

# Sind kommunale Bevölkerungsbefragungen als Onlinebefragung möglich?

## Ein Methodenexperiment zur Nutzung von schriftlich-postalischen und Onlinefragebögen in Repräsentativbefragungen

Text, Tabellen und Grafiken: **Michael Hanslmaier**

### Einleitung

Viele große Städte in Deutschland – wie München, Leipzig, Stuttgart oder Hannover – führen in regelmäßigen Abständen allgemeine Bevölkerungsbefragungen durch (Landeshauptstadt Hannover, 2021; Landeshauptstadt München, 2017, 2022; Schön, 2020; Stadt Leipzig, 2021). Die Ergebnisse dieser multithematischen Befragungen ergänzen die vorhandenen Daten der amtlichen Statistik sowohl um differenziertere Angaben zur Lebenssituation der Bürger\*innen als auch um Einschätzungen, Meinungen und Bewertungen zu Themen der Stadtentwicklung, etwa zur Zufriedenheit mit Angeboten und Dienstleistungen oder darüber, welche Probleme aus Sicht der Bürger\*innen besonders bedeutend sind. Damit liefern diese Befragungen wertvolle Hinweise für Verwaltung und Politik.

*Hoher Kostenaufwand für schriftlich-postalische Befragungen*

Allgemeine Bevölkerungsbefragungen mit größeren Stichproben erfordern aber auch in einem nicht zu vernachlässigenden Umfang finanzielle und personelle Ressourcen. Insbesondere der Druck der Fragebögen, Anschreiben und Erinnerungsschreiben sowie das Porto sind nicht unerhebliche Kostenfaktoren im Rahmen von schriftlich-postalischen Erhebungen. Zudem verursacht die Erfassung der schriftlichen Fragebögen zusätzlichen Aufwand. Online-Befragungen haben hier den Vorteil, dass – auch wenn die Kontaktaufnahme postalisch erfolgt – erhebliche Einsparungen im Bereich Druck und Versand möglich sind und auch die Dateneingabe entfällt. Darüber hinaus muss vor dem Hintergrund des Klimawandels auch berücksichtigt werden, dass Online-Befragungen im Vergleich zu schriftlich-postalischen Befragungen CO<sub>2</sub> einsparen und so einen Beitrag zur Bekämpfung des Klimawandels leisten können.

Gleichwohl müssen die Kostenvorteile von Online-Befragungen gegenüber schriftlich-postalischen Befragungen auch an methodischen Gütekriterien gemessen werden, d.h., es gilt zu betrachten, ob sich die Datenqualität von Online-Befragungen von der „klassischen“ Methode unterscheidet. In erster Linie ist zu fragen, ob die Teilnahmebereitschaft sich zwischen schriftlich-postalischen und Online-Befragungen unterscheidet. Gerade vor dem Hintergrund allgemein sinkender Ausschöpfungsquoten bei allgemeinen Bevölkerungsbefragungen in westlichen Demokratien (Neller, 2005) ist dies ein wichtiger Aspekt der Datenqualität.

*Bisherige Studien zeigen geringere Rücklaufquoten in Onlinebefragungen*

Die Frage nach dem Effekt der Befragungsart (schriftlich-postalische Befragung vs. Online-Erhebung) wurde bereits in zahlreichen Studien untersucht, die selbst wiederum Gegenstand verschiedener Meta-Analysen waren. In einer Meta-Analyse aus dem Jahre 2008 kommen die Autor\*innen zu dem Schluss, dass in schriftlich-postalischen Surveys höhere Antwortraten erzielt werden, wobei die Art der Grundgesamtheit und die Anzahl der Erinnerungen einen Effekt auf die Differenz der Rücklaufquote zwischen den Befragungsarten

haben (Shih & Fan, 2008). Eine neuere Meta-Analyse (Weigold, Weigold & Natera, 2019) bestätigt den Vorteil der schriftlich-postalischen Befragung in Bezug auf die Teilnahmebereitschaft.

Gleichwohl ist festzuhalten, dass die Ergebnisse auf der Ebene einzelner Studien inkonsistente Ergebnisse erzielt haben. So verweisen Weigold et al. (2019) auf Arbeiten, die eine höhere Teilnahmebereitschaft für die schriftliche Variante finden als auch auf solche, die in der Online- bzw. Computervariante höhere Rücklaufquoten finden – sowie auf Studien, die kaum Differenzen ausmachen können. Darüber hinaus sind die Unterschiede auch von verschiedenen methodischen Moderatoren (u.a. Zielpopulation, Instrument, Ort, Anzahl der Kontaktversuche) abhängig.

Darüber hinaus stammen die in den o.g. Meta-Analysen betrachteten Studien aus dem Zeitraum von 1992 bis 2013. Im Zuge der fortschreitenden Digitalisierung, die immer weitere Bereiche des täglichen Lebens umfasst, ist zu fragen, ob Hürden, die einer Online-Teilnahme an einer Befragung entgegenstehen, wie mangelnde digitale Kompetenzen oder fehlende Ausstattung mit entsprechenden Geräten oder einem Internetzugang, nicht im Zeitverlauf abgenommen haben. So zeigt etwa die ARD-ZDF-Onlinestudie, dass der Anteil der Online-nutzer seit 2009 von 67 % auf 94 % der Bevölkerung angestiegen ist<sup>1)</sup>. Im Zuge der Corona-Pandemie ist die Nutzung digitaler Technologien für viele Menschen noch einmal selbstverständlicher geworden, wie etwa die starke Zunahme des Home Office zeigt (Flüter-Hoffmann & Stettes, 2022).

Angesichts der gestiegenen Internetnutzung ist so zu fragen, ob es sinnvoll und praktikabel ist, kommunale Bürgerbefragungen als Onlinestudien durchzuführen und so Kosten zu sparen. Der vorliegende Aufsatz geht dieser Frage nach, indem im Rahmen eines Experimentes der Effekt der angebotenen Teilnahmemöglichkeiten auf die Rücklaufquote untersucht wird. Da die Intensität der Onlinenutzung mit dem Alter variiert – so liegt die Tagesreichweite des Internets bei den 14 bis 29-Jährigen bei 99 % während es bei den über 70-Jährigen nur 42 % sind<sup>2)</sup> – soll auch untersucht werden, inwiefern, die Differenzen in der Rücklaufquote zwischen den Teilnahmearten mit soziodemografischen Merkmalen zusammenhängen.

Neben der generellen Teilnahmebereitschaft ist für die Datenqualität auch entscheidend, zu welchem Grad die Fragen beantwortet werden. Der Begriff des Item-Non-Response bezeichnet die Tatsache, dass Befragte zwar teilnehmen, aber einzelne Fragen nicht beantworten. Dies kann analog zum Komplettausfall einer\* eines Befragten (Unit-Non-Response) auch zu Verzerrungen führen, wenn bestimmte Gruppen unter- oder überrepräsentiert sind (Čehovin, Bosnjak & Lozar Manfreda, 2022). Im Hinblick darauf, inwiefern sich die Rate des Item-Non-Response zwischen schriftlich-postalischen und Onlinebefragungen unterscheidet, liegen jedoch keine klaren Evidenzen vor: so gibt es Arbeiten, die eine niedrigere Item-Non-Response Rate in Onlinebefragungen aufweisen als auch Studien, die gegenteilige Effekte gefunden haben (sh. Čehovin et al., 2022 mit Verweisen).

Der vorliegende Aufsatz widmet sich der Frage, wie sich die Erhebungsart auf die Teilnahmebereitschaft an kommunalen Bürgerbefragungen und die Datenqualität auswirkt. Im Fokus stand dabei weniger, streng eine Onlinevariante mit einer Papiervariante zu vergleichen. Vielmehr sollten Kontroll- und Experimentalbedingung Varianten der tatsächlichen Umfragepraxis entsprechen. Da in der sozialwissenschaftlichen Umfragepraxis häufig mixed-mode Studien zum Einsatz kommen, d.h. Studien, bei denen die Befragten die Wahl zwischen mehreren Teilnahmearten haben, erhielten die Befragten in der Kontrollbedin-

*Onlinenutzung hat in den letzten Jahren stark zugenommen – zusätzlicher Digitalisierungsschub in der Corona Pandemie*

*Ist es sinnvoll und praktikabel, kommunale Bürgerbefragungen als Onlinestudien durchzuführen?*

*Der Aufsatz widmet sich der Frage, wie sich die Erhebungsart auf die Teilnahmebereitschaft an kommunalen Bürgerbefragungen und die Datenqualität auswirkt*

<sup>1)</sup> <https://www.ard-zdf-onlinestudie.de/onlinenutzung/entwicklung-der-onlinenutzung/> (Abruf am 22.03.2022).

