

Olsen, Jerome; van Hoorde, Joël; Varade, Neha; Yzern, Baptiste

Article

Mode-Analysen zur Zeitverwendungserhebung 2022

WISTA - Wirtschaft und Statistik

Provided in Cooperation with:

Statistisches Bundesamt (Destatis), Wiesbaden

Suggested Citation: Olsen, Jerome; van Hoorde, Joël; Varade, Neha; Yzern, Baptiste (2025) : Mode-Analysen zur Zeitverwendungserhebung 2022, WISTA - Wirtschaft und Statistik, ISSN 1619-2907, Statistisches Bundesamt (Destatis), Wiesbaden, Vol. 77, Iss. 5, pp. 22-33

This Version is available at:

<https://hdl.handle.net/10419/330762>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Dr. Jerome Olsen

ist Psychologe und promovierte an der Universität Wien. Anschließend war er als Senior Research Fellow am Max-Planck-Institut zur Erforschung von Gemeinschaftsgütern tätig und wechselte 2021 zum Statistischen Bundesamt. Seit 2022 leitet er das Referat „Freiwillige Haushaltserhebungen – Methodik“.

Joël Van Hoorde

ist Soziologe und arbeitet als Referent im Referat „Freiwillige Haushaltserhebungen – Methodik“ des Statistischen Bundesamtes. Seit 2022 betreut er als Product Owner die Umsetzung der IT-Projekte rund um die in den tagebuchgestützten Haushaltserhebungen eingesetzten Apps.

Neha Varade

studiert Volkswirtschaftslehre an der Universität Bonn und war im Zeitraum Mai bis Juli 2025 als Praktikantin im Referat „Freiwillige Haushaltserhebungen – Methodik“ des Statistischen Bundesamtes tätig.

Baptiste Yzern

studiert Statistik und Sozialwissenschaften an der ENS de Lyon. Für sein Abschlusspraktikum war er von April bis Juli 2025 als Praktikant im Referat „Freiwillige Haushaltserhebungen – Methodik“ des Statistischen Bundesamtes tätig. Seit September 2025 wird er als künftiger INSEE-Mitarbeiter an der ENSAE Paris ausgebildet.

MODE-ANALYSEN ZUR ZEIT- VERWENDUNGSERHEBUNG 2022

Jerome Olsen, Joël Van Hoorde, Neha Varade, Baptiste Yzern

➤ **Schlüsselwörter:** Zeitverwendungserhebung – ZVE – Mode-Effekte – Digitalisierung

ZUSAMMENFASSUNG

Erstmals kam eine mobile App als ergänzender Erhebungsmodus in der Zeitverwendungserhebung (ZVE) 2022 zum Einsatz. Ziel dieses Beitrags ist, potenzielle Unterschiede zwischen App- und Papiererhebung im Hinblick auf Teilnahmebereitschaft, Quotenerfüllung, Ausfallraten und Datenqualität zu analysieren. Ein Schwerpunkt liegt auf der Frage, inwiefern beobachtete Unterschiede auf Messeffekte oder auf Selbstselektionseffekte zurückzuführen sind. Hierzu wird ein Matching-Verfahren eingesetzt, um Vergleichsgruppen mit hoher statistischer Ähnlichkeit herzustellen. Die Ergebnisse liefern wichtige methodische Erkenntnisse sowie Ansatzpunkte für Verbesserungen der ZVE und betonen die zentrale Bedeutung eines Mixed-Mode-Ansatzes in der amtlichen Statistik.

➤ **Keywords:** time use survey – TUS – mode effects – digitalisation

ABSTRACT

The 2022 Time Use Survey (TUS) was the first to incorporate a mobile app as an additional data collection mode. The aim of this article is to analyse potential differences between app-based and paper-based data collection in terms of willingness to participate, quota attainment, dropout rates and data quality. A central focus is to determine the extent to which observed differences are attributable to measurement effects or self-selection effects. For this purpose, a matching technique is applied to create comparable groups with a high degree of statistical similarity. The findings provide valuable methodological insights, show possible ways to improve the TUS, and emphasise the importance of a mixed-mode approach in official statistics.

1

Einleitung

Die Zeitverwendungserhebung (ZVE) ist eine freiwillige Befragung von Haushalten in Deutschland, die die Statistischen Ämter des Bundes und der Länder alle zehn Jahre als dezentrale Statistik durchführen. Mit den Daten kann untersucht werden, wie verschiedene Bevölkerungsgruppen in Deutschland ihre Zeit nutzen. In der ZVE wird die Verteilung der Zeit auf Aktivitäten wie Erwerbsarbeit, Haushaltsaufgaben und Freizeit tagebuchgestützt detailliert erfasst. Besonderes Interesse gilt dem Umfang der „Care-Arbeit“, das heißt der unbezahlten Arbeit wie Hausarbeit, Kinderbetreuung, Ehrenamt und freiwilliges Engagement. In den teilnehmenden Haushalten führen alle Personen ab einem Alter von 10 Jahren für drei Tage (zwei Wochentage und ein Wochenende) ein Zeittagebuch. Zusätzlich werden personenbezogene und haushaltsbezogene Informationen mittels Fragebogen erhoben. Die erhobenen Daten sind von großer Bedeutung für die Analyse von gesellschaftlichen Strukturen und dienen als Grundlage für politische Entscheidungen, insbesondere im Bereich der Familien-, Arbeitsmarkt- und Gleichstellungspolitik. Anhand der ZVE-Ergebnisse lassen sich beispielsweise konkrete gesellschaftliche Herausforderungen quantifizieren, wie etwa der Gender Care Gap¹ (43,4 % im Jahr 2022) oder das Phänomen der Einsamkeit, wonach sich jede sechste Person oft einsam fühlt.²

Erstmals wurde für die ZVE 2022 ein digitales Erhebungsinstrument in Form einer mobilen App und Web-App (im Folgenden kurz „App“) eingeführt (Bitzer und andere, 2022). Die App „Zeitverwendung“ stand den Teilnehmenden über Googles Play Store und Apples App Store sowie als Webanwendung über alle gängigen Browser zur Verfügung. Sie basierte auf der Anwendung „Modular Online Time Use Survey“ (MOTUS), die von der belgischen Firma hbits CV entwickelt wird. Ziel war, den Erhebungsprozess zu modernisieren und die Teilnahme für die Haushalte zu erleichtern und attraktiver zu gestalten,

um der sinkenden Teilnahmebereitschaft in den freiwilligen Haushaltserhebungen entgegenzuwirken. Durch die Nutzung der App konnten die Teilnehmenden zudem ihre Tagebucheinträge bequem und zeitnah von unterwegs vornehmen, was Erinnerungslücken minimieren soll. Die Teilnahme über Papierunterlagen blieb weiterhin möglich, jedoch wurde im Rahmen einer Online-first-Strategie die App als bevorzugter Weg beworben.

Zunächst blicken die Kapitel 2 und 3 auf den Forschungsstand und die Durchführung der ZVE 2022. Danach stellt Kapitel 4 die mode-vergleichenden Analysen vor, welche nach Abschluss der Datenaufbereitung zur Bewertung des App-Einsatzes durchgeführt wurden. Hier liegt der Fokus vor allem auf Teilnehmeraten, Quotenerfüllung, Abbruchraten und Datenqualität. Zudem werden im Kapitel 5 die Vorteile eines modernen digitalen Erhebungsinstruments für die Feldarbeit einer dezentralen Statistik sowie für die Statistikproduktion im Gesamten aufgezeigt.

