

BIBB/BAuA-Erwerbstätigenbefragung 2024

Methodenbericht

Forschungsprojekt im Auftrag des
Bundesinstituts für Berufsbildung (BIBB) und der
Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA)

Durchgeführt von:
Verian (ehemals: Kantar Public)

Miriam Gensicke
† +49 (0)89 231221011
miriam.gensicke@veriangroup.com

Alexandra Strauß
† +49 (0)89 231221045
alexandra.strauss@veriangroup.com

Nikolai Tschersich
† +49 (0)89 231221035
nikolai.tschersich@veriangroup.com

Sophie Tschersich
† +49 (0)89 231221052
sophie.tschersich@veriangroup.com

München, Dezember 2024
MB_334500045

Inhaltsverzeichnis

1	Ziele und Untersuchungsanlage des Gesamtprojekts	3
2	Erhebungsinstrument	4
3	Grundgesamtheit und Anlage der Stichprobe	6
3.1	Auswahlgrundlage und Stichprobenziehung	6
3.1.1	Generierung der Auswahlrahmen für Festnetz- und Mobilfunknummern	7
3.1.2	Stichprobenrealisierung	9
3.2	Auswahl von Erwerbstätigen ab 15 Jahren	11
3.3	Selektives Teilnahmeverhalten niedrig gebildeter Erwerbstätiger	12
4	Durchführung und Erhebung	14
4.1	Interviewschulung	14
4.2	Einsatz der Interviewenden und Durchführung der Feldarbeit	15
4.3	Maßnahmen zur Qualitätssicherung	16
4.4	Ergebnisse der Feldarbeit	19
4.5	Ergebnisse der Maßnahme zur Verbesserung der Bildungsstruktur	23
4.6	Zustimmung zur Wiederbefragung und zur Datenverknüpfung	24
4.7	Kurzbefragung von Nichtteilnahmebereiten	25
5	Datenprüfung und -aufbereitung	26
6	Vercodung der offenen Angaben zur beruflichen Tätigkeit und zum Wirtschaftszweig	28
6.1	Das grundsätzliche Verfahren bei Verian	29
6.2	Berufevercodung	30
6.2.1	Vorgehensweise und allgemeine Regeln bei der manuellen Vercodung	30
6.2.2	Spezielle Regeln und Besonderheiten der jeweiligen Klassifikationssysteme	33
6.3	Branchenvercodung	34
6.3.1	Vorgehensweise und allgemeine Regeln bei der manuellen Vercodung	34
6.3.2	Spezielle Regeln und Besonderheiten der Klassifikation nach NACE	35
7	Gewichtung	37
7.1	Designgewichtung	37
7.2	Ausfallgewichtung über Strukturanpassung	39
7.3	Qualitätsindikatoren der Gewichtung	42
8	Literaturverzeichnis	47
Anhang		

1 Ziele und Untersuchungsanlage des Gesamtprojekts

Das Bundesinstitut für Berufsbildung (BIBB) und die Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA) haben im Mai 2023 Verian mit der Durchführung der „BIBB/BAuA-Erwerbstätigenbefragung 2024“ beauftragt. Ziel der Erhebung ist es, die Arbeitsbedingungen und -anforderungen, den Erwerb und die Verwertung beruflicher Kenntnisse sowie die diesbezüglichen Entwicklungen und Veränderungen im Zuge technischen und organisatorischen Wandels der Arbeitswelt zu erfassen.

Im Rahmen der Untersuchung war vorgesehen, von Oktober 2023 bis Anfang April 2024 rund 20.000 Erwerbstätige ab 15 Jahren telefonisch und computerunterstützt, d. h. per CATI (Computer Assisted Telephone Interview) zu befragen. Die Telefonerhebung erfolgte anhand einer Dual-Frame-Stichprobe, d. h. es wurden zu etwa gleichen Teilen sowohl Interviews über Festnetznummern als auch über Mobilfunknummern realisiert. Die durchschnittliche Interviewlänge war mit 40 Minuten geplant. Themen der Befragung waren die Arbeitsanforderungen und Tätigkeiten von Erwerbstätigen, ihre Arbeitsbedingungen, gesundheitliche Beschwerden, sowie der Erwerb und die Verwertung beruflicher Qualifikationen.

Der vorliegende Methodenbericht beschreibt zunächst in Abschnitt 2 die Entwicklung des Erhebungsinstruments. Anschließend wird in Kapitel 3 die Definition der Grundgesamtheit und die Anlage der Stichprobe dargestellt. Die Schulung, der Einsatz und die Kontrolle der Interviewenden sowie die Ergebnisse der Feldarbeit werden in Abschnitt 4 erläutert. Informationen zur Datenprüfung und -aufbereitung gibt Kapitel 5. In Abschnitt 6 wird das Vorgehen bei der Vercodung der beruflichen Tätigkeit sowie des Wirtschaftszweigs dokumentiert. Abschließend werden das Gewichtungsmodell sowie die Ergebnisse der Gewichtung dargestellt.

Das Projekt wurde bei Verian von einem Projektteam bearbeitet. Die Projektleitung lag bei Miriam Gensicke, Alexandra Strauß und Nikolai Tschersich. Für die Vercodung war Sophie Tschersich verantwortlich. Das Projektteam wurde beratend unterstützt durch Dr. Josef Hartmann. Für die Durchführung der CATI-Erhebung, inklusive Stichprobenziehung und Fragebogenprogrammierung, war die Kantar GmbH als Unterauftragnehmer von Verian (ehemals Kantar Public) verantwortlich. Verian und Kantar waren bis zur Ausgliederung von Verian ein gemeinsames Unternehmen und arbeiten nach wie vor eng zusammen. Zuständig für die CATI-Projektkoordination war bei Kantar Sören Schleypen, für die Fragebogenprogrammierung Gerd Döring, für die Stichprobensteuerung Ringo Uhlemann. Die Gewichtung erfolgte bei Verian durch Rita Reimer.

2 Erhebungsinstrument

Das Erhebungsinstrument der Befragung wurde vom BIBB und der BAuA auf Basis der Erhebung 2017/2018 entwickelt und am 16. Juni 2023 in einem persönlichen Treffen mit der Projektgruppe in den Räumlichkeiten des BIBB diskutiert. Wesentliche Bestandteile des Erhebungsinstruments sind dabei identisch zur letzten Erhebung. An einigen Stellen wurden Fragen herausgenommen und durch neue Fragestellungen ersetzt. Die Reihenfolge der Fragen wurde an einigen Stellen verändert. Der Frageblock zur beruflichen Ausbildung wurde dahingehend vereinfacht, dass statt bis zu fünf nur noch bis zu zwei berufliche Ausbildungsabschlüsse erfasst werden, nämlich der höchste und – im Fall von mehreren Abschlüssen – der chronologisch erste Abschluss.

Die Zufallsauswahl der Zielperson erfolgte im Rahmen der Festnetzstichprobe unter allen deutschsprachigen Personen ab 15 Jahren. Dies entsprach dem Vorgehen in der Erhebungswelle 2012, während 2018 in der Festnetzstichprobe die Zufallsauswahl unter den Erwerbstätigen ab 15 Jahren erfolgte. Da dieses Vorgehen mit einer deutlich geringeren Inzidenz verbunden war, die vermutlich auf versteckte Verweigerungen zurückzuführen war, erfolgte die Zufallsauswahl in der Erhebung 2024 unter allen Personen ab 15 Jahren. Verian überarbeitete in Abstimmung mit den Auftraggebern die Fragen zur Zufallsauswahl der Zielperson im Intervieweinstieg.

Das Erhebungsinstrument wurde bei Verian in ein Programm für computergestützte telefonische Interviews (CATI) umgesetzt. Als Befragungssoftware wurde dabei NIPO 32 genutzt.

Das BIBB und die BAuA erhielten einen Testzugang zum CATI-Fragebogen, um selbst Testinterviews durchführen zu können. Vor dem Start der Haupterhebung wurden zwei Pretests durchgeführt.

Die Feldarbeit des ersten Pretests fand vom 24. Juli 2023 bis zum 5. August 2023 statt, der zweite Pretest wurde in der Zeit vom 4. bis zum 11. September 2023 durchgeführt. Die Änderungen, die auf Basis des ersten Pretests vorgenommen wurden, wurden im Rahmen einer Videokonferenz am 16. August mit der Auftraggeberseite abgestimmt. Direkt vor Start der Feldarbeit des ersten Pretests fand die Einweisung für die Interviewenden als Videokonferenz durch die Projektleitung von Verian und unter Anwesenheit von BIBB und BAuA statt. Die Schulung wurde aufgezeichnet und stand damit auch Interviewern und Interviewerinnen zur Verfügung, die am Schulungstermin selbst nicht teilnehmen konnten. Vor dem zweiten Pretest fand ebenfalls direkt vor Start der Feldarbeit eine Einweisung per Videokonferenz durch die Projektleitung bei Verian statt. Der Fokus lag vor allem auf den Änderungen, die für den zweiten Pretest vorgenommen wurden, sowie auf Hinweisen für die Interviewenden, die sich aus den beiden Mithörterminen im ersten Pretest ergaben. Zudem wurden die Erläuterungen für die Fragen zur beruflichen Ausbildung überarbeitet und ein Fact Sheet für die Interviewenden vom BIBB zur Verfügung gestellt.

Im Rahmen des ersten Pretests wurden 200 Interviews mit Erwerbstätigen durchgeführt, beim zweiten Pretest waren es 100 Interviews.

Im ersten Pretest wurden bei ausgewählten neuen Fragen kognitive Nachfragen eingesetzt, um zu ermitteln, ob bestimmte Begriffe verständlich waren oder ob die Beantwortung schwergefallen ist und warum.

Um Probleme mit dem Instrument aufzudecken, nutzte die Auftraggeberseite sowie die Projektleitung während beider Pretests die Möglichkeit, bei den Telefoninterviews mitzuhören. Darüber hinaus wurden im Anschluss an jedes Interview den Interviewenden einige Fragen zum eben durchgeführten Interview gestellt. Mit dieser Nachbefragung sollten alle Schwierigkeiten, die sich im Interviewverlauf gezeigt haben, erfasst werden. Jeweils nach Abschluss der Feldarbeit der beiden Pretests fand ein Debriefing-Termin mit den Interviewenden und der Supervision statt, bei dem ebenfalls Probleme und Verbesserungsvorschläge gesammelt wurden.

Änderungsbedarf zeigte sich in erster Linie aufgrund der zu hohen Interviewdauer. Darüber hinaus gab es einzelne Fragen, bei denen sich Verbesserungsbedarf zeigte. Das methodische Vorgehen sowie die Ergebnisse dieser Pretests sind in jeweils eigenen Berichten dokumentiert.

Das Erhebungsinstrument beinhaltete unterschiedlich lange Themenblöcke mit folgenden Frage-nummern, die Übersicht 1 zu entnehmen sind:

Übersicht 1: Inhalte des Fragebogens

Frageblock	Inhalt
F100ff:	Aktuelle Berufstätigkeit
F200ff:	Arbeitszeit und Arbeitsort
F300ff + F503:	Ausgeübte Tätigkeiten
F400ff + F327:	Berufliche Anforderungen
F500ff:	Arbeitsverhältnis
F600ff/F700ff/F800ff:	Arbeitsbedingungen
F900ff:	Berufs- und Lebensziele
F1000ff:	Veränderungen in den letzten zwei Jahren
F1100ff + S3:	Schulbildung
F1200ff:	Berufsausbildung (in bis zu 2 Schleifen)
F1225ff:	Zusammenhang (letzte) Ausbildung und Beschäftigung
F1300ff:	Weiterbildung
F1400ff:	Berufsverlauf
F1450ff:	Arbeitszufriedenheit
F1500ff:	Gesundheit
F1600ff:	Zur Person

Die Informationen, die erhoben wurden, bezogen sich größtenteils auf die befragte Person und ihre aktuelle berufliche Tätigkeit. Hierbei wurden auch einige Informationen zum Betrieb, in dem die Person aktuell tätig ist, erhoben. Retrospektive Informationen wurden zur schulischen und beruflichen Ausbildung erhoben. Für den Zeitraum der letzten zwei Jahre wurden Informationen zur beruflichen Weiterbildung sowie zu Veränderungen im Arbeitsumfeld erfragt.

3 Grundgesamtheit und Anlage der Stichprobe

Grundgesamtheit der BIBB/BAuA-Erwerbstätigenbefragung 2024 waren deutschsprachige Erwerbstätige in Deutschland ab 15 Jahren mit einer Arbeitszeit von mindestens zehn Stunden pro Woche. Als erwerbstätig wurde definiert, wer zum Zeitpunkt der Befragung eine bezahlte bzw. eine mit einem Einkommen verbundene Tätigkeit ausübte. Diese Definition wurde an verschiedenen Stellen präzisiert:

- Personen, die ihre Tätigkeit kurzzeitig, d. h. bis zu einer Dauer von maximal drei Monaten, unterbrochen hatten, wurden einbezogen. Hierbei konnte es sich um Mutterschutz oder Elternzeit sowie um Sonderurlaub aus einem anderen Grund handeln.
- Ebenso eingeschlossen wurden Personen, die einer vergüteten Tätigkeit neben einer Ausbildung oder einem Studium oder im Rahmen eines Referendariats oder einer Facharztausbildung nachgingen.
- Auch mithelfende Familienangehörige wurden befragt. Sie zählen als erwerbstätig, obwohl sie u. a. darüber definiert sind, dass sie ohne feste Vergütung im Betrieb oder Geschäft eines Familienmitglieds mitarbeiten.
- Ausländer und Ausländerinnen wurden bei der Untersuchung einbezogen, wenn sie ausreichend deutsch sprachen.
- Dagegen wurden Personen, die der bezahlten Tätigkeit ehrenamtlich nachgehen, nicht befragt.
- Ausgeschlossen wurden auch Personen in einem Beschäftigungsverhältnis, das notwendiger Bestandteil einer Ausbildung ist, sowie Tätigkeiten im Rahmen eines Praktikums.
- Ebenso ausgeschlossen wurden Wehr- oder Freiwilligendienstleistende sowie Personen, die sich in einem freiwilligen sozialen oder ökologischen Jahr befanden.

3.1 Auswahlgrundlage und Stichprobenziehung

Die Verallgemeinerbarkeit von der Stichprobe auf die entsprechende Grundgesamtheit wurde über eine mehrstufige Zufallsauswahl sichergestellt. Im Rahmen einer Screenerhebung wurden Personen ab 15 Jahren in Privathaushalten in deutscher Sprache, die im Sinne der Untersuchung erwerbstätig sind, identifiziert und befragt.

Die Entwicklung, dass es Personen gibt, die ausschließlich per Mobiltelefon zu erreichen sind und keinen Festnetzanschluss mehr besitzen („Mobile-Only“), darf bei der Stichprobenanlage nicht vernachlässigt werden (Non-Coverage). Im Jahr 2022 verfügten 17 % der Haushalte nicht über ein Festnetztelefon und nur 2 % nicht über ein Mobiltelefon. Hinzu kommt, dass es sich bei den sogenannten Mobile-Onlys um Personen mit spezifischen Merkmalen handelt. So sind diese in Deutschland „häufiger jünger, eher männlich, stammen eher aus Ostdeutschland und leben eher als Singles“ (Häder, Sand 2019, S. 61). Gleichzeitig gibt es nach wie vor Personen, die telefonisch ausschließlich über einen Festnetzanschluss erreichbar sind (s. auch Strippel, Emmerich 2021, S. 5). Bei Letzteren handelt es sich laut Häder und Sand (2019, S. 62) „häufiger um ältere, auf dem Land in Westdeutschland lebende Personen“. Daher wurde eine Dual-Frame-Stichprobe konzipiert, bei der der Anteil der Interviews mit Erwerbstätigen über Mobilfunk auf 50 % festgelegt wurde.

Der Auswahlrahmen für die Stichprobenziehung wurde zunächst von dem Institut BIK über ein Random-Digit-Dialing-Verfahren erstellt, das im nächsten Abschnitt näher erläutert wird. Bei diesem

Verfahren liegen in der Bruttostichprobe unter den Mobilfunknummern deutlich mehr ungültige bzw. nicht geschaltete Nummern vor als in der Festnetzstichprobe. Aufgrund dieses erhöhten Anteils an ungültigen Nummern unter den Mobilfunknummern waren mehr Nummern in der Bruttostichprobe nötig, um die gewünschte Interviewzahl zu realisieren, als dies in der Festnetzstichprobe der Fall war. Das Ziel des Dual-Frame-Ansatzes war es, Mobile-Onlys in ausreichendem Umfang zu erfassen. Aus diesem Grund bezieht sich der Anteil der Mobilfunknummern auf den Anteil der realisierten Interviews aus der Mobilfunkstichprobe. Der Anteil von 50 % Interviews aus der Mobilfunkstichprobe wurde dabei über eine Nettosteuerung realisiert, wobei der benötigte Bruttoansatz höher als beim Festnetzanteil war. Grundsätzlich war es auch möglich, dass die über Mobilfunk kontaktierten Personen das Interview über Festnetz führen konnten, sofern sie das wünschten und über einen Festnetzanschluss verfügten. In diesen Fällen zählten die Interviews zur Mobilfunkstichprobe, da sie über Mobilfunknummern kontaktiert wurden.

Um der bekannten geringeren Teilnahmebereitschaft von Personen mit niedriger Bildung entgegenzuwirken, kam für einen begrenzten Teil der Festnetzstichprobe ein zweiter Auswahlrahmen zum Einsatz (letztlich bei rd. 7% der insgesamt realisierten Nettostichprobe). Dabei handelte es sich um Telefonnummern, für die aus vergangenen telefonischen Befragungen u. a. Informationen zur Bildung vorlagen. Das Zufallsverfahren dieser vorherigen Telefonbefragungen entsprach dabei sowohl für die Auswahl der Telefonnummern als auch für die Auswahl der Zielpersonen dem der BIBB/BAuA-Erwerbstätigenbefragung. Der zweite Auswahlrahmen wurde nur bei der Festnetzstichprobe und damit auf Haushaltsebene genutzt. Ausgehend von der Annahme der Bildungshomogamie im Haushalt wurde die vorliegende Telefonnummer nur für den Erstkontakt genutzt. Die Auswahl der Zielperson für die BIBB/BAuA-Erwerbstätigenbefragung erfolgte dann über die gleiche Zufallsauswahl, den Schwedenschlüssel (bzw. Kish-Methode), wie bei den frei generierten Telefonnummern. Unterschiede in den Teilnahmewahrscheinlichkeiten wurden im Rahmen der Designgewichtung berücksichtigt. Näheres zum Stichprobenverfahren wird in Abschnitt 3.3 und zum Gewichtungsverfahren in Kapitel 7.1 erläutert.

3.1.1 Generierung der Auswahlrahmen für Festnetz- und Mobilfunknummern

Der Auswahlrahmen für beide Stichproben wurde über ein Random-Digit-Dialing erstellt. Die Stichprobenziehung der Festnetznummern und Mobilfunknummern basiert auf dem Telefon-Master-Sample von Kantar. Die Dual-Frame-Befragung wurde mit zwei vergleichbaren Auswahlrahmen als kombinierte Festnetz- und Mobilfunkstichprobe durchgeführt. Beide Sampling-Frames bezogen sich auf die gleiche Grundgesamtheit und konnten ex-post auf Basis einer Designgewichtung so zusammengeführt werden, dass sie gemeinsam ein repräsentatives Abbild der Grundgesamtheit bilden. Das Dual-Frame-Verfahren führt zu verzerrungsfreien Stichproben ohne Klumpeneffekte. Sie entsprechen den Anforderungen an Zufallsstichproben, die auf der Wahrscheinlichkeitstheorie aufbauen (probability sampling).