<sup>2)</sup> [https://www.ard-zdf-onlinestudie.de/files/2021/ARD\\_ZDF\\_Onlinestudie\\_2021\\_Publikationscharts\\_final.pdf](https://www.ard-zdf-onlinestudie.de/files/2021/ARD_ZDF_Onlinestudie_2021_Publikationscharts_final.pdf) (Abruf am 20.04.2022).

gung ein postalisches Anschreiben mit einem schriftlichen Fragebogen und einem Zugangscode für die Onlinevariante. In der Experimentalbedingung wurden die Befragten ebenfalls postalisch kontaktiert und gebeten, an der Onlinebefragung teilzunehmen. Zudem bestand die Möglichkeit, den schriftlichen Fragebogen gesondert anzufordern.

Da das Methodenexperiment im Rahmen einer multithematischen, großen Repräsentativbefragung durchgeführt wurde und die verglichenen Methoden in der Praxis häufig anzutreffen sind, ist eine Verallgemeinerung der Ergebnisse gut möglich.

## Forschungsstand

*Nutzung von Online-Befragungen als Methode der Datenerhebung hat in den letzten Dekaden stark zugenommen*

Die Nutzung von Online-Befragungen als Methode der Datenerhebung hat in den letzten Dekaden stark zugenommen, so Daikeler, Bosnjak und Manfreda (2020) mit Verweisen auf entsprechende empirische Daten. Online-Surveys weisen dabei eine Reihe von Vorteilen gegenüber klassischen schriftlich-postalischen Erhebungen auf. So entfallen Kosten für Druck und Versand der Fragebögen, Filterführungen können problemlos angepasst werden, die Auswahl verschiedener Sprachen ist einfach möglich und es können zusätzliche Informationen wie etwa die Ausfülldauer erhoben werden. Zudem ist es möglich, auch innovative Formen der Datenerhebung (z.B. interaktive Kartentools) zu verwenden (Boyle, Morrison, MacDonald, Duncan, & Rose, 2016).

*Varianz methodischer Aspekte innerhalb von Online-Surveys*

Beachtet werden muss allerdings, dass innerhalb von Online-Surveys eine gewisse Varianz in Bezug auf methodische Aspekte besteht. Dies betrifft zum einen die Art der (Erst-)Kontaktaufnahme mit den Befragten. Diese kann „klassisch“ schriftlich-postalisch erfolgen, Zugangsdaten für die Onlinebefragung können in so einem Falle per URL oder QR-Code angegeben werden. Alternative Kontaktwege sind E-Mail oder auch die Möglichkeit, auf Websites oder aus Apps heraus direkt per Klick an einer Befragung teilzunehmen. Die Wahl der Kontaktmöglichkeiten ist durch die Verfügbarkeit entsprechender Daten eingeschränkt: Liegt eine Liste der Grundgesamtheit vor? Welche Kontaktdaten sind dort gespeichert? Zum anderen kann auch die Befragungssituation selbst zwischen Online-Surveys variieren. Die Teilnahme kann auf verschiedenen Arten von Geräten (Computer, Tablet oder Smartphone) durchgeführt werden. Darüber hinaus ist zu differenzieren, ob eigene Geräte genutzt werden oder die Befragungen etwa am Arbeitsplatz durchgeführt werden. Zudem besteht auch die Möglichkeit, dass die Geräte für den Zweck der Befragung zur Verfügung gestellt werden (z.B. Baier, 2018). Auch ist in Bezug auf die Art der Befragungssituation zu differenzieren: sind andere Personen anwesend (z.B. Klassenzimmerbefragung, Arbeitsplatz)? Kann die Situation von den Forschenden kontrolliert werden? Alle diese genannten methodischen Aspekte können sich auf generelle Teilnahmbereitschaft, die Bereitschaft (sensible) Fragen zu beantworten und generell auf die Antworten auswirken. Dementsprechend sind Ergebnisse aus vergleichenden Methodenstudien nicht ohne Weiteres auf jede Situation übertragbar.

*Rücklaufquoten von Onlinebefragungen im Vergleich zu schriftlichen Befragungen*

Was ist nun im Hinblick auf die Teilnahmbereitschaft für Onlinebefragungen im Vergleich zu schriftlichen Befragungen zu erwarten? Manfreda, Bosnjak, Berzelak, Haas und Vehovar (2008) verweisen auf die Diskussion in der Literatur, die eine Reihe von Gründen auflistet, die darauf hindeuten, dass die Rücklaufquoten in Onlinebefragungen niedriger sind.

*Gründe für die niedrigere Teilnahmbereitschaft in Onlinebefragungen*

Als Argument für eine niedrigere Teilnahmbereitschaft werden Sicherheits- und Datenschutzbedenken angeführt, die gegenüber der Datenübertragung im Internet bestehen (vgl. u.a. Sax, Gilmartin, & Bryant, 2003). Diese Bedenken können in den letzten Jahren ggf. auch zugenommen haben, wie Daikeler et al. (2020) argumentieren. Zudem kann die Zunahme an Spam-Mails und Onlinebefragungen in den letzten Jahren die Teilnahmbereitschaft an solchen Befragungen

weiter reduziert haben (Daikeler et al., 2020). Darüber hinaus werden technische Probleme bei der Durchführung von Onlinebefragungen auf Grund von nicht kompatibler Software, fehlender Geräte, falschen grafischen Darstellungen oder langen Ladezeiten als Faktoren angeführt, die die Teilnahmebereitschaft<sup>3)</sup> reduzieren können. Gleiches gilt für fehlende digitale Kompetenzen (Manfreda et al. 2008). Allerdings verweist eine neuere Studie (Daikeler et al., 2020) darauf, dass digitale Kompetenzen zugenommen haben, Internetzugänge schneller, weiter verbreitet und billiger geworden sind und eine große Bandbreite an nutzerfreundlichen Geräten (z.B. Tablets, Smartphones) verfügbar ist.

Außerdem stehen einige Strategien, die zur Erhöhung der Rücklaufquote eingesetzt werden, wie etwa monetäre Anreize<sup>4)</sup>, online nicht zur Verfügung (Manfreda et al., 2008). Allerdings hängt die Möglichkeit des Rückgriffs auf „traditionelle“ Verfahren, um die Rücklaufquote zu steigern (z.B. Incentives), auch von der Art der Kontaktaufnahme ab. Weiterhin wird von verschiedenen Autor\*innen angeführt, dass ein schriftlicher Fragebogen durch seine Präsenz als eine Art Erinnerung fungiert (z.B. Manfreda et al., 2008), die deutlich präsenter sein kann als eine E-Mail. Zudem kann die Legitimität von Online-Befragungen im Vergleich zu anderen Formen reduziert sein, da die Befragten einen geringeren Aufwand auf Seiten der Forschenden wahrnehmen. Dies kann auch ein Faktor zur Erklärung der niedrigeren Teilnehmeraten in online durchgeführten Studien sein (Daikeler et al., 2020).

Die Höhe der Unterschiede in den Rücklaufquoten zwischen Online-Befragungen und anderen Methoden hängt dabei von verschiedenen Charakteristika der Studien ab. So verweisen Daikeler et al. (2020) in ihrer Meta-Analyse auf eine Reihe wichtiger Moderatoren<sup>5)</sup>: Die Verwendung von existierenden Listen oder Panelteilnehmer\*innen verringert etwa die Differenz in der Rücklaufquote gegenüber der Verwendung einer neu rekrutierten Stichprobe. Zudem sind die Unterschiede in den Rücklaufquoten größer, wenn die Kontaktaufnahme per E-Mail erfolgt. Eine Vorankündigung der Befragung verstärkt ebenfalls die Differenz in der Rücklaufquote zuungunsten der Online-Befragung. Auch die Zahl der Kontaktversuche erhöht den Vorteil der schriftlichen Befragung gegenüber der Onlinevariante. Die Zielpopulation spielt ebenfalls eine Rolle, der Nachteil der Onlinebefragungen ist v.a. bei allgemeinen Bevölkerungsbefragungen größer. Die Autor\*innen schreiben weiterhin, dass sich die Differenz zwischen Online-Befragungen und schriftlichen Befragungen seit 2002 nicht verändert hat.

