

Diskurs zum künftigen Mixed-Mode-Design im Mikrozensus

**Primär Face-to-Face- oder Online-Befragungen –
Was raten Expertinnen und Experten aus der Survey-Methodologie?**

Nina Brunel-Geuder, B.A., Dipl.Soz.Univ. Dr. Britta Knauer, Dipl.Soz.Univ. Jochen Knöller und
Dipl.Soz.Univ. Dr. Christina Wübbeke

Derzeit wird im statistischen Verbund eine lebhafte Reformdebatte über die künftige Priorisierung der Erhebungswege im Mikrozensus geführt. Angestoßen wurde sie durch die Corona-Pandemie, die seit gut eineinhalb Jahren erhebliche Einschränkungen in der Datenerhebung mit sich bringt. Im Zentrum der Diskussion steht die Frage, ob statt des bisher dominierenden Face-to-Face-Interviews künftig die Online-Befragung den primären Befragungsmodus bilden soll, ergänzt um weitere nachrangige Erhebungswege. Die Kontroverse um das Für und Wider eines solchen Wechsels und dessen wissenschaftliche Absicherung veranlassten den Ausschuss für Regionalstatistik der Deutschen Statistischen Gesellschaft (DStatG) in Kooperation mit dem Bayerischen Landesamt für Statistik, ein digitales Expertengespräch zu diesem Thema zu organisieren.

Am 18. Mai 2021 trafen sich fünf namhafte Surveyexpertinnen und -experten zu der virtuellen Veranstaltung mit dem Titel:

Die Zukunft des Mikrozensus: Face-to-Face- oder Online-Befragung als primärer Erhebungsweg? – Expertengespräch zu den Effekten von Befragungsmodes auf die Ergebnisqualität von Haushaltserhebungen

Im Rahmen von jeweils etwa zwanzigminütigen Vorträgen präsentierten die Expertinnen und Experten Frau Prof. Annette Jäckle (Understanding Society¹), Herr Prof. Christof Wolf (GESIS Panel²), Herr Prof. Mark Trappmann (PASS³), Herr Dr. Michael Bergmann (SHARE⁴) und Herr Prof. Stefan Liebig (SOEP⁵) Erkenntnisse aus den von ihnen vertretenen Haushaltsbefragungen. Dabei erfuhren die Zuhörerinnen und Zuhörer unter anderem, aus welchen Gründen in den einzelnen Befragungen bestimmte Erhebungsmodi gewählt worden sind, was man dort jeweils über die spezifischen Effekte der Erhebungswege auf die Ergebnisse weiß und ob und welche Reformen des Modus-Konzepts bereits durchgeführt worden oder geplant sind. Im

Eröffnungsvortrag von Prof. Jäckle ging es zudem ausführlich um die Frage, welche Verfahren dazu geeignet sind, die Effekte von Veränderungen im Mixed-Mode-Design⁶ einer Haushaltserhebung wissenschaftlich zu untersuchen.

In der anschließenden etwa einstündigen Podiumsdiskussion nahmen die Wissenschaftlerinnen und Wissenschaftler zu ausgewählten Fragen Stellung, die ein möglicher Wechsel von einem CAPI first- zu einem CAWI first-Ansatz⁷ im Mikrozensus aufwirft. Dabei wurden die spezifischen Vor- und Nachteile der einzelnen Erhebungswege in Beziehung gesetzt zu zentralen Charakteristika des Mikrozensus – wie zum Beispiel der Auskunftspflicht, der besonderen Länge einzelner Fragenprogramme sowie der Komplexität und Sensibilität einiger seiner Themen. Angesichts der herausgehobenen wissenschaftlichen Bedeutung des Mikrozensus, der unter anderem den Referenzrahmen für viele andere Haushaltsbefragungen in Deutschland bildet, betonten die Diskussionsteilnehmerinnen und -teilnehmer die Notwendigkeit, die geplante Reform des Mixed-Mode-Designs

Vortrag

Einführung in das Veranstaltungsthema:

Kurzportrait des Mikrozensus ab 2020

Prof. Michael Fürnrohr

Deutsche Statistische Gesellschaft



Michael Fürnrohr ist Vorstandsmitglied der Deutschen Statistischen Gesellschaft (DStatG) und Vorsitzender des Ausschusses für Regionalstatistik. Im Bayerischen Landesamt für Statistik leitet er die Abteilung für Bevölkerungs-, Haushalts- und Finanzstatistiken. Seit 2014 lehrt er an der Otto-Friedrich-Universität Bamberg und ist dort seit 2019 Honorarprofessor für das Fachgebiet „Amtliche Statistik“.

Der Mikrozensus ist seit mehr als 60 Jahren die größte jährliche amtliche Haushaltsstatistik in Deutschland. Er umfasst seit 1968 auch die EU-Arbeitskräfteerhebung (Labour Force Survey). Seit 2020 sind außerdem die europäische Erhebung zu Einkommen und Lebensbedingungen in Europa (EU-SILC) sowie seit 2021 zusätzlich die europäische Befragung zur privaten Nutzung von Informa-

tions- und Kommunikationstechnologien (IKT) integriert. Deutschlandweit wird etwa ein Prozent der Bevölkerung von den Statistischen Landesämtern mit Auskunftspflicht zu einem breiten Spektrum soziodemographischer und sozioökonomischer Themen befragt. Damit spielt der Mikrozensus in der deutschen Dateninfrastruktur eine bedeutende Rolle – nicht zuletzt als Hochrechnungsrahmen für viele andere Haushaltssurveys. Hinsichtlich der Datenerhebung galt bisher das CAPI first-Konzept, das seit Beginn der Corona-Pandemie jedoch nicht mehr umgesetzt werden kann. Stattdessen erfolgt die Datengewinnung durch einen Mode-Mix aus telefonischen Befragungen und der selbstadministrierten Auskunftserteilung via Online-Befragung oder Papierfragebogen. Diese aktuelle Situation hat im Mikrozensus eine Debatte über zwei konkurrierende Mixed-Mode-Konzepte ausgelöst: CAPI first versus CAWI first (siehe Abbildung 1). Das digitale Expertengespräch zur Zukunft des Mikrozensus möchte einen Beitrag zu dieser – nicht nur für die amtliche Statistik – wichtigen Fragestellung leisten.

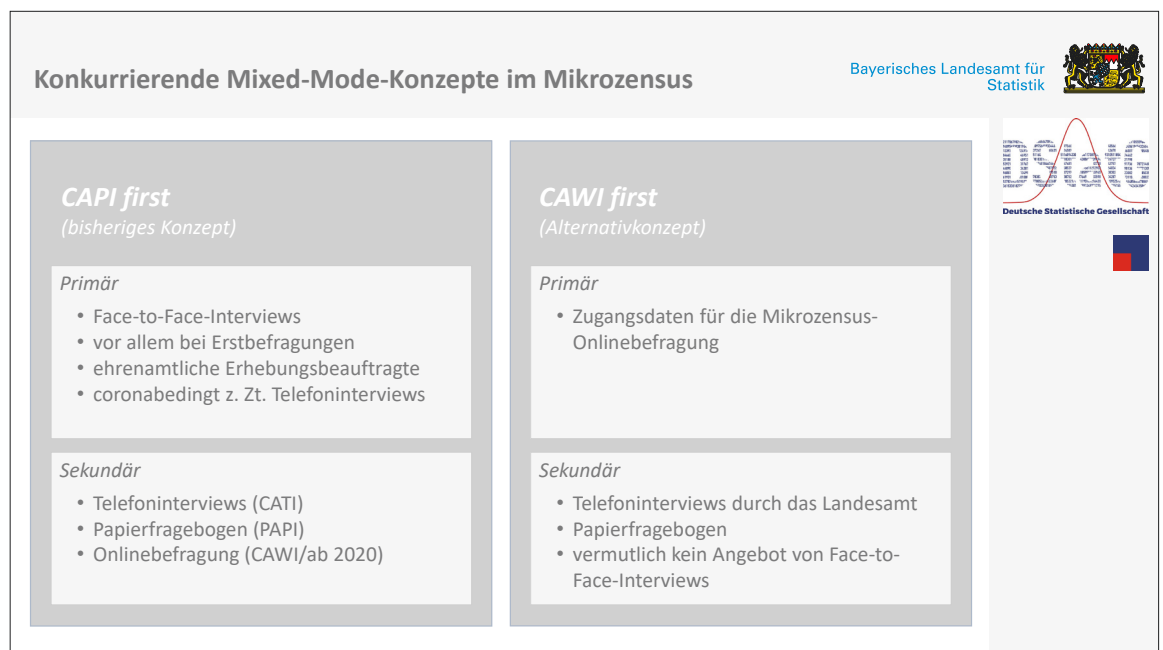


Abb. 1 – Quelle: Vortrag von Prof. Michael Fürnrohr im Rahmen des digitalen Expertengesprächs.