2

Aktueller Stand der Forschung

Bisherige Studien zum Vergleich papierbasierter und digitaler Erhebungsmethoden haben uneinheitliche Befunde bezüglich der Unterschiede festgestellt. Während einige Untersuchungen zeigen, dass digitale Methoden, insbesondere mobile Apps, durch geringere Raten fehlender Werte, reduzierte Messfehler und zeitnähere Datenerfassung zu einer höheren Datenqualität beitragen können (Stone und andere, 2003; Marcano Belisario und andere, 2015), fanden andere Arbeiten nur geringe oder keine signifikanten Qualitätsunterschiede zwischen den Modes (Muehlhausen und andere, 2015; Hohwü und andere, 2013). Allerdings ist bekannt, dass Unterschiede zwischen Erhebungsmodes nicht nur aus tatsächlichen Messeffekten resultieren, sondern häufig mit Selektionsprozessen vermischt sind. Mehrere methodische Untersuchungen haben versucht, diese komplexe Vermischung von Messeffekten und Selektionseffekten mithilfe statistischer Modellierungen (zum Beispiel Panel-Designs, Propensity-Score-Methoden oder Instrumentalvariablen-Ansätzen) zu entwirren (Lutig und andere, 2011; Vannieuwenhuyze und andere, 2010; Vannieuwenhuyze/Loosveldt, 2013), was trotz methodischer Fortschritte anspruchsvoll bleibt.

1 Der Gender Care Gap beschreibt auf Basis der Daten der Zeitverwendungserhebung den Unterschied zwischen dem durchschnittlichen Zeitaufwand für unbezahlte Arbeit von Frauen und Männern in Prozent. Dabei wird die Differenz beim Zeitaufwand für unbezahlte Arbeit von Frauen und Männern ins Verhältnis zum Zeitaufwand für unbezahlte Arbeit der Männer gesetzt.

2 Siehe www.zve2022.de für weitere Ergebnisse.

Zudem wird in der Literatur wiederholt betont, dass die Vorteile digitaler Methoden von Kontextbedingungen wie Nutzerfreundlichkeit, Zielgruppenaffinität und Erhebungsdesign abhängig sind (Dennison und andere, 2013; Byrom und andere, 2019). Insbesondere wenn ältere Personen die Zielgruppen bilden oder bei Bevölkerungsumfragen können Papierfragebogen weiterhin hohe Rücklaufquoten und Datenqualität gewährleisten (Kelfve und andere, 2020; Janssen, 2006; Khare, 2016). Insgesamt verdeutlicht die bisherige Forschung, dass Unterschiede zwischen Papier- und digitalen Modes sowohl messmethodischer Natur sein können als auch stark durch Selektionsmechanismen beeinflusst werden.

3

Durchführung der Zeitverwendungserhebung 2022

Die ZVE wird als Quotenstichprobe umgesetzt. Der Stichprobenplan wird hierfür entlang von zwei verschiedenen Quotenmerkmalen (Haushaltstyp und soziale Stellung) sowie einem Richtmerkmal (Gemeindegrößenklasse) erstellt. Die Teilnahme an der Erhebung ist freiwillig und Haushalte werden über verschiedene Anwerbemaßnahmen und aktive Rekrutierung sowie finanzielle Anreize zur Teilnahme ermutigt. Der Erhebungsmode (Papier oder App) wird durch die Haushalte bei der Erklärung ihrer Teilnahmebereitschaft selbst ausgewählt.¹³

Die Einführung einer App führte im Kontext der Werbemaßnahmen zur Möglichkeit, Haushalte über neue Kanäle anzuwerben. Verstärkte Werbung über soziale Medien (beispielsweise Instagram) und der prominente Hinweis auf eine vereinfachte Teilnahme per App führten dazu, dass vor allem neue Haushalte¹⁴ und jene mit einer jüngeren Altersstruktur eher per App teilnahmen. Konkret stammten 64 % der Haushalte, die per Papier teilnahmen, aus traditionellen Werbewegen¹⁵, während 67 % der App-Haushalte über neue Werbe-

quellen¹⁶ gewonnen wurden. Wie bereits anhand der Literatur dargestellt wurde, ist somit davon auszugehen, dass auch den Daten der ZVE 2022 ein starker Selektionseffekt zugrunde liegt. Der Mode ist somit mit weiteren – sowohl beobachteten als auch unbeobachteten – Drittvariablen konfundiert (Bethlehem, 2010; Lugtig und andere, 2016; Vannieuwenhuijze/Loosveldt, 2013). Diese Konfundierung erschwert es methodisch, valide Aussagen darüber zu treffen, in welchem Maße Unterschiede zwischen den Erhebungsmodes tatsächlich auf Messeffekte oder auf durch die Modewahl verursachte Selektionsprozesse zurückzuführen sind (Lugtig und andere, 2011; Vannieuwenhuijze und andere, 2010).

4

Mode-Analysen

Der empirische Teil gliedert sich in zwei Abschnitte: Zunächst werden Analysen zur Teilnahmebereitschaft und Quotenerfüllung dargestellt, gefolgt von Analysen zu Mode-Unterschieden. Zur Teilnahmebereitschaft werden die Teilnahmeraten nach Erhebungsmode für verschiedene Gruppen, die Quotenerfüllung sowie die Abbruchraten untersucht. Zu Mode-Effekten wurden die Anzahl der Aktivitätenwechsel im Tagebuch sowie die Struktur von Tagesverläufen zwischen den beiden Erhebungsmodes verglichen.¹⁷

4.1 Teilnahmebereitschaft und Quotenerfüllung

Insgesamt haben 9 764 Haushalte erfolgreich an der ZVE 2022 teilgenommen. Dem zugrunde liegen Daten von 19 407 Individuen mit einem ausgefüllten und verwertbaren Zeittagebuch. Hinsichtlich des Modes haben 5 178 Personen per Papier und 14 229 per App teilgenommen, was einer App-Rate auf Personenebene von 74,5 % entspricht. Somit ist eine hohe Akzeptanz der App als Erhebungsmode zu beobachten.

3 Für eine Übersicht zur Durchführung der ZVE 2022 siehe auch Bank und andere (2025).

4 Die Stichproben der freiwilligen Haushaltserhebungen bestehen teilweise aus Haushalten, die bereits an einer vorherigen freiwilligen Haushaltserhebung teilgenommen haben.

5 Hierzu gehören vor allem postalische Anschreiben an Haushalte, die an anderen amtlichen Erhebungen teilgenommen haben.

6 Hierzu gehören vor allem Maßnahmen wie Pressemitteilungen oder Internetwerbung über die sozialen Medien und weitere Kanäle.

7 Alle quantitativen Analysen basieren auf den ungewichteten Auswertungsdaten nach vollständiger Datenaufbereitung.

Als Ausdruck spezifischer Reichweiten- oder Akzeptanzunterschiede eines digitalen Meldewegs wurden Unterschiede in der App-Rate im Hinblick auf verschiedene Merkmale untersucht. Analysiert wurden dabei die beiden Quotenmerkmale (Haushaltstyp und soziale Stellung), das Richtmerkmal Gemeindegrößenklasse sowie verschiedene soziodemografische Variablen wie Bildung, Alter und Geschlecht. Im Folgenden sind exemplarisch die Merkmale mit den größten Differenzen dargestellt.