*Höhe der Unterschiede in den Rücklaufquoten zwischen Online-Befragungen und anderen Methoden hängt von Charakteristika der Studien ab*

Neben der kompletten Verweigerung der Teilnahme an Befragungen kommt es auch vor, dass Befragte auf einzelne Fragen nicht antworten. In diesem Fall spricht man von Item-Non-Response. Item-Non-Response kann dabei, genau wie die generelle Weigerung, an der Befragung teilzunehmen (Unit-Non-Response), zu einer Verzerrung der Ergebnisse beitragen, wenn es zu einem systematischen Ausfall bestimmter Gruppen kommt (Faulbaum, 2019). Ein Faktor, der zu unterschiedlichen Item-Non-Response Raten zwischen Online- und schriftlichen Befragungen führen kann, ist die technische Programmierung des Fragebogens, die Befragte auf fehlende Antworten oder Inkonsistenzen hinweisen kann (Čehovin et al., 2022).

*Item-Non-Response*

Čehovin et al. (2022) untersuchen in ihrer Metastudie, ob sich die Item-Non-Response Raten von Websurveys von anderen Erhebungsmethoden unterscheiden und finden im Allgemeinen keine Unterschiede „Thus, overall, there seems to be no difference in the item nonresponse rate when web surveys are compared to other survey modes“ (Čehovin et al., 2022, p. 5). In der Meta-Analyse wurde die Onlinevariante allerdings nicht nur mit schriftlichen Varianten

<sup>3)</sup> Technische Probleme haben nicht unbedingt einen Einfluss darauf, dass man bereit ist, an der Befragung teilzunehmen, können aber einer tatsächlichen Teilnahme entgegenstehen oder dazu führen, dass man abbricht, wenn im Verlauf des Ausfüllens technische Probleme auftauchen.

<sup>4)</sup> So haben Studien gezeigt, dass das Beilegen von kleinen Geldscheinen beim Versand des Fragebogens die Rücklaufquote erhöht (vgl. z.B. Baier, 2015).

<sup>5)</sup> Damit bestätigen sich weitestgehend die Ergebnisse einer früheren Metastudie (Manfreda et al., 2008).

verglichen. Betrachtet man Studien, die sich auf den Vergleich von Online- und schriftlichen Befragungen beziehen so finden einige Autor\*innen (Hsu & McFall, 2015; Kwak & Radler, 2002; Messer, Edwards, & Dillman, 2012) niedrigere Item-Non-Responderaten, d.h. bessere Datenqualität, in Onlinebefragungen.

## Hypothesen

In der vorliegenden Untersuchung sollte der Effekt der Art der Durchführung der Befragung auf die Teilnahmebereitschaft und die Datenqualität (Item-Non-Response) im Rahmen eines Studiendesigns geprüft werden, das reale Rahmenbedingungen abbildet, wie sie in kommunalen Bevölkerungsbefragungen vorkommen: Die Kontaktaufnahme musste für die Bevölkerungsbefragung zur Stadtentwicklung postalisch erfolgen, da zwar eine Liste der Grundgesamtheit vorliegt (Einwohnermelderegister) aber keine Liste mit Email-Adressen. Bei der Grundgesamtheit handelt es sich um keine bestehende Gruppe von Befragten, sondern um eine neu rekrutierte Zielpopulation. Es wurde eine Erinnerung (postalisch, ohne Fragebogen) versandt.

*Geringere Rücklaufquote in der Onlinevariante erwartet*

Diese Charakteristika der Befragung lassen vor dem Hintergrund der Ergebnisse der Meta-Analysen eine (deutlich) geringere Rücklaufquote in der Experimentalstichprobe erwarten: Daikeler et al. (2020) berichten eine im Durchschnitt um 12 Prozentpunkte geringere Rücklaufquote für Onlinesurveys. Diese Differenz wird noch vergrößert, da die beschriebenen Charakteristika (Zielpopulation, Kontaktaufnahme, Erinnerung) die Differenzen in der Rücklaufquote erhöhen (Daikeler et al., 2020).

*Gestiegene Digitalisierung könnte Rücklauf in Onlinevariante erhöhen*

Zwei Faktoren könnten allerdings dazu beitragen, die Differenz in der Rücklaufquote zu verringern. Zum einen war die Studie konsequent in einem Mixed-mode Design angelegt.<sup>6)</sup> In jeder der beiden Stichproben hatten die Befragten die Möglichkeit, sowohl schriftlich als auch online teilzunehmen – wengleich natürlich Personen in der Experimentalstichprobe den schriftlichen Fragebogen erst hätten anfordern müssen. Zum anderen fand die Befragung Anfang des Jahres 2021 statt. Auf Grund der Corona-Pandemie und den zu ihrer Bekämpfung ergriffenen Maßnahmen hat sich die Digitalisierung der Bevölkerung nochmals stark verstärkt. Die Nutzung digitaler Onlinedienste, wie Videokonferenzen<sup>7)</sup>, Streamingangeboten<sup>8)</sup> oder Onlineshopping (HDE Handelsverband Deutschland, 2021), haben massiv zugenommen. Diese Entwicklung kann dazu geführt haben, dass Vorbehalte und Barrieren (u.a. Kompetenzen, technische Ausstattung) gegenüber Onlinebefragungen weiter zurückgegangen sind. Dies könnte sich positiv auf die Teilnahme an der Onlinebefragung auswirken.

*Weniger digitalaffine Personen werden in der schriftlichen Variante besser erreicht*

Neben der absoluten Höhe der Rücklaufquote bzw. der Differenz der beiden Stichproben ist auch von Interesse, inwiefern eine niedrigere Rücklaufquote zu einer Verzerrung führt (Non-Response-Error bzw. -Bias). Zwar muss eine niedrigere Rücklaufquote nicht zwingend zu einer Verzerrung führen – diese tritt auf, wenn selektiv bestimmte Personengruppen auf Grund des Modus auf eine Teilnahme verzichten (Daikeler et al., 2020). In der vorliegenden Studie könnte es etwa der Fall sein, dass weniger digitalaffine Personen auf eine Teilnahme in der Experimentalgruppe verzichten, da der schriftliche Fragebogen extra angefordert werden musste.

<sup>6)</sup> Hierbei ist allerdings zu beachten, dass der Aspekt, dass verschiedene Modes gleichzeitig angeboten werden, ebenfalls bereits einen Effekt auf die Rücklaufquote haben kann. So finden Medway und Fulton (2012) in ihrer Meta-Analyse, dass das Anbieten verschiedener Teilnahmemöglichkeiten die Teilnahmebereitschaft insgesamt senkt – wengleich die Autor\*innen auch auf Studien verweisen, die keinen oder einen positiven Effekt auf die Rücklaufquote finden.

<sup>7)</sup> So hat z. B. die Nutzung von Videokonferenzen zugenommen. [https://www.destatis.de/DE/Presse/Pressemitteilungen/2021/11/PD21\\_537\\_52911.html](https://www.destatis.de/DE/Presse/Pressemitteilungen/2021/11/PD21_537_52911.html) (Abruf am 05.08.2022)

<sup>8)</sup> So ist z. B. die Anzahl der weltweiten Netflix Abonnent\*innen stark gestiegen. (<https://de.statista.com/statistik/daten/studie/196642/umfrage/abonnenten-von-netflix-quartalszahlen/#professional>) (Abruf am 05.08.2022)

Darüber hinaus soll mit Blick auf die Datenqualität untersucht werden, inwiefern sich die beiden Bedingungen im Hinblick auf Item-Non-Response auswirken. Die Ergebnisse der Metaanalyse (Čehovin et al., 2022) lassen keine größeren Unterschiede erwarten.