Vortrag

Mode-Effekte in Face-to-Face- und Online-Befragungen: Selektion oder Antwortverhalten?

Prof. Annette Jäckle

Institute for Social and Economic Research,
University of Essex

Annette Jäckle ist Professorin für Survey Methodology am Institute for Social and Economic Research der University of Essex in England. Sie ist Associate Director für das englische Haushaltspanel Understanding Society und leitet ein Forschungsprogramm zu Innovationen in Datenerhebungsmethoden. Sie untersucht unter anderem Effekte der Erhebungswege, die Nutzung von mobilen Apps für die Datenerhebung und die Zustimmung zur Zuspiegung von administrativen Daten.

Für das britische Haushaltspanel Understanding Society (vergleichbar mit dem deutschen SOEP) markiert der Übergang von einer reinen Face-to-Face- zu einer Online first-Erhebung im Jahr 2016 einen erheblichen methodischen Einschnitt. Dieser wurde aufgrund seiner Tragweite seit 2012 wissenschaftlich intensiv vorbereitet. Es stellte sich damals die Frage, ob durch die Einführung des neuen Mixed-Mode-Designs in Form von Online als primären und Face-to-Face als sekundären Erhebungsmodus Erhebungskosten eingespart werden können, ohne gleichzeitig die Panelmortalität zu erhöhen oder Messfehler zu verursachen, die mit jenen vor der Umstellung nicht mehr vergleichbar sind. Dazu war unter anderem zu klären, ob Befragte in einer Web-Befragung in ähnlicher Weise wie in einem Face-to-Face-Interview antworten. Wäre dies nicht der Fall, würden Längsschnittanalysen durch die Effekte der Umstellung beeinträchtigt. Besondere Risiken der Online-Erhebung wurden damals in einer erhöhten Fehlerquote bei der Erfassung des Haushaltsrasters (insbesondere der zu- und fortgezogenen Haushaltsmitglieder), einem möglichen

Anstieg fehlender Angaben bei bestimmten Fragen (Item-Nonresponse) und einer Zunahme von Befragungsabbrüchen gesehen.

Wie testet man Effekte des Befragungsmodus? Modus-Effekte liegen dann vor, wenn der Total Survey Error (TSE) zwischen den Befragungsmodi abweicht. Der TSE umfasst dabei alle Fehler im Prozess der Datengewinnung, die zu Verzerrungen der Schätzwerte im Vergleich zu den wahren Werten in der Grundgesamtheit führen. Der TSE setzt sich zusammen aus Fehlern der Nichtbeobachtung (u.a. Unit-Nonresponse), Fehlern der Beobachtung (fehlerhafte Antworten – also Messfehler) sowie Fehlern im Prozess der Dateneingabe und -aufbereitung. Um Effekte des Modus auf den TSE zu identifizieren, müssen alle anderen Erklärungen ausgeschlossen werden. Am besten eignen sich dazu Experimente, bei denen das Studiendesign zwischen den Modi konstant gehalten wird und die Teilnehmenden zufällig auf die Befragungsmodi verteilt werden (siehe Abbildungen 2 und 3). Doch da auch in einem Experiment das Problem des Unit-Nonresponse auftritt, überlagern sich typischerweise die Auswirkungen des Modus auf die Antworten mit den Auswirkungen der unterschiedlichen Gruppenzusammensetzungen, die durch einen modusspezifisch differierenden Unit-Nonresponse bedingt sind (Selektionseffekte). Um beide Effekte voneinander zu trennen, können sogenannte Fixed-Effects-Panelmodelle geschätzt werden – wobei nur Befragungsdaten geeignet sind, bei denen es bei den einzelnen Befragten jeweils eine Variation des Modus im Zeitverlauf gegeben hat. Eine weitere Möglichkeit ist der Vergleich der Befragungsdaten aus den unterschiedlichen Modus-Gruppen jeweils mit dem Goldstandard prozessproduzierter administrativer Daten, die annahmegemäß die wahren Werte in der Grundgesamtheit abbilden. Eine andere Option besteht darin, Variablen zu finden, welche die Selektion in die unterschiedlichen Gruppen vollkommen erklären können und deren Messung gleichzeitig nicht vom Modus beeinflusst wird. Ob die ausgewählten Variablen die beiden Bedingungen erfüllen, kann nur angenommen werden, wobei diese Annahmen sehr stark und nicht überprüfbar sind. Die gefundenen Variablen

können dann in ein Regressionsmodell aufgenommen oder zur Gewichtung verwendet werden, um die Modus-Gruppen einander anzugleichen. Ein weiterer Ansatz ist die Intention-to-Treat-Analyse, bei der in dem präsentierten Beispiel die (1) Face-to-Face-Gruppe und die (2) Mixed-Mode-Gruppe (Online first und Face-to-Face second) hinsichtlich des Antwortverhaltens miteinander verglichen werden. Die Teilnehmerinnen und Teilnehmer wurden beiden Gruppen zufällig zugeteilt. Dabei wird ausgenutzt, dass kein Selektionseffekt die Zusammensetzung beider Gruppen und somit die Antwortverteilungen beeinflusst. Denn die Face-to-Face-Befragten aus Gruppe 1 unterscheiden sich zwar von den Online-Befragten aus Gruppe 2, nicht aber von den Online- und den Face-to-Face-Befragten aus Gruppe 2 zusammengekommen. Der Effekt des Modus auf die Messung wird jedoch unterschätzt, da die Mixed-Mode-Gruppe auch Face-to-Face-Befragte enthält. Schließlich besteht die Möglichkeit, instrumentelle Variablen zu nutzen. Dazu werden die Schätzungen der Intention-to-Treat-Analyse zunächst als Basis genommen. Dann wird jedoch der Unterschied zwischen der Face-to-Face- und der Mixed-Mode-Gruppe neu skaliert mit dem Anteil der Online-Befragten aus der Mixed-Mode-Gruppe. Ergebnis ist der kausale Effekt des Modus auf die Antworten der Online-Teilnehmenden. Die Randomisierung – also die zufällige Zuteilung der Personen zu Gruppe 1 oder Gruppe 2 – wird dabei als Instrument für den Modus verwendet. Voraussetzung für die Validität dieses Instruments ist, dass die Zuweisung mit dem tatsächlichen Modus korrelieren muss (ist in diesem Fall gegeben); gleichzeitig darf die Zuweisung weder mit den wahren Werten der jeweils interessierenden Variablen noch mit den unbeobachteten Faktoren korreliert sein (beides trifft wegen der randomisierten Zuweisung zu den Gruppen 1 oder 2 zu).

