Antwortverhalten zwischen CATI und CAWI wird jedoch bei der Aussage „Ich persönlich profitiere von Wissenschaft und Forschung.“ erkennbar. CATI-Befragte stimmten dieser Aussage deutlich öfter zu, die Effektstärken sind hier mit einem  $r$  von bis zu 0,13 im Jahr 2025 am höchsten – wenngleich qua Definition immer noch schwach. Doch dass die verstärkte Zustimmung zu dieser Aussage bei den CATI-Befragten

nicht nur bei der äußerten Zustimmungskategorie erkennbar ist, sondern auch bei der aggregierten Betrachtung beider Zustimmungskategorien (Top2), ist eine nennenswerte Besonderheit.

Betrachtet man alle Parallelerhebungen von 2021 bis 2025, zeigt sich, dass sowohl beobachtete Unterschiede als auch Ähnlichkeiten beim Antwortverhalten über alle Jahre hinweg konsistent auftreten. Dies spricht zum einen dafür, dass die oben beschriebenen Unterschiede für einzelne Fragen tatsächlich durch die Erhebungsmethode bedingt scheinen. Zum anderen lässt sich für die meisten anderen Fragen feststellen, dass auch bei einer Parallelerhebung über mehrere Jahre hinweg keine nennenswerten Unterschiede im Antwortverhalten auftreten, weshalb vor allem für diese Fragen – wie bspw. das Vertrauen – ein Vergleich der kommenden CAWI-Daten mit den früheren CATI-Daten weiterhin möglich ist (siehe Abbildung 2).

## Diskussion

Über den Zeitraum von 2021 bis 2024 konnte festgestellt werden, dass sowohl CATI- als auch CAWI-Erhebungen überwiegend ähnliche Ergebnisse produzieren. Zwar lassen sich Unterschiede zwischen beiden Erhebungsmethoden feststellen, die sich einerseits auf die gezogene Stichprobe (Sampling-Effekte) beziehen und andererseits auf das Antwortverhalten, welches möglicherweise durch die andersartige Befragungsart verursacht wird (Modus-Ef-