## Empirisches Vorgehen

Das Referat für Stadtplanung und Bauordnung führt in regelmäßigen Abständen Bevölkerungsbefragungen zur Stadtentwicklung<sup>9)</sup> durch. Grundgesamtheit der im Jahr 2021 durchgeführten Welle waren – wie in den Vorgängerstudien – alle Münchner\*innen mit Hauptwohnsitz ab 18 Jahren. Die Stichprobenziehung erfolgt über das Einwohnermelderegister nach einem uneingeschränkten Zufallsverfahren. Mit der Durchführung der Befragung war die Weeber und Partner GmbH aus Stuttgart betraut.

Neben der Hauptstichprobe von 20 000 Adressen wurde auch eine Experimentalstichprobe mit einem Umfang von 2 000 Adressen gezogen. Die Kontaktaufnahme erfolgte in beiden Stichproben per Brief. Die Befragten der Hauptstichprobe erhielten ein Anschreiben mit dem Zugangscode für die Onlinebefragung und den schriftlichen Fragebogen. An die Experimentalstichprobe wurde nur ein Anschreiben verschickt, das den Zugangscode für die Onlinebefragung enthielt und den Hinweis, dass auch ein schriftlicher Fragebogen angefordert werden kann. Die Feldphase der Befragung lief von Mitte Januar bis Ende Februar 2021. Etwa zwei Wochen nach Erhalt des Fragebogens wurde eine Erinnerung an alle Befragten verschickt.

Um mögliche Sprachbarrieren zu reduzieren, wurden Anschreiben und Fragebogen in insgesamt zwölf verschiedenen Sprachen<sup>10)</sup> angeboten. Basierend auf der im Melderegister erfassten (ersten) Staatsangehörigkeit der ausgewählten Personen, erhielten diese Anschreiben und Fragebogen zweisprachig in Deutsch und der sich aus der jeweiligen Staatsangehörigkeit ergebenden Sprache. An Personen mit einer Staatsangehörigkeit, für die keine eigene Sprachversion zur Verfügung stand, wurde der Fragebogen in Deutsch und Englisch versandt. Darüber hinaus wurde auf die verfügbaren Sprachen im Anschreiben hingewiesen und online standen alle zwölf Sprachen zur Verfügung.

## Ergebnisse

Von den insgesamt 22 000 verschickten Fragebögen waren 5,0 % nicht zustellbar, so dass die bereinigte Bruttostichprobe bei 19 011 (Hauptstichprobe) bzw. 1 897 (Experimentalstichprobe) Adressen liegt. Zurückgesandt bzw. online beantwortet wurden 6 428 bzw. 571 Fragebögen, so dass sich eine Rücklaufquote in der Hauptstichprobe von 33,8 % ergibt und in der Experimentalgruppe von 30,1 %. 74 Papierfragebögen konnten keiner Stichprobe zugeordnet werden, da der auf den Fragebogen gedruckte Zugangscode entfernt wurde<sup>11)</sup>. Ein Chi<sup>2</sup> Test zeigt einen signifikanten Unterschied in der Rücklaufquote zwischen beiden Bedingungen ( $p \leq 0,001$ ,  $n = 20\,908$ ). Somit bleibt festzuhalten, dass der Rücklauf in der Experimentalstichprobe unter dem Rücklauf in der Hauptstichprobe liegt. Bezogen auf die Rücklaufquote ist die „klassische“ Variante daher überlegen.

*Bevölkerungsbefragung zur Stadtentwicklung als Grundlage*

*Insgesamt 22 000 Befragte aus dem Einwohnermelderegister gezogen*

*Fragebogen in zwölf verschiedenen Sprachen angeboten*

*Rücklaufquote in Abhängigkeit der Variante*

<sup>9)</sup> Weitere Informationen zur Methodik finden sich in Landeshauptstadt München (2022).

<sup>10)</sup> Der Fragebogen wurde neben Deutsch in insgesamt elf weiteren Sprachen angeboten (Englisch, Polnisch, Französisch, Italienisch, Kroatisch, Türkisch, Griechisch, Russisch, Arabisch, Spanisch und Rumänisch). Die Auswahl der Sprachen erfolgte in Anlehnung an die Vorgängerbefragung und einer Analyse der häufigsten Staatsangehörigkeiten in München.

<sup>11)</sup> Auf Grund der Tatsache, dass in der Experimentalstichprobe 24 Fragebögen angefordert wurden, von denen 19 zurückgesandt wurden, sind mindestens 69 dieser Fragebögen aus der Hauptstichprobe. Eine individuelle Zuordnung ist aber nicht möglich.

Tabelle 1: Rücklaufquote und Teilnahmeart nach Stichprobe

Merkmal	Form	Netto-SP bereinigt	Brutto-SP unbereinigt	Rücklauf unbereinigt	Brutto-SP angepasst	Rücklauf angepasst
Hauptstichprobe	Papier	4 096		20,5%		
	Online	2 332		11,7%		
	<b>gesamt</b>	<b>6 428</b>	<b>20 000</b>	<b>32,1%</b>	<b>19 011</b>	<b>33,8%</b>
Experimentalstichprobe	Papier	19		1,0%		
	Online	552		27,6%		
	<b>gesamt</b>	<b>571</b>	<b>2 000</b>	<b>28,6%</b>	<b>1 897</b>	<b>30,1%</b>
keine Angabe	Papier	74				
Gesamtstichprobe	Papier	4 189		19,0%		
	Online	2 884		13,1%		
	<b>gesamt</b>	<b>7 073</b>	<b>22 000</b>	<b>32,2%</b>	<b>20 908</b>	<b>33,8%</b>

Quelle: Münchner Bevölkerungsbefragung zur Stadtentwicklung 2021, eigene Berechnungen.

Modepräferenzen in Abhängigkeit der Variante

Interessant ist auch der Blick auf die Teilnahmepräferenz der Befragten. Wird ein schriftlicher Fragebogen mitgesandt, so entscheiden sich 63,7% für die schriftliche Variante. Nur ein gutes Drittel nimmt online an der Befragung teil. Wird der Fragebogen nicht mitgesandt, so entscheidet sich nur ein Bruchteil der Befragten, diesen extra anzufordern. In der Experimentalgruppe haben nur 24 Personen den Fragebogen extra angefordert. Dies entspricht 1,3% bezogen auf die angepasste Bruttostichprobe. Insgesamt haben sich 3,3% der tatsächlich Teilnehmenden in der Experimentalstichprobe für die Papiervariante entschieden.

Aus dem Vergleich der beiden Stichproben lässt sich demnach ableiten, dass ein großer Teil der Befragten mit Präferenz für den schriftlichen Fragebogen auf die Online-Variante wechselt, wenn der Fragebogen nicht mitgeschickt wird. Ein kleiner Teil fordert den schriftlichen Fragebogen tatsächlich an – jedoch gibt es auch einen Teil an Personen, die auf Grund der fehlenden Möglichkeit, den schriftlichen Fragebogen auszufüllen, auf eine Teilnahme verzichten. Auf den Anteil dieser Gruppe kann aus der Differenz der beiden Rücklaufquoten geschlossen werden – also etwa 3,7% aller Befragten.

Modepräferenzen in Abhängigkeit soziodemografischer Merkmale

Interessant ist ein Blick auf die Modepräferenzen, wenn die schriftliche und Onlinevariante „gleichberechtigt“ angeboten werden, d.h., die Befragten können ohne weitere Schritte (sofort) nach Erhalt des Anschreibens mit der Beantwortung beginnen. Die Betrachtung der Modepräferenzen macht in der vorliegenden Studie freilich nur für die Hauptstichprobe Sinn, da in der Experimentalstichprobe nur sehr wenige Befragte den schriftlichen Fragebogen überhaupt angefordert haben.

Insgesamt haben 36,3% der Befragten der Hauptstichprobe online teilgenommen. Betrachtet man die Modepräferenzen in Abhängigkeit soziodemografischer Merkmale, siehe Tabelle 2, so zeigt sich, dass Männer die Onlinevariante deutlich häufiger wählen als Frauen (44,8% zu 27,8%). Je älter die Befragten sind, desto ausgeprägter ist die Präferenz für die schriftliche Variante. In der jüngsten Altersgruppe nimmt jede\*r zweite schriftlich teil, bei den Personen ab 60 Jahren sind es mehr als vier von fünf Befragten. Deutsche ohne Migrationshintergrund<sup>12)</sup> nehmen im Vergleich häufiger schriftlich an der Befragung Teil als Deutsche mit Migrationshintergrund und Ausländer\*innen.