Das Telefon-Master-Sample ist als Dual-Frame-Stichprobe konzipiert. Das bedeutet, dass sowohl Festnetz- als auch Mobilfunknummern für die Stichprobe ausgewählt werden. Das Random-Digit-Dialing wurde gemäß dem ADM-Standard (ADM = Arbeitsgemeinschaft Deutscher Markt- und Sozialforschungsinstitute) durchgeführt, eine Erweiterung des sog. „Gabler-Häder-Verfahrens“.¹ Datenbasis

¹ Das „Gabler-Häder-Verfahren“ (vgl. z. B. Gabler, Häder 1998, für eine einfache Zusammenfassung s. auch Strippel, Emmerich 2021, S. 4) stellt sicher, dass innerhalb eines Ortsnetzes Rufnummern verzerrungsfrei gezogen werden können – unabhängig davon, ob eine Rufnummer eingetragen ist oder nicht, und unabhängig von der Dichte und Häufigkeit der Einträge im Telefonbuch. Dazu werden in einem ersten Schritt auf Basis der Rufnummernstammliste der Bundesnetzagentur (BNA) und aktueller Telefonverzeichnisse die sogenannten Nummernstämme identifiziert. Bei den Nummernstämmen handelt es sich um Telefonnummern ohne die beiden letzten Ziffern. Im zweiten Schritt werden diese dann um alle Zahlen von 00 bis 99 ergänzt. Die entscheidende Erweiterung gegenüber dem „Gabler-Häder-Verfahren“ ist, dass durch die

war die Auswahlgrundlage für Fest- und Mobilfunknetznummern der Arbeitsgemeinschaft ADM-Telefonstichproben.

Erstellungsbasis der ADM-Auswahlgrundlage waren die Rufnummernstammlisten der Bundesnetzagentur (BNA) für Mobil- und Festnetz sowie das aktuelle Telefonverzeichnis. Die BNA weist Nummernblöcke aus, die sie den Netzbetreibergesellschaften zugeteilt hat. Der BNA-Bestand ist durch 10er-, 100er-, 1.000er- und 10.000er-Blöcke gekennzeichnet. Im Rahmen eines Abgleichs mit dem aktuellen Telefonbuch wurde ermittelt, ob die generierte Nummer im Telefonbuch eingetragen ist und ob – soweit erkennbar – es sich dabei um einen gewerblichen Anschluss handelt. Wenn Teile in den Einträgen offensichtlich für einen gewerblichen Anschluss sprechen (bspw. „GmbH“ oder „Ltd.“), werden diese Telefonnummern für Personenbefragungen nicht berücksichtigt.

Dieses erweiterte Gabler-Häder-Verfahren sichert – auch durch die jährliche Aktualisierung – die Inklusion aller vergebenen Rufnummern mit berechenbaren Auswahlwahrscheinlichkeiten. Es garantiert die Selektion von Rufnummern ohne Eintrag und ebenso die Auswahl von Nummern, die in Blöcken ohne Eintrag vorliegen. Dies gilt für Mobilfunk- und Festnetznummern. Für die Praxis bedeutet dies, dass erstens die wachsende Zahl von Anschlüssen bei den neueren Netzbetreibergesellschaften enthalten ist. Zweitens erhalten auch Personen, die über keinen Festnetzanschluss, dafür aber über Mobilfunk verfügen, eine positive Auswahlchance.

Regionalisierung der Festnetzstichprobe

Jede Festnetznummer des ADM-Systems wurde auf Gemeindeebene regionalisiert. Die Festnetzstichprobe wurde so gezogen, dass sich eine ausgewogene regionale Verteilung ergab. Eingetragene Rufnummern wurden mit den Informationen des Verzeichnisses direkt verortet. Nichteingetragene Nummern waren nur dann eindeutig verortbar, wenn sich alle eingetragenen Rufnummern des betreffenden Blockes in einer einzigen Gemeinde befanden. War dies nicht der Fall, wurde im Telefon-Master-Sample-System bei den generierten Rufnummern mit mehreren möglichen Gemeindekennziffern eine per Zufall innerhalb des betreffenden Blockes ausgewählt. Diese Zufallsauswahl war mit Bedeutungsgewicht gesteuert; die Häufigkeitsverteilung der Gemeindekennziffern der nichteingetragenen Nummern im jeweiligen Block entsprach der der eingetragenen.

Bei den Festnetznummernblöcken ohne einen einzigen Eintrag war die Gemeindezuordnung komplexer. Es wurden im Falle eines Nummernblockes der Telekom die Gemeindekennziffern der Nachbarblöcke mit Einträgen angesetzt. Bei anderen nationalen und überregionalen Anbietern erfolgte diese Zuordnung pro Ortsnetzvorwahl gemäß der Verteilung der Einträge dieser Gruppe von Netzbetreibergesellschaften. Im Fall von regionalen Anbietern war die Verteilung der Gemeindekennziffer pro Vorwahl des jeweils einzelnen Anbieters entscheidend.

Die Schichtung der Festnetzstichprobe erfolgte zum einen anhand von Kriterien der amtlichen Gebietseinteilung, zum anderen anhand der BIK-Gemeindetypen. Für die Allokationsrechnung wurde das Cox-Verfahren verwendet. Die allozierte Sollverteilung des Schichtungstableaus wurde haushaltsproportional auf die jeweiligen schichtangehörigen Gemeinden verteilt und so die Ziehungsvorgabe berechnet. Die Ziehung der Rufnummern erfolgte pro Gemeinde per reiner Zufallsauswahl.

Die Stichprobe mit Festnetznummern stellt eine Haushaltstichprobe dar, bei der ein zweistufiges Auswahlverfahren verwendet wurde. Im Interview wurde mit dem Schwedenschlüssel (Kish-Selection-Grid) zufällig eine Zielpersonenauswahl innerhalb des jeweiligen Haushaltes getroffen (vgl. dazu Abschnitt 3.2).

Nutzung der Rufnummernstammliste der BNA auch Nummernstämme in die Auswahlgrundlage einbezogen werden, zu denen keine Telefonnummer in einem aktuellen Telefonverzeichnis gelistet ist.

Keine Ex-ante-Regionalisierung der Mobilfunkstichprobe

Mobilfunknummern wurden mit einfacher Zufallsauswahl gezogen. Die Auswahl der Personen erfolgte mit der Selektion der Nummern in der Mobilfunkstichprobe als einfache Zufallswahl: Der kontaktierte Nutzer des Gerätes war eine Zielperson, sofern sie 15 Jahre oder älter und Hauptnutzer des Gerätes war. Für die Mobilfunknummernblöcke konnte keine regionale Verortung vorgenommen werden, da die Eintragsdichte mit ca. 0,5 % generell zu niedrig und ein regionaler Rückschluss einer eingetragenen Nummer auf andere Nummern im Nummernblock nicht möglich war. Eine Ex-ante-Schichtung und regionale Filterung entfiel somit für dieses Subsample. Im Telefon-Master-Sample wurde daher eine Ex-post-Schichtung im Verlauf der Feldarbeit nach Abfrage der Postleitzahl und des Wohnorts im Interview verwendet (siehe dazu Abschnitt 3.1.2).

Die Auswahlgrundlage der Festnetz- sowie der Mobilfunkstichprobe enthält lediglich Telefonnummern und qualifizierende Sekundärmerkmale, jedoch keine Namen und Adressen. Die Nummern aller Personen, die auf keinen Fall im Zusammenhang mit einer Befragung angerufen werden wollten, werden in einem zentralen Register gesperrt. Alle Telefonnummern der Bruttostichprobe werden vor dem ersten Anruf mit dieser Liste abgeglichen und gesperrte Nummern werden nicht kontaktiert.

3.1.2 Stichprobenrealisierung

Beim vorliegenden Dual-Frame-Design wurde im Vorfeld festgelegt, welches Mischungsverhältnis Festnetz zu Mobilfunk für die Nettofallzahl erreicht werden sollte. Ausschlaggebend war dabei, ob der erste Kontakt über eine Mobilfunk- bzw. Festnetznummer stattfand, und nicht, ob das Interview vollständig über Mobilfunk realisiert wurde oder ob für das Interview oder Teile des Interviews auf eine andere Nummer umgestiegen wurde. Die Bruttostichproben wurden separat alloziert, getrennt gezogen, aber gemeinsam in der Feldarbeit abgearbeitet. Bei beiden Ziehungen waren – soweit erkennbar – nicht-private Einträge, bereits gezogene sowie generell gesperrte Rufnummern nicht auswählbar.

Die Stichprobenrealisierung erfolgte auf regionaler Ebene nach dem Konzept der Nettosteuerung per Sample-Management-System, das zu mikrostratifizierten, ungeklumpten Stichproben führt, die sich bevölkerungsproportional auf die Mikrozellen aufteilen. Multistratifikation und Aufteilung der Stichprobe auf die Zellen erfolgten vollautomatisch. Das Schichtungstableau der Allokationsrechnung ging als Sollstruktur in die Steuerung ein. Mobilfunk- und Festnetznummern wurden gemeinsam in einem System gesteuert, das gewährleistete, dass in jeder Zelle die erforderliche Zahl von Interviews durchgeführt wurde. Innerhalb jeder Zelle waren die Festnetznummern nach Zufallszahlen sortiert. Mobilfunknummern wurden gemäß ihrem angestrebten Anteil zufällig in die Kontakte insgesamt eingemischt, wobei die Zellenzuordnung im Interview durch Abfrage von Postleitzahl und Wohnort erfolgte. Dieses Verfahren stellte sicher, dass eine Zellenübererfüllung kaum vorkam und die regionalen Strukturen durch Gewichtung gut korrigierbar waren. Die beiden folgenden Grafiken veranschaulichen das Verfahren.

Abbildung 1: Ablauf einer Dual-Frame-Studie

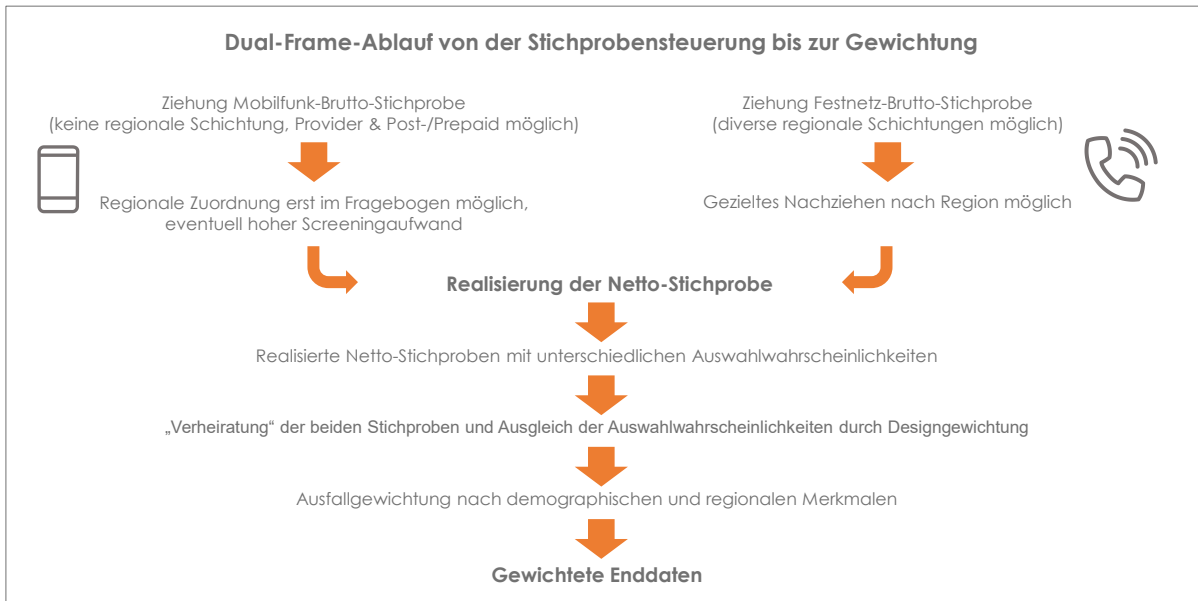
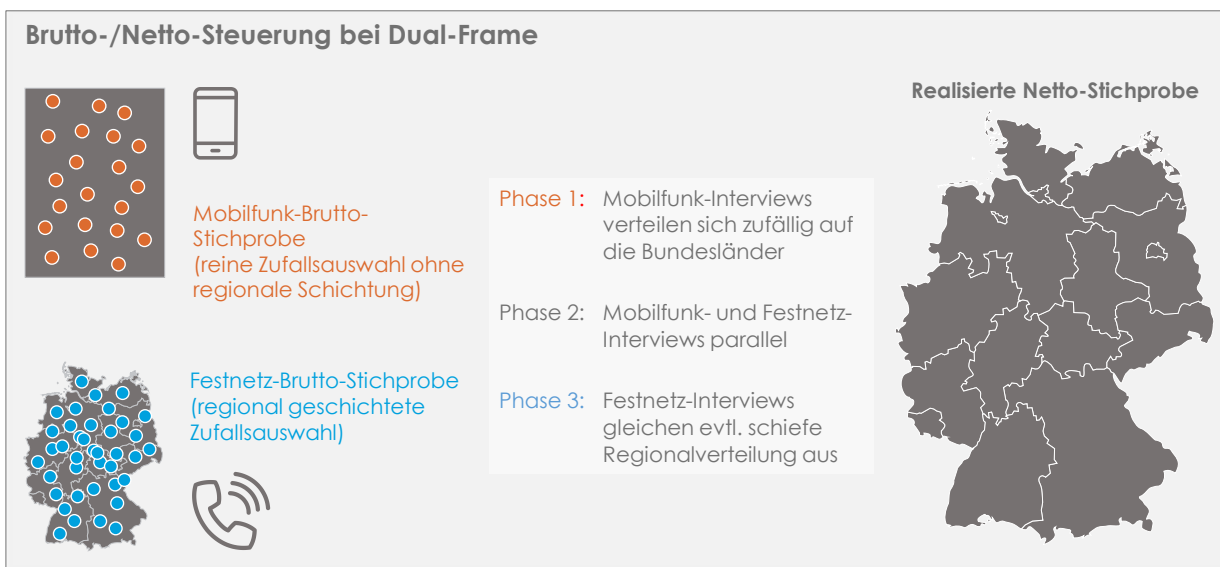


Abbildung 2: Stichprobenrealisierung



Generell gibt es im Mobilfunksample neben den bereits beschriebenen Unterschieden zur Festnetzstichprobe, die sich auf die regionale Auswahl und die Personen- statt die Haushaltsebene beziehen, weitere technische Unterschiede. Dies betrifft den bereits erwähnten höheren Anteil an ungültigen bzw. nicht geschalteten Nummern. Dieser begründet sich darin, dass Mobilfunknummern nur zu einem sehr geringen Anteil in Telefonbüchern registriert sind und deshalb der Anteil an generierten Nummern in der Bruttostichprobe deutlich höher als bei der Festnetzstichprobe ist. Zudem kann der Autodialer bei Mobilfunknummern seltener erkennen, ob es sich um eine geschaltete Nummer handelt oder nicht. Dies liegt wiederum daran, dass bei nicht geschalteten Mobilfunknummern – je nach Netzbetreiber – auch Bandansagen oder Freizeichen als Codes zurückgegeben werden, die nicht als falsche Nummer erkannt werden können. Hinzukommen seit einiger Zeit auch vermehrt Machine-

to-Machine-Nummern, die – je nach Anbieter – vom Autodialer nicht erkannt werden können, da sie von den Netzbetreibern meist nicht als Machine-to-Machine-Nummern – z. B. über eigene Nummernkreise – klassifiziert werden. Nicht geschaltete Telefonnummern, die kein Signal aussenden, können teilweise vom Autodialer erkannt und nach dem ersten Kontakt als falsche Telefonnummer klassifiziert werden. Insgesamt besteht bei Mobilfunknummern ein erhöhter Anteil an Fehlkontakten, der eine höhere Übersetzung an Telefonnummern bei Mobilfunk nötig macht.

Durch das sog. Home-Location-Register-(HLR)-Lookup-Verfahren kann dieser Aufwand etwas reduziert werden. Allerdings haben v.a. die großen Telefonanbieter mittlerweile das Verfahren, das über einen „stillen“ Versand von SMS durchgeführt wurde, gesperrt, so dass nur noch ein kleiner Teil der Mobilfunknummer über dieses Verfahren im Vorfeld geprüft werden kann.

Während der Feldarbeit nicht erreichte Rufnummern wurden im Sample-Management-System zurückgelegt und kamen nach größerem zeitlichem Abstand zu anderen Tageszeiten zur Wiedervorlage. Die an einem bestimmten Tag nicht Erreichten wurden durch solche substituiert, die an anderen Tagen nicht erreicht wurden. Damit entfiel der sogenannte Not-at-home-Bias weitgehend. Nach ADM-Standard liegt die empfohlene maximale Zahl an Kontaktversuchen bei zehn. Demnach wurden Festnetznummern, bei denen in zehn Anrufen kein Kontakt zu einer Person im Haushalt hergestellt werden konnte, also immer entweder das Freizeichen ertönte oder ein Anrufbeantworter geschaltet war, ausgesteuert. Bei Mobilfunknummern war dies nach dem dritten Anruf ohne Kontakt zu einer Person der Fall, da es sich dabei im Regelfall um (noch) nicht genutzte Rufnummern handelt. Telefonnummern, bei denen in einem der ersten zehn bzw. drei Versuche ein Kontakt zu einer Person hergestellt werden konnte, wurden so lange kontaktiert, bis ein Interview oder ein finaler Ausfallgrund vorlag.

Vom Sample-Management-System wurde je nach Bedarf auch die Gleichverteilung der Interviews auf Befragungstage und Tageszeiten gesteuert. In der vorliegenden Untersuchung war hier insbesondere an eine gleichmäßige Wochenverteilung zu denken, um saisonale Effekte zu vermeiden. Die üblichen Telefonzeiten waren spätnachmittags bis abends und am Wochenende am Nachmittag, um Privatpersonen, insbesondere Erwerbstätige, auch erreichen zu können. Um mögliche Einflüsse auf Untersuchungsergebnisse von vornherein auszuschalten, wurde die Stichprobe nach einem Verfahren der dynamischen Repräsentativität bezüglich der Besetzung der Zellen des Multistratifikations-tableaus optimiert, so dass sich für jedes Stundenintervall proportionale Teilstichproben ergaben.

Es kann somit festgehalten werden: Telefon-Master-Sample-Stichproben von Kantar sind über das Bindeglied amtlicher Flächeneinheiten auf die Verteilung der Privathaushalte justiert. Bei der Stichprobenanlage wurde dabei sowohl auf der Ebene der Bruttoziehung (Schichtung à priori) als auch bei der Stichprobenrealisierung eine optimale Aussteuerung garantiert, so dass eine Gewichtung nach Haushaltszahlen pro Region zu einer äußerst kleinen Faktorenspanne führt und praktisch unterbleiben könnte. Das Telefon-Master-Sample-Stichprobendesign erforderte jedoch zwingend eine Designgewichtung (Schichtung à posteriori), da unterschiedliche (aber berechenbare) Auswahlchancen ausgeglichen werden müssen (vgl. hierzu Abschnitt 7.1).

3.2 Auswahl von Erwerbstätigen ab 15 Jahren

Festnetzstichprobe

Bei der Teilstichprobe mit Festnetznummern wurde auf der ersten Stufe zunächst ein Haushalt ausgewählt. In diesem wurde dann über einen Zufallsmechanismus eine zu befragende Zielperson ausgewählt. Hierfür musste zunächst geklärt werden, ob und wie viele Personen ab 15 Jahren im Haushalt leben. Im Anschluss an die Ermittlung der Zielperson schloss sich ein kurzes Screeninginterview an,

über das ermittelt wurde, ob die Zielperson erwerbstätig im Sinne der Untersuchung ist und die Tätigkeit z. B. nicht ehrenamtlich oder im Rahmen einer Ausbildung ausübt.