Für die empirische Untersuchung des Modus-Effekts in Understanding Society konnte das kleinere sogenannte Innovation Panel (1 500 Haushalte) zur großen Hauptstudie verwendet werden. Es wurden alle genannten Untersuchungsansätze angewandt mit dem übergreifenden Ergebnis, dass sich zwar große Unterschiede im Antwortverhalten

zwischen Face-to-Face- und Online-Befragten zeigten. Jedoch waren diese überwiegend auf Selektionseffekte zurückzuführen – also darauf, dass sich beide Gruppen in ihrer Zusammensetzung voneinander unterscheiden haben. Allerdings wäre der Modus-Effekt auf die Messung wohl stärker ausgefallen, wenn in den Face-to-Face-Befragungen nicht die besonders sensiblen Fragen ausgeklammert und schriftlich erhoben worden wären. Der Instrumental-Variablen-Ansatz hat sich methodisch als der beste Weg herausgestellt, um kausale Aussagen über die Effekte des Modus auf das Antwortverhalten unter Kontrolle von Selektionseffekten zu gewinnen. Zudem war zu beobachten, dass sich die Panelmortalität durch den Übergang zum Mixed-Mode-Design nicht erhöht hat und dass finanzielle Anreize sich sehr positiv auf die Rücklaufquote auswirken. Eine Schlussfolgerung der Studie lautet außerdem, dass Face-to-Face-Interviews als sekundärer Modus für das britische Haushaltspanel unverzichtbar sind, da die Online-Befragten von den Face-to-Face-Befragten systematisch differieren. Relevante Kosteneinsparungen wurden deshalb durch den Umstieg nur in geringem Maße erzielt.

Causal effect of mode on measurement

- Aim: rule out alternative explanations
- Would like to test: $y_i^{m1} = y_i^{m2}$?
- Problem: only observe response y_i in one mode
- Solution: experiments
 - Alternative explanations* ruled out by keeping survey design the same for both modes
 - Counterfactual problem* solved by allocating modes to ex ante identical groups through randomisation
 - Test: $\overline{Y^{m1}} = \overline{Y^{m2}}$?
 - Difference in group means = effect of mode

Conclusion

Selection or measurement?

- Difficult to identify causal effect of mode on measurement
- Rule out alternative causes of differences in measurement between modes
 - Keep survey design same across modes
- Plan how to identify measurement versus selection
 - Repeated measures including “mode switchers”?
 - Good covariates to explain selection or explain measurement effects – that are themselves not affected by mode / selection?
 - Hall test experiment?**
 - Comparison with “gold standard” data?**
 - Combine single and mixed modes data to use Instrumental Variable methods?**

Abb. 2 und 3 – Quelle: Vortrag von Prof. Annette Jäckle im Rahmen des digitalen Expertengesprächs.

Vortrag

Selbstadministriert statt Face-to-Face? Ergebnisse eines großen Feldexperiments zur Machbarkeit und Datenqualität

Prof. Christof Wolf

GESIS – Leibniz-Institut für Sozialwissenschaften



Christof Wolf ist Präsident der GESIS – Leibniz-Institut für Sozialwissenschaften und Professor für Soziologie an der Universität Mannheim. Seine Forschungsschwerpunkte liegen im Bereich Umfragemethodik, Sozialstrukturanalyse und Gesundheitsforschung. Er veröffentlichte unter anderem „The Sage Handbook of Survey Methodology“ (gemeinsam mit Dominique Joye, Tom W. Smith, Yang-chi Fu, 2016).

Sozialwissenschaftliche Umfragen sind eine unverzichtbare Quelle für Informationen über die Gesellschaft und haben eine hohe Bedeutung für Wissenschaft und Politik. Bisher gilt die Face-to-Face-Befragung mit geschulten Interviewerinnen und Interviewern bei der Datenerhebung immer noch als der Goldstandard. Es lässt sich jedoch beobachten, dass die Rücklaufquoten seit geraumer Zeit deutlich abnehmen und die Durchführungskosten im Gegensatz dazu stark ansteigen. Mit einem experimentellen Design auf Basis des Fragebogens der European Values Study (EVS) wurde deshalb untersucht, welche Auswirkungen der Wechsel von einer Face-to-Face-Erhebung zu einem Mixed-Mode-Design mit Papier- und Webbefragung auf den Rücklauf und die Qualität der Datenerhebung sowie die Durchführungskosten hätte. Miteinander verglichen wurden (a) die Face-to-Face-Befragung mit dem regulären Fragenprogramm der EVS und (b) das Mixed-Mode-Design mit den selbstadministrierten Erhebungswegen in den Varianten der regulären und einer gekürzten

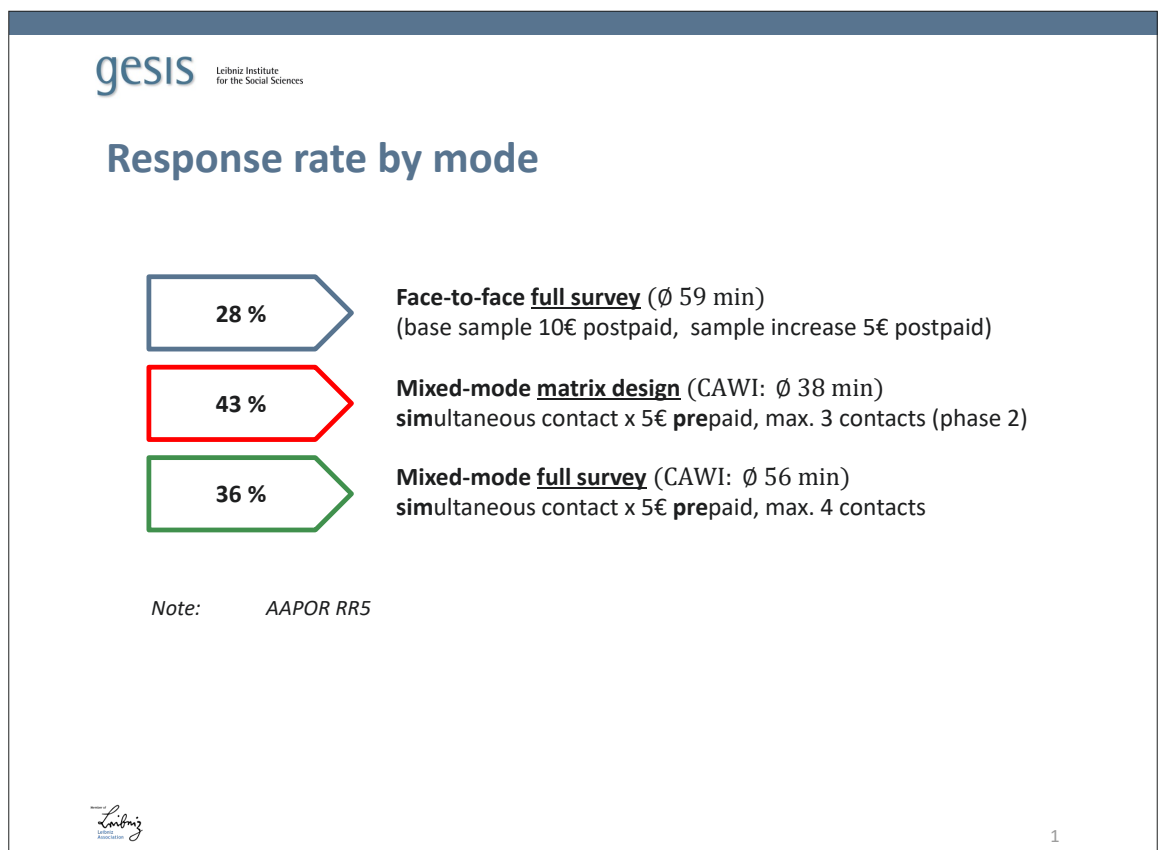


Abb. 4 – Quelle: Vortrag von Prof. Christof Wolf im Rahmen des digitalen Expertengesprächs.

Version (Matrix-Design) des Fragebogens – jeweils in Kombination mit unterschiedlichen Incentive-Strategien. Alle Erhebungen basierten auf dem gleichen Stichprobenrahmen einer Einwohnermeldeamtsstichprobe.