Quotenmerkmale

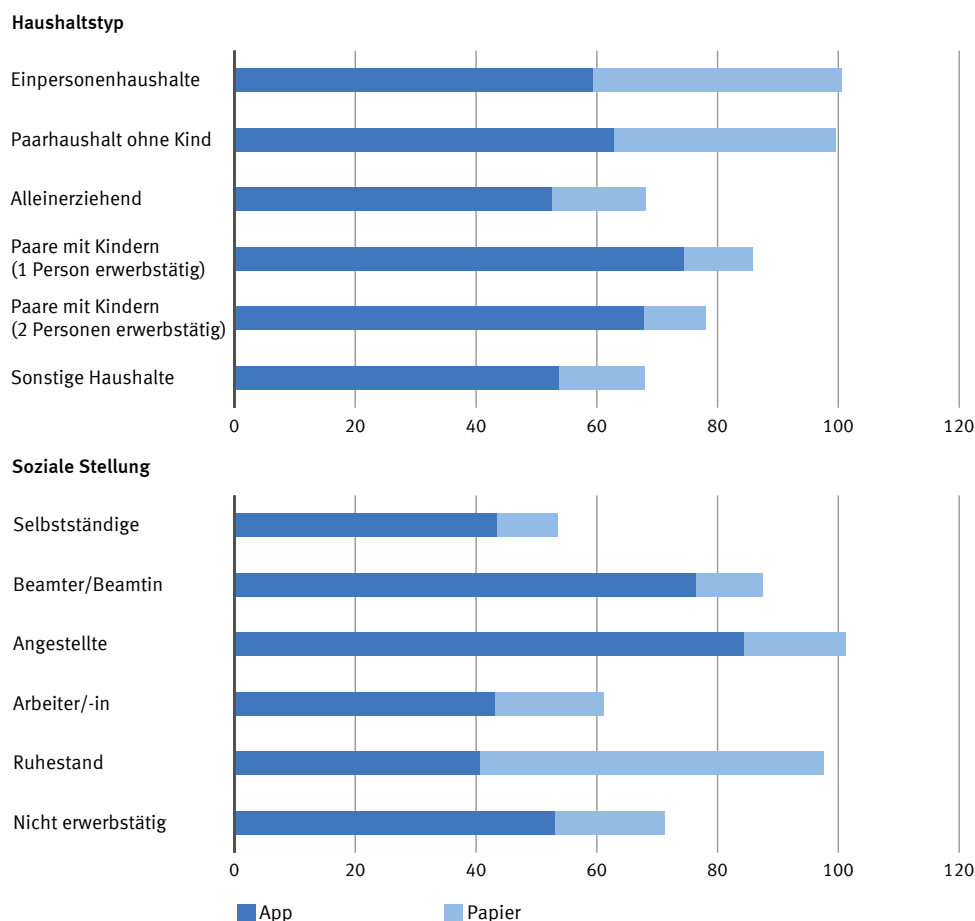
Das Quotenmerkmal Haushaltstyp unterscheidet sechs verschiedene Arten von Haushalten.⁸ Der obere Teil von [Grafik 1](#) zeigt die prozentuale Quotenerfüllung je Haushaltstyp, differenziert nach Erhebungsmode. Es fällt auf, dass die Papierquote insbesondere bei Einpersonenhaushalten sowie Paarhaushalten ohne Kind hoch ist. Demgegenüber weisen die traditionell schwerer für eine Teilnahme zu gewinnenden Haushalte – vor allem Haushalte mit Kindern sowie sonstige Haushalte⁹ –

8 Das sind (1) Einpersonenhaushalte, (2) Paarhaushalte ohne Kind, (3) Alleinerziehende, (4) Paare mit Kindern und einer erwerbstätigen Person, (5) Paare mit Kindern und zwei erwerbstätigen Personen sowie (6) sonstige Haushalte.

9 Hierzu gehören beispielsweise Mehrgenerationenhaushalte und Haushalte mit erwachsenen Kindern.

Grafik 1

Zeitverwendungserhebung 2022: Quotenerfüllung und Mode-Anteile nach Haushaltstyp und sozialer Stellung
in %



jeweils besonders hohe App-Quoten auf. Insgesamt erreichen Einpersonenhaushalte und Paare ohne Kind höhere Werte bei der Quotenerfüllung. Gleichwohl stellt sich die Frage, ob die Teilnahmebereitschaft bei den herausfordernden Haushaltstypen ohne das digitale Angebot ähnlich erfolgreich gewesen wäre; angesichts der hohen App-Quote erscheint dies eher unwahrscheinlich. Ein Pearson-Chi-Quadrat-Test bestätigt den Zusammenhang: Die Verteilung der Erhebungsmodes unterscheidet sich signifikant zwischen den Haushaltstypen, $\chi^2(5, N = 9\,729) = 674,05, p < 0,001, V = 0,26$. Dies deutet darauf hin, dass der bevorzugte Mode eng mit dem Haushaltstyp verknüpft ist.

Auch beim Quotenmerkmal Soziale Stellung, das sechs Ausprägungen¹⁰ differenziert, zeigen sich deutliche Unterschiede. Grafik 1 stellt im unteren Teil für das Quotenmerkmal Soziale Stellung die Quotenerfüllung sowie den relativen Anteil der Erhebungsmodes je Ausprägung dar. Besonders auffallend ist die hohe Papierquote bei Personen im Ruhestand (58 %), was zugleich die hohe Papierquote bei den zuvor beschriebenen Einpersonenhaushalten und Paaren ohne Kind teilweise erklärt. Zugleich wird deutlich, dass unterschiedliche Modeverteilungen zu einer erfolgreichen Quotenerfüllung führen können. So erreichen Personen im Ruhestand mit einer Papierquote von 58 % nahezu vollständig die vorgegebenen Quoten, während Angestellte mit einer deutlich niedrigeren Papierquote von nur 17 % die Quotenvorgaben ebenfalls vollständig erfüllen. Dieses Ergebnis unterstreicht, dass das Anbieten beider Erhebungsmodes („Mode-Mix“) eine effektive Strategie darstellt, um unterschiedliche Zielgruppen zu erreichen. Bestimmte Gruppen nutzen trotz digitaler Optionen weiterhin bevorzugt Papierunterlagen. Auch hier bestätigt der Pearson-Chi-Quadrat-Test einen signifikanten Unterschied zwischen den Erhebungsmodes in Bezug auf die soziale Stellung, $\chi^2(5, N = 9\,729) = 1.388,20, p < 0,001, V = 0,38$.