Mobilfunkstichprobe

Die Ermittlung der Zielperson war bei der Teilstichprobe mit Mobilfunknummern insofern einfacher, da es sich bereits um eine Personenstichprobe handelt. Bei dem Erstkontakt musste mit der Person am Telefon deshalb zunächst nur geklärt werden, ob sie Hauptnutzer des Handys und mindestens 15 Jahre alt ist. War dies der Fall, wurde anhand des Screeninginterviews ermittelt, ob die Person erwerbstätig im Sinne der Untersuchung ist. Um die regionale Verortung des Interviews vornehmen zu können, war es zudem zwingend, dass die Person uns die Postleitzahl ihres Wohnortes nannte. War sie dazu nicht bereit, wurden im Rahmen der Kurzbefragung an Nichtteilnahmebereite (siehe Kapitel 0) wenige Fragen gestellt ansonsten war das Gespräch an dieser Stelle beendet.

Um später die Auswahlwahrscheinlichkeiten berechnen zu können, wurde in allen Interviews der Festnetz- und Mobilfunkstichprobe die Zahl der Personen im Haushalt insgesamt erfasst. Ebenso wurde die Zahl der Telefonnummern (Handy und Festnetz) erhoben, unter denen die Zielpersonen zu erreichen sind.

3.3 Selektives Teilnahmeverhalten niedrig gebildeter Erwerbstätiger

Die Erreichbarkeit und Teilnahmebereitschaft ist nicht in allen Teilgruppen gleich groß. Erwerbstätige mit niedriger Bildung und jüngere Erwerbstätige sind über eine telefonische Ansprache schwieriger zu erreichen als Erwerbstätige mit mittlerer oder höherer Bildung und als ältere Erwerbstätige.

Für die Erwerbstätigenerhebung 2024 wurde deshalb eine Maßnahme auf Ebene der Stichprobenziehung entwickelt, um dem zu erwartenden Bias bei der Bildung entgegenzuwirken. Kantar verfügt über einen zweiten Auswahlrahmen von Telefonnummern, für die bestimmte Informationen, wie z. B. zur Bildung, vorliegen (Auswahlrahmen 2). Diese Informationen basieren auf vergangenen telefonischen Befragungen, bei denen die Zielperson sowohl die Erlaubnis zur Speicherung als auch dazu, dass die Telefonnummer wieder angerufen werden darf, gegeben hat. Diese Nummern und die Zielpersonen wurden in den ursprünglichen Befragungen nach dem gleichen Zufallsverfahren ausgewählt, wie es in der BIBB/BAuA-Erwerbstätigenbefragung angewandt wurde. Im Vorfeld der Planungen für die aktuelle Erhebungswelle der Erwerbstätigenerhebung wurde auf Basis von Logitmodellen geprüft, ob es in Abhängigkeit verschiedener Merkmale eine unterschiedliche Bereitschaft der Zielpersonen gab, noch einmal kontaktiert zu werden. Die erwartbaren Unterschiede in den Teilnahmewahrscheinlichkeiten bezüglich Alter und Bildung wurden auf Basis der aus vorherigen Telefonbefragungen vorliegenden Informationen im Rahmen der Designgewichtung berücksichtigt (vgl. dazu Abschnitt 7.1).

Im Grundsatz handelt es sich hierbei um eine zeitversetzt nach dem gleichen Zufallsprinzip gezogene Stichprobe, für die auf Rufnummerebene bestimmte Merkmale bekannt sind, die als Auswahlrahmen für eine Schichtung genutzt werden können. Theoretisch kann man den gesamten Auswahlrahmen in zwei Auswahlrahmen aufteilen: Die Rufnummern, für die keine Information vorliegt (Auswahlrahmen 1), und die Rufnummern, für die Merkmalsinformationen bekannt sind (Auswahlrahmen 2). Bei Letzterem definiert man auf Basis der bekannten Merkmale und deren Ausprägungen Schichtungszellen und wählt zufällig Nummern innerhalb dieser Zellen aus. In dem Teil, für den keine Informationen vorliegen, wird frei gezogen.

Über den Auswahlrahmen 2 wurden bei der Festnetzstichprobe in begrenztem Umfang Nummern eingesetzt, für die bekannt ist, dass unter ihr eine niedrig gebildete Person erreichbar ist. Auswahlrahmen 2 wurde nur bei der Festnetzstichprobe genutzt, da es sich um eine Stichprobe auf Haushaltsebene handelt und die Auswahl der Zielperson über die gleiche Zufallsauswahl, den Schwedenschlüssel (bzw. Kish-Methode), erfolgte wie bei den frei generierten Telefonnummern des Auswahlrahmens 1. Das heißt, es wurde nicht nach der Person, deren Merkmalsausprägungen bekannt sind, gefragt, sondern es wurde unter potenziellen Zielpersonen (Personen ab 15 Jahren) unter dieser Nummer zufällig eine ausgewählt und für diese dann in einem Screeninginterview ermittelt, ob sie erwerbstätig im Sinne der Untersuchung ist, um sie anschließend als Zielperson zu befragen. Da die Information über die Merkmalsausprägungen nur für eine Person im Haushalt unter dieser Rufnummer vorlag, kam es also durchaus vor, dass die tatsächlich ausgewählte Zielperson andere Ausprägungen aufweist. Diese Personen wurden dennoch interviewt. Die Erwartung war, dass auf Haushaltsebene eine höhere Homogenität hinsichtlich der Bildung besteht.

Der Wahrscheinlichkeit der Homogenität auf Haushaltsebene kann man sich über Schätzungen auf Basis des Mikrozensus annähern. So zeigen Daten für 2021 zum Beispiel, dass in Mehrpersonenhaushalten, in denen die Bezugsperson eine niedrige Bildung aufweist, weitere 1,06 Personen eine niedrige Bildung haben. Selbst wenn die umgekehrte Folgerung von einer beliebigen anderen Person im Haushalt auf Basis der veröffentlichten Mikrozensus-Analysen nicht möglich ist, so zeigt es doch, dass die Wahrscheinlichkeit, Personen mit niedriger Bildung auszuwählen, über das vorgesehene Verfahren gesteigert werden kann.

Im Rahmen der Festnetzstichprobe sollten maximal 2.000 Nettointerviews mit Erwerbstätigen über Telefonnummern des Auswahlrahmens 2 realisiert werden. Für die Ziehung wurde nach Bildung geschichtet und es war für die Bruttostichprobe eine Verteilung von 80 % mit niedriger Bildung (maximal Hauptschulabschluss) und 20 % mit mittlerer Bildung (maximal Realschulabschluss) vorgesehen. Der Einsatz der Telefonnummern aus dem Auswahlrahmen 2 erfolgte in monatlichen Tranchen, um die realisierte Nettofallzahl monitoren zu können und eine Streuung über die Feldlaufzeit zu haben.

Vor dem Einfließen der aus Auswahlrahmen 2 realisierten Interviews in die Strukturanpassungsgewichtung im Rahmen der Kalibrierung wurde geprüft, ob eine auf den im Auswahlrahmen vorliegenden Merkmalen basierende Ausfallanalyse zu einer Qualitätsverbesserung führt (s. auch Abschnitt 7.2). Die Kehrwerte der in der Ausfallanalyse für Auswahlrahmen 2 ermittelten Teilnahmewahrscheinlichkeiten gingen dann als Eingangsgewichte in die Designgewichtung ein. Die umgesetzten Gewichtungsschritte gewährleisteten, dass die Kombination der beiden Teilstichproben eine erwartungstreue Abbildung der Grundgesamtheit darstellt.

4 Durchführung und Erhebung

4.1 Interviewschulung

Kurz vor dem Start der Haupterhebung wurde eine projektspezifische Schulung der Interviewenden durch die Projektleitung in Kooperation mit dem BIBB durchgeführt. Die Schulung fand als Videokonferenz am 28. September statt und wurde aufgezeichnet, so dass sie für alle weiteren Nachschulungen zur Verfügung stand. Jeder Interviewer und jede Interviewerin, der bzw. die bei der Studie zum Einsatz kam, wurde vorab mit dem Schulungsvideo auf das Interview vorbereitet. Der Feldstart war am Mittwoch, den 4. Oktober, da der 2. Oktober von sehr vielen Interviewerinnen und Interviewern als Brückentag vor dem Feiertag am 3. Oktober genutzt wurde. Etwa eine Woche nach Feldstart gab es nochmals die Möglichkeit in einem Live-Termin für alle CATI-Interviewenden direkt Fragen an die Projektleitung zu stellen.

Bei der Schulung wurde auf folgende Punkte eingegangen:

- Vorstellung der Kundenseite und des Projekts: Inhalte und Ziele, Hintergrund, Methodisches Vorgehen und Besonderheiten
- Allgemeine Hinweise zur Interviewführung
- Einstieg und Ermittlung der Zielperson
- Informationsmöglichkeiten für die Befragten
- Erläuterungen zu ausgewählten Fragen, v.a. der Schul- und Berufsausbildung
- Abschluss des Interviews: Frage nach Bereitschaft zur Wiederbefragung und Datenverknüpfung
- Begleitmaterialien: Informationsschreiben, Datenschutzblatt und weitere Informationsmöglichkeiten für die Befragten

Die Interviewer erhielten zudem schriftliches Schulungsmaterial, das ihnen während des Telefonierens zur Verfügung stand und folgende Materialien enthielt:

- die Schulungspräsentation,
- eine von BIBB und BAuA erstellte Übersicht zu den beruflichen Ausbildungsabschlüssen mit Erläuterungen (Factsheet),
- eine von BIBB und BAuA erstellte Übersicht mit Hinweisen „was tun, wenn...“ zu verschiedenen Gesprächssituationen.

Im Anschluss an die Schulung gingen alle Interviewenden ein Testinterview durch, bis sie mit allen Fragen vertraut waren, bevor sie das erste Interview durchführten. Die Projektverantwortlichen an den verschiedenen Standorten standen für Rückfragen zur Verfügung. Dies gewährleistete eine sichere Interviewführung von Beginn an. Alle eingewiesenen Interviewenden erhielten in der Projektdatenbank einen Vermerk, dass sie dieses Projekt bearbeiten können.

Wenn in einem Telefonstudio Fragen auftraten, wurden die Antworten auch an die anderen Standorte und Interviewenden kommuniziert. Zu Beginn der Feldarbeit gab es die Rückmeldung, dass der Einstiegstext im Kontakt recht lang ist. Daraufhin wurde der Text in Absprache mit den Auftraggebern noch etwas gekürzt. Außerdem wurde die Frage, ob die Zielperson erwerbstätig ist, mit dem Interviewendenhinweis, dass auch Krankheit und Urlaub von weniger als drei Monaten als Erwerbstätigkeit zählt, konkretisiert.

Die Schulungspräsentation wurde im Vorfeld mit dem Auftraggeber abgestimmt. Die Schulungsunterlagen sind im Anhang dokumentiert.

4.2 Einsatz der Interviewenden und Durchführung der Feldarbeit

Nach Abschluss der beiden Pretests und der Einarbeitung des dort festgestellten Anpassungsbedarfs startete die Feldarbeit der Haupterhebung wie geplant am 4. Oktober 2023. Die letzten Interviews wurden am 9. April 2024 durchgeführt.² Insgesamt wurden 20.006 Erwerbstätige befragt.

Die Telefoninterviewenden sind ausnahmslos im Angestelltenverhältnis oder in geringfügiger Beschäftigung angestellt. Zum Zeitpunkt der Befragung betrug die durchschnittliche Verweildauer im Interviewendenstab rund 59 Monate. Über die gesamte Feldzeit kamen 473 Interviewende zum Einsatz. Im Durchschnitt wurden 42 Erwerbstätige pro Interviewendem befragt.

Tabelle 1: Zahl der Interviews pro Interviewendem

Zahl der Interviews	Zahl der Interviewenden	Anteil in %
1 bis 10 Interviews	103	22%
11 bis 20 Interviews	71	15%
21 bis 30 Interviews	97	21%
31 bis 50 Interviews	73	15%
51 bis 90 Interviews	59	12%
91 oder mehr Interviews	70	15%
Insgesamt	473	100%

Der Interviewendenstab der BIBB/BAuA-Erwerbstätigenbefragung setzt sich wie folgt zusammen. Das Durchschnittsalter der Interviewenden liegt bei 42 Jahren. 34 % sind männlich und 66 % weiblich. Nach höchstem Schulabschluss verteilen sich die Interviewenden so: 8 % haben einen niedrigen Schulabschluss (inklusive kein Schulabschluss), 34 % einen mittleren und 51 % einen hohen (7 % der Interviewenden haben keine Angabe zum Schulabschluss gemacht).

Zur Sicherstellung einer möglichst hohen Erreichbarkeit wurde vorwiegend nachmittags und abends bzw. an den Wochenenden kontaktiert. Die Kontaktzeiten wurden dabei über das Sampling Management System nach Uhrzeit und Wochentag variiert (siehe Abschnitt 3.1.2).

Um die Interviewenden bei der Gewinnung der Zielpersonen zu unterstützen, wurden verschiedene vertrauensbildende Maßnahmen vorgesehen.

- Informationsschreiben

Zunächst stand den Interviewenden ein Informationsschreiben der Auftraggeber zur Verfügung, das sie auf Nachfrage per E-Mail als PDF-Dokument an potentielle Zielpersonen versenden konnten. Zusammen mit diesem Informationsschreiben wurde ein Datenschutzblatt versendet. Das Informationsschreiben und das Datenschutzblatt sind im Anhang dokumentiert.

- Hinweis auf Homepages

Des Weiteren existiert bereits eine Projekthomepage zur BIBB/BAuA-Erwerbstätigenbefragung, deren Adresse im Informationsschreiben genannt wurde. Auch während der Einstiegsphase des Interviews konnten die Interviewenden die potentiellen Befragten auf diese Projekthomepage hinweisen. Sowohl auf der Projekthomepage als auch auf der Homepage von Verian (während der Feldphase

² 20 Interviews wurden im November 2024 nacherhoben, da bei der Qualitätskontrolle unplausible Angaben aufgefallen waren.

Kantar Public) wurde während der Feldphase ein Hinweis auf die laufenden Feldarbeiten aufgenommen, so dass sich die Befragten von der Korrektheit des Anrufs überzeugen konnten.

■ Auftraggeber und Thema der Befragung

Eine weitere wichtige Unterstützung ist es für die Interviewenden, wenn sie zu Beginn des Gesprächs den Auftraggeber nennen können, zumal wenn es sich – wie es bei der vorliegenden Studie der Fall ist – um öffentliche und seriöse Institute handelt. Auch das Thema der Befragung – die eigene berufliche Tätigkeit – ist für die Befragten überwiegend ein wichtiges und interessantes Thema.

■ Hinweise zur Studie bei Nachfragen

Der Einleitungstext wurde so kurz wie möglich gehalten, um mit den Zielpersonen möglichst rasch ins Gespräch zu kommen. Für Nachfragen zu verschiedenen Aspekten, wie Zielsetzung der Studie, Länge des Interviews, Datenschutz u. ä. standen den Interviewenden Texte zur Verfügung, die sie je nach Bedarf und Nachfrage der Befragten aufrufen konnten (siehe Anhang A – Screeningfragebogen).

Angesichts der Länge des Interviews hatten die Interviewenden verschiedene Möglichkeiten, um auf die Befragten einzugehen. So konnten Termine zu jeder Uhrzeit und an allen Wochentagen – außer Sonntag – vereinbart werden. Begonnene Interviews konnten jederzeit unterbrochen werden und zu einem anderen Zeitpunkt fortgeführt werden. Befragte, die über Mobilfunk kontaktiert wurden, hatten die Möglichkeit, - sofern vorhanden – eine Festnetznummer anzugeben und unter dieser das Interview zu führen.

Um die termingerechte Realisierung der Fallzahl von 8.000 Interviews zum Jahreswechsel 2023/2024 sowie die Fallzahl von 20.000 zu Anfang April 2024 sicherzustellen, fand ein zweiwöchiger Jour-Fix mit dem CATI-Produktionsteam statt, in der die aktuelle Fallzahl, die Feldkapazitäten sowie Fragen zur Erhebung oder zum Instrument besprochen wurden. Die Auftraggeber erhielten wöchentliche Updates des Feldstands, darin enthalten war die Anzahl der Interviews, eine Auszählung über wesentliche soziodemographische Merkmale der Befragten sowie die Anzahl und Art der Screenout-Fälle. Insgesamt wurde 26-mal ein Update zum Feldstand berichtet.

4.3 Maßnahmen zur Qualitätssicherung

Ein wesentlicher Vorteil von CATI-Umfragen ist die Möglichkeit der zeitnahen Qualitätskontrolle während der Feldzeit. Dazu wurden Prüfprogramme im Rahmen der Datenprüfung sowie folgende Monitoring-Techniken bei der Interviewdurchführung eingesetzt:

- (1) Frühzeitige Datenprüfung anhand von Zwischendaten
 - (2) Regelmäßig Kurzberichte aus den Telefonstudios zu Schwierigkeiten oder Fragen
 - (3) Monitoring von Interviews durch das Qualitätsmanagement
 - (4) Prüfung der Interviewzeit durch das Qualitätsmanagement
- Zu (1): Frühzeitig und regelmäßig wurden Zwischendaten gezogen, die die Projektleitung auf korrekte Filterführung, korrekte Erfassung der Postleitzahl des Wohnorts bei Mobilfunk wie des Betriebs, Plausibilität sowie Verwertbarkeit der offenen Tätigkeitsbezeichnung prüfte. Die Auftraggeber erhielten fünfmal Zwischendaten (Näheres dazu in Abschnitt 5).
 - Zu (2): Die Projektleitung stand über einen 14-tägigen Jour-Fix in engem Kontakt zu dem CATI-Produktionsteam, an dem auch die Hauptansprechpartnerin für die Studie in den Telefonstudios teilnahm, und erhielt regelmäßig Kurzberichte über den Feldverlauf und konnte Fragen klären.
 - Zu (3): Grundsätzlich werden für die Gewährleistung einer hohen Qualität bei der Interviewdurchführung die eingesetzten Interviewenden durch Monitoringmaßnahmen beobachtet, die auch bei der BIBB/BAuA-Erwerbstätigenbefragung 2024 zum Einsatz kamen. Das Monitoring (Mithören

und -sehen) dient zur Überprüfung der Richtlinien zur korrekten und optimalen Datenerfassung und wird auf dem Qualitätsformular dokumentiert.

Durch ein PC-gestütztes Programm zum Mithören und -sehen kann die Supervision erkennen, in welchem Programmabschnitt sich der Interviewer oder die Interviewerin befindet (Einstieg, Maske xx, etc.) und zu welchen Zeiten und wie lange er oder sie nicht aktiv ist.

Mit diesem Instrument werden in der Kontaktphase folgende Aspekte geprüft:

- Wird die richtige Zielperson ausgewählt und befragt?
- Wie ist das Argumentations- und Motivationsverhalten?
- Werden Ausfälle richtig eingestuft?
- Werden Terminvereinbarungen korrekt vorgenommen?