Insgesamt zeigen sich für alle untersuchten Varianten der selbstadministrierten Befragungen im Vergleich zur Face-to-Face-Erhebung höhere Rücklaufquoten (siehe Abbildung 4). Der höchste Rücklauf konnte beim gleichzeitigen Versand des Papierfragebogens und der Online-Kennung in Verbindung mit einem Prepaid-Incentive erzielt werden. Gleichzeitig führte dieses Design aber im Vergleich zu Face-to-Face-Interviews zu einem höheren Unit-Nonresponse-Bias (siehe Abbildung 5). So war die Teilnahmequote von Ausländerinnen und Ausländern sowie von Personen mit niedrigeren Bildungsabschlüssen bei den selbstadministrierten Befragungen in allen untersuchten Varianten etwas geringer als bei den Face-to-Face-Befragungen (siehe Abbildung 6). Bei einer ersten

Analyse des Messfehler-Bias zeigen sich zum Teil zwar signifikante Unterschiede zwischen den Erhebungsmodi, allerdings sind diese relativ gering. Für eine abschließende Beurteilung sind hier noch weitergehende Untersuchungen nötig. Es lässt sich aber festhalten, dass in Deutschland der Einsatz selbstadministrierter Erhebungsformen bei sozialwissenschaftlichen Befragungen im Hinblick auf das Kosten-Nutzen-Verhältnis Vorteile bieten kann. Neben der höheren Rücklaufquote zeigen sich gegenüber Face-to-Face-Befragungen bedeutende Kostenvorteile und durch kürzere Feldzeiten eine schnellere Verfügbarkeit der Daten.

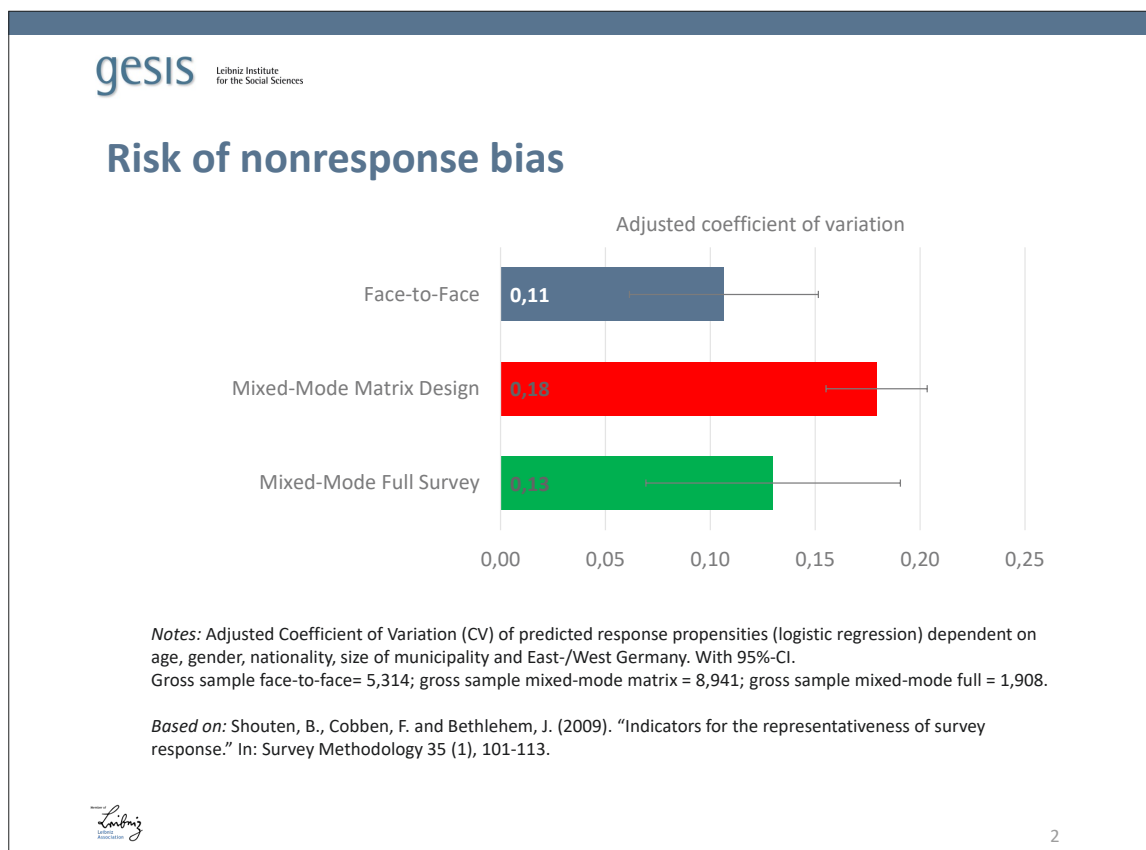


Abb. 5 – Quelle: Vortrag von Prof. Christof Wolf im Rahmen des digitalen Expertengesprächs.

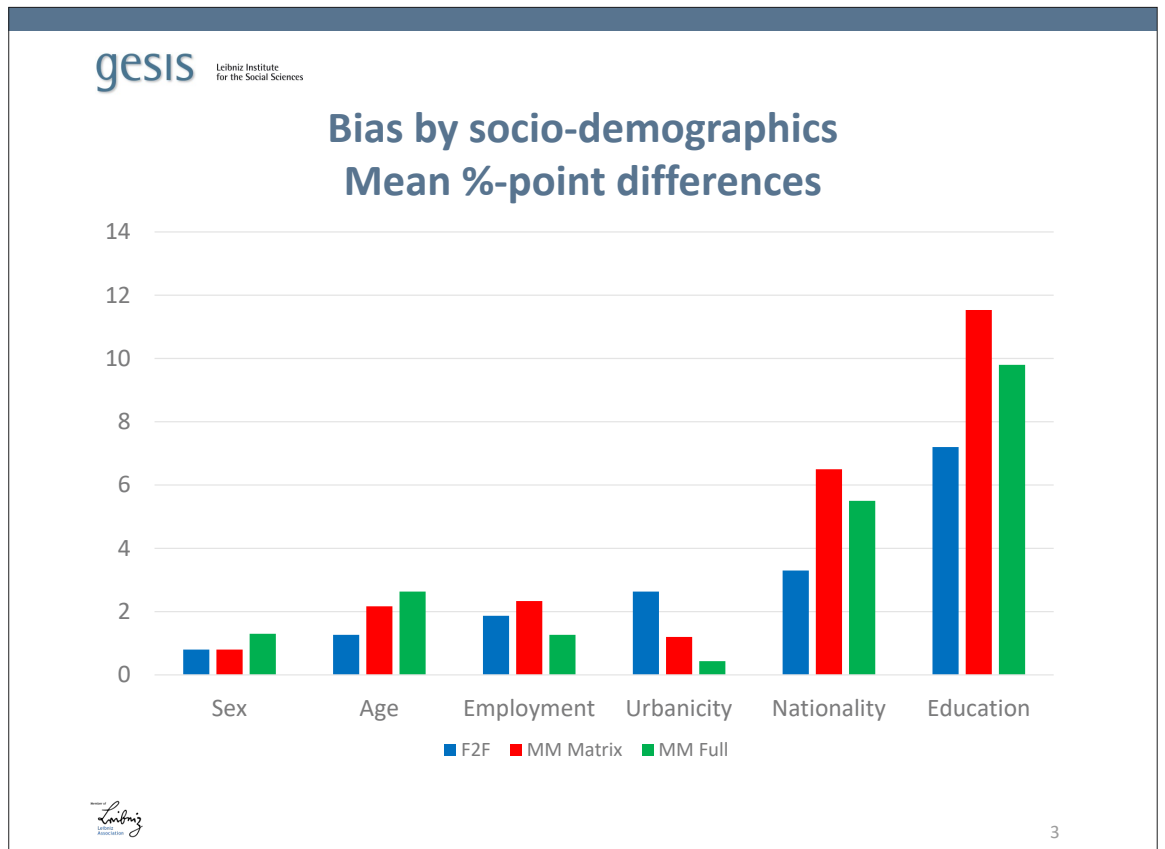


Abb. 6 – Quelle: Vortrag von Prof. Christof Wolf im Rahmen des digitalen Expertengesprächs.

Vortrag

Untersuchung von Effekten des Erhebungsmodus im Panel Arbeitsmarkt und soziale Sicherung (PASS)

Prof. Mark Trappmann

Institut für Arbeitsmarkt- und Berufsforschung



Mark Trappmann leitet seit 2007 den Forschungsbereich Panel Arbeitsmarkt und soziale Sicherung am IAB und ist seit 2012 zudem Inhaber des Lehrstuhls für Soziologie an der Otto-Friedrich-Universität Bamberg. Er ist Mitherausgeber der Zeitschrift „Soziale Welt“ und vertritt die Bundesagentur für Arbeit und das IAB im Rat für Sozial- und Wirtschaftsdaten. In seiner methodischen Forschung beschäftigt sich Prof. Trappmann insbesondere mit Fehlerquellen in Surveys wie Nonresponse oder Messfehlern und ihrer Korrektur.