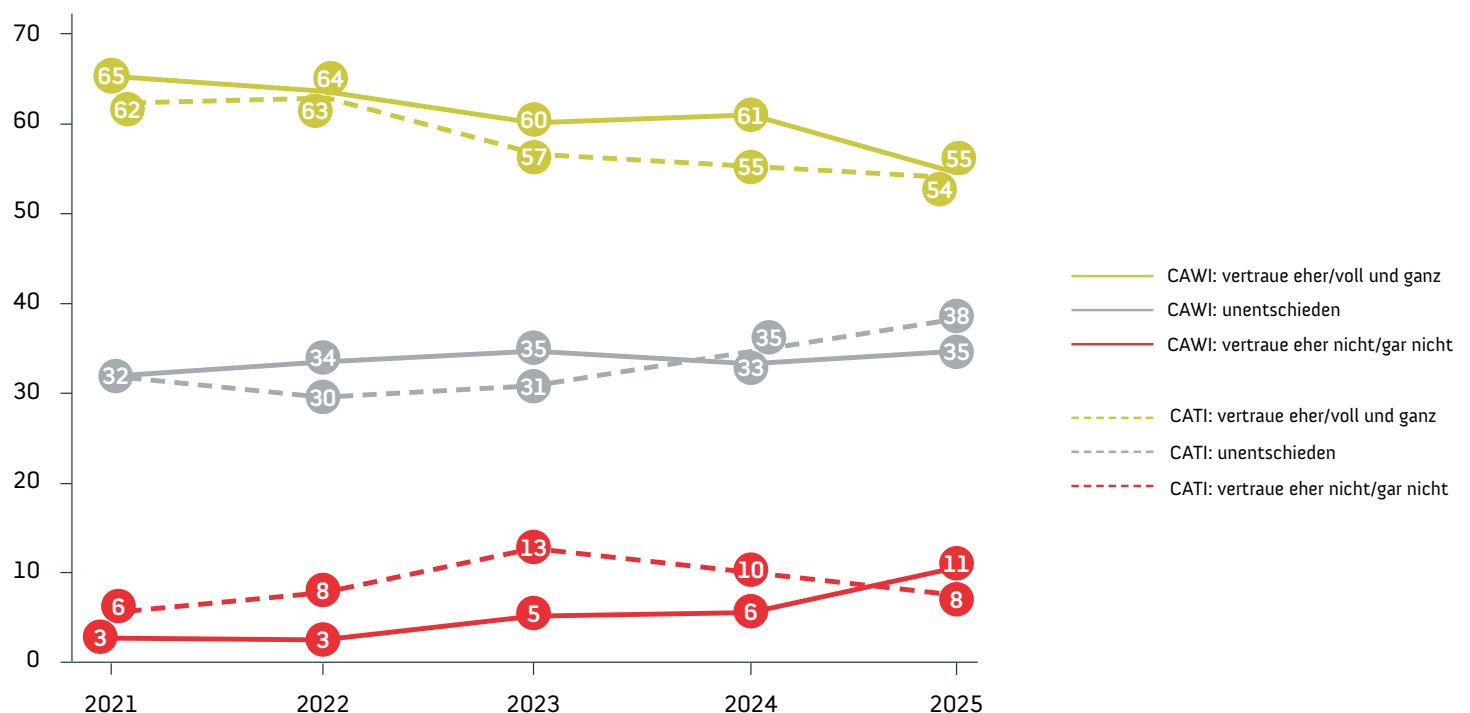


Abbildung 2: Vertrauen in Wissenschaft bei CAWI- und CATI-Befragten von 2021 bis 2025 (in %)

fekte). Dank der nachträglichen Gewichtung der Stichproben und aufgrund überwiegend schwacher Effektstärken bei den vereinzelt beobachteten Unterschieden im Antwortverhalten bleibt ein Vergleich der Daten des *Wissenschaftsbarometer* vor 2025 (CATI) und ab 2025 (CAWI) jedoch größtenteils weiterhin möglich. Trotz dieser nur schwachen Effektstärken sollten die hier identifizierten Abweichungen zwischen CATI und CAWI aber dennoch berücksichtigt werden:

- **Verstärkte Nutzung von Extremkategorien vs. Mittelkategorien:** Bei fast allen Fragen des *Wissenschaftsbarometer* zeigt sich eine deutlich stärkere Tendenz der CATI-Befragten, die Extremkategorien – z. B. „stimme voll und ganz zu“ – auszuwählen. CAWI-Befragte dagegen machen oft stärkeren Gebrauch von der Mittelkategorie. Bei einer reinen Betrachtung des Mittelwerts oder der aggregierten Zustimmung (Top2) ist dieser Unterschied zwischen CATI und CAWI zwar kaum noch erkennbar, trotzdem ist dies eine wichtige Erkenntnis vor allem für deskriptive Betrachtungen des Antwortverhaltens vor und nach 2025.
- **Divergierendes Antwortverhalten bei bestimmten Fragen:** Am auffälligsten ist die unterschiedliche Zustimmung zu der Aussage „Ich persönlich profitiere von Wissenschaft und Forschung“, die bei CAWI-Befragten deutlich niedriger ausfällt. Auch zeigen CAWI-Befragte eine etwas geringere Zustimmung bei Fragen zum Verhältnis von Wissenschaft und Politik. Dies verkompliziert zum einen Längsschnittbetrachtungen zu diesen Fragen, zum anderen ist bei Querschnittsbetrachtungen zu berücksichtigen, dass die Erhebungsmethode scheinbar einen Einfluss auf die Messung dahingehender Einstellungen hat.
- **Leicht geringere Nähe zur Wissenschaft bei CAWI-Befragten:** Trotz der Korrektur durch die Nachgewichtung ist der Anteil an Befragten, die angeben, in Wissenschaft und Forschung zu arbeiten oder jemanden zu kennen, der dies tut, in der CATI-Stichprobe höher als in der CAWI-Stichprobe. Zwar spiegelt sich diese geringe Nähe zur Wissenschaft in der CAWI-Stichprobe nicht wieder in bspw. einem geringeren Vertrauen in Wissenschaft und Forschung oder anderweitig negativerer Bewertung von Wissenschaft. Doch gerade bei Auswertungen, die die Nähe zur Wissenschaft als Faktor mit einbeziehen, muss dieser Unterschied zwischen den CAWI- und CATI-Daten beachtet werden.