Soziodemografie

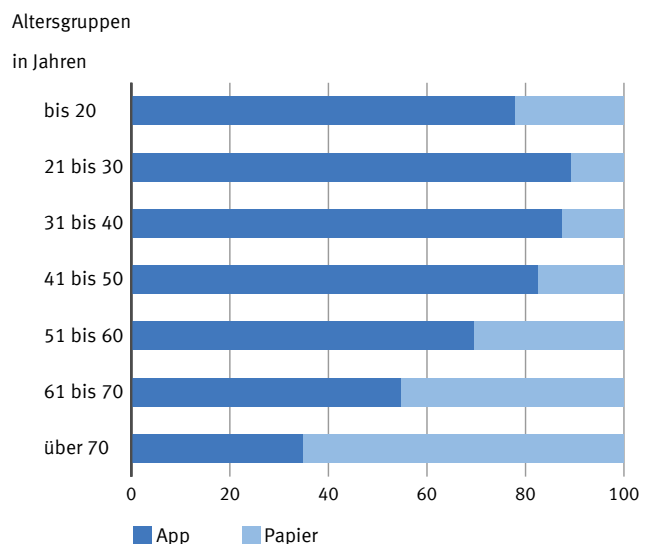
Mit Blick auf die Soziodemografie unterscheiden sich die App-Raten besonders zwischen den Altersgruppen der Teilnehmenden. Es zeigt sich ein deutli-

cher Zusammenhang zwischen Alter und Modewahl, $\chi^2(6, N = 19\,407) = 2\,514,80, p < 0,001, V = 0,36$. Jüngere Personen nutzten deutlich häufiger die App (App-Quoten um 90 % bei 21- bis 40-Jährigen), während ältere Personen eher Papier bevorzugten (nur 35 % App-Nutzung in der Altersgruppe ab 70 Jahren). Die Auswertung erfolgte auf Personenebene. Dabei ist zu berücksichtigen, dass der gewählte Mode in der Regel für alle Haushaltsmitglieder verbindlich war und von der Person festgelegt wurde, die den Haushalt anmeldete („Haushaltsvorstand“). Daher haben Kinder die Mode-Präferenz der Eltern übernommen, was die etwas abweichende Quote in der jüngsten Altersgruppe erklärt. Ohne diese Abhängigkeit wäre der App-Anteil bei jüngeren Personen womöglich höher.

➤ Grafik 2

Grafik 2

Zeitverwendungserhebung 2022: Mode-Anteile nach Altersgruppen in %



Auch beim Geschlecht und nach Bildungsstand zeigten sich Unterschiede in den App-Raten, die jedoch etwas weniger stark ausgeprägt waren als beim Alter. Zusammengefasst wählten Männer etwas häufiger die App als Frauen und formal höher gebildete Personen entschieden sich ebenfalls etwas häufiger für die digitale Teilnahme.

10 Das sind (1) Selbstständige, (2) Beamtinnen/Beamte, (3) Angestellte, (4) Arbeiter/-innen, (5) Personen im Ruhestand sowie (6) Nicht-erwerbstätige.

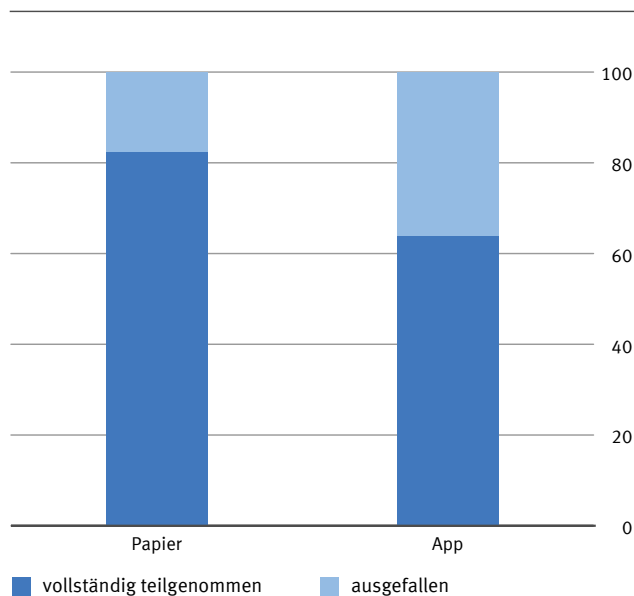
Ausfallraten

Insgesamt wurden über das Jahr 2022 (einschließlich Nachbesetzungen von Ausfällen) 14 573 Haushalte in die Stichprobe gezogen. Davon wollten 11 724 per App und 2 849 mittels Papier teilnehmen.¹¹ Im Jahresverlauf sind im Mode App 36 % der Haushalte ausgefallen, dagegen im Mode Papier mit 18 % deutlich weniger, $\chi^2(1, N = 14\,573) = 358,14, p < 0,001, V = 0,16$. Ungeklärt auf Basis der Erhebungsdaten bleibt die Frage, wie viele der Haushalte, die die App als Erhebungsmode ausgewählt haben, überhaupt aktiv mit der Erhebung begonnen haben. Ein erheblicher Teil wird möglicherweise nie aktiv mit der Teilnahme begonnen haben.¹² Es kann somit nicht quantifiziert werden, ob und wie viele dieser Haushalte aufgrund der Interaktion mit der App oder aus anderen Gründen trotz Teilnahmewunsch die Erhebung nicht erfolgreich abgeschlossen haben.

↘ Grafik 3

Grafik 3

Zeitverwendungserhebung 2022: Teilnahmeausfälle von Haushalten nach Erhebungsmode in %



11 Leichte Abweichungen in den Zahlen der Haushalte zu bisherigen Angaben im Text sind damit zu erklären, dass diese Auswertung aus einem Export der Fachanwendung generiert wurde und nicht auf Basis der aufbereiteten Erhebungsdaten.

12 Vorläufige Analysen zur Durchführung der Einkommens- und Verbrauchsstichprobe (EVS) 2023 zeigen, dass etwa 50 % der im Mode App ausgefallenen Haushalte sich gar nicht erst in die App einloggen. Nach Abschluss der Datenaufbereitung zur EVS 2023 soll zum Thema Mode-Unterschiede ein weiterer Artikel in dieser Zeitschrift folgen.

4.2 Mode-Unterschiede

Analysestrategie

Zur Untersuchung von Mode-Effekten wurden zwei Indikatoren betrachtet. Als erster Indikator wurde die Anzahl von Aktivitätenwechseln im Zeittagebuch analysiert, also die Anzahl eingetragener Aktivitäten im Tagesverlauf zwischen den beiden Erhebungsmodes verglichen. Grundsätzlich kann bei tagebuchgestützten Erhebungen davon ausgegangen werden, dass der Messfehler asymmetrische Eigenschaften aufweist: Teilnehmende vergessen eher, eine Aktivität zu dokumentieren, als dass sie eine Aktivität erfinden. Unter der Annahme eines gegebenen Messfehlers ist daher von einer Tendenz zur Unterschätzung der tatsächlichen Aktivität auszugehen. Je mehr eingetragene Aktivitäten im Tagebuch vorliegen, desto höher sollte im Durchschnitt auch die Qualität der Daten sein. Als zweiter Indikator wurde die inhaltliche Struktur der Tagesaktivitäten untersucht. Die Messung der Aktivitäten erfolgt in Intervallen von 10 Minuten, wodurch ein Tag aus 144 Zeittakten besteht, die unterschiedlichen inhaltlichen Kategorien zugeordnet sein können. Es wurde analysiert, inwieweit sich diese Zuordnungen zwischen den Erhebungsmodes unterscheiden.

Wie bereits erläutert ist bei der Bewertung potenzieller Mode-Effekte die Vermischung von Mess- und Selektionseffekten ein zentrales methodisches Problem (Bethlehem, 2010; Vannieuwenhuyze/Loosveldt, 2013). Da Teilnehmende den Erhebungsmode selbst gewählt haben, ist nicht auszuschließen, dass beobachtete Unterschiede teilweise auf systematische Unterschiede in der Zusammensetzung der Gruppen (zum Beispiel Alter, Bildung, Technikaffinität) zurückzuführen sind und nicht ausschließlich auf Messeffekte des Instruments. Zur statistischen Kontrolle dieser Selektionsprozesse wurde daher ein Propensity-Score-Matching auf Personenebene durchgeführt. Dieses Verfahren ermöglicht es, Teilnehmende aus beiden Gruppen auf Basis relevanter Kovariaten (zum Beispiel Alter, Geschlecht, Bildung, Haushaltstyp) einander zuzuordnen, um Unterschiede zwischen den Modes unter möglichst vergleichbaren Bedingungen zu schätzen (Rosenbaum/Rubin, 1983). Das Matching erhöht die Vergleichbarkeit, jedoch können Effekte unbeobachteter Variablen auch hierbei nicht vollständig ausgeschlossen werden.