<sup>12)</sup> Deutsche mit Migrationshintergrund sind Personen, die die deutsche Staatsangehörigkeit besitzen und die entweder selbst oder deren Eltern nach 1955 in das Gebiet der heutigen Bundesrepublik zugewandert sind. Ausländer\*innen sind Personen ohne deutsche Staatsangehörigkeit.

**Table 2: Soziodemografische Merkmale nach Ausfüllart (nur Hauptstichprobe)**

Variable	Geschlecht	Online	Schriftlich	Signifikanz (Chi <sup>2</sup> -Test)
Geschlecht	Männlich	44,8%	55,2%	p ≤ 0,000
	Weiblich	27,8%	72,2%	
	N	2 094	3 786	
Alter	18–29 Jahre	49,9%	50,1%	p ≤ 0,000
	30–59 Jahre	42,2%	57,8%	
	60 und älter	16,3%	83,7%	
	N	2 095	3 705	
Migrationshintergrund	deutsch ohne MH	33,0%	67,0%	p ≤ 0,000
	deutsch mit MH	36,7%	63,3%	
	ausländisch	41,3%	58,7%	
	N	2 142	4 043	
Einkommen	Arme Haushalte	24,9%	75,1%	p ≤ 0,000
	Untere Mitte	32,7%	67,3%	
	Obere Mitte	42,3%	57,7%	
	Reiche Haushalte	48,3%	51,7%	
	N	1 783	3 113	
Bildung	Kein Abitur	22,3%	77,7%	p ≤ 0,000
	Abitur	41,3%	58,7%	
	N	2 159	4 018	

Quelle: Münchner Bevölkerungsbefragung zur Stadtentwicklung 2021, eigene Berechnungen.

Befragte mit höherer Bildung präferieren im Vergleich zu Befragten ohne Abitur fast doppelt so häufig die Onlinevariante. Ein höheres Einkommen<sup>13)</sup> geht ebenfalls mit einer größeren Präferenz für die Onlinevariante einher. Eine logistische Regressionsanalyse (nicht dargestellt) bestätigt, dass Ältere, Deutsche (mit und ohne Migrationshintergrund), Befragte mit Abitur und Wohlhabendere häufiger zur Onlinevariante greifen.

Insgesamt kann so festgehalten werden, dass die Präferenz für eine bestimmte Teilnahmeart – im vorliegenden Fall schriftliche Teilnahme vs. Onlineteilnahme – mit den soziodemografischen Charakteristika zusammenhängt. Die Präferenz für die schriftliche Variante scheint allerdings nicht dergestalt ausgeprägt zu sein, dass Befragte im Falle eines nicht vorliegenden Fragebogens diesen häufig anfordern oder auf die Teilnahme verzichten. Insgesamt haben, wie oben dargestellt, nur 24 Befragte aus der Experimentalgruppe einen schriftlichen Fragebogen angefordert. Auch läge die Rücklaufquote in der Experimentalgruppe deutlich niedriger, wenn alle Befragten mit schriftlicher Präferenz tatsächlich auf eine Teilnahme verzichtet hätten, wenn kein schriftlicher Fragebogen mitgeschickt wird.

In der Experimentalgruppe stehen Befragten mit einer Präferenz für die schriftliche Variante drei Möglichkeiten zur Verfügung: Onlineteilnahme, Anforderung des schriftlichen Fragebogens oder Nichtteilnahme. Angesichts der Abhängigkeit der Modepräferenz von soziodemografischen Merkmalen, kann es sein, dass diese Entscheidung ebenfalls durch die Soziodemografie der

*Präferenz für Teilnahmeart hängt mit soziodemografischen Merkmalen zusammen*

*Soziodemografische Merkmale in Abhängigkeit der Stichprobe*

<sup>13)</sup> Die Einteilung in die Einkommensklassen erfolgte auf Basis des bedarfsgewichteten Äquivalenzzetoeinkommens am Median der gewichteten Daten. Arme Haushalte verfügen über weniger als 60 % des Medianeinkommens, die untere Mitte über 60 bis unter 120 %, die obere Mitte über 120 % bis unter 200 % und Reiche Haushalte verfügen über 200 % und mehr (sh. hierzu (Landeshauptstadt München, 2022)



Befragten beeinflusst ist. Dies könnte dazu führen, dass es in der Experimentalstichprobe zu systematischen Ausfällen kommt (im Vergleich zur Hauptstichprobe). Dies wäre der Fall, wenn z.B. jüngere und besser gebildete Befragte sich eher für eine Onlineteilnahme entscheiden, während ältere Befragte auf eine Teilnahme verzichten. Um zu analysieren, inwiefern es tatsächlich zu einer Verzerrung kommt, werden im Folgenden die soziodemografischen Charakteristika beider Stichproben miteinander verglichen, siehe Tabelle 3.

In der Hauptstichprobe ist der Anteil der Frauen<sup>14)</sup> um fast 4 Prozentpunkte höher als in der Experimentalstichprobe, wenngleich die Differenz nicht signifikant ist. Damit ist die Experimentalstichprobe im Vergleich zur Grundgesamtheit jedoch weniger verzerrt. Der Anteil der Frauen in der Grundgesamtheit beträgt 50,9% (Landeshauptstadt München, 2022). Die Differenzen in der Alterszusammensetzung sind sehr gering. Die mittlere Altersgruppe der 30- bis 59-Jährigen ist in der Experimentalgruppe auf Kosten der beiden anderen Altersgruppen um 0,7 Prozentpunkte größer. Diese Unterschiede sind aber nicht signifikant. Ein zusätzlicher T-Test zeigt, dass die Mittelwertsunterschiede (48,9 Jahre in der Hauptstichprobe und 48,4 Jahre in der Experimentalstichprobe) ebenfalls nicht signifikant sind. Auch in Bezug auf den Migrationshintergrund ergeben sich keine signifikanten Unterschiede zwischen den beiden Varianten. Wenngleich in der Experimentalvariante im Vergleich etwas mehr Ausländer\*innen und etwas weniger Deutsche mit Migrationshintergrund teilnehmen.

<sup>14)</sup> Insgesamt haben sechs Befragte als Geschlecht „divers“ angegeben. Auf Grund der geringen Fallzahl wurden diese Personen bei der Analyse der Geschlechterzusammensetzung nicht berücksichtigt.

Tabelle 3: **Soziodemografische Merkmale nach Stichprobe**

Variable	Geschlecht	Hauptstichprobe	Experimentalstichprobe	Signifikanz (Chi <sup>2</sup> -Test)
Geschlecht	Männlich	45,9%	49,6%	n.s.
	Weiblich	54,1%	50,4%	
	N	5 880	472	
Alter	18–29 Jahre	13,6%	13,3%	n.s.
	30–59 Jahre	58,9%	59,6%	
	60 und älter	27,6%	27,2%	
	N	5 800	475	
Migrationshintergrund	deutsch ohne MH	73,2%	73,4%	n.s.
	deutsch mit MH	12,4%	9,6%	
	ausländisch	14,4%	17,0%	
	N	6 185	488	
Einkommen	Arme Haushalte	12,6%	10,6%	n.s.
	Untere Mitte	44,2%	45,5%	
	Obere Mitte	34,4%	34,0%	
	Reiche Haushalte	8,8%	9,8%	
	N	4 896	376	
Bildung	Kein Abitur	33,6%	30,6%	n.s.
	Abitur	66,4%	69,4%	
	N	6 177	490	

n.s. = nicht signifikant für  $p \leq 0,05$

Quelle: Münchner Bevölkerungsbefragung zur Stadtentwicklung 2021, eigene Berechnungen.