Bei der Durchführung großer Panelstudien stehen Forschende vor allem vor zwei zentralen Herausfor-

derungen: Sie müssen rückläufigen Responseraten entgegenwirken und die Kosten für die Erhebung in einem möglichst moderaten Rahmen halten (siehe Abbildung 7). Beides kann durch den Einsatz eines Mixed-Mode-Designs erreicht werden, das den Respondierenden den jeweils erfolgversprechendsten Erhebungsmodus anbietet. Das Panel Arbeitsmarkt und soziale Sicherung (PASS) führt seit 2011 ein CAPI-CATI-Mixed-Mode-Paneldesign durch. Im Rahmen dieses sequenziellen Designs wird den Haushalten zunächst ein Face-to-Face-Interview (CAPI) angeboten, in bestimmten Konstellationen jedoch zu einer telefonischen Befragung (CATI) gewechselt (siehe Abbildung 8). Im Rahmen der methodischen Begleitforschung wurde jeweils der Einfluss des Mixed-Mode-Designs auf Nonresponse-, Mess- und Gesamtfehler in der Ursprungs- und in Folgewellen untersucht. Zur Bestimmung der Fehler wurden die Paradata der Erhebung mit den Survey-Daten der PASS-Erhebung sowie administrativen Daten aus den Sozialversicherungsmeldungen der Arbeitgeber sowie der Arbeitslosenversicherung und Grundsicherung zusammengespielt (siehe Abbildung 9). Es zeigt sich, dass die Befragungsergebnisse durch Nonresponse stärker verzerrt werden als durch Messfehler. Zudem hat der Erhebungsmodus einen

HINTERGRUND / AUSGANGSLAGE

- Große Panelstudien in Deutschland verwenden inzwischen alle irgendeine Art von Mixed-Mode-Designs, um abnehmende Responseraten zu bekämpfen oder Kosten zu sparen
- Fundamentalproblem aller Mixed-Mode-Studien:
 - Mit den unterschiedlichen Modi sollen verschiedene Typen von Respondenten erreicht werden (intendierter Selektionseffekt der Modi)
 - Der Preis dafür sind Messunterschiede zwischen den Modi (unintendierter Messeffekt der Modi)
 - Diese zwei Effekte sind konfundiert
 - Große Herausforderung bei der statistischen Korrektur/Separierung: Ein Set von Variablen zu finden, das die Selektion vollständig erklärt, aber selbst nicht von Modus-spezifischen Messeffekten beeinflusst ist (erfordert starke Annahmen)

// Seite 3

Abb. 7 – Quelle: Vortrag von Prof. Mark Trappmann im Rahmen des digitalen Expertengesprächs.

stärkeren Einfluss auf den Nonresponse-Bias als auf den Messfehler-Bias. Durch den Einsatz von CATI als zweitem Modus verstärkt sich sowohl in der Ursprungswelle als auch in den fünf Folgewellen der relative Nonresponse-Bias für den Arbeitslosengeld-II-Bezug und die ausländische Nationalität. Für die Erwerbstätigkeit und das Geschlecht reduziert sich der Nonresponse-Bias hingegen.

Der Messfehler-Bias verändert sich durch den Einsatz des nachgelagerten CATI-Modus kaum. Deshalb lassen sich die Modus-Effekte auf den Nonresponse-Fehler auch auf den Gesamtfehler übertragen.

SURVEYDATEN

- Panel Arbeitsmarkt und soziale Sicherung (PASS) bietet die seltene Gelegenheit unabhängiger Validierungsdaten für Respondenten und Nonrespondenten
- PASS
 - Jährliches Haushaltspanels des IAB seit 2007
 - Liefert Daten zur Forschung zur Grundsicherung für Arbeitsuchende
 - Kombiniert dazu eine Stichprobe der allgemeinen Bevölkerung mit einer Stichprobe von Leistungsbeziehenden der Grundsicherung
- Mixed-Mode-Design (ab Welle 5)
 - CAPI als Default
 - Wechsel zu CATI, falls
 - nicht kontaktierbar und Telefonnummer verfügbar
 - Verständigung auf Deutsch nicht möglich (in CATI russische und türkische Fassung)
 - Zielperson Präferenz für CATI äußert
 - Der erfolgreiche Erhebungsmodus der Vorwelle wird dann zum Default für die Folgewelle

// Seite 4

ADMINISTRATIVE DATEN (VALIDIERUNGSDATEN) UND VERKNÜPFUNG

- Integrierte Erwerbsbiographien stammen aus unterschiedlichen administrativen Verfahren
 - Sozialversicherungsmeldungen der Arbeitgeber
 - aus der Administration der Arbeitslosenversicherung und der Grundsicherung
- Eine individuelle Verknüpfung von Befragungsdaten und administrativen Daten erfordert informierte Einwilligung
- Paradata der Befragung können (als Daten über die Aufgabenerledigung des IAB im Rahmen seines gesetzlichen Auftrags) ohne Einwilligung der Befragten sowohl den Surveydaten als auch den administrativen Daten zugespielt werden

// Seite 5

Abb. 8 und 9 – Quelle: Vortrag von Prof. Mark Trappmann im Rahmen des digitalen Expertengesprächs.

Vortrag

Die Gewinnung länderübergreifender Umfragedaten in Zeiten von COVID-19: Herausforderungen und Implikationen für die Datenerhebung am Beispiel von „SHARE“ („Survey of Health, Ageing and Retirement in Europe“)

Dr. Michael Bergmann

Max-Planck-Institut für Sozialrecht und Sozialpolitik



Michael Bergmann ist wissenschaftlicher Mitarbeiter an der Technischen Universität München und am Munich Center for the Economics of Aging des Max-Planck-Instituts für Sozialrecht und Sozialpolitik. Er leitet dort die Abteilung Survey Methodology des Survey of Health, Ageing and Retirement in Europe (SHARE). Seine Forschungsinteressen liegen im Bereich der methodischen

Umfrageforschung, vor allem in der Untersuchung von Einflussfaktoren auf die Datenqualität in Längsschnittstudien.

Es lief gerade die achte Welle des Survey of Health, Ageing and Retirement in Europe (SHARE), einer länderübergreifenden Panel-Befragung zum Leben von Menschen über 50 Jahren in 27 europäischen Ländern und Israel, als die Corona-Pandemie Europa erreichte. Mitte März 2020 musste daraufhin die Befragung, die bis dahin ausschließlich mit Face-to-Face-Interviews durchgeführt wurde, gestoppt werden. Um der Wissenschaft trotzdem verlässliche Daten während der Pandemie zur Verfügung stellen zu können, wurde kurzfristig beschlossen, ein verkürztes Fragenprogramm zu entwickeln, das in allen Ländern telefonisch durchgeführt werden sollte. Da es sich hierbei auch um einen interviewergestützten Modus handelt, versprach man sich die geringsten negativen Auswirkungen auf die Datenqualität. Der Moduswechsel hatte aber zur Folge, dass die Komplexität des Fragebogens deutlich reduziert

Lessons learned & outlook II



- The future of SHARE
 - switch from pure face-to-face interviewing to a mix with other survey modes (CATI, CAWI)
 - maintain SHARE's high standards in response rates & data quality
 - adapt mode to survey content and survey population
 - develop tailored designs for specific subgroups (e.g. nursing homes)
 - investigate and document the impact of mode (selection & measurement) effects
- Interviewers still play an important role in SHARE
 - to collect physical data (e.g. handgrip strength)
 - to motivate respondents (recruitment!) and build-up of rapport
 - to help with complex/cognitively demanding questions (esp. oldest old!) → data quality!

© 2020 SHARE-ERIC - The content of this presentation is governed by the copyright laws of Germany. Any duplication, processing, distribution or any form of utilisation beyond the scope of copyright law shall require the prior written consent of SHARE-ERIC.