Konkret wird beim Propensity-Score-Matching für alle Papier-Teilnehmenden – die hier kleinere Vergleichsgruppe – in der Gruppe der App-Teilnehmenden jeweils diejenige Person gesucht, die hinsichtlich einer Vielzahl relevanter Merkmale möglichst ähnlich ist. Für dieses Vorgehen wird zunächst für alle Fälle ein Propensity Score geschätzt, der die Wahrscheinlichkeit der Teilnahme per App in Abhängigkeit der Kovariaten beschreibt (Rosenbaum/Rubin, 1983). Anschließend erfolgt die Auswahl der am besten passenden Vergleichsfälle anhand des minimalen Abstands im Propensity Score. Dabei kann es vorkommen, dass für einzelne Papier-Teilnehmende keine passende Vergleichsgröße unter den App-Teilnehmenden gefunden wird, sodass auch diese Fälle aus der Analyse ausscheiden. Der resultierende Datensatz ist daher in der Anzahl der Beobachtungen reduziert, zeichnet sich aber durch eine bessere Vergleichbarkeit der beiden Gruppen aus.


Zur weiteren Verbesserung der Passgenauigkeit wurde das Propensity-Score-Matching um die Mahalanobis-Distanz erweitert. Dieses Verfahren berücksichtigt zusätzlich den multivariaten Abstand zwischen den Kovariatenvektoren und gewichtet Unterschiede entlang Variablen mit geringerer Streuung stärker (Rubin, 1980; Stuart, 2010). So können auch bei ähnlichen Propensity Scores noch feiner abgestimmte Paare gebildet werden. Die Kombination beider Ansätze – Propensity-Score-Matching in Verbindung mit der Mahalanobis-Distanz – wird in der methodischen Literatur als bewährte Strategie empfohlen, um die Qualität des Matchings zu optimieren (Gu/Rosenbaum, 1993; Stuart, 2010).

Im Folgenden werden die beiden genannten Indikatoren – die Anzahl der Aktivitätenwechsel sowie die inhaltliche Struktur der Aktivitäten im Tagesverlauf – betrachtet. Es werden stets die Unterschiede zwischen den App- und den Papier-Teilnehmenden in der Gesamtstichprobe neben den Ergebnissen des Vergleichs der beiden Matching-Gruppen gezeigt. Auf diese Weise lässt sich prüfen, in welchem Umfang sich die beobachteten Differenzen zwischen den Erhebungsmodes auch nach statistischer Kontrolle potenzieller Selektionseffekte zeigen.

Ergebnisse

Die beiden Ausgangsgruppen in der Gesamtstichprobe ($N = 19\,407$) umfassten $n = 5\,178$ Personen im Mode Papier und $n = 14\,229$ Teilnehmende in der App. Nach

dem Matching ($N = 7\,652$) reduzierten sich diese Gruppen auf jeweils $n = 3\,826$ statistische Paare aus den beiden Mode-Gruppen. Für 1 352 Papier-Teilnehmende konnte kein adäquater statistischer Zwilling unter den App-Teilnehmenden identifiziert werden. Insgesamt wurden somit 74 % der per Papier teilnehmenden Personen einer per App teilnehmenden Person zugeordnet. Bei der Konstruktion der beiden Gruppen wurden insgesamt 21 Kovariaten berücksichtigt.¹³

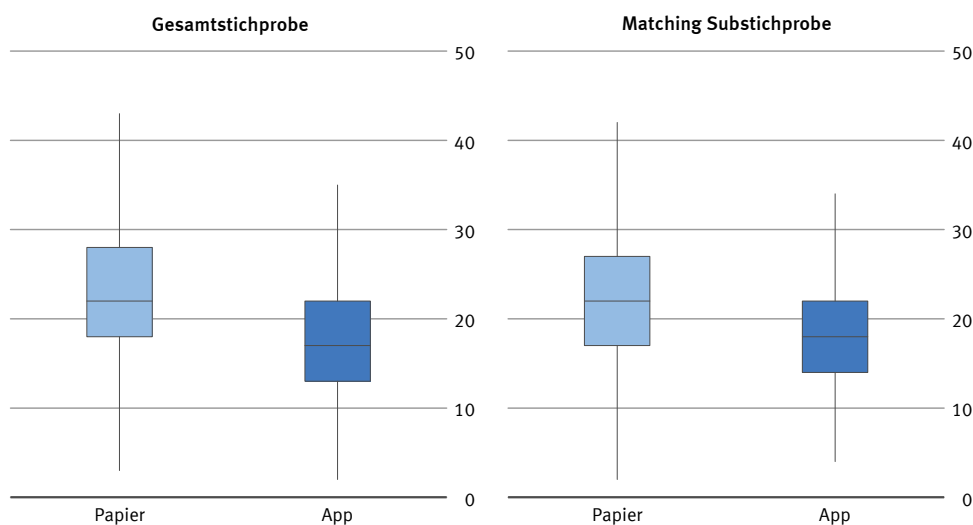
Bezüglich der Anzahl von Aktivitätenwechseln im Zeit-tagebuch zeigt  **Grafik 4** die Unterschiede zwischen den Erhebungsmodes in der Gesamtstichprobe sowie in der Substichprobe nach dem Matching. In der Gesamtstichprobe verzeichneten Papier-Teilnehmende erkennbar mehr Aktivitätenwechsel ($Mdn = 22$) als App-Teilnehmende ($Mdn = 17$). Ein Wilcoxon-Rangsummentest ergab, dass dieser Unterschied statistisch signifikant war, $W = 53\,743\,865$, $p < .001$, $r = .35$. Die Effektstärke deutet auf einen moderaten Effekt hin. Nach dem Matching blieb der Unterschied bestehen, $W = 10\,051\,732$, $p < .001$, $r = .32$. Die Größenordnung war damit nur leicht reduziert (Papier: $Mdn = 22$; App: $Mdn = 18$). Auch die Effektstärke ist weiterhin als moderat zu interpretieren. Zusammenfassend zeigt sich, dass der Unterschied in der Anzahl der Aktivitätenwechsel zwischen den Modes auch nach statistischer Anpassung bestehen bleibt und weiterhin von substanzieller Größe ist. Dies spricht dafür, dass der Effekt zumindest teilweise auf messmethodische Unterschiede und nicht ausschließlich auf Selektionseffekte zurückzuführen ist.

Der zweite Qualitätsindikator betrifft die inhaltliche Struktur der Tagesaktivitäten, also wie sich die Tagesverläufe zwischen den beiden Erhebungsmodes unterscheiden. Alle erfassten Aktivitäten wurden anhand einer Aktivitätenliste vereinheitlicht. Diese differenzierte zwischen 174 Aktivitäten, welche hierarchisch geordnet waren. Auf höchster Ebene gab es dabei neun Hauptkategorien, die im Folgenden betrachtet werden.