Vergleicht man die Verteilung der Befragten auf die relativen Einkommenspositionen<sup>15)</sup> zwischen beiden Stichproben, so finden sich hier auch kaum Unterschiede. Befragte aus reichen Haushalten sind in der Experimentalgruppe etwas stärker vertreten, Personen aus ärmeren Haushalten in der klassischen Variante, die Unterschiede sind aber nicht signifikant. Betrachtet man die Unterschiede im Bildungsniveau zwischen beiden Stichproben, so zeigen sich ebenfalls keine signifikanten Unterschiede, wenngleich in der Experimentalgruppe der Anteil der Personen mit (Fach-)Abitur etwas höher liegt (69,4 % zu 66,4 %).

Neben der generellen Teilnahmebereitschaft spielt in Umfragen auch das Thema Item-Non-Response, d.h., die Nichtbeantwortung von einzelnen Fragen, eine Rolle und soll deshalb hier auch als Merkmal der Datenqualität betrachtet werden. Grundsätzlich kann bei Item-Non-Response zwar zwischen dem selektiven Nichtbeantworten einzelner Fragen (Item-Non-Response) und einem Abbruch (Drop-Out) unterschieden werden. In der Praxis der vorliegenden Studie ist dies aber nicht so einfach zu differenzieren. So werden Personen, die beim Ausfüllen des schriftlichen Fragebogens abbrechen, diesen eher nicht zurücksenden und zählen daher als Unit-Non-Response. Im Gegensatz hierzu konnte der Onlinefragebogen über die ganze Feldzeit hinweg mit dem individuellen Zugangscode bearbeitet werden und wurde am Ende des Feldzeitraumes automatisch abgegeben. Eine feine Differenzierung der Antwortmuster, wie sie z.B. Faulbaum (2019) vornimmt, ist hier nicht möglich und auch nicht zielführend, da ein Vergleich der Antwortmuster über die verschiedenen Modes hinweg auch nicht sinnvoll ist.

Die folgenden Analysen beruhen daher auf der Anzahl der nicht beantworteten Items der jeweiligen Befragten. Bei der Zählung der Item-Non-Response wurden aus Gründen der Vergleichbarkeit Filterfragen nicht berücksichtigt. Insgesamt bezieht sich die Analyse zum Item-Non-Response auf 248 Items.

Vergleicht man die Zahl der fehlenden Werte zwischen der Haupt- und der Experimentalstichprobe, so zeigen sich im Mittel deutliche Unterschiede. Liegt der Mittelwert in der Hauptstichprobe bei 15,6 fehlenden Werten, so sind es in der Experimentalstichprobe 29,5. Diese Unterschiede sind statistisch signifikant ( $p \leq 0,000$ ). Dieser Befund steht nicht im Einklang mit den oben skizzierten Erwartungen. Deshalb soll in einem nächsten Schritt ein genauerer Blick auf die Antwortmuster geworfen werden.

Grafik 1, auf Seite 44, zeigt die Anzahl der nichtbeantworteten Fragen (Items) getrennt für die beiden Stichproben. In der vorliegenden Studie wurden alle Fragebögen in den Datensatz aufgenommen, die mindestens eine Frage beantwortet haben. Im Ergebnis zeigt sich, dass mehr als 80 % der Befragten in beiden Stichproben 20 nicht beantwortete Fragen aufweisen. Der Anteil der sehr gut ausgefüllten Bögen mit keinem oder maximal zehn fehlenden Werten ist sogar in der Experimentalgruppe höher. Allerdings ist auch der Anteil Fragebögen in der Experimentalstichprobe die sehr schlecht ausgefüllt sind (mehr als 100 fehlende Werte) deutlich größer (11,9 % zu 2,7 %). Dementsprechend kann man nicht von einer generell schlechteren Ausfülldisziplin in der Experimentalstichprobe ausgehen, sondern der Anteil der „schlechten Ausfüller\*innen“ ist nur höher.

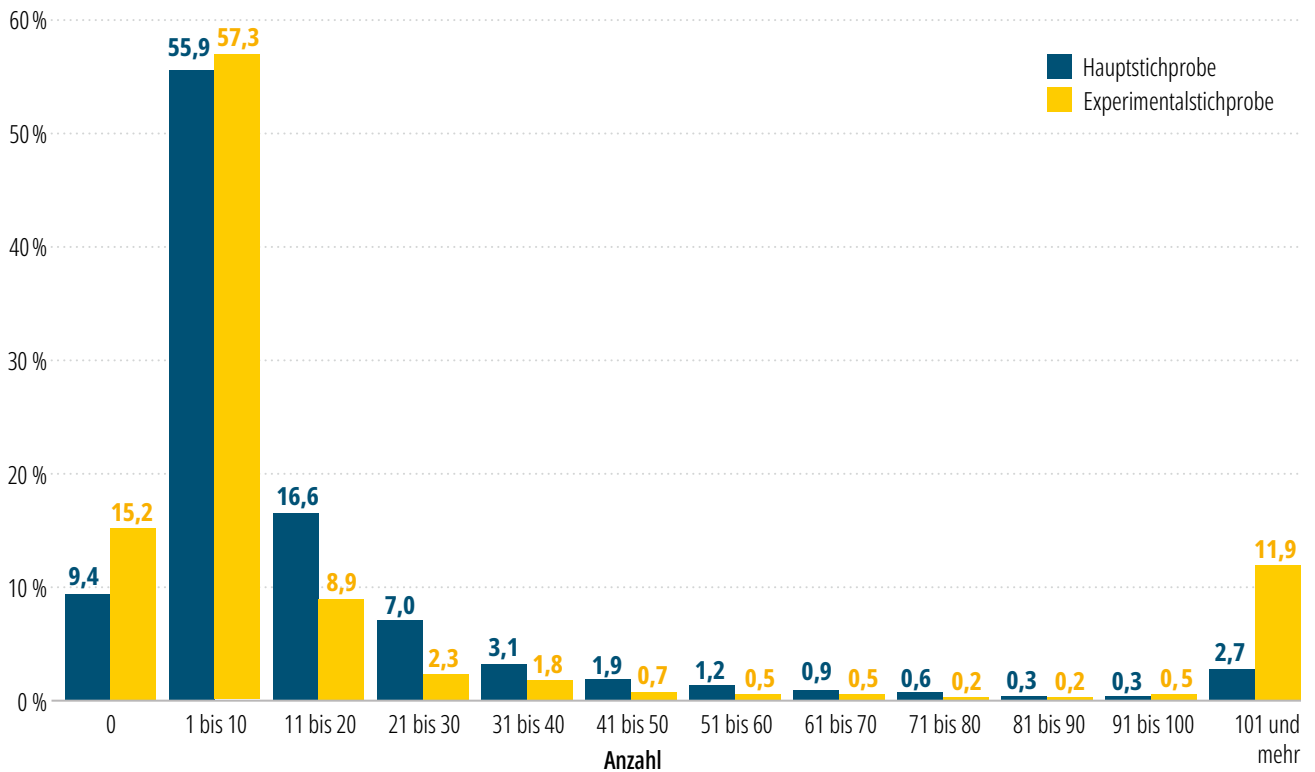
Ein Grund hierfür könnte sein, dass bei online ausgefüllten Fragebögen nach Ende des Feldzeitraumes automatisch „abgegeben“ wurde. Befragte, die den Fragebogen erst in Teilen ausgefüllt haben und geplant hatten, dies zu einem späteren Zeitpunkt fortzusetzen, das aber nicht in die Tat umgesetzt haben, gehen so in die Stichprobe mit ein. Im Falle einer schriftlichen Teilnahme muss ein teilweise ausgefüllter Fragebogen aktiv zurückgesandt werden. Bei dieser Gelegenheit werden entweder noch Fragen ergänzt oder der\*die Befragte verzichtet ggf. ganz auf die Teilnahme.

*Item-Non-Response*

*Durchschnittliche Anzahl nicht beantworteter Items in der Experimentalstichprobe höher*

<sup>15)</sup> Siehe Fußnote 13.

Grafik 1: Item-Non-Response nach Stichprobe – Anzahl der nicht beantworteten Items



Quelle: Münchner Bevölkerungsbefragung zur Stadtentwicklung 2021, eigene Berechnungen.