11

mea | MAX PLANCK INSTITUTE FOR
SOCIAL LAW AND SOCIAL POLICY
Munich Center for the Economics of Aging



This project has received funding from the European Union under grant agreement 101019213 and the European Union's Horizon 2020 research and innovation programme under grant agreements No 870626, No 101019214



Abb. 10 – Quelle: Vortrag von Dr. Michael Bergmann im Rahmen des digitalen Expertengesprächs.

werden musste. Da zudem coronaspezifische Fragen aufgenommen wurden, ist die Vergleichbarkeit zu den Vorwellen nur eingeschränkt möglich. Die Datenerhebung der telefonischen Befragung fand hauptsächlich im Juni und Juli 2020 statt.

Aufgrund des Paneldesigns lagen von vielen Haushalten bereits Telefonnummern vor und es konnte insgesamt eine sehr hohe Rücklaufquote erzielt werden (ca. 80 % im Durchschnitt über alle Länder; 84 % in Deutschland). Auf die Aufnahme neuer Haushalte (Panel Refreshment) wurde bewusst verzichtet.

Erste vorläufige Analysen der Ergebnisse zeigen beim Vergleich mit der Vorwelle zum Teil signifikante Unterschiede zwischen der persönlichen (CAPI) und telefonischen (CATI) Befragung. Hier spielen vermutlich sowohl Modus- als auch Pandemieeffekte eine Rolle. Um die Unterschiede genauer beurteilen zu können, sind jedoch weitere Analysen notwendig. Insgesamt hat sich

der pandemiebedingte Moduswechsel zu telefonischen Befragungen in der Praxis jedoch bewährt.

Um in Zukunft flexibler auf gesellschaftliche Veränderungen reagieren zu können, wird SHARE zu einem Multi-Mode-Einsatz wechseln (siehe Abbildung 10). Neben einem weiter mit persönlichen Interviews erhobenen Kernprogramm soll es zukünftig zwischen den Hauptbefragungen dynamische Befragungsmodule geben, die je nach Zielpopulation und Befragungsschwerpunkt in unterschiedlichen Modi (CAPI, CATI, CAWI) erhoben werden (siehe Abbildung 11). Aufgrund der komplexen Befragungsinhalte und der Befragung einer älteren Zielgruppe, die zum Teil kognitiv eingeschränkt ist, kann auf persönliche Interviews jedoch mittelfristig nicht vollständig verzichtet werden. Durch die persönliche Befragung kann außerdem ein Vertrauensverhältnis zu den Studienteilnehmerinnen und -teilnehmern aufgebaut werden, wie es telefonisch nur schwer möglich ist.

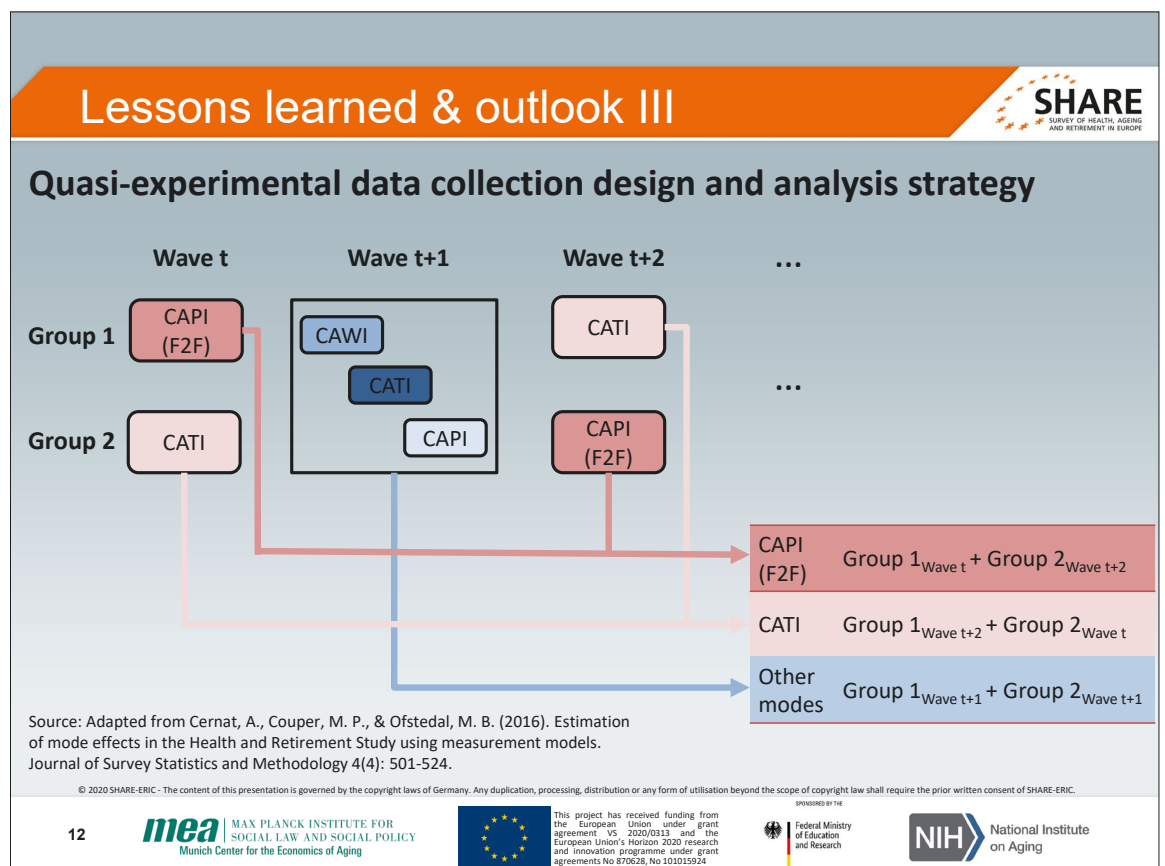


Abb. 11 – Quelle: Vortrag von Dr. Michael Bergmann im Rahmen des digitalen Expertengesprächs.

Vortrag

**Multi-Mode in Haushaltspanels –
Erfahrungen und aktuelle Weiterentwicklungen
im Sozio-oekonomischen Panel (SOEP)**

Prof. Stefan Liebig

Deutsches Institut für Wirtschaftsforschung



Stefan Liebig ist seit dem Jahr 2018 Direktor des Sozio-oekonomischen Panels (SOEP) am DIW Berlin und seit 2019 Professor für Sozialstrukturanalyse und Survey-Methodologie an der Freien Universität Berlin. In seiner Forschung beschäftigt er sich mit der Wahrnehmung und Bewertung sozialer Ungleichheiten und den methodischen Aspekten der Einstellungsmessung.

Das Sozio-oekonomische Panel (SOEP) hat bereits eine lange Tradition: Seit 1984 beschreibt es die Situation der Privathaushalte in Deutschland mit Blick auf ein breites Themenspektrum. Bisher setzte das SOEP primär auf interviewergestützte Befragungsmodi und dabei vorrangig auf das Face-to-Face-Interview (Abbildung 12 und 13).

Verschiedene Faktoren führten jedoch zu der Entscheidung, mehr Vielfalt in der Datenerhebung zu wagen. Dazu gehören die steigenden Kosten von Face-to-Face-Befragungen, steigender Unit-Nonresponse oder Partial-Unit-Nonresponse, mehr Flexibilität in den Inhalten und Messungen sowie schnellere Reaktionszeiten bei externen Ereignissen (wie z. B. die Corona-Pandemie). Vor diesem Hintergrund wurde im SOEP 2021 ein neues Konzept entwickelt, das nun ein Multi-Mode- und Multi-Device-Design in den Mittelpunkt stellt. Dabei hält das SOEP aber an seiner bisherigen Philosophie fest und betont weiterhin die zentrale Funktion der Interviewerinnen und Interviewer für die Ausschöpfung und Stabilität des Panels. Zukünftig übernehmen sie als Fallmanagerinnen und Fallmanager die Organisation des Interviewablaufs (siehe Abbildung 14). Sie entscheiden vor Ort, welche Befragungsmodi und Endgeräte für die Beantwortung des Haushaltsfragebogens sowie der Personenfragebögen eingesetzt werden können. Dafür steht ein Mix aus interviewer-, aber auch selbst-administrierten Erhebungswegen zur Verfügung: CAPI, CAWI, PAPI (Paper and Pencil Interview) sowie CASI (Computer Assisted Selfadministered Interview).