13 Auflistung der berücksichtigten Variablen: Alter, Lebenssituation, Erwerbstätigkeit, Ausmaß der Erwerbstätigkeit, spitzes Bruttoeinkommen, gruppiertes Bruttoeinkommen, berufliche Stellung, Haushaltsgröße, Familienstand, Nationalität, Region, Haushaltstyp, soziale Stellung der Haupteinkommensperson, Art der Anwerbung, Häufigkeit Internetaktivität für Kommunikation, Häufigkeit Internetaktivität für Ansehen von Filmen/Serien/Videos, Häufigkeit Internetaktivität für Sonstiges, Besitz von Smartphones oder Tablets, Besitz von Computern, Anzahl von Smartphones oder Tablets, Computeranzahl.

Grafik 4

Zeitverwendungserhebung 2022: Aktivitätenwechsel nach Erhebungsmode und Stichprobe
Anzahl

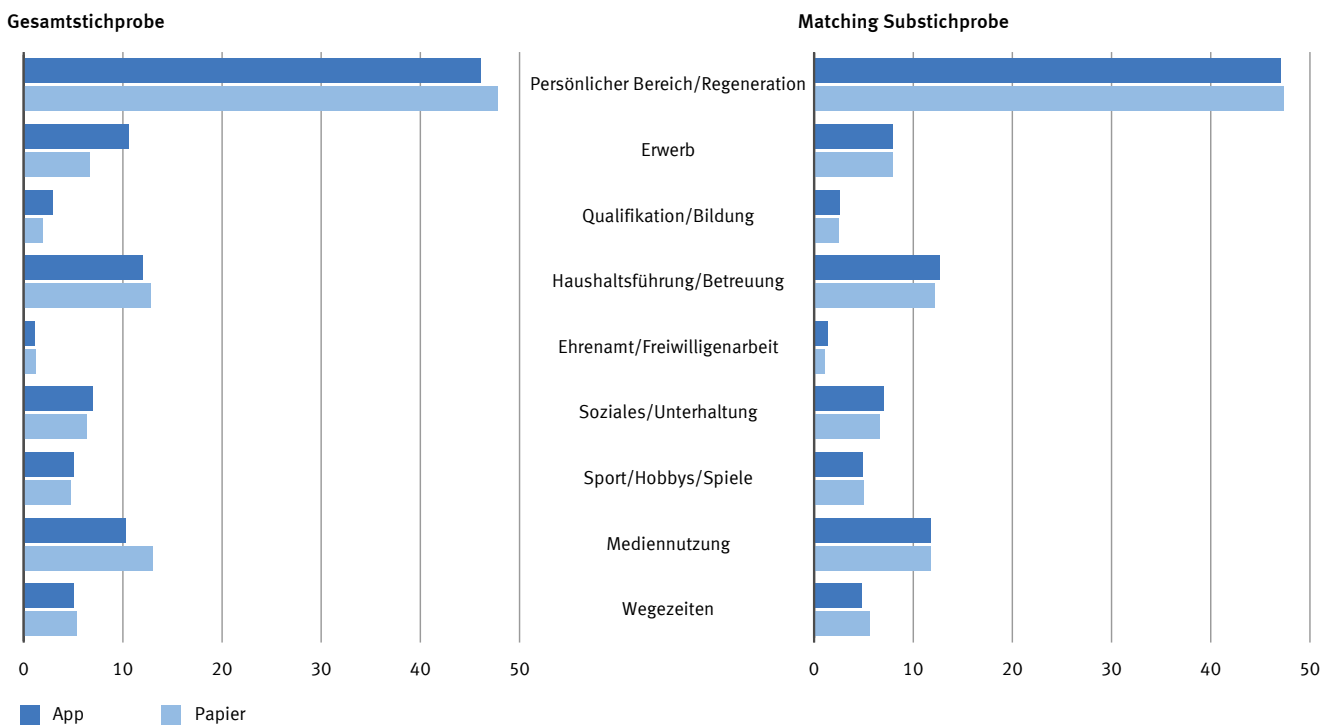


➤ Grafik 5 zeigt die relativen Anteile der einzelnen Hauptkategorien an der Gesamtzeitdokumentation, ge-

trennt nach Erhebungsmode und Stichprobe. In der Gesamtstichprobe sind deutliche Unterschiede sicht-

Grafik 5

Zeitverwendungserhebung 2022: Dokumentation der Tagesaktivitäten im Tagesverlauf nach Hauptkategorien
in %



bar: So fällt beispielsweise der Anteil der Erwerbstätigkeit in der App-Gruppe höher aus (10,6%) als in der Papiergruppe (6,6%), während bei der Mediennutzung der umgekehrte Trend zu beobachten ist (App: 10,3%, Papier: 13,0%). Auch in weiteren Kategorien zeigen sich kleinere Abweichungen.

Nach dem Matching gleichen sich die Strukturen der Tagesaktivitäten jedoch deutlich an. Hier liegen die relativen Anteile in allen Hauptkategorien sehr nah beieinander, beispielsweise bei der Haushaltsführung und Betreuung der Familie (Papier: 12,2%, App: 12,6%) oder der Mediennutzung (Papier: 11,8%, App: 11,8%). Zwar sprechen die Ergebnisse zur Anzahl der Aktivitätenwechsel dafür, dass Messeffekte bestehen, diese scheinen jedoch in letzter Konsequenz keinen Einfluss auf die aggregierte Tagesstruktur auf Ebene der Hauptkategorien zu haben.

5

Schlussfolgerung

Die vorliegenden Analysen haben es ermöglicht, den erstmaligen Einsatz einer App zur Datenerhebung in der Zeitverwendungserhebung 2022 umfassend zu untersuchen und zentrale methodische Erkenntnisse zu Unterschieden zwischen app- und papierbasierten Erhebungsmodes herauszustellen. Die Ergebnisse zeigen zunächst eine insgesamt hohe Akzeptanz der App. Ein Großteil der Teilnehmenden entschied sich für die digitale Teilnahme, insbesondere jüngere sowie über neue Werbequellen gewonnene Haushalte. Gleichzeitig wurde deutlich, dass die Ausfallquoten bei App-Teilnehmenden etwa doppelt so hoch waren wie bei Papier-Teilnehmenden. Es spricht vieles dafür, dass ein erheblicher Anteil dieser Ausfälle weniger auf Probleme im Nutzungserlebnis der App als vielmehr auf Selektionsprozesse in Verbindung mit der Werbung und den Erwartungen neu rekrutierter Haushalte zurückzuführen ist.

Mit Blick auf die Reichweite und Quotenerfüllung unterstreichen die Ergebnisse die Bedeutung eines Mode-Mix. Unterschiedliche Bevölkerungsgruppen präferieren klar unterschiedliche Meldewege, sodass eine breite Repräsentation (besonders bei einer freiwilligen Quotenstichprobe) durch die Kombination aus Papier und App besser gelingt. Die gezielte Ansprache über neue

Kommunikationskanäle und die Bereitstellung einer modernen digitalen Teilnahmeoption scheinen dabei die Reichweite insbesondere bei Haushalten mit Kindern und sonstigen schwer zu rekrutierenden Gruppen erhöht zu haben. Gleichzeitig zeigt sich in der Zusammensetzung der Teilnehmenden, dass der Erhebungsmode eng mit soziodemografischen Faktoren verknüpft ist. Dies spricht für einen kontextsensiblen Einsatz verschiedener Modes.