Im Rahmen einer Kurz-Qualitätskontrolle werden durch zufälliges Aufschalten über einen kurzen Zeitraum Fragen oder Fragebogenteile verfolgt. Darüber hinaus wird spezielles Augenmerk auf von der Projektleitung als wichtig definierte Fragen gelegt. Bei den zudem stichprobenartig durchgeführten Voll-Qualitätskontrollen wird über die Kontaktphase hinaus das vollständige Interview anhand des folgenden Leitfadens beobachtet:

- Verhalten während des Vorlesens der Fragen
- Beherrschung der Fragetechnik (wörtlich, korrekte Eingabe)
- Motivationsfähigkeit, Tempo der Gesprächsführung
- Verhalten bei Nachfragen
- Wahrung der Neutralität
- Verabschiedung höflich, Hinweis auf Wiederholungsanrufe, Datenschutz

Aus den sich hierbei ergebenden Hinweisen erstellt die Supervision eine Gesamtbeurteilung des Interviewers oder der Interviewerin. Diese drückt sich in seiner Benotung für „Qualität“ aus:

- 1 = Sehr gutes Interview
- 2 = Gutes Interview
- 3 = Leichte Mängel
- 4 = Interview mit großen Mängeln

Eine Bewertung mit dem Wert 4 erfolgt beispielsweise, wenn eine falsche Zielpersonenauswahl oder eine suggestive Fragestellung erfolgte. Die Note 3 wird z.B. vergeben, wenn Frage- oder Antworttexte verkürzt vorgelesen werden oder Antwortskalen nicht ausreichend häufig vorgelesen wurden. Die Bewertung wird anschließend in das Informationssystem eingegeben. Es werden jeweils die aktuellsten zehn Bewertungen gespeichert, um ein ausreichendes Abbild der gelieferten Qualität des Interviewers oder der Interviewerin zu haben. Bei neuen Interviewern werden die ersten drei Einsätze voll kontrolliert. Bei den übrigen Interviewenden werden in regelmäßigen Abständen Voll-Qualitätskontrollen und Kurz-Qualitätskontrollen in verschiedenen Studien vorgenommen.

Bei der vorliegenden Untersuchung wurden insgesamt 732 Monitoring-Kontrollen durchgeführt, darunter ein Drittel während der Kontaktphase, etwa jede fünfte Kontrolle war eine sog. Kurzkontrolle und in gut der Hälfte der Kontrollen wurde eine sog. Vollkontrolle durchgeführt. Die genaue Anzahl der durchgeführten Kontrollen und das Ergebnis der Gesamtbeurteilung wird in der folgenden Tabelle dargestellt.

Tabelle 2: Durchgeführte Mithörkontrollen und ihr Ergebnis – Interviews und Kontakte

Note	Anzahl Kontrollen	Art der Kontrolle in %			Kontrollen in %
		Kontaktphase	Kurzkontrolle	Vollkontrolle	
1	311	6%	10%	26%	42%
2	349	21%	6%	20%	48%
3	63	4%	2%	2%	9%
4	9	1%	0,3%	0,4%	1%
Insgesamt	732	32%	18%	49%	100%

Der Anteil an Kontrollen, bei denen nicht korrektes Interviewendenverhalten konstatiert werden musste, ist gering. Die meisten der festgestellten Mängel wurden dabei während der Kontaktphase beobachtet und betreffen damit nicht unbedingt Interviews, sondern auch nicht erfolgreiche Kontakte und Ausfälle. Je nach Bewertung erhielten die Interviewenden im Anschluss an die Kontrolle ein Lob, Hinweise auf die Punkte, die sie verbessern sollen, eine gezielte Nachschulung oder es erfolgten eine Verwarnung und der Abzug vom Projekt. Interviews mit der Note 4 wurden ausnahmslos als nicht verwertbar aus dem Datensatz entfernt, bei der Note 3 erfolgte eine Nachschulung durch die Supervision.

- Zu (4): Ziel dieser Prüfung ist es, zu bewerten, ob die durchgeführten Interviews anhand gespeicherter Datensätze und Merkmale hinsichtlich der Telefonzeiten plausibel erscheinen. Zu dem Zweck werden die Daten aus zwei Quellen verglichen: der CATI-Software³ und der Dialer-Software⁴. Es wird das Vorkommen in beiden Datenquellen überprüft und sichergestellt, dass zu jedem Interview in der CATI-Software ein zeitlich plausibler telefonischer Kontakt in der Dialer-Datenbank existiert (Prüfung von 995 Interviews). Die Interviewdauer wird zur Prüfung herangezogen. Bei diesem Prüfschritt wurden insbesondere sehr kurze Interviews mit einer Dauer von weniger als 18 Minuten sowie unterbrochene Interviews untersucht (Prüfung von 230 Interviews). Für alle diese Fälle wurde die Verweildauer pro Frage geprüft. Hierbei gab es keine Auffälligkeiten, d.h. keine Verweildauern unter vier Sekunden pro Frage. Es konnten alle bezüglich der Interviewdauer auffälligen Interviews durch diese Prüfungen validiert werden.

Insgesamt wurden über das Monitoring von Interviews und Kontakten sowie die Prüfung der Interviewzeit durch das Qualitätsmanagement insgesamt 1.760 Kontrollen durchgeführt. Tabelle 3 gibt einen Überblick über die Anzahl und die Art der durchgeführten Kontrollen.

Tabelle 3: Überblick über die Anzahl und die Art der durchgeführten Kontrollen

Anzahl Kontrollen	Art der Kontrolle		
	Mithören	Interviewzeit - CATI Verbindungszeit vs. Dialer Verbindungszeit	Interviewzeit - unterbrochenes Interview
1.760	732	995	230

³ Enthält Datensätze zu allen Kontakten, die von Interviewenden als Interviews vercodet wurden. Merkmale: Startzeitpunkt, Dauer des Kontakts/Interviews, Interviewendennummer, eindeutiges Adresskennzeichen (Interviewnummer), Telefonnummer des Adressdatensatzes. Nicht als Interviews vercodete Kontakte werden ebenfalls erfasst (z. B. Termine, Absagen, Ausfälle).

⁴ Speichert Daten zu jedem Anruf, einschließlich Dauer des telefonischen Kontakts. Verwendet dasselbe eindeutige Adresskennzeichen wie die CATI-Software. Speichert Start- und Endzeitpunkte sowie die Dauer eines Anrufs.

4.4 Ergebnisse der Feldarbeit

Insgesamt wurden 20.006 auswertbare Interviews mit Erwerbstätigen durchgeführt, davon 9.565 Interviews, die über Festnetz kontaktiert wurden, und 10.441 Fälle, die über eine Mobilfunknummer kontaktiert wurden. Zum 31.12.2023 lagen 9.261 Interviews vor.

Die Interviewdauer war mit durchschnittlich 40 Minuten geplant, wobei hier die Nettodauer gemeint ist, d. h. ohne Kontaktphase, Screeninginterview, GewichtungsvARIABLEN sowie Fragen nach Einwilligung zu Wiederbefragungen und Datenverknüpfung. Die realisierte durchschnittliche Nettointerviewdauer lag mit 42,5 Minuten um gut zwei Minuten höher als geplant. Betrachtet man den Median liegt die Interviewdauer bei 41 Minuten. Hinzu kommt die Zeit, die für die Kontaktphase, das Screening, die gewichtungsrelevanten Fragen sowie die Erfassung des Einverständnisses zur Wiederbefragung und zur Datenverknüpfung benötigt wird. Betrachtet man die Gesamtdauer also die komplette Dauer des Anrufs, liegt diese im Durchschnitt bei 49 Minuten, der Median bei 48 Minuten. Vergleicht man die Netto-Dauern der Interviews der Festnetz- und der Mobilfunkstichprobe, so dauerten die Interviews in der Mobilfunkstichprobe eineinhalb Minuten länger als in der Festnetzstichprobe. Diesen Unterschied gab es auch 2018 in ähnlicher Größenordnung und er liegt im üblichen Rahmen. Hintergrund dafür dürften Verbindungsprobleme oder kurze Unterbrechungen in der Gesprächssituation sein, die bei Mobilfunkgesprächen häufiger auftreten als bei Festnetztelefonaten. Die folgende Tabelle gibt einen Überblick über die jeweiligen Interviewdauern.

Tabelle 4: Interviewdauer in Minuten⁵

	Gesamtdauer	Kontaktphase	Screening	Nettodauer
Durchschnitt Festnetz	48,0	2,1	1,4	41,7
Median Festnetz	46,3	1,3	1,3	40,3
Durchschnitt Mobilfunk	50,3	2,5	1,6	43,2
Median Mobilfunk	48,7	1,6	1,5	41,7
Durchschnitt Gesamt	49,2	2,3	1,5	42,5
Median Gesamt	47,6	1,5	1,4	41,0

⁵ Von der Analyse der Interviewdauer wurden 40 Interviews ausgeschlossen, die eine Nettodauer unter 18 Minuten aufweisen. Es handelt sich dabei um Interviews mit Unterbrechungen, wodurch die Berechnung der Dauer nicht korrekt vorgenommen werden konnte, da Zeitstempel die Basis für die Berechnung der Dauer darstellen.

Tabelle 5 gibt einen Überblick über die verschiedenen Ausfallgründe sowie die Ausschöpfungs- und Kooperationsrate nach AAPOR.⁶ Grundsätzlich gibt es verschiedene Varianten in der Berechnung der Ausschöpfungs- und Kooperationsraten, die auf Unterschiede in der Zuordnung zu neutralen Ausfälle (not eligible) und nicht neutraler Ausfälle, Zielgruppen-Zugehörigkeit ungeklärt (unknown eligibility) beruhen, und zu unterschiedlichen Ergebnissen bei der Ausschöpfungs- und Kooperationsrate führen. Wir verwenden die Ausschöpfungsrate RR2 und die Kooperationsrate COOP4.

⁶ The American Association for Public Opinion Research. 2023 Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 10th edition. AAPOR. <https://aapor.org/standards-and-ethics/standard-definitions>

Tabelle 5: Ausschöpfungs- und Kooperationsrate nach AAPOR

A Bruttoansatz	Festnetz	Mobilfunk	Insgesamt
	1.068.679	1.241.457	2.310.136
Fax/Modem	0,8%	0,5%	0,6%
ZP: nicht in der Lage/krank	0,0%	0,0%	0,0%
ZP: Sprachprobleme	0,0%	0,0%	0,0%
Kein Privathaushalt	0,5%	0,1%	0,3%
Falsche Telefonnummer	43,3%	49,7%	46,7%
Doppeladresse	0,0%	0,0%	0,0%
System error	0,1%	0,0%	0,0%
Abandoned call (Leitungsproblem ohne Kontakt)	0,0%	0,0%	0,0%
Stratifikation erreicht (Regionalzelle voll)	28,4%	7,7%	17,3%
max. Kontaktanzahl erreicht	0,2%	9,7%	5,3%
Firmen Anrufbeantworter	0,7%	1,8%	1,3%
KP: Sprachprobleme	0,3%	0,4%	0,4%
ZP wohnt im Ausland	0,0%	0,0%	0,0%
Summe neutraler Ausfälle (nicht Zielgruppe – not eligible)	74,3%	69,9%	72,0%
B Bereinigtes Brutto	274.503	373.193	647.696
Teilnehmer nimmt nicht ab	8,7%	43,3%	28,7%
Privater Anrufbeantworter	2,7%	22,7%	14,3%
Anschluss besetzt	4,2%	7,3%	6,0%
Angerufene Person nicht mehr in Leitung	4,8%	2,3%	3,3%
Summe nicht neutraler Ausfälle (ZG-Zugehörigkeit ungeklärt – unknown eligibility)	20,4%	75,6%	52,2%
C Verbleibende Adressen	218.594	91.125	309.719
Kein Termin während der Feldzeit möglich	1,5%	2,0%	1,7%
KP: ungefährender Termin konnte nicht realisiert werden	0,2%	0,2%	0,2%
KP: fester Termin konnte nicht realisiert werden	0,0%	0,0%	0,0%
KP: kein Interesse am Thema	20,1%	19,5%	19,9%
KP: keine Auskunft/sonstiger Grund	40,8%	34,8%	39,0%
KP: Absolutverweigerer	0,6%	0,9%	0,7%
KP: wortlos aufgelegt	15,0%	14,9%	15,0%
KP: Verweigerung im Namen der Zielperson	2,6%	1,4%	2,3%
KP verweigert aufgrund von Bedenken wegen Datenschutz	0,1%	0,1%	0,1%
KP verweigert wegen Länge des Interviews	1,9%	3,6%	2,4%
ZP nicht mehr im Haushalt/konnte unter der Nummer nicht ermittelt werden	0,2%	0,1%	0,2%
ZP: ungefährender Termin konnte nicht realisiert werden	0,1%	0,0%	0,1%
ZP: fester Termin konnte nicht realisiert werden	0,0%	0,0%	0,0%
ZP: Verweigerung aus Zeitgründen	0,2%	0,1%	0,2%
ZP: kein Interesse am Thema	0,6%	0,4%	0,5%
ZP: keine Auskunft/sonstiger Grund	1,1%	0,4%	0,9%
ZP: Absolutverweigerer	0,0%	0,0%	0,0%
ZP verweigert aufgrund von Bedenken wegen Datenschutz	0,0%	0,0%	0,0%
ZP verweigert wegen Länge des Interviews	0,4%	0,3%	0,3%
ZP: wortlos aufgelegt	0,4%	0,1%	0,3%
Verweigerung wg. Mithören, d.h. Abbruch nach Mithörfolge	0,1%	0,1%	0,1%
Keine Angabe bei PLZ (MF)	0,0%	0,2%	0,1%
Keine Tätigkeitsbezeichnung in F101	0,0%	0,0%	0,0%
Abbruch im Interview	0,7%	1,7%	1,0%
Summe nicht neutraler Ausfälle (Zielgruppe - eligible)	86,8%	81,0%	85,1%
D Durchgeführte Interviews insgesamt	28.855	17.335	46.190
darunter Interviews mit Erwerbstätigen	9.565	10.441	20.006
Interview mit erwerbstätiger Person	33,1%	60,2%	43,3%
Person unter 15 Jahren (MF)	-	1,0%	-
Person laut Screeninginterview nicht erwerbstätig (MF)	-	38,8%	-
Person laut Screeninginterview nicht erwerbstätig (FN)	66,8%	-	-
Alle Personen im Haushalt unter 15 Jahren (FN)	0,0%	-	-
Ausschöpfungsrate RR2 (D in % von B, d.h. bereinigtes Brutto)	10,5%	4,6%	7,1%
Kooperationsrate COOP4 (D in % von C, d.h. verbleibende Adressen)	13,2%	19,0%	14,9%

Für die Bewertung der Ausschöpfungs- und Kooperationsrate ist wichtig, dass diese Zuordnung transparent erfolgt und inhaltlich nachvollziehbar ist. Wir erläutern im Folgenden, wie dies bei der BIBB/BAuA-Erwerbstätigenbefragung 2024 erfolgte. Betrachtet man die Ausschöpfungsübersicht, fällt der hohe Anteil an neutralen Ausfällen auf. Dieser ist begründet in der Tatsache, dass die ausgewählten Telefonnummern zufallsgeneriert sind und es damit entsprechend häufig keine real existierenden bzw. vergebenen Telefonnummern sind. Die unterschiedlich hohen Anteile von „falschen Telefonnummern“ sowie „maximale Kontaktzahl erreicht“ bei der Festnetzstichprobe und der Mobilfunkstichprobe ergeben sich daraus, dass nicht vergebene Festnetznummern i.d.R. als ungültige Nummern mit einem entsprechenden Ton oder einer Ansage über den Autodialer zu identifizieren sind und Mobilfunknummern im Gegensatz dazu nicht. Es gab 2022 eine Technikumstellung (von ISDN auf Internet - VoiP) sowohl bei Festnetz als auch bei Mobilfunk. Der Anteil falscher Nummern bei Festnetz liegt bei 43,3%. Der Autodialer erkennt bei Festnetz in der Regel die nicht vergebenen Nummern und kann sie als falsche Nummern klassifizieren. Mittlerweile kann der Autodialer auch bei Mobilfunk den Teil der Nummern, bei denen kein Signal gesendet wird, erkennen. Dies betrifft auch Machine-to-Machine-Nummern, z.B. Kühlschränke mit Internetzugang. Diese Nummern werden bereits nach einem Kontakt als falsche Nummer deklariert. Der Anteil an falschen Telefonnummern bei Mobilfunk ist seit der Technikumstellung 2022 deutlich höher und umfasst auch Nummern, die zuvor bei „maximale Kontaktzahl erreicht“ eingruppiert waren.

Auch bei nicht vergebenen Mobilfunknummern kommt bei einem Anruf ein Freizeichen oder eine Mailboxansage. In diesen Fällen kann der Autodialer nicht identifizieren, ob die Telefonnummer vergeben ist oder nicht. Entsprechend ist davon auszugehen, dass im Mobilfunkteil der Stichprobe ein Großteil der Telefonnummern, bei denen trotz maximaler Kontaktzahl niemand erreicht werden konnte, nicht vergebene Telefonnummern sind (vgl. Abschnitt 3.1.2). Bei Festnetznummern, unter denen, während der Feldzeit kein Kontakt hergestellt werden konnte, konnte nicht festgestellt werden, ob es sich um einen Privathaushalt handelt. Aus diesem Grund sind diese Ausfälle den neutralen Ausfällen zugeordnet.

Unter den nicht neutralen Ausfällen („unknown eligible“) ist der mit Abstand größte Anteil an Ausfällen in den Kategorien „Teilnehmer nimmt nicht ab“ und „privater Anrufbeantworter“. Dies betrifft vor allem die Mobilfunkstichprobe, bei der der Anteil in diesen Kategorien deutlich höher ist als in der Festnetzstichprobe. Bei der Mobilfunkstichprobe dürften darunter zum einen nicht vergebene Nummern sein (vergleiche vorherigen Absatz). Zum anderen scheint es bei Mobilfunk deutlich häufiger als bei Festnetz der Fall zu sein, dass Anrufe von unbekannten Telefonnummern nicht beantwortet werden.

Unter den „unknown eligible“ Adressen gibt es auch den Ausfallgrund „Angerufene nicht mehr in der Leitung“. Hier wurde über den Autodialer ein Kontakt hergestellt, dieser konnte jedoch zu keinem Interviewenden durchgestellt werden, da alle Interviewenden zu diesem Zeitpunkt im Gespräch waren. In diesen Fällen wird eine Bandansage eingespielt. Personen, die auflegen, bevor ein Interviewer oder eine Interviewerin das Gespräch aufnimmt, werden hier eingestuft. Dies ist ein notwendiges und übliches Verfahren, um die Masse an Telefonnummern effizient abarbeiten zu können. Gemessen an den insgesamt eingesetzten Telefonnummern ist dieser Anteil sehr gering. Nach den ADM-Richtlinien sind diese Fälle als Ausfälle zu werten und werden nicht erneut angerufen.

Bei den nicht neutralen Ausfällen (eligible) wird unterschieden zwischen Ausfällen auf Ebene der Kontaktperson (KP) und der Zielperson (ZP). Bei Ausfällen durch eine Kontaktperson ist i.d.R. keine Zufallsauswahl der Zielperson erfolgt, es hat jedoch ein Kontakt zu einer Person stattgefunden. Es konnte also zumindest geklärt werden, dass es sich um eine gültige Telefonnummer handelt, die zu einem Privathaushalt oder einer Privatperson gehört. Deshalb werden diese Ausfälle in der vorliegenden

Berechnung den „eligible“ Adressen zugerechnet.⁷ Die höchsten Anteile an Ausfällen sind auf der Ebene der Kontaktperson zu verzeichnen. Bei den Festnetznummern konnte bei den Verweigerungen durch die Kontaktperson i.d.R. keine Information zur Anzahl der Personen ab 15 Jahren im Haushalt und damit auch keine Zielperson ermittelt werden. Bei den Mobilfunknummern konnte bei diesen Ausfällen nicht geklärt werden, ob der Hauptnutzer des Handys am Apparat ist bzw. die Person 15 Jahre oder älter ist. Die Verweigerungen durch die Kontaktpersonen wurden also ausgesprochen, ohne dass die Personen am Telefon Informationen preisgaben. Deren Anteil ist daher vergleichsweise hoch, der Anteil an Verweigerungen durch Zielpersonen dagegen relativ gering.