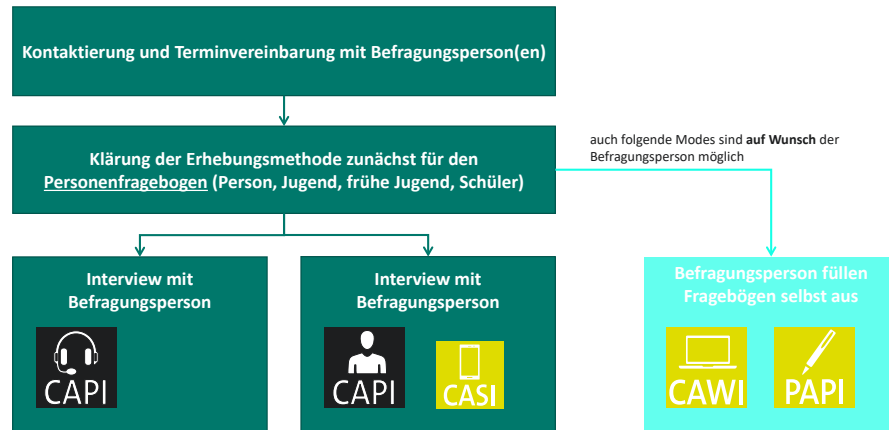


Abb. 12 – Quelle: Vortrag von Prof. Stefan Liebig im Rahmen des digitalen Expertengesprächs.

Überblick über Interviewablauf

Interviews mit den Befragungspersonen im Haushalt

infas

Durchführung der Interviews mit den weiteren Befragungspersonen:

2

**Weitere Planungen für 2022ff.****Erfahrungen mit der Nutzung der modes/devices?** (ab 2022 erste Ergebnisse)**Erfahrungen mit „InterviewerIn als FallmanagerIn“?****Engerer Kontakt zu befragten Haushalten & Personen über eine SOEP-App**

- Zielpersonen zugeschnittenes Feedback ermöglicht
- Auslagerung von Befragungsinhalten in kurze Online-Befragungen für spezielle Zielgruppen (z.B. Jugendliche)
- Incentive? Punktesystem?
- Konzept in 2021, Anwendung in 2022
- **ACHTUNG:** App muss spannend genug sein, damit sie dauerhaft genutzt wird *oder* Nutzung hoch incentiviert



12

Abb. 13 und 14 – Quelle: Vortrag von Prof. Stefan Liebig im Rahmen des digitalen Expertengesprächs.

Podiumsdiskussion: Was empfehlen die Expertinnen und Experten für den Mikrozensus?

Ein Umstieg von einem CAPI first- auf ein CAWI first-Design könne grundsätzlich funktionieren, wie Erfahrungen aus dem britischen Haushaltspanel Understanding Society oder dem US-amerikanischen Pendant zum Mikrozensus, dem American Community Survey, zeigten. Jedoch gelte es, dabei wichtige Aspekte zu beachten⁸:

Kein Verzicht auf Face-to-Face-Interviews

Die Wissenschaftlerinnen und Wissenschaftler stimmen darin überein, dass ein Verzicht auf Face-to-Face-Interviews im Mikrozensus erhebliche Risiken für die Datenqualität beinhalte. Sie raten dazu, CAPI nicht aufzugeben und weiterhin zumindest als ergänzenden Modus anzubieten. Nur so ließen sich eine vollständige Ausschöpfung der Stichprobe erreichen und Messfehler vermeiden. Dabei stützen sie sich auf folgende Argumente:

- **Responsequote.** Es sei zu vermuten, dass CAPI in Hinblick auf die Rücklaufquote im Mikrozensus besser abschneide als CATI, da die Anschriften der zu befragenden Haushalte vorab bekannt seien, nicht jedoch deren Telefonnummern. Zudem gebe es Haushalte, die beim Anruf einer Interviewerin oder eines Interviewers direkt auflegten, Anrufe mit unbekannter Nummer ignorierten oder Anrufbeantworter als Gatekeeper nutzten. Generell greife eine Bewertung der Befragungswege anhand ihrer jeweiligen globalen Response-Quote jedoch zu kurz, wenn nicht zugleich auch die nachfolgenden Aspekte berücksichtigt würden:
- **Zielgruppenspezifisch unterschiedliche Eignung von Modi**
 - **Bildungsbias.** Selbstadministrierte Erhebungsinstrumente zeigten einen deutlich größeren Bildungsbias als Face-to-Face-Interviews – selbst dann, wenn sie in Surveys eine höhere

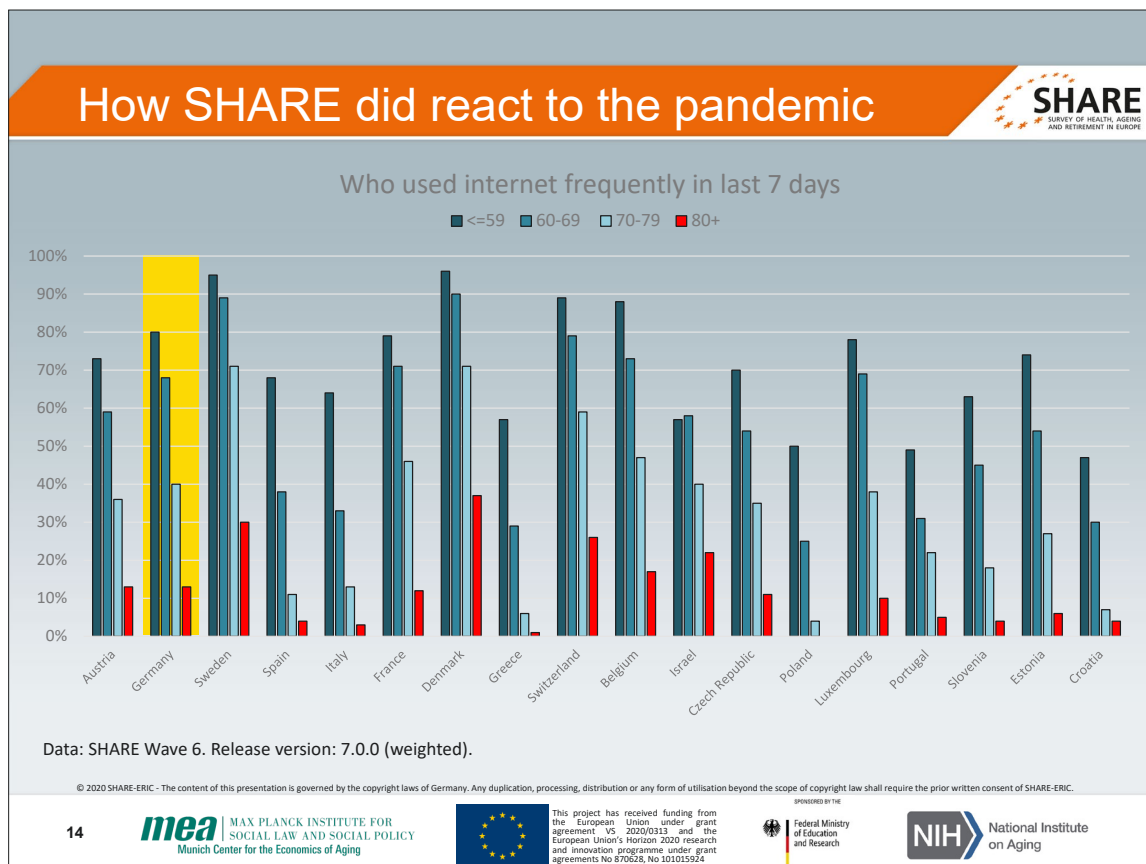


Abb. 15 – Quelle: Vortrag von Dr. Michael Bergmann im Rahmen des digitalen Expertengesprächs.