Diese Unterschiede gilt es – trotz der geringen Effektstärken – zu berücksichtigen, will man künftige CAWI-Daten des *Wissenschaftsbarometer* mit den bisherigen CATI-Daten vergleichen. Gerade bei Längsschnittbetrachtungen künftiger *Wissenschafts-*

*barometer* zu diesen Aspekten wird auch in den kommenden Jahren immer wieder auf die identifizierten Unterschiede hingewiesen, sodass sie transparent sind. Zudem wird der in 2025 erfolgte Wechsel der Erhebungsmethode bei solchen Längsschnittbetrachtungen stets kenntlich gemacht.

Die Auswertung der parallel erhobenen CATI- und CAWI-Daten von 2021 bis 2025 hat jedoch auch gezeigt: Für fast alle erhobenen Einstellungen lassen sich bei der aggregierten Betrachtung der Zustimmung zu den untersuchten Fragen, des Vertrauens in Wissenschaft und Forschung und des Interesses daran keine nennenswerten Unterschiede zwischen CATI und CAWI feststellen. Mit Blick auf die im Payback Online Access Panel deutlich bessere Response-Rate von über 50%, die wesentlich gleichmäßiger verteilte ungewichtete Stichprobe hinsichtlich Alter und formaler Bildung sowie die erheblich gesteigerte Kosteneffizienz computergestützter Web-Interviews spricht dies klar für einen Wechsel der Erhebungsmethode des *Wissenschaftsbarometer* von CATI zu CAWI.

Mit künftigen Erhebungen via Online-Panel können so nicht nur Längsschnittbetrachtungen für die meisten in der Vergangenheit bereits erhobenen Einstellungen aufrecht erhalten werden, auch kann mit dieser Erhebungsmethode in Zukunft ein realistischer Stimmungsbild der deutschen Bevölkerung hinsichtlich ihrer Einstellungen gegenüber Wissenschaft und Forschung gezeichnet werden.

## Autor\*innen

Sascha Schneider

Bastian Kremer

Liliann Fischer

Das *Wissenschaftsbarometer* 2025 wird gefördert und unterstützt von



Klaus Tschira  
Stiftung



Fraunhofer

## Zitation

Schneider, Sascha; Kremer, Bastian; Fischer, Liliann (2025): *Von Telefon- zu Online-Erhebungen in bevölkerungsrepräsentativen Umfragen. Chancen und Herausforderungen des Methodenwechsels im Wissenschaftsbarometer*. Berlin: Wissenschaft im Dialog gGmbH.