Um festzustellen, ob unter der kontaktierten Telefonnummer eine Zielperson ab 15 Jahren zu erreichen ist, war bei der Festnetzstichprobe zunächst die Frage, wie viele Personen ab 15 Jahren im Haushalt leben, durch die Kontaktperson zu beantworten. Im Anschluss wurde mit der zufällig ausgewählten Zielperson das Screeninginterview geführt, um zu ermitteln, ob sie zur Zielgruppe der Erwerbstätigen gehört. Bei der Mobilfunkstichprobe wurde ein Screeninginterview geführt, sofern die Person Hauptnutzer des Handys und 15 Jahre oder älter war. Da in all diesen Konstellationen Informationen erfragt wurden, werden diese Anrufe hinsichtlich der Rücklaufberechnungen als Interviews gewertet.⁸ Im Vergleich zur Erhebung 2018 liegen Ausschöpfungs- und Kooperationsrate in einer vergleichbaren Größenordnung.

4.5 Ergebnisse der Maßnahme zur Verbesserung der Bildungsstruktur

Wie in Abschnitt 3.3 dargestellt, kam in der Stichprobe ein Auswahlrahmen 2 zum Einsatz, um dem für Telefonerhebungen bekannten Bildungs-Bias entgegenzuwirken. Der Auswahlrahmen 2 kam nur bei der Festnetzstichprobe zum Einsatz, da hier im Haushalt bei der Zielpersonenauswahl dasselbe Verfahren angewendet werden konnte, wie bei Auswahlrahmen 1. Insgesamt wurden 32.452 Bruttoadressen aus dem Auswahlrahmen 2 eingesetzt. 80 % der Adressen waren aus einem Haushalt, in dem ursprünglich ein Interview mit einer Person mit niedriger Bildung geführt wurde, 20 % aus Haushalten mit Interview mit einer Person mit mittlerer Bildung. Es wurden 2.713 Interviews geführt, darunter 1.314 Screeninginterviews mit Nicht-Erwerbstätigen und 1.399 Interviews mit Erwerbstätigen.

Tabelle 6 weist die Verteilung des höchsten Schulabschlusses in der realisierten Stichprobe aus. Die Interviews im Auswahlrahmen 2 weisen im Vergleich zu Auswahlrahmen 1 und zur Mobilfunkstichprobe einen höheren Anteil an niedrig Gebildeten aus. Allerdings fand mehr als die Hälfte der Interviews im Auswahlrahmen 2 mit erwerbstätigen Personen mit mittlerer Bildung statt und gut ein Viertel der Interviews wurde mit Personen mit höherer Bildung geführt. Dies ist damit zu erklären, dass über die Festnetznummern lediglich bekannt war, dass eine Person mit niedriger Bildung im Haushalt lebt. Unter der Annahme, dass die Bildungsstruktur im Haushalt homogen ist, war die Erwartung, dass überproportional oft eine Person mit niedriger Bildung ausgewählt wird. Dies hat sich nicht in dem erwarteten Umfang realisiert, d.h. – ausgehend von der Verteilung in der Bruttostichprobe von 80 % Haushalten mit niedriger und 20 % mit mittlerer Bildung gab es relativ viele Wechsel bzgl. der Bildungsschicht in der aus Auswahlrahmen 2 realisierten Stichprobe.

⁷ In anderen Varianten werden Ausfälle auf Ebene der Kontaktperson den „unknown eligibility“-Adressen zugeordnet, wodurch sich eine höhere Kooperationsrate ergibt.

⁸ In anderen Varianten werden die Ausfälle „Person unter 15 Jahre“ und „Alle Personen im Haushalt unter 15 Jahre“ zu den neutralen Ausfällen (not eligible) gerechnet. Da dies hier sehr kleine Anteile sind, würde sich durch eine solche Zuordnung keine nennenswerte Veränderung der Ausschöpfungs- oder Kooperationsrate ergeben.

Zusammenfassend lässt sich festhalten, dass die Bildungsstruktur in der realisierten Stichprobe des Auswahlrahmens 2 zwar einen höheren Anteil an niedrig Gebildeten enthält als die Stichprobe des Auswahlrahmens 1, jedoch hinter den Erwartungen zurückgeblieben ist. Bei einer Fallzahl von $n=1.399$ ist der Einfluss auf die Bildungsstruktur der Gesamtstichprobe zudem eher gering. Der Anteil der niedrig Gebildeten konnte durch den Auswahlrahmen 2 um rund einen Prozentpunkt in der Gesamtstichprobe erhöht werden (8,2 % in der Gesamtstichprobe vs. 7,2 % in der Stichprobe des Auswahlrahmens 1, Tabelle 6). Der Anteil an Personen mit mittlerer Bildung liegt um 1,9 Prozentpunkte höher und der Anteil an Personen mit Abitur oder (Fach-)Hochschulreife sinkt aufgrund des Auswahlrahmens 2 um 2,8 Prozentpunkte.

Tabelle 6: Verteilung nach höchstem Schulabschluss – Gesamtstichprobe, Auswahlrahmen 1 und 2 im Vergleich

	Gesamt	Auswahlrahmen 1	Auswahlrahmen 2
Fallzahl	20.006	18.607	1.399
Hauptschulabschluss oder gleichwertiger Abschluss	8,2%	7,2%	21,8%
Realschulabschluss, mittlere Reife oder gleichwertiger Abschluss	27,9%	26,0%	51,9%
Abitur, Hochschul- oder Fachhochschulreife	63,4%	66,2%	25,6%
keinen Schulabschluss	0,3%	0,3%	0,4%
anderer Schulabschluss	0,2%	0,2%	0,1%
keine Angabe	0,1%	0,1%	0,1%
Insgesamt	100%	100%	100%

4.6 Zustimmung zur Wiederbefragung und zur Datenverknüpfung

Am Ende des Interviews wurden die Befragten um ihre Zustimmung zu zwei Punkten gebeten:

- einer erneuten – deutlich kürzeren – Befragung zu ähnlichen Themen sowie
- zur Zuspiegelung von Daten des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) in Nürnberg.

Im Falle des Einverständnisses wurde der Name der Zielperson sowie ihre E-Mailadresse erhoben.

85 % aller Befragten gaben ihr Einverständnis zu einer erneuten Kontaktierung für eine Befragung, von ihnen gaben 52 % eine E-Mailadresse an.

Nach dem Einverständnis zur Datenverknüpfung wurden alle abhängig Beschäftigten gefragt. Mit einer Datenverknüpfung von Befragungsdaten und Daten des IAB waren 72 % einverstanden, allerdings gab ein Teil von ihnen keinen oder nur einen unvollständigen Namen an. Einen vollständigen Namen gaben 83 % der Befragten mit erklärtem Einverständnis zur Datenverknüpfung an.

Die folgende Tabelle gibt einen Überblick über die Ergebnisse der beiden Einverständnisfragen.

Tabelle 7: Ergebnis der Zustimmungsfragen zur Wiederbefragung und zur Datenverknüpfung

	Wiederbefragung	Datenverknüpfung
Fallzahl	20.006	15.743
einverstanden	85%	72%
nicht einverstanden	15%	28%

4.7 Kurzbefragung von Nichtteilnahmebereiten

Alle nichtteilnahmebereiten Zielpersonen wurden gebeten, zumindest einige wenige Angaben zu machen. Ziel dabei war es, Ausfallanalysen erstellen zu können und nach Abschluss der Feldarbeit Hinweise auf soziodemographische Strukturen der Verweigerer zu erhalten. Die Verweigerer wurden gebeten, ihr Geburtsjahr und ihren Schulabschluss zu nennen. Das Geschlecht der Person wurde von dem Interviewer oder der Interviewerin eingestuft.

Voraussetzung für die Fragen an die Verweigernden war, dass es vorher bereits gelungen war, die Zielperson zu ermitteln. Für den Fall, dass die Kontaktperson verweigerte, was in den meisten Fällen passierte, wurden keine Angaben erhoben.

Insgesamt wurde 6.957-mal versucht, von den verweigernden Zielpersonen Informationen zu erhalten. Das Anliegen wurde wie folgt eingeleitet: „Auch wenn Sie nicht an der Studie teilnehmen möchten, würden Sie mir dennoch Ihr Geburtsjahr und Ihren Schulabschluss nennen?“. Hierzu erklärten sich lediglich 360 Personen bereit (5,3 %). In der Gruppe der verweigernden Zielpersonen sind auch 191 Personen enthalten, die entweder keine Berufsangabe machen wollten oder ihre Postleitzahl (bei Mobilfunkinterviews) nicht nennen wollten. Da beide Angaben unerlässlich sind für das Interview, wurden diese Personen zum Ausfall deklariert. Diese Fälle wurden dann im Rahmen der Kurzbefragung ohne weitere Einleitungsfrage nach ihrer Schulbildung gefragt, da Informationen zum Geburtsjahr und Geschlecht bereits im Rahmen der Hauptbefragung im Screening abgefragt wurden.

Zum Geburtsjahr liegen von Nichtteilnahmebereiten insgesamt 529 Angaben und 22-mal mal „keine Angabe“ vor. Die Frage nach dem Schulabschluss beantworteten 477 Personen, „keine Angabe“ war die Antwort von 74 Verweigerern. Eine Information zum Geschlecht liegt für 6.957 Personen vor, in 672 Fällen konnte das Geschlecht nicht von den Interviewenden eingeschätzt werden. Für das Geschlecht gab es mit 53 % nichtteilnahmebereiten Männern praktisch keine Abweichung gegenüber den Zahlen des Mikrozensus. Für Alter und Bildung sind die Fallzahlen sehr gering, so dass keine gesicherte Aussage über das Teilnahmeverhalten möglich ist.

5 Datenprüfung und -aufbereitung

Der Prozess der Datenprüfung wird bei computerunterstützten Befragungen bereits in die Entwicklungsphase der Erhebung verlagert. Prinzipiell ist hier das Ziel, ein Erhebungsinstrument zu entwickeln, welches fehlerfreie, konsistente und plausible Daten liefert: Das Fragenprogramm läuft unter Berücksichtigung der richtigen Filterführung auf dem Bildschirm ab; die Antworten werden von den Interviewenden direkt eingegeben. In das Programm implementierte Prüfroutinen und integrierte Verfahren der Datenprüfung und -edition stellen die Plausibilität und Konsistenz der Antworten sicher. Sollten sich während des Interviews Unstimmigkeiten ergeben, so können diese sofort mit dem Befragten geklärt und korrigiert werden. Vor Feldstart finden ausgiebige technische Tests des Erhebungsinstruments statt.

Vor dem Hintergrund, dass das Fragenprogramm unter Berücksichtigung der Filterführung auf dem Bildschirm abläuft, ist eine Datenbereinigung bei computerunterstützten Befragungen eigentlich nicht notwendig. Voraussetzung hierfür ist allerdings ein fehlerfreies Erhebungsinstrument. Dennoch kann es sich bei den Auswertungen zeigen, dass bei einem komplexen Erhebungsinstrument gewisse Inkonsistenzen bestehen – dies beispielsweise deshalb, weil man auf eine entsprechende Prüfung und Nachfrage verzichtet hat, um an einer bestimmten Stelle nicht den Gesprächsfluss zu stören, um keinen Interviewabbruch zu provozieren oder weil der Befragte die Inkonsistenz nicht korrigieren wollte oder konnte.

Da trotz intensiven Prüfungen des computergestützten Erhebungsinstruments vor Beginn der Erhebung nicht ausgeschlossen werden kann, dass die Filterführung völlig fehlerfrei abgelaufen ist, wurde eine intensive Prüfung der Filter vorgenommen. Verian hat hierfür aufwendige Prüfprogramme in SPSS erstellt. Diese Programme wurden auf Syntaxbasis programmiert, so dass der komplette Prozess der Prüfung und Fehlerbereinigung nachvollziehbar ist und Änderungen dokumentiert sind. In einem ersten Schritt wurde die korrekte Filterführung geprüft. In einem weiteren Schritt wurden Plausibilitätsprüfungen zwischen einzelnen Fragen durchgeführt.

In allen Fällen, in denen Auffälligkeiten entdeckt wurden, wurde die Notwendigkeit einer Datenbereinigung sowie die Möglichkeiten einer sinnvollen Bereinigung geprüft, dem Auftraggeber Vorschläge unterbreitet und die Bereinigung anschließend in Abstimmung mit ihm vorgenommen. Eine enge Abstimmung bei der Erstellung von Entscheidungsregeln muss auch wegen der Vergleichbarkeit zwischen den verschiedenen Erhebungswellen gewährleistet sein: Die Tests waren deshalb möglichst analog zu denen aus den Vorwellen aufgebaut, so dass die Entscheidungen zwischen den Wellen konsistent ausfallen konnten. An folgenden Variablen wurden zu Beginn der Feldphase Probleme sichtbar, woraufhin Korrekturen während der Feldphase vorgenommen wurden:

- Frage B1219: Bei der Frage nach der Nutzungsintensität der erworbenen Kenntnisse in der letzten beruflichen Ausbildung gab es zunächst einen Filterfehler. Dieser fiel in der ersten Datenprüfung nach Feldstart auf, sodass nur für zwei Fälle eine Antwort auf diese Frage fehlt. Es wurde ein Sondercode mit -9 mit dem Wertelabel „Filterfehler“ vergeben.
- Frage A1203: Bei der Frage nach der genauen Fach- oder Berufsbezeichnung der ersten Ausbildung gab es einen Fehler in der Fragebogenkonzeption, so dass für manche Konstellationen die Frage erstmal nicht gestellt wurde. Insgesamt waren 10 Fälle davon betroffen, auch hier wurde der Sondercode -9 mit dem Wertelabel „Filterfehler“ vergeben.

Die Datenprüfung fand mehrfach anhand von verschiedenen (Zwischen-)Datensätzen statt:

- Im Anschluss an jeden Pretest wurde ein Datensatz generiert, der aufbereitet und auf Filterfehler und Plausibilität geprüft wurde

- Jeder Zwischendatensatz wurde vor Auslieferung an den Kunden geprüft (nach 1.000, 8.000, 10.000 und 15.000 Interviews)
- Intensive Prüfung der endgültigen Befragungsdaten

Direkt nach Ende der Befragung stellt die verwendete Befragungssoftware NIPO die Befragungsdaten als ASCII-Datensatz bereit. Aus technischen Gründen entsprachen die Variablennamen der unaufbereiteten Daten nicht immer denen der Fragebogenvorlagen. Dies kommt beispielsweise bei Verschleifungen vor. Verian hat die Variablen entsprechend den Bezeichnungen in den Fragebogenvorlagen umbenannt, so dass die endgültigen Daten den Variablennamen der Fragebogenvorlage entsprechen. Darüber hinaus wurden Variablen für einfachere inhaltliche Auswertungen gebildet (z. B. Umrechnen des Geburtsdatums in das aktuelle Alter als eigene, zusätzliche Variable). Dies geschah mit einer eigenen SPSS-Syntax, so dass der Auftraggeber diese auch im Nachhinein einfach anpassen kann. Alle Datensätze wurden mit Variablen- und Wertelabels ausgeliefert.

Sämtliche Schritte der Datenprüfung und Datenaufbereitung wurden auf SPSS-Syntaxbasis durchgeführt. Damit ist sichergestellt, dass jede Veränderung, von den Rohdaten bis hin zum endgültigen Datensatz dokumentiert und nachvollziehbar ist.

Folgende Datensätze wurden an das BIBB und die BAuA geliefert:

- 25.10.2023: Zwischendaten mit 1.038 Fällen
- 14.12.2023: Zwischendaten mit 6.635 Fällen
- 20.12.2023: Zwischendaten mit 8.342 Fällen
- 12.01.2024: Zwischendaten mit 11.641 Fällen
- 08.02.2024: Zwischendaten mit 15.025 Fällen
- 15.04.2024: vorläufige finale Daten mit 20.006 Fällen
- 19.04.2024 Datensatz mit allen Zeitdauervariablen (20.006 Fälle)
- 07.05.2024: Datensatz zur Kontaktdokumentation mit allen telefonischen Kontakten mit Datum und Kontaktergebnis (4.548.888 Fälle)
- 08.05.2024: Bruttoband mit allen eingesetzten Telefonnummern und dem jeweils dazugehörigen finalen Kontaktergebnis (2.306.162 Fälle)
- 08.05.2024: Datensatz mit Verweigerern (6.957 Fälle)
- 08.05.2024: Datensatz mit Screeninginterviews (46.167 Fälle)
- 15.05.2024: Datensatz der Qualitätssicherung mit durchgeführten Kontrollen der Interviewenden (1.760 Fälle)
- 27.06.2024: finale Gewichtungsfaktoren
- 02.07.2024: Datensatz mit der Berufe- und Branchenvercodung

6 Vercodung der offenen Angaben zur beruflichen Tätigkeit und zum Wirtschaftszweig

Im Erhebungsinstrument gab es an verschiedenen Stellen offene Textangaben zum Wirtschaftszweig sowie zur beruflichen Tätigkeit der Zielperson und ihrer Eltern zu dem Zeitpunkt, als sie 15 Jahre alt war. Ebenso wurden Angaben zu Ausbildungsberufen erhoben. Sämtliche offenen Textangaben wurden im Anschluss an die Datenerhebung in verschiedene Klassifikationssysteme vercodet. Im Folgenden wird die Anzahl der zu vercodenden Angaben dargestellt, nach welchen Systemen die Angaben vercodet wurden und wie hierbei vorgegangen wurde.

Zur Ordnung von Berufsinformationen und, darauf aufbauend, zur Beschreibung von Struktur und Verteilung beruflicher Tätigkeiten kommen in Deutschland derzeit mehrere Klassifikationssysteme zum Einsatz. Das erste System ist die nationale „Klassifikation der Berufe 2010“ sowie deren überarbeitete Fassung 2020 (kurz KldB 2010 bzw. KldB 2020). Bei der KldB 2020 handelt es sich um eine Überarbeitung der Codierung bei einem Teil der Berufsgruppen. Für die Vercodung wurde auf die letzte verfügbare Überarbeitung der Bundesagentur für Arbeit zurückgegriffen (siehe Abschnitt 6.2.1). Beim anderen System handelt es sich um die Internationale Standardklassifikation der Berufe („International Standard Classification of Occupations“) des „International Labour Office“ (ILO) aus dem Jahr 2008 (ISCO 08). Demnach wurden die beruflichen Tätigkeiten nach drei Klassifikationssystemen vercodet.

Die Angaben zu den Wirtschaftszweigen wurden nach der neusten Wirtschaftszweigklassifikation der Europäischen Union („Nomenclature des statistiques des activités économiques de la Communauté Européenne“: NACE Rev. 2) bzw. der deutschen Klassifikation der Wirtschaftszweige 2008 (WZ 2008) kodiert. Bei der Wirtschaftszweigklassifikation nach NACE handelt es sich um das System der Europäischen Union, das System WZ stellt das der Bundesrepublik Deutschland dar. Die beiden Klassifikationssysteme stimmen jeweils bis zur Ebene der Klassen überein, die letzte Ebene der Unterklassen stellt ein nationales Spezifikum dar.

Im Folgenden werden die Mengengerüste der offenen Angaben, die durch Verian vercodet wurden, dargestellt und im Anschluss das genaue Vorgehen bei Verian erläutert.