⁸ Nicht jeder Teilnehmende der Podiumsdiskussion hat sich zu allen gestellten Fragen geäußert. Dennoch wird im Folgenden auf Namensnennungen zugunsten der Übersichtlichkeit verzichtet.

Responsequote erzielen. Einer der Gründe dafür sei, dass textbasierte Befragungsformen für Personen, die Schrift im Alltag nicht verwenden könnten oder darin zumindest ungeübt seien, keine geeignete Befragungsform darstelle⁹. So liefere das Panel Arbeitsmarkt und soziale Sicherung (PASS) eine klare Evidenz dafür, dass selbstadministrierte Erhebungsformen zu einem erheblich höheren Unit-Non-response derjenigen führten, die arbeitslos, von Mindestsicherung abhängig oder depriviert seien. Würde man gering gebildete Personen unter Androhung von Zwangs- oder Bußgeld zu einer Auskunft mittels eines selbstadministrierten Fragebogens nötigen, erhielte man Ergebnisse, die invalide seien. Somit würde sich dieses Problem im Mikrozensus weniger in einem Bildungsbias infolge von Unit-Non-response äußern als vielmehr in unkontrollierbaren Messfehlern.

– *Technikbias*. Interviewergestützte Befragungswege sollten denjenigen Personen als Alternative zur Webbefragung angeboten werden, die keinen Internetzugang hätten oder im Gebrauch des Internets wenig geübt seien. Dies treffe nach Erkenntnissen aus dem Survey of Health, Ageing and Retirement in Europe (SHARE) nach wie vor auf einen bedeutenden Teil der Älteren zu (siehe Abbildung 15).

- **Lange Fragenprogramme.** CAPI eigne sich besser als CAWI und CATI für lange Fragenprogramme. Lange Online-Befragungen könnten allerdings bei attraktiver Incentivierung (wie im britischen Survey Understanding Society) ebenfalls funktionieren und böten grundsätzlich den Vorteil, dass die Fragen in Etappen beantwortet werden könnten.
- **Schwierige Fragen.** CAPI zeige Vorteile bei der Beantwortung schwieriger Fragen. So wirke das Conversational Interviewing in bestimmten Bereichen positiv auf die Datenqualität, wenn es zum Beispiel um die Erklärung komplexer Begriffe, wie die Definition des Haushalts, gehe. Die Kehrseite seien allerdings Abweichungen von der standardisierten Erhebung und die kaum mögliche Überwachung der Interviewsituation.

- **Vermeidung sozial erwünschter Antworten.** CAPI könne um einen Drop-off-Einsatz von selbst-administrierten Modi für sensible Fragen ergänzt werden.

Notwendigkeit einer wissenschaftlichen Begleitung

Aus dem hohen Stellenwert des Mikrozensus für die Wissenschaft resultiere die Verantwortung der amtlichen Statistik, Entscheidungen über ein neues Mixed-Mode-Design durch eine eigene Studie für den Mikrozensus wissenschaftlich abzusichern. Dies gelte insbesondere deshalb, weil die Ergebnisse aus freiwilligen Haushaltsbefragungen auf den auskunftspflichtigen Mikrozensus nicht unbedingt übertragbar seien. Die Effekte der verschiedenen Befragungswege auf die Ergebnisse müssten dabei aus zwei Gründen identifiziert werden:

- als empirische Grundlage für die Konzipierung eines geeigneten Mixed-Mode-Designs und
- zur Kontrolle der modusspezifischen Messfehler, damit nach dem Umstieg methodische Effekte von realen Veränderungen getrennt werden könnten.

Art der erforderlichen Untersuchung

Für den Mikrozensus böte sich der Weg der experimentellen Modus-Zuweisung an, weil so Modus-Effekte eindeutig identifiziert werden könnten. In der Forschung zum Survey Understanding Society habe sich gezeigt, dass die genauesten kausalen Aussagen resultierten, wenn auf Basis experimenteller Daten die randomisierte Zuweisung zu den Modus-Gruppen als Instrument für den Erhebungsweg genutzt werde. Weitere Möglichkeiten böten: Fixed-Effects-Panelmodelle, der Vergleich der modusspezifischen Befragungsdaten mit dem Goldstandard prozessproduzierter Registerdaten (ggf. in Kombination mit einem experimentellen Design), die Erweiterung von Regressionsmodellen um Kovariaten, welche zwar die Selektion in die unterschiedlichen Modus-Gruppen vollkommen erklärten, deren Messung jedoch selbst nicht vom Modus beeinflusst werde (nicht überprüfbare sehr starke Annahmen) oder eine Intention-to-Treat-Analyse im Rahmen eines experimentellen Designs.

⁹ Gemäß der Grundbildungsstudie „LEO 2018 – Leben mit geringer Literalität“ gab es 2018 in Deutschland 10,6 Millionen Menschen, die nicht gut lesen und schreiben konnten, sowie rund 6,2 Millionen funktionale Analphabetinnen und Analphabeten, zusammen also rund ein Fünftel der Bevölkerung. Vor acht Jahren lagen diese Anteile allerdings noch deutlich höher. Quelle: <https://www.bmbf.de/bmbf/shareddocs/pressemitteilungen/de/universitaet-hamburg-und-bunde-ichen-neue-grundbildungsstudie.html>

Orientierung an den Erkenntnissen aus anderen Ländern

Es sei ratsam, die Begleitforschung aus denjenigen europäischen Ländern auszuwerten, die bei LFS und SILC jeweils den CAPI first-Ansatz zugunsten eines neuen Mixed-Mode-Designs aufgegeben hätten. Relevante Forschung dazu gebe es insbesondere von der amtlichen Statistik in den Niederlanden. Positiv hervorzuheben sei zudem die wissenschaftliche Begleitung der Methodenumstellung beim britischen Haushaltspanel Understanding Society oder dem US-amerikanischen American Community Survey, das wie der Mikrozensus eine auskunftspflichtige Befragung sei.

Fazit

Der Mikrozensus besitzt für die Wissenschaft einen hohen Stellenwert – als Datenquelle für inhaltliche Analysen ebenso wie als Referenzrahmen für die Gewichtung und Hochrechnung vieler anderer Haushaltssurveys in Deutschland. Methodische Entscheidungen im Mikrozensus müssen daher gut begründet, gut vorbereitet und ihre Effekte auf die erhobenen Daten gut dokumentiert werden. Entsprechend rieten die Expertinnen und Experten eindringlich dazu, den diskutierten Umstieg auf ein Online first-Konzept im Mikrozensus vorab auf seine Effekte hin wissenschaftlich zu untersuchen und die gewonnenen Befunde für Nutzerinnen und

Nutzer der Daten transparent zu machen, damit diese in ihren Analysen um die methodischen Effekte wissen. Ebenso mahnten die Methodikerinnen und Methodiker zur Vorsicht, was einen vollständigen Verzicht auf Face-to-Face-Interviews im Mikrozensus betrifft. Zumindest als nachrangiger Modus sollte das persönliche Interview weiterhin denjenigen Befragtengruppen angeboten werden, die auf anderem Wege nicht oder nur unzureichend Auskunft geben könnten. Würden diese Hinweise beachtet, könne ein primär auf Online-Befragungen gestützter Mikrozensus durchaus ohne Qualitätseinbußen funktionieren.

Allerdings ist es ein herausforderndes Unterfangen, die Effekte unterschiedlicher Erhebungswege auf die Datenqualität des Mikrozensus zu identifizieren: In Befragungsdaten vermischen sich üblicherweise viele unterschiedliche Einflüsse, von denen die Wirkung des Befragungsweges nur einer ist. Um Modus-Effekte zu separieren, eignet sich ein Feldexperiment am besten. Weil der Mikrozensus jedoch eine mit Auskunftspflicht verbundene amtliche Erhebung ist, müsste für valide Befunde auch die Auskunft im Rahmen des Feldexperiments für die Teilnehmenden verpflichtend sein. Die benötigte rechtliche Grundlage für ein solches experimentelles Design zu schaffen und dieses umzusetzen, stellt eine enorm hohe Hürde dar.