## Literaturverzeichnis

- Allen, J. & Schmich, P. (2017). *Methodische Probleme von Mixed-Mode-Ansätzen in der Umfrageforschung. Gesundheit in Deutschland aktuell (GEDA)*. Wiesbaden, 161-182.
- Ansolabehere, S., & Schaffner, B. F. (2014). Does Survey Mode Still Matter? Findings from a 2010 Multi-Mode Comparison. *Political Analysis*, 22(3), 285–303. <https://doi.org/10.1093/pan/mpt025>
- Berinsky, A. J. (2017). Measuring Public Opinion with Surveys. *Annual Review of Political Science*, 20(1), 309–329. <https://doi.org/10.1146/annurev-polisci-101513-113724>
- Blom, A. G., Herzing, J. M. E., Cornesse, C., Sakshaug, J. W., Krieger, U., & Bossert, D. (2017). Does the Recruitment of Offline Households Increase the Sample Representativeness of Probability-Based Online Panels? Evidence From the German Internet Panel. *Social Science Computer Review*, 35(4), 498–520. <https://doi.org/10.1177/0894439316651584>
- Braunsberger, K., Wybenga, H., & Gates, R. (2007). A comparison of reliability between telephone and web-based surveys. *Journal of Business Research*, 60(7), Article 7. <https://doi.org/10.1016/j.jbusres.2007.02.015>
- Brick, J. M. (2011). The Future of Survey Sampling. *Public Opinion Quarterly*, 75(5), 872–888. <https://doi.org/10.1093/poq/nfr045>
- Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone interviewing versus the Internet. *Public Opinion Quarterly*, 73(4), 641–678. <https://doi.org/10.1093/poq/nfp075>
- D21-Digital-Index 2023/24 Jährliches Lagebild zur Digital Gesellschaft (1. Auflage). (2024). Initiative D21.
- Di Franco, G. (2024). The return of non-probability sample: The electoral polls at the time of internet and social media. *Quality & Quantity*, 58(4), 3811–3830. <https://doi.org/10.1007/s11135-024-01835-8>
- Dillman, D. A. (2020). Three decades of advancing survey methodology. In T. W. Smith (Hrsg.), *A meeting place and more...: A history of the American Association for Public Opinion Research* (S. 95–116).
- Fricker, S. (2005). An Experimental Comparison of Web and Telephone Surveys. *Public Opinion Quarterly*, 69(3), 370–392. <https://doi.org/10.1093/poq/nfi027>
- Grewenig, E., Lergetporer, P., Simon, L., Werner, K., & Woessmann, L. (2023). Can internet surveys represent the entire population? A practitioners' analysis. *European Journal of Political Economy*, 78, 102382. <https://doi.org/10.1016/j.ejpoleco.2023.102382>
- Kennedy, C., & Hartig, H. (2019, Februar 27). Response rates in telephone surveys have resumed their decline. *Pew Research Center*. <https://www.pewresearch.org/short-reads/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/>
- Kohler, U., & Kreuter, F. (2008). *Datenanalyse mit Stata: Allgemeine Konzepte der Datenanalyse und ihre praktische Anwendung* (3., aktualisierte und überarb. Aufl.). Oldenbourg.
- Liedl, B., & Steiber, N. (2023). Führen Online-Befragungen zu anderen Ergebnissen als persönliche Interviews? Eine Schätzung von Moduseffekten am Beispiel eines Mixed-Mode Surveys. *Österreichische Zeitschrift für Soziologie*, 49(1), 1–22. <https://doi.org/10.1007/s11614-023-00532-4>
- Mohorko, A., De Leeuw, E., & Hox, J. (2013). *Coverage Bias in European Telephone Surveys: Developments of Landline and Mobile Phone Coverage across Countries and over Time*. <https://doi.org/10.13094/SMIF-2013-00002>
- Mohorko, A., Leeuw, E. D., & Hox, J. (2013). Internet Coverage and Coverage Bias in Europe: Developments Across Countries and Over Time. *Journal of Official Statistics*, 29(4), 609–622. <https://doi.org/10.2478/jos-2013-0042>
- Nathan, G. (2001). Telesurvey methodologies for household surveys: A review and some thoughts for the future. *Survey Methodology*, 27(1), 7–31.
- Porst, R., & Schneidel, M. (1991). *Software-Anforderungen an computergestützte Befragungssysteme*.
- Sarracino, F., Riillo, C. F. A., & Mikucka, M. (2017). Comparability of web and telephone survey modes for the measurement of subjective well-being. *Survey Research Methods*, Vol 11, 141–169 Pages. <https://doi.org/10.18148/SRM/2017.V11I2.6740>
- Tourangeau, R. (2004). Survey Research and Societal Change. *Annual Review of Psychology*, 55(1), 775–801. <https://doi.org/10.1146/annurev.psych.55.090902.142040>
- Yun, G. W., & Trumbo, C. W. (2006). Comparative Response to a Survey Executed by Post, E-mail, & Web Form. *Journal of Computer-Mediated Communication*, 6(1), 0–0. <https://doi.org/10.1111/j.1083-6101.2000.tb00112.x>



## Anhang

Tabelle 1: Gewichtete und ungewichtete (in Klammern) Verteilung der Soziodemographie bei CATI- und CAWI-Befragten