Die Qualität der erhobenen Tagebuchdaten wurde anhand der Anzahl der Aktivitätenwechsel sowie der inhaltlichen Tagesstrukturen bewertet. Papier-Teilnehmende berichteten im Median eine höhere Zahl an Aktivitätenwechseln, was auf ein detaillierteres Antwortverhalten schließen lässt. Dieser Unterschied blieb auch nach Durchführung eines Propensity-Score-Matchings mit Mahalanobis-Distanz bestehen und deutet zunächst auf das Vorliegen tatsächlicher Messeffekte hin, die nicht ausschließlich auf unterschiedliche Zusammensetzungen der Stichproben zurückgeführt werden können. Gleichwohl zeigte sich in der Auswertung der aggregierten Tagesstrukturen, dass diese Unterschiede keinen nennenswerten Einfluss auf die Verteilung der Hauptaktivitäten hatten. Dieses Ergebnis spricht dafür, dass etwaige Messeffekte der App zwar die Granularität der Angaben, nicht jedoch die grundsätzliche Allokation der Zeit auf Aktivitätsbereiche wesentlich verändern. Dennoch geben diese Ergebnisse Anlass zu einer kritischen Prüfung, wie das bestehende Tagebuchmodul der App weiter verbessert werden kann.


Eine Limitation der vorliegenden Analysen besteht allerdings darin, dass auch das eingesetzte Matching-Verfahren nur auf beobachteten Merkmalen basiert und somit verbleibende Verzerrungen durch unbeobachtete Drittvariablen nicht auszuschließen sind. So könnten etwa Unterschiede in der Motivation, der Technikaffinität oder der Genauigkeit der Selbstdokumentation zwischen den Gruppen bestehen, die in den erhobenen Variablen und in der Folge auch nicht im Matching vollständig abgebildet sind. Entsprechend bleibt eine gewisse Unsicherheit, in welchem Umfang die beobachteten Differenzen tatsächlich auf Messeffekte des Erhebungsinstruments oder doch auf unbeobachtete Selektionsmechanismen zurückzuführen sind. Diese Problematik ist in der Literatur zu Mixed-Mode-Erhebungen gut dokumentiert und stellt eine zentrale methodische Herausforderung dar (Vannieuwenhuyze/Loosveldt, 2013).

Unter der Annahme, dass es sich bei den Unterschieden in der Anzahl der Aktivitätenwechsel tatsächlich um genuine Messeffekte handelt, lassen sich mehrere plausible Erklärungsansätze anführen. Zum einen ist das papiergestützte Verfahren seit Jahrzehnten etabliert und gut erprobt. Insbesondere die Beispielseiten im Papierfragebogen vermitteln sehr anschaulich, wie detailliert die Eintragungen idealerweise erfolgen sollen. Zudem ermöglicht das Papierformat ein freies Eintragen und unkompliziertes Korrigieren von Angaben. Im Unterschied dazu wurde die App erstmals im Rahmen der ZVE 2022 eingesetzt. Die Eingabe erfolgte hier über einen Suchalgorithmus, der möglicherweise nicht für alle Nutzenden intuitiv verständlich war. Es könnte zudem unklar geblieben sein, wie fein granular die Angaben tatsächlich gewünscht waren. Insgesamt lassen diese Befunde darauf schließen, dass es verschiedene Ansatzpunkte gibt, um das Tagebuchmodul in der App zu optimieren. Diese sollen in künftigen Erhebungen geprüft und umgesetzt werden.

Dieselbe App wurde bei der Einkommens- und Verbrauchsstichprobe (EVS) 2023 genutzt; auch in den laufenden Wirtschaftsrechnungen (LWR) ist sie seit 2024 im Einsatz. Im Unterschied zur ZVE wird hier jedoch das Ausgabenverhalten erfasst. Erste Analysen zur EVS 2023 deuten darauf hin, dass sich die hier berichteten Effekte nicht generalisieren lassen. So liegen die Niveaus der per App erhobenen Konsumausgaben über jenen aus der Papiererhebung. Erste Matching-Analysen zeigen zudem, dass sich die Ausgaben-niveaus nach weitgehender Kontrolle von Selektionseffekten angleichen. Dazu soll nach Abschluss der Datenaufbereitung der EVS 2023 ein weiterer Artikel in dieser Zeitschrift folgen. Bereits die vorläufigen Ergebnisse legen jedoch nahe, dass die für die ZVE 2022 festgestellten Unterschiede nicht zu verallgemeinern sind; vielmehr scheinen sie erhebungsspezifische Ursachen zu haben.

Die Einführung der App als Erhebungsinstrument erforderte aufgrund der dezentralen Organisation der amtlichen Statistik eine enge Abstimmung zwischen dem Statistischen Bundesamt und den Statistischen Ämtern der Länder. Bestehende Abläufe waren dafür grundlegend anzupassen und neue Prozesse, etwa zur Verteilung der Erhebungsunterlagen oder zur Betreuung der Teilnehmenden, zu entwickeln. Aufgrund der App war es erstmals möglich, detaillierte prozessbegleitende Meta-informationen über den Erhebungsverlauf bereitzustellen.

So wurde sichtbar, an welchen Punkten Haushalte ihre Teilnahme abbrechen und welche Phasen besonders betreuungsintensiv sind. Diese Transparenz eröffnet neue Potenziale für eine proaktive Unterstützung der Teilnehmenden und eine fortlaufende Qualitätssicherung. Der Einsatz digitaler Erhebungsinstrumente führt somit nicht automatisch zu einer operativen Entlastung, da neue technische und organisatorische Chancen und Herausforderungen entstehen können. An dieser Stelle war die erfolgreiche Zusammenarbeit aller Akteure die Grundlage dafür, dass die ZVE 2022 nicht nur methodisch, sondern auch organisatorisch einen bedeutenden Meilenstein im Bereich der tagebuchgestützten Haushaltserhebungen darstellt, an den künftige Erhebungen gezielt anknüpfen können. Diese Transformation wurde intensiv vom Statistischen Amt der Europäischen Union (Eurostat) begleitet, welches die Vorreiterrolle der tagebuchgestützten Haushaltserhebungen in Deutschland wiederholt hervorgehoben hat.

Der Einsatz einer App hat die Modernisierung der ZVE substanziell vorangebracht. Die hohe App-Quote und die positiven Effekte auf die Reichweite legen nahe, dass die amtliche Statistik mit dem Einsatz mobiler Erhebungsinstrumente einen weiteren wichtigen Schritt gegangen ist, um den Zugang zu schwer erreichbaren Bevölkerungsgruppen zu verbessern und die Teilnahmebereitschaft zu erhöhen. Messmethodische Herausforderungen wurden identifiziert und können durch Weiterentwicklungen der App optimiert werden. Zugleich wird der parallele Einsatz papierbasierter Verfahren weiterhin notwendig bleiben, um eine hohe Abdeckung und Akzeptanz in allen Bevölkerungsgruppen sicherzustellen. Künftige Optimierungen sollten sich daher sowohl auf die Benutzerfreundlichkeit der App als auch auf die Ausgestaltung einer passgenauen Ansprache und Betreuung der Haushalte konzentrieren. 