Um diesen Aspekt noch einmal genauer zu betrachten, sind in Tabelle 4 die Stichproben nach Art der Teilnahme untergliedert. Es zeigt sich, dass der Anteil Fragebögen, bei denen mehr als die Hälfte nicht beantwortet wurde, bei den online Teilnehmenden der Hauptstichprobe mit 5,4% deutlich über der Rate in der schriftlichen Variante liegt. In der Experimentalstichprobe liegt der Anteil der sehr schlecht ausgefüllten Fragebögen noch höher. Gleichwohl füllt der große Teil der Befragten die Fragebögen online oder schriftlich sehr gut aus, wie die niedrigen Medianwerte zeigen.

Tabelle 4: Item-Non-Response nach Stichprobe und Mode

Merkmale	Hauptstichprobe			Experimentalstichprobe	
	gesamt	online	schriftlich	gesamt	online
N	6 428	2 332	4 096	571	552
fehlende Werte					
Mittelwert	15,6	15,7	15,6	29,5	29,7
Median	6	2	9	4	4
Anteil Befragte mit mehr als 50% Missings	2,3%	5,4%	0,4%	11,0%	11,2%

Quelle: Münchner Bevölkerungsbefragung zur Stadtentwicklung 2021, eigene Berechnungen.

Als Fazit kann hier gezogen werden, dass Onlinebefragungen im vorliegenden Studiendesign zu mehr Item-Non-Response führen, v.a. im Hinblick auf den Anteil an schlecht ausgefüllten Fragebögen. Dieser Effekt verstärkt sich, wenn man die Befragten stark in Richtung Onlinebefragung „schubst“, wie in der Experimentalbedingung geschehen, wo die Hürden für die schriftliche Teilnahme höher liegen.

Auch wenn sich die beiden Stichproben hinsichtlich soziodemografischer Variablen nur wenig voneinander unterscheiden, so ist zu fragen, ob das Antwortverhalten in Bezug auf Fragen variiert, die sich speziell mit digitalen Inhalten auseinandersetzen. Dabei stehen nicht die Präferenzen für Onlineteilnahme bzw. schriftlichen Fragebogen im Vordergrund. Die Analysen sollten vielmehr aufzeigen, ob in der Experimentalstichprobe systematisch weniger „digital affine“ Personen teilnehmen, da die Onlinevariante abgelehnt wird und das Anfordern des Fragebogens als zu aufwändig empfunden wird.

*„Digitale Affinität“ in Abhängigkeit der Variante*

In der Befragung wurde erhoben, ob der Haushalt, in dem die\*der Befragte lebt, über ein Smartphone, einen Computer, Laptop oder Tablet sowie über einen Internetanschluss verfügt. Die Antwortkategorien waren „ja“, „nein, kann sich der Haushalt nicht leisten“ und „nein, hat der Haushalt aus anderen Gründen nicht.“ Für die Analysen wurden die beiden nein-Kategorien zusammengefasst. Signifikante Unterschiede zeigen sich in Bezug auf die Verfügbarkeit von Computer, Laptop oder Tablet sowie Internetanschluss: Der Anteil an Befragten, die in Haushalten ohne diese Geräte bzw. Dienste leben, ist in der Hauptstichprobe etwas höher, siehe Tabelle 5.

Neben dem Gerätebesitz wurde auch erhoben, ob die Befragten schon einmal digitale Dienste der Landeshauptstadt München genutzt haben. In der Abfrage wurde nach sechs unterschiedlichen Diensten bzw. Services differenziert: Kostenfreies öffentliches Münchner WLAN, München App, Offizielle Social-Media-Kanäle der Landeshauptstadt München, Bildungsangebote für Computer und Internet, Onlinedienste der Landeshauptstadt München (z. B. Terminvereinbarung, KitaFinder) oder öffentliche Bürgerterminals zur Nutzung dieser Onlinedienste. Betrachtet man den Anteil der Personen, die mindestens einen Dienst genutzt haben, so zeigt sich eine etwas höhere Nutzung in der Methodstichprobe (86,5 % zu 82,9 %; Chi<sup>2</sup>-Test;  $p \leq 0,05$ ).

**Tabelle 5: Gerätebesitz nach Stichprobe**

Variable		Hauptstichprobe	Experimentalstichprobe	Signifikanz (Chi <sup>2</sup> )
Haushalt verfügt über Smartphone	Ja	93,3 %	94,6 %	n.s.
	Nein	6,7 %	5,4 %	
	N	6 049	482	
Haushalt verfügt über Computer, Laptop oder Tablet	Ja	94,4 %	96,9 %	$p \leq 0,05$
	Nein	5,6 %	3,1 %	
	N	6 119	488	
Haushalt verfügt über Internetanschluss	Ja	95,1 %	97,1 %	$p \leq 0,05$
	Nein	4,9 %	2,9 %	
	N	6 129	489	

n. s. = nicht signifikant für  $p \leq 0,05$

Quelle: Münchner Bevölkerungsbefragung zur Stadtentwicklung 2021, eigene Berechnungen.

## Zusammenfassung und Fazit

### *Methodenexperiment zum Vergleich realer Mixed-Mode Designs*

Ziel des vorliegenden Aufsatzes war es, zu analysieren, ob kommunale Bürgerbefragungen als Sonderfall repräsentativer Surveys als Online-Befragungen konzipiert werden können. Hierfür wurde im Rahmen der Bevölkerungsbefragung zur Stadtentwicklung 2021 ein entsprechendes Experimentaldesign angewendet. Verglichen wurden eine Hauptstichprobe (N=20 000), die den schriftlichen Fragebogen postalisch zugesandt bekam und die Möglichkeit hatte, auch online teilzunehmen, und eine Experimentalstichprobe (N=2 000), die nur ein Anschreiben mit den Zugangsdaten zur Onlinebefragung erhielt und die Möglichkeit hatte, den schriftlichen Fragebogen gesondert anzufordern. Im Fokus des durchgeführten Methodenexperiments stand dabei nicht der Vergleich „reiner“ schriftlicher und Onlinevarianten, vielmehr wurden reale Mixed-Mode Designs verglichen, wie sie in der Praxis (kommunaler) Umfrageforschung zum Einsatz kommen.

### *Höhere Rücklaufquote in der schriftlichen Befragung*

Insgesamt zeigte sich, dass die Rücklaufquote in der Hauptstichprobe mit 33,8 % etwa 3,7 Prozentpunkte über der Rücklaufquote in der Experimentalstichprobe liegt. Das bestätigt auch die aus den bisherigen Forschungsbefunden abgeleiteten Annahmen, wonach schriftliche Befragungen mit höheren Rücklaufquoten einhergehen (vgl. u.a. Weigold et al., 2019).

Interessant ist vor diesem Hintergrund auch ein Blick auf die Modepräferenzen: so wählen zwei Drittel der Befragten den Papierfragebogen, wenn dieser angeboten wird. Muss der Fragebogen extra angefordert werden, so machen nur sehr wenige von der Möglichkeit Gebrauch. Zudem hat sich gezeigt, dass Modepräferenzen nach soziodemografischen Merkmalen variieren: Frauen, Ältere, Befragte mit höherer formaler Bildung und höherem Einkommen sowie Deutsche ohne Migrationshintergrund nehmen häufiger schriftlich teil. Dies deutet darauf hin, dass zwar eine deutliche Präferenz für den schriftlichen Fragebogen zu erkennen ist – wenn dieser vorliegt. Gleichzeitig sind aber viele Befragte bereit online teilzunehmen, wenn der Fragebogen extra bestellt werden müsste. Die Modepräferenzen wirken sich auch nicht auf eine selektive Teilnahme bestimmter Gruppen in der Experimentalbedingung aus: Tendenziell haben zwar in der Experimentalstichprobe mehr Männer, mehr Ausländer\*innen sowie wohlhabendere und höher formal gebildete Personen teilgenommen, die Unterschiede sind aber nicht signifikant.

Neben der allgemeinen Teilnahmebereitschaft ist auch die Ausfülldisziplin eine relevante Kennzahl der Datenqualität. Die Analysen zum Unit-Non-Response haben gezeigt, dass die Ausfüllbereitschaft in der Hauptstichprobe größer ist. Die Unterschiede sind dabei zum Teil auf die Teilnahmeart zurückzuführen. Schriftliche Fragebögen werden zu einem höheren Grad ausgefüllt als Onlinefragebögen – dies liegt sicher auch daran, dass letztere in der Studie nach Ablauf des Feldzeitraums „automatisch abgegeben“ wurden. Es zeigt sich aber auch, dass Onlineteilnehmende der Hauptstichprobe die Fragebögen vollständiger ausfüllen als Onlineteilnehmende der Experimentalstichprobe. Ein Grund hierfür könnte sein, dass Befragte, die in der Hauptstichprobe online teilnehmen ihre tatsächliche Modepräferenz wählen.