Übersicht über die beruflichen Angaben:

- KldB 2010 80.745 Angaben
- KldB 2020 80.745 Angaben
- ISCO 08 80.745 Angaben

- Übersicht über die Angaben zum Wirtschaftszweig:
 - WZ 2008 19.788 Angaben

6.1 Das grundsätzliche Verfahren bei Verian

Wir beschreiben im Folgenden die Verfahren zur Vercodung der offenen Angaben zur beruflichen Tätigkeit nach KldB (2010 und 2020), ISCO08 sowie der offenen Angaben zu den Wirtschaftszweigen nach der Klassifikation der Wirtschaftszweige WZ2008 genauer (vgl. Abschnitt 6.2 und 6.3). Die Vorgehen beruhen grundsätzlich jeweils auf demselben mehrstufigen Verfahren:

1. **Automatik:** Grundlage der automatischen Zuordnung der Codes sind die elektronisch verfügbaren Verzeichnisse der Klassifikationssysteme.⁹ Wenn ein Begriff mit dem in der Befragung angegebenen Begriff genau übereinstimmt, wird der zugehörige Code verwendet und die Vercodung ist damit abgeschlossen. Von der automatischen Vercodung ausgenommen werden allgemeine Berufsbezeichnungen, sofern zusätzliche Angaben in den Befragungsdaten, z. B. zur Branche, vorliegen, die für eine differenzierte (manuelle) Vercodung herangezogen werden können (bspw. „Verkäufer/-in“ mit Angabe zur Branche).
2. **Code-Vorschläge durch Machine Learning:** Die auf der Grundlage von Algorithmen weiterentwickelte computergestützte Berufs- und Branchenklassifizierung (s. den folgenden kurzen Exkurs) erhält die unveränderten Nennungen aus der Befragung als Eingabe und klassifiziert diese in entsprechende Codes aus den Klassifikationsschemata für Berufe oder Branchen. Diese Klassifizierung enthält neben dem wahrscheinlichsten Code ggf. auch alternative Codes und jeweils die entsprechende Konfidenz (Confidence-Score) der Klassifizierung. Diese Vorschläge sind erfahrungsgemäß (mittlerweile) sehr gut, Verian sieht dennoch bis auf weiteres davon ab, diese Codes ungeprüft zu übernehmen.
3. Eine **manuelle Kodierung** der Leftovers, also der Fälle, die im Rahmen der automatischen Kodierung in Schritt 1 nicht kodiert werden konnten, findet unter Berücksichtigung der Vorschläge aus Schritt 2 statt. Die bei der manuellen Kodierung angewendeten Regeln sind eindeutig formuliert und dokumentiert (siehe folgende Abschnitte).

Exkurs: Generierung der Code-Vorschläge für die manuelle Kodierung

Die Datengrundlage, auf der die Software-Komponente arbeitet bzw. trainiert wurde und laufend weiter trainiert wird, besteht aus mehreren Hunderttausend bereits von menschlichen Experten in früheren Projekten durch Verian durchgeführten Klassifizierungen. Die offenen Angaben werden einem Neuronalen Netzwerk einer Multi-Layer-Perceptron-Architektur übergeben, welches auf der genannten Datenmenge trainiert und speziell für die Klassifizierung von Berufen bzw. Branchen optimiert wurde. Dieses Neuronale Netz ist in der Lage eine Klassifizierung durchzuführen, auch wenn die Berufs- oder Branchenbezeichnungen von bereits „gesehenen“ Daten abweichen.

Um die höchste Genauigkeit liefern zu können, wurden diverse Verfahren des Maschinellen Lernens intensiv geprüft und über Kreuzvalidierungsverfahren auf Genauigkeit getestet. Das im zweiten Schritt verwendete Neuronale Netzwerk war den anderen Methoden signifikant überlegen und lieferte konsistent eine sehr gute Genauigkeit, die über viele Iterationen von Optimierung und Kreuzvalidierung bis zum Status quo hin weiter verbessert wurde. Die Ausgabe der Software sind bis zu vier vorgeschlagenen Klassifizierungs-Codes inklusive der Trefferwahrscheinlichkeit in Form eines Confidence-Scores.

⁹ KldB2010: Bundesagentur für Arbeit, 2011: „Klassifikation der Berufe 2010 – alphabetisches Verzeichnis der Berufsbenennungen“ <https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010-erste-Fassung/Generische-Publikationen/Systematiken/Alphabetisches-Verzeichnis-Berufsbenennungen-Einfuehrung-KldB-2010.xls>
KldB2020: Bundesagentur für Arbeit, 2024: „Klassifikation der Berufe 2010 – überarbeitete Fassung 2020 - alphabetisches Verzeichnis der Berufsbenennungen“ <https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010-Fassung2020/Systematik-Verzeichnisse/Generische-Publikationen/Alphabetisches-Verzeichnis-Berufsbenennungen.xlsx>
ISCO08: „STATISTIK AUSTRIA – Die Informationsmanager“, 2010: „ISCO 08 – gemeinsame deutschsprachige Titel und Erläuterungen auf Basis der englischsprachigen Version von Juni 2009“
WZ08: Statistisches Bundesamt, 2008: Klassifikation der Wirtschaftszweige, Ausgabe 2008

6.2 Berufeencoding

Die Encoding der offenen beruflichen Angaben ist ein komplexes Verfahren. Dies liegt zum einen an den teilweise sehr unspezifischen Angaben der Befragten, zum anderen beinhalten die Klassifikationen sehr viele verfügbare Codes. Während bei KldB 2010 und KldB 2020 der fünfstellige Code die feinste Untergliederung hat, ist es beim ISCO 08 ein vierstelliger Code. Die jeweils feinste Gliederung wurde von Verian encoded. Der KldB 2010 hat demnach 1.286 verschiedene Codes (1.300 in der Version 2020) und der ISCO 08 beinhaltet 436 Codes.

Da häufig Berufsbezeichnungen angegeben werden, die nicht spezifisch genug sind, um diese eindeutig encode zu können (z. B. nur „Mechaniker“ statt „KFZ-Mechaniker“ oder nur „Lehrer“ statt „Gymnasiallehrer“), wurde im CATI-Programm eine Liste mit zu allgemeinen Berufsangaben hinterlegt. Wenn ein solcher Begriff genannt wurde, kam eine Nachfrage, ob es noch eine genauere Bezeichnung für diesen Beruf gibt. Mit diesem Verfahren war es möglich, typische – für eine Encoding zu allgemein gehaltene – Berufsbezeichnungen im Vorfeld zu identifizieren und die Befragten auffordern zu können, ihre Berufsbezeichnung detaillierter zu nennen.

Doppelkodierung aller Berufe

Um eine besondere Qualität bei den Kodierungen zu gewährleisten, haben wir – wie schon im Jahr 2018 – eine Doppelkodierung durchgeführt. Alle manuell kodierten Berufsangaben wurden zwei Mal von jeweils unterschiedlichen Personen kodiert. Bei Abweichungen zwischen den zugewiesenen Codes wurde in einem „Schiedsverfahren“ durch die bei Verian Verantwortliche für die Encoding entschieden, welcher Code der passendere war oder ob sogar ein anderer, dritter Code zu vergeben war. Durch dieses Verfahren wurde für größtmögliche Qualität der Kodierungsergebnisse gesorgt.

6.2.1 Vorgehensweise und allgemeine Regeln bei der manuellen Encoding

Grundsätzlich wird die Angabe encoded, die als berufliche Tätigkeit genannt wird. Liegen weitere Angaben in anderen Variablen (z. B. Branche, Stellung, für den Beruf erforderliche Ausbildung) vor, können diese zur genaueren Encoding herangezogen werden. Um die Encoder durch die zusätzlichen Angaben nicht sofort zu beeinflussen, wurde in der Bearbeitungsoberfläche nur angezeigt, ob zusätzliche Informationen vorliegen – was konkret in diesen Angaben enthalten ist, wird erst angezeigt, wenn die Encoder diese aufrufen. Hintergrund dieses Vorgehens ist, dass die Angabe encoded werden soll, die genannt wurde. Sofern dies eindeutig möglich ist, werden keine zusätzlichen Angaben verwendet.

Angaben zu Berufsausbildungen werden so behandelt, als wäre die Ausbildung abgeschlossen, soweit eine Zuordnung zu einem Beruf möglich ist (z. B. Schreinerlehre als Schreiner). Bei Studiengängen wird der Beruf encoded, der mit höchster Wahrscheinlichkeit mit diesem Abschluss ausgeübt wird.

Typischerweise enthalten die offenen Angaben Nennungen, die so allgemein sind, dass die Zuordnung von Codes schwierig oder nicht eindeutig möglich ist. Um dennoch möglichst viele Codes zuweisen zu können, mussten Entscheidungen für schwierige Fälle getroffen werden. Dies erscheint dann allerdings wenig problematisch, wenn die Entscheidungsregeln dokumentiert und diese Fälle entsprechend in einer Variable „XXX_prob“ (Name vom Klassifizierungssystem abhängig) gekennzeichnet werden.

Die allgemeinen Regeln werden an dieser Stelle kurz dargestellt:

Regel 1:

Vercodet wird grundsätzlich der angegebene Begriff, wenn er in der jeweiligen Klassifizierung der Berufe aufgeführt ist.

Regel 2:

Bei mehreren Nennungen in einer Klartextangabe wird die erste vercodet, es sei denn, die weiteren Angaben dienen einer näheren Spezifizierung. Bei Doppelangaben, bei denen die eine Angabe die andere näher spezifiziert, wird die jeweils spezifischere Angabe vercodet. Bei zusammengesetzten Begriffen, die so nicht vercodet werden können, wird, wenn möglich, der übergeordnete Begriff vercodet (z. B. Hellermonteur als Monteur).

Regel 3:

Wo eine Vercodung auf der Basis der offenen Angabe allein nicht möglich ist, werden zusätzliche Informationen – soweit vorhanden – hinzugezogen. Falls diese zusätzlichen Informationen für die Klassifizierung ausschlaggebend sind, wird dies in einer Variable „XXX_info“ (Name vom Klassifizierungssystem abhängig) dokumentiert.

Regel 4:

Ist die offene Angabe so unspezifisch, dass sie mehreren Kennziffern zugeordnet werden kann (z. B. Händler – hier ist nicht klar, ob es sich um Einzel- oder Großhändler handelt), wird sie derjenigen zugeordnet, für die angenommen werden kann, dass sie empirisch am häufigsten vorkommt. Diese Fälle werden in der Variable „XXX_prob“ (Name vom Klassifizierungssystem abhängig) gekennzeichnet (s. u.).

Regel 5:

Wenn die Angabe zur beruflichen Tätigkeit keinen Aufschluss über die für die Tätigkeit erforderliche Qualifikation gibt und keine zusätzlichen Informationen hierzu vorliegen (siehe auch Regel 3), wird der Code der Tätigkeit zugewiesen, für die die geringste Qualifikation notwendig ist. Diese Fälle werden in der Variable „XXX_prob“ (s. u.) gekennzeichnet.

Problematische Angaben, bei denen auf die Regeln 4 oder 5 zurückgegriffen wird, werden in einer Zusatzvariablen „XXX_prob“ als solche gekennzeichnet. Dadurch besteht die Möglichkeit, Auswertungen mit diesen Fällen und ohne sie vorzunehmen und somit Sensitivitätstests durchzuführen. Um darüber hinaus eine Hilfestellung zur Beantwortung der Frage zu geben, inwiefern ein solches Vorgehen inhaltlich notwendig ist, werden im Folgenden die konkreten Sachverhalte, bei denen die Entscheidungsregeln angewendet werden, detailliert dokumentiert.

Die Variable „XXX_prob“ hat damit folgende Ausprägungen:

- | | | |
|------------------|----|--|
| XXX_prob: | 1: | Qualifikationsstufe nicht bekannt, niedrigste vercodet |
| | 2: | Mehrere Codes möglich, Entscheidung getroffen |

Bei der Doppelvercodung kann es vorkommen, dass zwar die beiden vergebenen Codes identisch sind, aber von einem Kodierer ein XXX_prob oder XXX_info Code vergeben wurde, vom Zweitbearbeiter jedoch nicht. In diesen Fällen wird die „sicherere“ Variante übernommen. Das bedeutet, sollten bei einer Bearbeitung Zusatzinformationen verwendet worden sein, so wird der XXX_info Code auch final vergeben. Mit XXX_prob verhält es sich ebenso, wobei hier bei unterschiedlicher Vergabe die Ausprägung 2 („Mehrere Codes möglich, Entscheidung getroffen“) sticht. Wird ein Fall im Rahmen des Schiedsverfahrens in der Drittbearbeitung entschieden, so werden die Ausprägungen von

XXX_prob und XXX_info aus dieser Bearbeitung übernommen, unabhängig von den vorangegangenen zwei Bearbeitungen.

Im Unterschied zu den beiden älteren Systemen (KldB 1992 und ISCO 88), in denen für bestimmte Berufe Allgemeincodes vorgesehen waren (z. B. für Angestellter; Ingenieur, Selbstständiger usw.), gibt es diese Codes in den neueren Systemen des KldB 2010 und ISCO 08 nicht mehr. Verian codiert solche Angaben mit Sondercodes im Minusbereich (-10, -11 usw.), die im nächsten Abschnitt erläutert werden.

Sondercodes

Trotz der in eben erläuterten Entscheidungsregeln beim Vorliegen nicht eindeutiger Angaben ist es nicht möglich, allen offenen Angaben einen Code zuzuordnen. Um die vorliegende Information zur beruflichen Tätigkeit dennoch zu verwerten und um den Forschenden Hinweise auf deren Inhalt zu geben, werden in solchen Fällen, wo möglich, Sondercodes vergeben. Die Bedeutung der Sondercodes wird im Folgenden dokumentiert.

- 10: Aushilfe
- 11: Auszubildende, Azubi, Berufsvorbereitung
- 12: Arbeiter, Facharbeiter, Hilfsarbeiter, Produktionsarbeiter, Werker, Industrieangestellter, Industriearbeiter, Industriefacharbeiter, Teilwerker
- 13: Industriemeister, Meister
- 14: Techniker, technischer Angestellter
- 15: Ingenieur, technischer Ingenieur
- 16: Praktikant, Trainee, Volontär
- 17: Freiberufler, Geschäftsmann/-frau, Selbstständiger
- 18: mithelfender Familienangehöriger (sowohl in als auch außerhalb der Landwirtschaft)
- 19: Bundesfreiwilligendienst, Freiwilliges soziales Jahr, Wehrdienst, Zivildienst
- 20: Dienstleister, Dienstleistung
- 21: Zeitarbeit
- 22: Fachkraft

- 30: studentische Angestellte/Hilfskraft, Werkstudent, wissenschaftliche Hilfskraft
- 31: Bachelor
- 32: Diplom, Magister, Master, 1. Staatsexamen, Staatsexamen
- 33: 2. Staatsexamen, Assessor, 2. Staatsprüfung, Referendariat
- 34: Studium, Studium an Hochschule, Universitätsstudium, Fachhochschulabschluss, Fachhochschule, Aufbaustudium, Fachdiplom

- 50: Berater, Consultant
- 51: Einsatzleiter, Leiter, leitender Angestellter
- 52: Betriebsleiter
- 53: Kaufmann/-frau
- 54: Gutachter, Sachverständiger
- 55: Referent

- 60: sozial, Sozialbereich, Sozialwesen
- 61: medizinisch, Medizinbereich, medizinischer Bereich
- 62: Therapeut

- 90: Hauptschulbesuch, Hauptschulabschluss, Berufseinstiegsjahr (BEJ), Berufsgrundbildungsjahr (BGJ), Berufsvorbereitungsjahr
- 91: Realschulbesuch, Realschulabschluss, mittlere Reife, Fachoberschulreife, Berufsgrundschuljahr, Wirtschaftsschule, Wirtschaftsschulabschluss
- 92: Berufsfachschulbesuch, Fachschulreife
- 93: Fachschulbesuch, Fachschulabschluss (nur wenn eine Fachrichtung nicht erkennbar ist; wenn es sich z. B. um eine soziale, medizinische, pflegerische oder erzieherische Fachschule handelt, wird der entsprechende Code vergeben)
- 94: Handelsschulbesuch, Abschluss einer Handelsschule
- 95: Fachoberschulbesuch (FOS), Fachhochschulreife, Berufsfachoberschule, Berufsoberschule
- 96: Gymnasium, Abitur, Hochschulreife, Fachabitur, fachbezogener Hochschulzugang
- 1: keine Zuordnung möglich

6.2.2 Spezielle Regeln und Besonderheiten der jeweiligen Klassifikationssysteme

Die Berufsangaben wurden zuerst nach KldB 2010 codiert, bevor die Vercodung nach KldB 2020 und ISCO 08 vorgenommen wurde.

Durch dieses Vorgehen konnte für die Klassifikation nach ISCO 08 – entsprechend der Empfehlung der Internationalen Arbeitsorganisation (International Labour Organisation, ILO)¹⁰ – auf der nationalen Klassifikation aufgebaut werden. Hierzu wurden, nach dem ersten Schritt der automatischen Codierung, mithilfe des vorhandenen Umstiegsschlüssels¹¹, bis zu zwei mögliche ISCO 08 Codes (Schwerpunkt und erste Alternative) vorgeschlagen und vorrangig vor den Codes des Machine Learning bei der manuellen Bearbeitung berücksichtigt.

Für die Vercodung nach der überarbeiteten Fassung 2020 der Klassifikation der Berufe 2010 konnten keine Vorschläge durch Machine Learning generiert werden, da hier die entsprechende Datenbasis noch nicht vorhanden ist. Die Änderungen, die zwischen den Fassungen 2010 und 2020 der Klassifikation vorgenommen wurden, sind dokumentiert und systematisch nachvollziehbar. Deshalb wurden in der Vercodung KldB 2010-Codes, für die keine Änderung in der Fassung 2020 vorgenommen wurde, für KldB 2020 übernommen. Hierzu hat Verian, auf der offiziellen Änderungsliste¹² aufbauend und unter Berücksichtigung der geänderten Zuordnungen von Berufsgruppen¹³, eine Liste der Berufe mit geänderten Berufsgattungen erstellt. Diese ist im Anhang E zu finden. In der manuellen

¹⁰ International Labour Organisation, 2009: International Standard Classification of Occupations 2008 (ISCO-08): Structure, group definitions and correspondence tables, Chapter 5 "Adapting ISCO-08 for National Use and for the Production of Internationally Comparable Data"

¹¹ Bundesagentur für Arbeit, 2011: "Umstiegsschlüssel von der Klassifikation der Berufe 2010 (5-Steller) zur ISCO-08 (4-Steller)" <https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010-erste-Fassung/Generische-Publikationen/Umstiegsschlüssel/Umstiegsschlüssel-KldB2010-ISCO-08.xls>

¹² Statistik der Bundesagentur für Arbeit, Änderungen des Berufs- und Tätigkeitsverzeichnis (Historienliste), Januar 2024: <https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010-Fassung2020/Arbeitsmittel/Generische-Publikationen/KldB-Aenderungshistorie-ueF2020.xlsx>

¹³ Statistik der Bundesagentur für Arbeit, Grundlagen: Hintergrundinfo – Geänderte Zuordnungen von Berufen in der Klassifikation der Berufe – Auswirkungen auf die Arbeitsmarktstatistiken, Juli 2021: <https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Methodik-Qualitaet/Methodenberichte/Uebergreifend/Generische-Publikationen/Hintergrundinfo-Geaenderte-Zuordnungen-von-Berufen-in-der-KldB-Auswirkungen-Arbeitsmarktstatistik.pdf>

Bearbeitung wurden alle Nennungen mit KldB 2010-Codes, die von Änderungen betroffen waren, auf Änderungsbedarf im Einzelfall überprüft und entsprechend der KldB 2010 Code übernommen oder ein anderer nach KldB 2020 zugeordnet.