Soziodemographische Verteilung (in %), ungewichtete Anteile in Klammern		Alter					formale Bildung			Geschlecht	
Welle	Erhebungsmethode	14–29	30–39	40–49	50–59	60+	niedrig	mittel	hoch	weiblich	männlich
2025	CAWI	19,3 (15,5)	15,1 (17,3)	14,2 (13,7)	16,9 (20,7)	34,5 (32,8)	33,2 (33,3)	28,8 (30,1)	37 (36,7)	51 (51,5)	49 (48,5)
	CATI	17,7 (4,9)	12 (5,1)	13,6 (9,3)	19,7 (16,4)	37 (64,3)	40,4 (13,2)	29,3 (26,8)	30,3 (60)	53,3 (47,5)	46,7 (52,5)
2024	CAWI	19,6 (20,2)	12,9 (13,1)	16,2 (15,9)	17,5 (14,4)	33,8 (36,5)	27,9 (34,9)	33,4 (31)	38,7 (34,1)	51 (51,1)	49 (48,9)
	CATI	19,8 (6,2)	14,9 (6,3)	14 (12,7)	18,6 (22,1)	32,7 (52,7)	34,1 (13,2)	31,1 (25,5)	34,7 (61,3)	51 (49,7)	49 (50,3)
2023	CAWI	19,3 (19,8)	13,2 (13,1)	15,6 (15,8)	18,1 (16,3)	33,8 (35,1)	30,9 (34,9)	32,1 (31)	37 (34,1)	51 (51,1)	49 (48,9)
	CATI	16 (8,1)	12,7 (8,3)	14 (11,1)	18,6 (17,6)	38,6 (55)	32,6 (13,5)	32,3 (25,2)	35,1 (61,2)	52,5 (47,8)	47,5 (52,2)
2022	CAWI	19,5 (20)	12,6 (12,5)	16,3 (16,1)	18,4 (14,7)	33,3 (36,6)	31,4 (35,4)	32 (30,8)	36,7 (33,9)	51 (51,2)	49 (48,4)
	CATI	20,4 (8,2)	15,5 (6,1)	13,3 (9,5)	18,7 (19,6)	32,7 (56,7)	33,5 (15,6)	31 (27,1)	35,4 (57,3)	49,9 (51,1)	50,1 (49,9)
2021	CAWI	19,9 (19,4)	12,3 (12,6)	16,6 (16,6)	18,5 (15,1)	32,7 (36,3)	34,1 (34,4)	31,1 (31,3)	34,7 (34,2)	51 (51,5)	49 (48,5)
	CATI	17,4 (9,3)	12,5 (9,3)	14,9 (13,1)	19,6 (23,4)	35,7 (45)	36,9 (13,3)	29,1 (27,3)	34 (59,3)	52,4 (49,1)	47,6 (50,9)

Tabelle 2: Effektstärken der Unterschiede zwischen gewichteten CATI- und CAWI-Daten

Parallel via CATI und CAWI erhobene Fragen	2021	2022	2023	2024	2025
Subjektives Interesse an Themen: Politik	0,07**				
Subjektives Interesse an Themen: Wissenschaft und Forschung	0,07**	0,05*			
Subjektives Interesse an Themen: Lokales aus meiner Umgebung	0,07**	0,06**			
Arbeiten Sie selbst in Wissenschaft und Forschung?	0,07**	0,06*	0,09***	0,07**	
Kennen Sie einen Wissenschaftler oder eine Wissenschaftlerin persönlich?	0,18***	0,26***	0,2***	0,15***	
Ich persönlich profitiere von Wissenschaft und Forschung.	0,08***			0,10***	0,13***
Grund für Vertrauen in Wissenschaftler: Weil Wissenschaftler Experten auf ihrem Feld sind.					0,04*
Grund für Vertrauen: Weil Wissenschaftler nach Regeln und Standards arbeiten.	0,06*				0,08***
Grund für Misstrauen in Wissenschaftler: Weil Wissenschaftler häufig Fehler machen.	0,05*		0,04		
Grund für Misstrauen: Weil Wissenschaftler oft Ergebnisse ihren eigenen Erwartungen anpassen.					0,04*
Politische Entscheidungen sollten auf wissenschaftlichen Erkenntnissen beruhen.	0,09***	0,05*	0,06*		0,07***
Es ist nicht Aufgabe von Wissenschaftlern, sich in die Politik einzumischen.		0,06*			
Es ist richtig, dass Wissenschaftler sich öffentlich äußern, wenn politische Entscheidungen Forschungsergebnisse nicht berücksichtigen.	0,12***	0,07***	0,05*		

r-Werte des Mann-Whitney-U-Tests, Kursiv: Cramer's V;  
nur signifikante Werte dargestellt;  
Grau: Frage wurde in dem Jahr nicht erhoben;  
Signifikanzniveaus: \*  $p < 0,05$ , \*\*  $p < 0,01$ , \*\*\*  $p < 0,001$