LITERATURVERZEICHNIS

- Bank, Kathrin/Irmer, Alexander/Losch, Ute/Sandig, Anna/Schöne, Max. *Die Zeitverwendungserhebung 2022 – ein Überblick*. In: WISTA Wirtschaft und Statistik. Ausgabe 1/2025, Seite 106 ff.
- Bethlehem, Jelke. *Selection Bias in Web Surveys*. In: International Statistical Review. Jahrgang 78. Ausgabe 2/2010, Seite 161 ff. DOI: [10.1111/j.1751-5823.2010.00112.x](https://doi.org/10.1111/j.1751-5823.2010.00112.x)
- Bitzer, Eric/Van Hoorde, Joël/Schöne, Max. *Erstmaliger App-Einsatz für die Zeitverwendungserhebung 2022*. In: WISTA Wirtschaft und Statistik. Ausgabe 4/2022, Seite 63 ff.
- Byrom, Bill/Gwaltney, Chad/Slagle, Ashley/Gnanasakthy, Ari/Muehlhausen, Willie. *Measurement Equivalence of Patient-Reported Outcome Measures Migrated to Electronic Formats: A Review of Evidence and Recommendations for Clinical Trials and Bring Your Own Device*. In: Therapeutic Innovation & Regulatory Science. Jahrgang 53. Ausgabe 4/2019, Seite 426 ff. DOI: [10.1177/2168479018793369](https://doi.org/10.1177/2168479018793369)
- Dennison, Laura/Morrison, Leanne/Conway, Gemma/Yardley, Lucy. *Opportunities and Challenges for Smartphone Applications in Supporting Health Behavior Change: Qualitative Study*. In: Journal of Medical Internet Research. Jahrgang 15. Ausgabe 4/2013, e86.
- Gu, Xiao/Rosenbaum, Paul R. *Comparison of Multivariate Matching Methods: Structures, Distances, and Algorithms*. In: Journal of Computational and Graphical Statistics. Jahrgang 2. Ausgabe 4/1993, Seite 405 ff.
- Hohwü, Lena/Lyshol, Heidi/Gissler, Mika/Hrafns Jonsson, Stefan/Petzold, Max/Obel, Carsten. *Web-Based Versus Traditional Paper Questionnaires: A Mixed-Mode Survey With a Nordic Perspective*. In: Journal of Medical Internet Research. Jahrgang 15. Ausgabe 8/2013, e173. DOI: [10.2196/jmir.2595](https://doi.org/10.2196/jmir.2595)
- Janssen, Björn. *Web data collection in a mixed mode approach: an experiment*. European Conference on Quality in Survey Statistics. 2006. [Zugriff am 13. August 2025]. Verfügbar unter: ec.europa.eu
- Kelfve, Susanne/Kivi, Marie/Johansson, Boo/Lindwall, Magnus. *Going web or staying paper? The use of web-surveys among older people*. In: BMD Medical Research Methodology. Ausgabe 20 (252)/2020. DOI: [10.1186/s12874-020-01138-0](https://doi.org/10.1186/s12874-020-01138-0)
- Khare, Meena. *Estimated Prevalence and Characteristics of Web Users: National Health Interview Survey, 2014-2015*. In: Joint Statistical Meetings. 2016. Seite 660 ff. [Zugriff am 14. August 2025]. Verfügbar unter: www.asasrms.org
- Lutig, Peter/Lensvelt-Mulders, Gerty J. L. M./Frerichs, Remco/Greven, Assyn. *Estimating Nonresponse Bias and Mode Effects in a Mixed-Mode Survey*. In: International Journal of Market Research. Jahrgang 53. Ausgabe 5/2011, Seite 669 ff. [Zugriff am 14. August 2025]. Verfügbar unter: research.uvh.nl

LITERATURVERZEICHNIS

Lugtig, Peter/Toepoel, Vera/Alerk, Amin. *Mobile-only web survey respondents*. In: Survey Practice. Jahrgang 9. Ausgabe 4/2016, Seite 1 ff. DOI: [10.29115/SP-2016-0020](https://doi.org/10.29115/SP-2016-0020)

Marcano Belisario, José S./Jamsek, Jan/Huckvale, Kit/O'Donoghue, John/Morrison, Cecily P./Car, Josip. *Comparison of Self-Administered Survey Questionnaire Responses Collected Using Mobile Apps Versus Other Methods*. In: Cochrane Database of Systematic Reviews. Ausgabe 7/2015, Artikelnummer MR000042.

DOI: [10.1002/14651858.MR000042.pub2](https://doi.org/10.1002/14651858.MR000042.pub2).

Muehlhausen, Willie/Doll, Hellen/Quadri, Nuz/Fordham, Benthany/O'Donohoe, Paul/Dogar, Nijda/Wild, Diane J. *Equivalence of Electronic and Paper Administration of Patient-Reported Outcome Measures: A Systematic Review and Meta-Analysis of Studies Conducted between 2007 and 2013*. In: Health and Quality of Life Outcomes. Ausgabe 13/2015, Artikelnummer 167. DOI: [10.1186/s12955-015-0362-x](https://doi.org/10.1186/s12955-015-0362-x)

Rosenbaum, Paul R./Rubin, Donald B. *The Central Role of the Propensity Score in Observational Studies for Causal Effects*. In: Biometrika. Jahrgang 70.

Ausgabe 1/1983, Seite 41. DOI: [10.1093/biomet/70.1.41](https://doi.org/10.1093/biomet/70.1.41)

Rubin, Donald B. *Bias Reduction Using Mahalanobis-Metric Matching*. In: Biometrics. Jahrgang 36. Ausgabe 2/1980, Seite 293 ff.

Stone, Arthur A./Shiffman, Saul/Schwartz, Joseph E./Broderick, Joan E./Hufford, Michael R. *Patient Compliance with Paper and Electronic Diaries*. In: Controlled Clinical Trials. Jahrgang 24. Ausgabe 2/2003, Seite 182 ff.

DOI: [10.1016/s0197-2456\(02\)00320-3](https://doi.org/10.1016/s0197-2456(02)00320-3)

Stuart, Elizabeth A. *Matching Methods for Causal Inference: A Review and a Look Forward*. In: Statistical Science. Jahrgang 25. Ausgabe 1/2010, Seite 1 ff.

DOI: [10.1214/09-STS313](https://doi.org/10.1214/09-STS313)

Vannieuwenhuyze, Jorre T. A./Loosveldt, Geert. *Evaluating Relative Mode Effects in Mixed-Mode Surveys: Three Methods to Disentangle Selection and Measurement Effects*. In: Sociological Methods & Research. Jahrgang 42. Ausgabe 1/2013,

Seite 82 ff. DOI: [10.1177/0049124112464868](https://doi.org/10.1177/0049124112464868)

Vannieuwenhuyze, Jorre T. A./Loosveldt, Geert/Molenberghs, Geert. *A Method for Evaluating Mode Effects in Mixed-Mode Surveys*. In: Public Opinion Quarterly. Jahrgang 74. Ausgabe 5/2010, Seite 1027 ff. DOI: [10.2307/40985415](https://doi.org/10.2307/40985415)

Herausgeber
Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung
Dr. Daniel Vorgrimler
Redaktion: Ellen Römer

Ihr Kontakt zu uns
www.destatis.de/kontakt

Erscheinungsfolge
zweimonatlich, erschienen im Oktober 2025
Ältere Ausgaben finden Sie unter www.destatis.de sowie in der [Statistischen Bibliothek](#).

Artikelnummer: 1010200-25005-4, ISSN 1619-2907

© Statistisches Bundesamt (Destatis), 2025
Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.