6.3 Branchenvercodung

Durch das Verfahren von Verian wird der zweistellige numerische Code für die Abteilungen zugewiesen. Typischerweise enthalten die offenen Angaben Nennungen, die so allgemein sind, dass die Zuordnung von Codes schwierig oder nicht eindeutig möglich ist. Um dennoch möglichst viele Codes zuweisen zu können, mussten Entscheidungen für schwierige Fälle getroffen werden. Dies erscheint dann allerdings wenig problematisch, wenn die Entscheidungsregeln dokumentiert werden. Verian hat alle bislang aufgetretenen Problemfälle erfasst und festgehalten, wie entschieden wurde. So entstand eine Sammlung sowohl von allgemeinen als auch speziellen Regeln, die genau festlegen, wie in solchen Fällen zu verfahren ist.

6.3.1 Vorgehensweise und allgemeine Regeln bei der manuellen Vercodung

Grundsätzlich wird die Angabe vercodet, die als Branche genannt wird. Liegen weitere Angaben in anderen Variablen (z. B. Angabe zu beruflicher Tätigkeit, Stellung im Beruf o. ä.) vor, können diese zur genaueren Vercodung herangezogen werden. Um die Vercoder durch die zusätzlichen Angaben nicht sofort zu beeinflussen, wurde in der Bearbeitungsoberfläche nur angezeigt, ob zusätzliche Informationen vorliegen – was konkret in diesen Angaben enthalten ist, wird erst angezeigt, wenn die Vercoder diese aufrufen. Hintergrund dieses Vorgehens ist, dass die Angabe vercodet werden soll, die genannt wurde. Sofern dies eindeutig möglich ist, werden keine zusätzlichen Angaben verwendet.

Die allgemeinen Regeln werden an dieser Stelle kurz dargestellt:

Regel 1:

Vercodet wird grundsätzlich der angegebene Begriff, wenn er in der „Klassifikation der Wirtschaftszweige“ des Statistischen Bundesamtes aufgeführt ist.

Regel 2:

Bei mehreren Nennungen in einer Klartextangabe wird die erste vercodet, es sei denn, die weiteren Angaben dienen einer näheren Spezifizierung. Bei Doppelangaben, bei denen die eine die andere näher spezifiziert, wird die jeweils spezifischere Angabe vercodet.

Regel 3:

Wo eine Vercodung auf der Basis der offenen Angabe allein nicht möglich ist, werden zusätzliche Informationen – soweit vorhanden – hinzugezogen; diese eventuell zudem in geeigneter Kombination. Falls dies der Fall ist, wird dies in einer Variable „WZ2008_info“ dokumentiert.

Regel 4:

Ist die offene Angabe so unspezifisch, dass sie mehreren Kennziffern zugeordnet werden kann, wird sie derjenigen zugeordnet, für die angenommen werden kann, dass sie empirisch am häufigsten vorkommt.

Regel 5:

Bei Angaben zum Wirtschaftszweig, die nicht erkennen lassen, ob Herstellung oder Handel gemeint ist, wird Herstellung vercodet.

Problematische Angaben, bei denen auf Regel 4 zurückgegriffen wird, werden in der Zusatzvariablen „WZ2008_prob“ mit Wert 2 als solche gekennzeichnet. Problematische Angaben, bei denen Regel 5 zur Verwendung kommt, werden in der Zusatzvariablen „WZ2008_prob“ mit Wert 1 gekennzeichnet. Dadurch besteht die Möglichkeit, Auswertungen mit diesen Fällen und ohne sie durchzuführen und somit Sensitivitätstests durchzuführen.

Die Variable „WZ2008_prob“ hat damit folgende Ausprägungen:

XXX_prob:	1:	Herstellung oder Handel unklar, Herstellung vercodet
	2:	Angabe unspezifisch, wahrscheinlichste vercodet

6.3.2 Spezielle Regeln und Besonderheiten der Klassifikation nach NACE

Die Befragten nennen als Branche oftmals den Namen der Firma, nochmals die berufliche Tätigkeit oder eine Bezeichnung, die nicht in das Klassifikationsschema passt bzw. unter mehrere der Wirtschaftszweig-Abteilungen fallen könnte, z. B. wenn der Betrieb eine Mischform darstellt. Grundsätzlich stellt die Vercodung nach NACE bei Letzterem auf den *überwiegenden* Geschäftsbereich ab, der von uns aber oft nicht feststellbar ist.

Vielfach kennen die Befragten ihre Branche offensichtlich auch gar nicht, wie Angaben wie Industrie, Handwerk, IG Metall, IG Phantasienamen, die den Bereich irgendwie umschreiben, zeigen.

Ebenso sind allgemeine Angaben wie Elektro, Elektriker, Elektrohandwerk, Elektrobranche, Elektrobetrieb usw., Metall, Metallbearbeitung, Metallbereich, Metallindustrie, Metallgewerbe nicht eindeutig vercodbar.

Auch Angaben wie z. B. Agrargenossenschaft, Kommune, öffentlicher Dienst oder ein Gewerkschaftsbereich beziehen sich, wie den zugeordneten Berufen zu entnehmen ist, vielfach nicht auf diese Branche, sondern dokumentieren offensichtlich das Zugehörigkeitsgefühl bzw. die Mitgliedschaft der Befragten zu, nicht aber ihr Beschäftigungsverhältnis in dieser Kategorie. Hier einfach die entsprechenden Codes (Kommune = Kommunalverwaltung, Agrargenossenschaft = Interessengemeinschaft, IG Metall oder IG Druck und Papier = Gewerkschaft = Interessenvertretung) zu vergeben, würde die Aussagekraft der Angaben deutlich verfälschen. In solchen Fällen wird – wo möglich – nach der beruflichen Tätigkeit der entsprechende Wirtschaftszweig ermittelt. Lässt die Angabe zur beruflichen Tätigkeit darauf schließen, dass der Befragte tatsächlich in dem fraglichen Bereich tätig ist oder sein könnte, also z. B. in der Kirchenverwaltung oder als Gewerkschaftssekretär, oder keine Berufsangabe vorliegt, wird der entsprechende Code vergeben.

Bei zu allgemeinen Angaben, wie z. B. Bergbau oder Fensterbau, werden Codes vergeben, die spekulativ erscheinen mögen, aber aufgrund von überwiegendem Vorkommen getroffen werden (z. B. Fenster-/Türen Bau = Holzverarbeitung, da Holz hierfür das überwiegende Material darstellt). Diese Codes werden in „WZ2008_prob“ mit 2 gekennzeichnet.

Bei mehreren Angaben (z. B. Metall-Kunststoff) wird in der Regel die erste vercodet, es sei denn, die weiteren Angaben dienen einer näheren Bestimmung der ersten.

Allgemeine Angaben ohne nähere Charakterisierung, die nicht erkennen lassen, ob Herstellung oder Handel gemeint ist, wie z. B. Kraftfahrzeuge/Lebensmittel o. ä., werden dem Bereich zugeordnet, der in der Wertschöpfungskette am weitesten vorne liegt, also beispielsweise dem Bereich Förderung, Erzeugung oder Herstellung und nicht dem Bereich Handel, Vermietung oder Reparatur. Auf der oben angesprochenen Zusatzvariablen „WZ2008_prob“ wird in diesen Fällen der Wert 1 gesetzt.

Eine Ausnahme von dieser Regel bildet die Angabe Computer, Computerbranche ohne nähere Spezifizierung. Hier wird nicht Computerherstellung vercodet, sondern Erbringung von Dienstleistungen der Informationstechnologie, da diesem Wirtschaftszweig im Vergleich zur Herstellung die größere Reichweite zugeordnet werden kann. Auch hier wird auf der Zusatzvariablen „WZ2008_prob“ der Wert 1 gesetzt.

Bei Metallbranchen wird die allgemeine Angabe Metall, Metallindustrie o.ä. wie *Verarbeitung*, also mit 25 („WZ2008_prob“ Wert 2), vercodet; bei den einzelnen Metallen (wie Aluminium, Stahl o. ä.) werden die analogen Angaben wie *Erzeugung*, also mit 24 („WZ2008_prob“ Wert 1), vercodet, da die zusätzlichen Berufsangaben weit überwiegend diese Einordnung nahelegen.

Außerdem werden je Klassifikationssystem vier weitere Codes für Angaben vergeben, die sonst nicht zuzuordnen wären:

WZ 2008:

201	Industrie
202	Handwerk
203	Dienstleistungen
204	Produzierendes Gewerbe
-1	keine Zuordnung möglich

7 Gewichtung

Ziel der Untersuchung ist es, eine repräsentative Stichprobe der Grundgesamtheit zu erhalten. Telefonbasierte Befragungen über Festnetz haben zwei Designkomponenten, die die Proportionalität der Abbildung beeinträchtigen:

- die von der Zahl der für Gespräche nutzbaren Anschlüsse abhängige Auswahlwahrscheinlichkeit für Telefonhaushalte und
- die von der Zahl der zur Zielgesamtheit gehörenden Personen im Haushalt abhängende Auswahlwahrscheinlichkeit für Personen.

Eine besondere Herausforderung entsteht, wenn zusätzlich eine Befragung über Mobilfunk durchgeführt wird, da dann eine Haushaltsstichprobe (Festnetz) und eine Personenstichprobe (Mobilfunk) zusammengeführt werden müssen. Darüber hinaus sollten die unterschiedlichen Teilnahmewahrscheinlichkeiten des Auswahlrahmens 2 berücksichtigt werden (siehe Abschnitt 3.3).

In der Mobilfunkstichprobe ist die Auswahl der Personen mit der Selektion der Nummern als einfache Zufallsauswahl bereits abgeschlossen: Der kontaktierte (Haupt-) Nutzer des Gerätes ist die Zielperson. Demgegenüber wird in der Festnetzstichprobe ein zweistufiges Auswahlverfahren angewandt. In der Kontaktphase wird zunächst die Anzahl der Zielpersonen, also der Personen ab 15 Jahren, ermittelt, auf der zweiten Stufe wird anschließend mit dem Schwedenschlüssel (Kish-Selection-Grid) zufällig eine Zielpersonenauswahl innerhalb des jeweiligen Haushaltes vorgenommen.

Darüber hinaus gibt es einen zweiten Ursachenkomplex für Abbildungsprobleme: Einerseits ist bei Zufallsstichproben ein Zufallsfehler unvermeidlich, andererseits sind bei jeder auf Freiwilligkeit basierenden Befragung nonresponse-bedingte Selektivitäten zu konstatieren.

Aufgrund dieser Stichprobenanlage sind die Befragungsdaten stets zu gewichten. Mit einer Designgewichtung werden die unterschiedlichen Auswahlchancen korrigiert. Zusätzlich ist eine Ausfallgewichtung nach demografischen Größen auf Basis amtlicher Werte ein bewährtes Mittel zur Verbesserung der Abbildung der Grundgesamtheit durch die Stichprobe. Beide Verfahren werden im Folgenden beschrieben.

7.1 Designgewichtung

Die designbedingten Unterschiede in den Auswahlwahrscheinlichkeiten sind theoretisch und praktisch leicht mittels Designgewichtung zu korrigieren. Unterschiedliche Auswahlwahrscheinlichkeiten ergeben sich durch den Dual-Frame-Ansatz, durch die unterschiedliche Anzahl an Telefonnummern, unter denen die Zielperson erreichbar ist und durch die unterschiedliche Anzahl an Zielpersonen im Haushalt.

Beim Dual-Frame-Design ist die Inklusionswahrscheinlichkeit insgesamt abhängig von der Wahrscheinlichkeit, per Festnetz oder per Mobilfunk befragt zu werden. Die Wahrscheinlichkeit, für die gleiche Untersuchung in beiden Modes ausgewählt zu werden, ist ausgesprochen klein und wird deshalb an dieser Stelle vernachlässigt.

Jede Telefonnummer hat innerhalb des Modes die gleiche Auswahlchance. Personen, die über mehrere Rufnummern verfügen, haben daher eine der Zahl dieser Anschlüsse entsprechende Mehrfachchance bei der Auswahl. Zu deren Bereinigung wird die realisierte Stichprobe mit der reziproken Zahl der Telefonnummern multipliziert. Gleichzeitig ist die Abhängigkeit der Auswahl von der Anzahl der Zielpersonen im Haushalt zu korrigieren. In jedem der Haushalte der Festnetzstichprobe wird nur eine Zielperson für das Interview per Zufall ausgewählt. Hierzu wird der Schwedenschlüssel verwendet.

Dieses Verfahren produziert innerhalb des einzelnen Haushaltes gleiche Auswahlchancen. Auf der Ebene der Gesamtstichprobe aber ist die Auswahlchance des Einzelnen umgekehrt proportional zur Zahl der potenziellen Zielpersonen im jeweiligen Haushalt. Daher ist die Stichprobe ex-post mathematisch mit dem reziproken Wert (Designgewicht) umzuformen.

Eine denkbare Abhängigkeit der Inklusionswahrscheinlichkeit von der Anzahl der möglichen Nutzer eines Handys kann derzeit vernachlässigt werden. In fast allen Fällen handelt es sich bei Mobilfunktelefonen um persönliche Geräte, die nicht gemeinsam genutzt werden.

Nach Gabler und Ayhan (2007) ergibt sich die Inklusionswahrscheinlichkeit¹⁴:

$$\pi_i \approx k_i^F \frac{m^F}{M^F} \cdot \frac{1}{z_i} + k_i^C \frac{m^C}{M^C}, i = 1, \dots, N$$

$m^{F/C}$ Anzahl Nummern in der Stichprobe (Festnetz / Mobilfunk)

$M^{F/C}$ Anzahl gültiger Rufnummern insg. (Festnetz / Mobilfunk)

$k_i^{F/C}$ Anzahl der Rufnummern (Festnetz / Mobilfunk) über die der Haushalt bzw. die Person erreicht werden kann

z_i Anzahl Zielpersonen im Haushalt

Folgende Komponenten werden im Interview erhoben: Anzahl der persönlichen Rufnummern in Fest- (k_i^F) und Mobilnetz (k_i^C) und die Anzahl der Personen ab 15 Jahren im Haushalt (z_i). Dies gilt für alle Befragten, unabhängig davon, in welchem Mode sie interviewt wurden.¹⁵ Darüber hinaus ist der unterschiedliche Auswahlatz (m^F/M^F bzw. m^C/M^C) der beiden Netzziehungen zu berücksichtigen. Die Werte m^F und m^C werden unter Verwendung des ersten Designfaktors gezählt. Die Werte für die Parameter M^F und M^C werden im Rahmen einer großen, unabhängigen Grundlagenstudie bestimmt. Hierzu kann auf eine regelmäßig (jährlich) durchgeführte F2F-Zufalls-Bevölkerungstichprobe in der Größe von rund 20.000 Befragten zurückgegriffen werden, für die vorliegende Studie auf die Daten aus 2023. In dieser Studie wird neben der Demografie unter anderem auch die Haushaltsausstattung und persönliche Ausstattung bezüglich Festnetztelefon und Mobilfunk erhoben. Damit ist man in der Lage, die Größen M^F respektive M^C für die Grundgesamtheit der Erwerbstätigen auf unabhängige Weise zu bestimmen.

Für die vorliegende Studie wurden M^F und M^C anhand der Auswertung der Mehrthemenbefragung mit Unterstützung folgender Basiswerte ermittelt, daraus resultieren die folgenden Werte:

M^F : 44,169 Mio.

M^C : 74,334 Mio.

Ab diesem Schritt werden die Festnetz- sowie die Mobilfunkstichprobe als eine gemeinsame Personenstichprobe behandelt.

Berücksichtigung unterschiedlicher Teilnahmewahrscheinlichkeiten innerhalb des Auswahlrahmen 2

Um evtl. unterschiedliche Teilnahmewahrscheinlichkeiten innerhalb der Auswahlrahmen 2-Stichprobe zu berücksichtigen, wurde darüber hinaus ein Logit-Modell für diese Teil-Stichprobe gerechnet. Die Zielvariable war „Teilnahme ja/nein“, als erklärende Variablen standen Alter, Geschlecht, Anzahl der Personen im Haushalt, Bildung und das Bundesland zur Verfügung. Darüber hinaus gingen Interaktionsvariablen aus Geschlecht und Alter in das Modell ein. Die Regression hatte mit einem Nagelkerke R^2 von knapp 0,15 eine geringe Erklärungskraft, was ein Hinweis darauf ist, dass sich die Teilnahmewahrscheinlichkeiten nicht stark unterscheiden haben. Die Teilnahmewahrscheinlichkeiten

¹⁴ Gabler, S. und Ayhan, Ö. (2007): Gewichtung bei Erhebungen im Festnetz und über Mobilfunk: Ein Dual-Frame-Ansatz, in: Mobilfunktelefonie – Eine Herausforderung für die Umfrageforschung, Gabler S., Häder S. (Hrsg.), GESIS – ZUMA-Nachrichten Spezial Band 13.

¹⁵ In der entsprechenden Frage werden zudem explizit Telefonnummern, die nur für Fax oder PC benutzt werden, ausgeschlossen.

wurden als Variablen weggeschrieben und der reziproke Wert wurde mit dem berechneten Designgewicht multipliziert. Dieses Designgewicht bildet den Eingangsfaktor der folgenden Ausfallgewichtung.

7.2 Ausfallgewichtung über Strukturanpassung

Über Designkomponenten hinaus gibt es einen zweiten Ursachenkomplex für Abbildungsprobleme: Fehlende Erreichbarkeit und Nonresponse sind Quellen für mögliche Selektivitäten. Diese werden im Rahmen einer Ausfallgewichtung über eine Strukturanpassung ausgeglichen.

Im Rahmen der Strukturanpassung werden Referenzstatistiken herangezogen, die (per Konvention) als Abbildung der Grundgesamtheit gelten und die interessierenden Strukturmerkmale in vergleichbarer Weise abbilden. Ausfallgewichtungen orientieren sich in diesen Fällen an Vergleichen der Strukturen der realisierten Stichprobe mit den Strukturen der Grundgesamtheit und passen dann die Stichprobenstrukturen denen der Grundgesamtheit mittels Gewichtungsfaktoren an (Kalibrierung).

Jede Art von Ausfallgewichtung basiert auf der Annahme, dass die Strukturmerkmale, für die im Rahmen der Ausfallgewichtung Anpassungen vorgenommen werden, mit den inhaltlich interessierenden Merkmalen korreliert sind und somit über die Gewichtung auch deren Strukturen besser an die in der Grundgesamtheit angepasst werden. Es empfiehlt sich, solche Strukturgewichtungen so behutsam wie möglich vorzunehmen und dabei Merkmale (und deren Kombinationen) zu verwenden, die möglichst hoch mit den Zielmerkmalen der Untersuchung korreliert sind (vgl. z. B. Heyde 2014, S. 58).¹⁶

Die Berechnung der Gewichtungsfaktoren erfolgt in einem iterativen Randsummenverfahren. Auf Basis der Merkmale oder Merkmalskombinationen wird das Modell als Gewichtung mit mehrdimensionalen Randverteilungen realisiert. Die Gewichtungsfaktoren werden dabei auf ein vorgegebenes geschlossenes Intervall begrenzt, die Faktorenspannbreite darf also ein bestimmtes vorher definiertes Maximum nicht überschreiten. Damit werden extrem hohe oder extrem niedrige Faktoren verhindert. Die Begrenzung der Gewichtungsfaktoren wurde bei der vorliegenden Untersuchung bei Minimum 0.08 und Maximum 9.9 vorgenommen. Die Priorität der Gewichtung gibt an, welcher Gewichtungsrand die anderen Gewichtungsråder dominiert, d. h. welcher Gewichtungsrand bevorzugt wird, wobei sich dadurch die nicht-dominanten Gewichtungsråder in ihrer Anpassung etwas verschlechtern können.