Schließlich wurde im Rahmen der Studie noch untersucht, ob in der Onlinevariante im Vergleich zur Hauptstichprobe verstärkt Personen teilnehmen, die eine größere „digitale Affinität“ aufweisen. Dies scheint nur in geringem Umfang der Fall zu sein. Zwar verfügen die Haushalte, in denen die Befragten leben, in der Experimentalstichprobe signifikant in größerem Umfang über Computer, Laptop oder Tablet und einen Internetanschluss, die Differenzen sind aber eher gering. Gleiches gilt für die Nutzung von Onlinediensten der Landeshauptstadt.

Kritisch zu reflektieren ist die Größe der Experimentalstichprobe: Deren Umfang ist zwar nicht unerheblich, es ist jedoch möglich, dass bei einem größeren Stichprobenumfang in der Experimentalbedingung die Differenzen durchaus signifikant sein könnten.

Wägt man die Vor- und Nachteile im Hinblick auf die Frage ab, ob man kommunale Bürgerbefragungen als Onlinebefragungen durchführen kann, so lassen die Ergebnisse dieses Methodenexperimentes den Schluss zu, dass die Datenqualität in der Onlinevariante reduziert ist: die Rücklaufquote ist geringer und die Fragebögen werden weniger gut ausgefüllt. Die Differenz in der Rücklaufquote von 3,7 Prozentpunkten berücksichtigt nicht die Tatsache, dass der Anteil der schlecht ausgefüllten Fragebögen in der Experimentalbedingung deutlich höher ist. Würde man alle Fragebögen mit mehr als 50 % fehlenden Werten als „nicht auswertbar“ zählen, so würde die Differenz in der Rücklaufquote bei 6,2 Prozentpunkten liegen. Dies bestätigt auch die Ergebnisse früherer Studien (Daikeler et al., 2020). Es werden auch tendenziell bestimmte Bevölkerungsgruppen schlechter erreicht. Allerdings sind diese Effekte nicht besonders groß.

Die Abstriche in der Datenqualität müssen aber den Vorteilen von Onlinebefragungen gegenübergestellt werden. Neben Kostenvorteilen durch Einsparung von Druck-, Porto- und Dateneingabekosten und der damit verbundenen CO<sup>2</sup>-Reduktion kann die Datenqualität durch den Wegfall der fehleranfälligen Dateneingabe und automatische Plausibilitätsprüfungen (z. B. Wertebereich, Filter) gesteigert werden. Darüber hinaus können in Onlinebefragungen neben komplexeren Filterfragen auch innovativere Datenerhebungstechniken wie Karten eingesetzt werden (Boyle et al., 2016).

Insgesamt kann man so zum Schluss kommen, dass kommunale Bürgerbefragungen als (reine) Onlinevariante durchgeführt werden können (dies wird in verschiedenen Städten auch bereits praktiziert), wenn man bereit ist, kleinere Abstriche in Sachen Datenqualität zu akzeptieren. Darüber hinaus hängt die Entscheidung auch immer von den verfügbaren Ressourcen ab.

## Autor

Dr. Michael Hanslmaier ist im Referat für Stadtplanung und Bauordnung, PLAN-HAI-21, Perspektive München und soziale Grundlagen tätig.

*Fazit: Rücklaufquote und Datenqualität in Onlinevariante schlechter*

*Onlinebefragung bieten ökonomische und methodische Vorteile*

## Literatur

- Baier, D. (2015). Sicherheit und Kriminalität in Niedersachsen. Ergebnisse einer Repräsentativbefragung. KFN-Forschungsbericht Nr. 127. Hannover: Kriminologisches Forschungsinstitut.
- Baier, D. (2018). Computer-Assisted Versus Paper-and-Pencil Self-Report Delinquency Surveys: Results of an Experimental Study. *European Journal of Criminology*, 15(4), 385–402.
- Boyle, K. J., Morrison, M., MacDonald, D. H., Duncan, R., & Rose, J. (2016). Investigating Internet and Mail Implementation of Stated-Preference Surveys While Controlling for Differences in Sample Frames. *Environmental and Resource Economics*, 64(3), 401–419.
- Čehovin, G., Bosnjak, M., & Manfreda, K. L. (2022). Item Nonresponse in Web Versus Other Survey Modes: A Systematic Review and Meta-Analysis. *Social Science Computer Review*, Online First.
- Daikeler, J., Bosnjak, M., & Manfreda, K. L. (2020). Web Versus Other Survey modes: An Updated and Extended Meta-Analysis Comparing Response Rates. *Journal of Survey Statistics and Methodology*, 8(3), 513–539.
- Faulbaum, F. (2019). *Methodische Grundlagen der Umfrageforschung*. Wiesbaden: Springer Fachmedien Wiesbaden.
- Flüter-Hoffmann, C., & Stettes, O. (2022). Homeoffice nach fast zwei Jahren Pandemie. *IW-Report (Vol. 2/2022)*.
- HDE Handelsverband Deutschland. (2021). *Online-Monitor 2021*. Köln.
- Hsu, J. W., & McFall, B. H. (2015). Mode Effects in Mixed-Mode Economic Surveys: Insights from a Randomized Experiment. *Finance and Economics Discussion Series*, 2015(008), 1–35.
- Kwak, N., & Radler, B. (2002). A Comparison Between Mail and Web Surveys: *Journal of Official Statistics*, 18(2), 257–273.
- Landeshauptstadt Hannover. (2021). *Repräsentativerhebung 2021. Innenstadt und Mobilität*. Hannover: Dezernat für Stadtentwicklung und Bauen.
- Landeshauptstadt München. (2017). *Bevölkerungsbefragung zur Stadtentwicklung 2016. Soziale Entwicklungen und Lebenssituation der Münchner Bürgerinnen und Bürger*. München: Referat für Stadtplanung und Bauordnung.
- Landeshauptstadt München. (2022). *Münchner Bevölkerungsbefragung zur Stadtentwicklung 2021*. München: Referat für Stadtplanung und Bauordnung.
- Manfreda, K. L., Bosnjak, M., Berzelak, J., Haas, I., & Vehovar, V. (2008). Web Surveys Versus Other Survey Modes: A Meta-Analysis Comparing Response Rates. *International Journal of Market Research*, 50(1), 79–104.
- Medway, R. L., & Fulton, J. (2012). When More Gets You Less: A Meta-Analysis of the Effect of Concurrent Web Options on Mail Survey Response Rates. *Public Opinion Quarterly*, 76(4), 733–746.
- Messer, B. L., Edwards, M. L., & Dillman, D. A. (2012). Determinants of Item Nonresponse to Web and Mail Respondents in Three Address-Based Mixed-Mode Surveys of the General Public. *Survey Practice*, 5(2), 1–9.

Neller, K. (2005). Kooperation und Verweigerung: eine Non- Response-Studie. ZUMA Nachrichten, 29(57), 9–36.

Sax, L. J., Gilmartin, S. K., & Bryant, A. N. (2003). Assessing Response Rates and Nonresponse Bias in Web and Paper Surveys. *Research in Higher Education*, 44(4), 409–432.

Schön, A. (2020). Die Ergebnisse der Stuttgarter Bürgerumfrage 2019 im Überblick. *Statistik und Informationsmanagement*, 1/2020, 8–27.

Shih, T.-H., & Fan, X. (2008). Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis. *Field Methods*, 20(3), 249–271.

Stadt Leipzig. (2021). Kommunale Bürgerumfrage 2020. Leipzig: Amt für Statistik und Wahlen.

Weigold, A., Weigold, I. K., & Natera, S. N. (2019). Response Rates for Surveys Completed with Paper-and-Pencil and Computers: Using Meta-Analysis to Assess Equivalence. *Social Science Computer Review*, 37(5), 649–668.