Ein integriertes Verfahren mit iterativem Vorgehen ermöglicht es, Verschiebungen der Strukturen von in einem vorangegangenen Iterationsschritt berücksichtigten Merkmalen in späteren Iterationsschritten, die sich auf diese Merkmale beziehen, wieder zu korrigieren. Dadurch wird die Anpassung an alle Ränder sukzessive so lange verbessert, bis ein vorher definiertes Konvergenzkriterium erreicht ist. Als Konvergenzkriterium dient die Vorgabe, dass für jede Randverteilung maximal 0,1 % Abweichung in der Summe der einzelnen Zellen toleriert wird. Die Iteration wird dann beendet, wenn die Anpassung an die einzelnen Randverteilungen nicht mehr verbessert wird.

Während der Gewichtung wird des Weiteren geprüft, ob die Faktoren in einzelnen Zellen zu hoch oder zu niedrig werden oder die Zellenbesetzung gleich Null ist. In diesen Fällen werden Zusammenfassungen mit einer geeigneten Nachbarzelle vorgenommen.

¹⁶ Heyde, C. von der (2014): Techniken und Möglichkeiten der Realisierung von Random-Stichproben. S. 37 – 65 in: ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V.: Stichproben-Verfahren in der Umfrageforschung. Eine Darstellung für die Praxis. 2. Auflage. Wiesbaden: Springer VS.

Im Folgenden wird das Vorgehen kurz genauer dargestellt. Die allgemeine Formel für die Gewichtung lautet (vgl. Deming, Stephan 1940; Cochran 1968): ¹⁷

$$g_j = \frac{1}{\frac{n_j}{n} * \frac{N}{N_j}} = \frac{n * N_j}{n_j * N}$$

g = Gewicht

j = Schicht, definiert durch das jeweilige Gewichtungsmerkmal

n_j = Anzahl (designgewichteter) Fälle in Schicht j in der Stichprobe

n = Anzahl der Fälle in der Stichprobe

N_j = Anzahl der Fälle in Schicht j in der Grundgesamtheit

N = Anzahl der Fälle in der Grundgesamtheit

Diese allgemeine Formel kann im iterativen Randsummenverfahren wiedergefunden werden. Wir illustrieren dies am Beispiel eines iterativen Randsummenverfahrens über zwei Schichten. Hierbei werden zwei Randverteilungen nacheinander iterativ angepasst. Zuerst werden die Gewichte gemäß der Verteilung von Schicht i berechnet. Darauf folgend werden die Gewichte gemäß der Verteilung von Schicht j berechnet, wobei das Ergebnis aus dem 1. Rand die Eingangsverteilung für diesen 2. Rand bedeutet.

$$n_{ij} \Rightarrow m'_{ij} \Rightarrow m''_{ij} \Rightarrow m'''_{ij} \Rightarrow m''''_{ij} \Rightarrow \text{usw.}$$

1. Rand über Schicht i (gemäß Gleichung 52 in Deming, Stephan (1940))

$$m'_{ij} = n_{ij} * \frac{n * N_{i.}}{n_{i.} * N}$$

2. Ev. Kappung von Faktoren höher als bestimmte Grenze

3. Rand über Schicht j (gemäß Gleichung 53 in Deming, Stephan (1940))

$$m''_{ij} = m'_{ij} * \frac{n * N_{.j}}{n_{.j} * N}$$

4. Ev. Kappung von Faktoren höher als bestimmte Grenze

¹⁷ Vgl. Cochran, W.G. (1968): The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*. 24. 295-313.

Deming, W.E., Stephan, F.F. (1940): On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*. 11. 427-444.

wobei:

n_{ij} = Anzahl der Fälle in Zelle ij (ev. designgewichtet) in der Stichprobe

m'_{ij} resp. m''_{ij} = Anzahl der Fälle in der Zelle ij nach jeweiliger Randanpassung in der Stichprobe

i bzw. j = Schicht i bzw. j , i und j laufen über die Ausprägungen in der jeweiligen Schicht: i kann zum Beispiel das Alter in 4 Ausprägungen und j die Bildung in 3 Ausprägungen sein.

n = Anzahl der Fälle insgesamt der Stichprobe

$n_{i.}$ = Anzahl der Fälle in Schicht i in der Stichprobe, ungeachtet von Schicht j (analog: $n_{.j}$)

N = Anzahl der Fälle in der Grundgesamtheit

$N_{i.}$ = Anzahl der Fälle in Schicht i in der Grundgesamtheit, ungeachtet von Schicht j (analog: $N_{.j}$)

Nun beginnt die Iteration: das Resultat aus der Anpassung an Schicht j (also Verteilung nach dem 2. Rand) gilt als Eingangsverteilung für eine erneute Berechnung gemäß der Verteilung von Schicht i und wiederum nachfolgend Schicht j usw. Die Iteration wird so lange wiederholt, bis die Anpassungen zufriedenstellend sind bzw. keine Verbesserung der Anpassung zu erkennen ist. Bei der Iteration über weitere Schichten wird jeweils analog verfahren.

Zusammenfassend ist festzuhalten: Im Rahmen der Kalibrierung zur Erstellung der Gewichtungsfaktoren wird systematisch geprüft, ob und in welchem Umfang disproportionale Ausfälle auftreten. Dazu wird die realisierte Nettostichprobe mit den Verteilungen der Grundgesamtheit verglichen. Nach Festlegung der in Bezug auf eine Anpassung relevanten Merkmale wird das Gewichtungsmodell mit mehrdimensionalen Randverteilungen realisiert. Es wird ein iterativer Gewichtungsprozess aufgesetzt, dessen Ergebnis Gewichtungsfaktoren sind, die die realisierte Stichprobe an alle vorgegebenen Sollverteilungen mit vordefinierter Genauigkeit und minimaler Varianz anpassen.

Die Ausfallgewichtung erfolgte weitgehend nach denselben Strukturmerkmalen wie bei der Gewichtung der Erwerbstätigenbefragung 2012, wurde aber aufgrund des Dual-Frame-Auswahlrahmens um das hierfür zu empfehlende zusätzliche Merkmal Haushaltsgröße ergänzt. 2018 erfolgte die Ausfallgewichtung aufgrund eines geänderten Screeningverfahrens direkt auf die Strukturen der Erwerbstätigen ab 15 Jahren, während 2012 und 2024 in einem vorgelagerten Schritt auf die Gesamtbevölkerung ab 15 Jahren gewichtet wurde.¹⁸

Zur Ermittlung des endgültigen Gewichtungsmodells wurden – wie in 2012 und 2018 – zwei Varianten gerechnet, wobei das Merkmal „Stellung im Beruf“ (StiB) bei Variante 1 nicht berücksichtigt und in Variante 2 berücksichtigt wurde. Die Referenzstrukturen bei der Gewichtung lieferten Sonderauswertungen des Mikrozensus 2023.¹⁹ Tabelle 8 gibt einen Überblick über die Ränderkombinationen für die Gewichtung und ob die Referenz die Gesamtbevölkerung ab 15 Jahren oder Erwerbstätige ab 15 Jahren waren. Bei der Gewichtung auf die Gesamtbevölkerung wurde auf alle Interviews, inkl. Screeninginterviews ($n=46.167$) gewichtet, bei der Gewichtung auf die Erwerbstätigen ab 15 Jahren auf die Interviews mit Erwerbstätigen ($n=20.006$).

¹⁸ 2018 erfolgte bei der Festnetzstichprobe die Zufallsauswahl unter allen Erwerbstätigen ab 15 Jahren. Der Anteil an Haushalten ohne Erwerbstätige war jedoch deutlich geringer als erwartet, vermutlich wegen verdeckten Verweigerungen. 2024 wurde bei der Festnetzstichprobe deshalb wieder unter allen Personen ab 15 Jahren ausgewählt.

¹⁹ An dieser Stelle möchten wir uns herzlich bei Destatis für die zur Verfügung gestellten Sonderauswertungen und beim BIBB für das Beschaffen der benötigten Zählungen bedanken.

Tabelle 8: Ränderkombinationen der Gewichtungsmodelle (Referenzstruktur: Mikrozensus 2023)

Ränderkombinationen ¹⁾	Variante 1: ohne StiB	Variante 2: mit StiB	Referenz: Personen ab 15 J./ Er- werbstätige ab 15 J.
Regierungsbezirk x BIK10	x	x	Pers. ab 15 J.
Bundesland x BIK10	x	x	Pers. ab 15 J.
Nielsengebiet x Haushaltsgröße	x	x	EWT ab 15 J.
Stellung im Beruf (StiB)		x	EWT ab 15 J.
West/Ost x Geschlecht x Nationalität	x	x	EWT ab 15 J.
Geschlecht x Altersgruppe x Familienstand	x	x	EWT ab 15 J.
West/Ost x Bildung x Alter (15-50, 51+) ²⁾	x	x	EWT ab 15 J.
Nielsengebiet x Geschlecht x Alter (6 Gruppen)	x	x	Pers. ab 15 J.
Geschlecht x Alter (6 Gruppen)	x	x	EWT ab 15 J.
Bundesland x Geschlecht	x	x	Pers. ab 15 J.
Fallzahlen			46.167(Pers. ab 15 J.)/ 20.006 (EWT ab 15 J.)

1) Die Merkmalsausprägungen der Variablen sind Tabelle 9zu entnehmen.

2) Hier wurde eine Dichotomisierung verwendet. Hintergrund ist, dass im Mikrozensus die Angabe zur Schulbildung freiwillig ist, wenn die befragte Person mindestens 51 Jahre alt ist. Dementsprechend ist für diese Befragten ein deutlich höherer Anteil fehlender Angaben festzustellen.

7.3 Qualitätsindikatoren der Gewichtung

Im Folgenden werden die Qualität der beiden Gewichtungsmodelle und die Güte der Stichprobe betrachtet. Zunächst werden die Merkmale, die in die Gewichtung einfließen, in ihrer Verteilung in der Referenzstatistik und in den beiden Gewichtungsmodellen dargestellt sowie die „Strukturinkongruenz“ betrachtet. Die Strukturinkongruenz wird hier gemessen als die Summe der absoluten Abweichungen zwischen Soll- und Stichprobenstruktur nach Gewichtung für die in Tabelle 9 dargestellten Merkmale. Anzumerken ist, dass die Abweichung noch geringer ausfallen würde, wenn man jeweils nur die gültigen Angaben heranziehen und darauf prozentuieren würde. Dies liegt am Vorgehen bei der Entwicklung der Gewichtungsfaktoren, welches darauf ausgelegt ist, die Verteilungen bezüglich der gültigen Angaben anzupassen.

Tabelle 9: Vergleich zwischen Stichproben- und Referenzstruktur für Erwerbstätige ab 15 Jahren (Mikrozensus 2023)²⁰

		Ist-Struktur gewichtet	
Strukturmerkmal	MZ 2023	Variante 1: ohne Stib	Variante 2: mit Stib
Geschlecht			
Männlich	54,1	53,8	53,8
Weiblich	45,9	46,1	46,1
Familienstand			
Ledig	36,6	36,6	36,6
Verheiratet	53,8	53,5	53,5
Verwitwet	1,4	1,4	1,4
Geschieden	8,3	8,3	8,3
Angabe fehlt		0,2	0,2
Deutsch - Nicht-Deutsch			
Nicht-deutsch	14,7	14,4	14,4
Deutsch	85,3	85,5	85,5
Angabe fehlt		0,1	0,1
Alter			
15 bis 20	1,5	0,7	0,7
21 bis 30	16,5	17,2	17,2
31 bis 40	23,4	23,4	23,4
41 bis 50	21,9	21,8	21,8
51 bis 64	33,9	33,8	33,8
65 und älter	2,7	2,7	2,7
Angabe fehlt		0,4	0,4
Stellung im Beruf (Stib)			
Arbeiter	11,3	10,8	11,3
Angestellte	74,5	73,0	74,4
Beamte	5,5	6,6	5,5
Selbständig	8,6	9,0	8,6
Mith. Familienangeh.	0,1	0,5	0,1
Angabe fehlt		0,1	0,1
Höchster Schulabschluss			
Maximal Volks-/Hauptschule	21,2	21,0	21,0
Mittlere Reife/POS	32,4	32,3	32,3
Fachhochschulreife, Abitur	46,4	46,4	46,4
Angabe fehlt		0,3	0,3

²⁰ Die Merkmale Alter, Geschlecht und Bundesland wurden auf Ebene der Gesamtbevölkerung gewichtet, deshalb kommt es hier zu geringen Abweichungen bei den gewichteten Strukturen der Erwerbstätigen.

		Ist-Struktur gewichtet	
Strukturmerkmal	MZ 2023	Variante 1: ohne StB	Variante 2: mit StB
Bundesland, Wohnort			
Schleswig-Holstein	3,4	3,4	3,4
Hamburg	2,4	2,3	2,3
Niedersachsen	9,5	9,5	9,5
Bremen	0,7	0,8	0,8
Nordrhein-Westfalen	20,6	20,7	20,7
Hessen	7,6	7,6	7,6
Rheinland-Pfalz	4,9	5,2	5,2
Baden-Württemberg	13,9	14,0	14,0
Bayern	16,9	16,9	16,9
Saarland	1,1	0,7	0,7
Berlin	4,6	4,6	4,6
Brandenburg	3,0	2,7	2,6
Mecklenburg-Vorpommern	1,8	2,1	2,1
Sachsen	4,7	4,6	4,5
Sachsen-Anhalt	2,4	2,5	2,5
Thüringen	2,4	2,5	2,5
Bundesland, Betriebsort²¹			
Schleswig-Holstein	3,0	2,8	2,8
Hamburg	2,9	3,2	3,2
Niedersachsen	9,0	8,8	8,7
Bremen	1,0	1,2	1,2
Nordrhein-Westfalen	20,6	20,4	20,4
Hessen	7,8	7,8	7,8
Rheinland-Pfalz	4,4	4,7	4,7
Baden-Württemberg	13,6	14,3	14,3
Bayern	16,4	17,0	17,0
Saarland	1,1	0,7	0,7
Berlin	4,7	5,1	5,1
Brandenburg	2,6	2,2	2,2
Mecklenburg-Vorpommern	1,9	2,0	2,0
Sachsen	4,9	4,3	4,3
Sachsen-Anhalt	2,6	2,2	2,2
Thüringen	2,6	2,4	2,4
Ausland	0,7	0,4	0,4
Angabe fehlt	0,2	0,4	0,4

²¹ Für das Bundesland des Betriebsorts liegen die Werte nur aus dem Mikrozensus 2022 vor.

Zur Beurteilung der Qualität der Gewichtungsmodelle werden die folgenden Indikatoren herangezogen:

- Effektivität:
 - Diese wird berechnet als das Verhältnis der Stichprobenvarianz unter Berücksichtigung des aktuellen komplexen Designs zur Stichprobenvarianz einer einfachen Zufallsstichprobe derselben Größe. Die Effektivität wird in der Regel mit der Designgewichtung berechnet und kann damit bei gegebener Stichprobengröße n zur Berechnung der effektiven Stichprobengröße n' herangezogen werden. Sie nimmt im optimalen Fall den Wert 1 bzw. 100% an. Je niedriger ihr Wert ist, umso stärker ist die Varianzzunahme durch die Gewichtung. Für die Folgerungen auf die Grundgesamtheit kann man trotz des tatsächlich realisierten Stichprobenumfangs n nur von der effektiven Stichprobengröße n' ausgehen. Dementsprechend größer sind die Konfidenzintervalle anzusetzen.
 - Die Effektivität der Stichprobe nach der Anpassungsgewichtung (Kalibration) ergibt zusätzlich ein Maß für die weitere Zunahme der Varianz der Gewichtungsfaktoren, allerdings ohne weitere zusätzliche Reduzierung des effektiven Stichprobenumfangs.
- Standardabweichung der Gewichtungsfaktoren:
 - Dies ist ein Maß für die Variation der Gewichtungsfaktoren. Sie sollte möglichst klein sein.
- Strukturinkongruenz:
 - Hier wird die Summe der absoluten Abweichungen zwischen Soll- und Stichprobenstruktur nach Gewichtung gemessen.

Tabelle 10: Qualitätsindikatoren der Gewichtungsmodelle für die Erwerbstätigenstichprobe (Referenzstruktur: Mikrozensus 2023)

Qualitätsindikatoren	Variante 1 - ohne StB	Variante 2 - mit StB
Fallzahl ungewichtet	20.006	20.006
Fallzahl gewichtet	20.006	20.006
Effektivität nach Designgewichtung inkl. AWR2	79,83%	79,83%
Effektivität nach Ausfallmodellierung AWR2	77,84%	77,84%
Effektivität insgesamt	30,77%	30,72%
Standardabweichung	1,19	1,26
Strukturinkongruenz gesamt ²²	15,1	11,3
Strukturinkongruenz im Durchschnitt	0,3	0,2

Die Designgewichtung wird vor der Kalibration durchgeführt und dient – wie beschrieben – als Eingangsgewicht für die weiteren Schritte. Demnach ist die Effektivität der Designgewichtung für beide Varianten gleich. Mit rund 31 % ist sie niedriger als in der Erhebung 2018. Zurückzuführen ist dies u. a. darauf, dass bestimmte Personengruppen noch schwerer zu erreichen sind bzw. mit einer noch

²² Die Sollstruktur bilden dabei die in Tabelle 9 aufgeführten Merkmale.

geringeren Wahrscheinlichkeit an Befragungen teilnehmen als vor sechs Jahren. Für diese Fälle muss im Rahmen der Strukturanpassung stärker eingegriffen werden, was bedeutet, dass die Varianz der Faktoren steigt. Dies führt dazu, dass die Effektivität niedriger als 2018 ist.

Die anschließende Kalibrationsgewichtung verringert in der Regel das Konfidenzintervall der Messvariablen und verbessert die Güte einer Stichprobe. Diese hat jedoch nur Einfluss auf die anschließende Varianz der Gewichtungsfaktoren, nicht auf den effektiven Stichprobenumfang. Eine „feiner“ angelegte Kalibrationsgewichtung erhöht in der Regel die Güte der Stichprobe, ohne jedoch den effektiven Stichprobenumfang weiter zu verringern. Aus diesem Grund sind bei der vorliegenden Untersuchung mit einer entsprechend hohen Fallzahl nicht nur ausgewählte Gewichtungsränder, sondern auch -zellen angepasst worden.

Variante 2 wurde in den vergangenen Erhebungen als die finale Gewichtungsvariante verwendet und ist auch 2024 unsere Empfehlung. Die Effektivität sowie die Standardabweichung unterscheiden sich bei beiden Gewichtungsmodellen nur geringfügig. Dagegen sind die Strukturen bei dem Merkmal „Stellung im Beruf“ deutlich verbessert und die Strukturinkongruenz dementsprechend niedriger.

8 Literaturverzeichnis

- ADM – Arbeitskreis deutscher Markt- und Sozialforschungsinstitute (2012): ADM-Forschungsprojekt „Dual-Frame-Ansätze“ 2011/2012. Forschungsbericht. 2012. URL: <https://www.adm-ev.de/wp-content/uploads/2018/07/Forschungsbericht-Dual-Frame-Ansätze.pdf>; letzter Zugriff: 12.01.2023.
- Biemer, P.B. (2010): Total Survey Error. Design, Implementation, and Evaluation. Public Opinion Quarterly, Vol. 74, No. 5, 2010, S. 817–848.
- Bundesagentur für Arbeit (2011): Umsteigeschlüssel von der Klassifikation der Berufe 2010 (5-Steller) zur ISCO-08 (4-Steller). Nürnberg
- Bundesagentur für Arbeit (2011): Klassifikation der Berufe 2010 – Band 1: Systematischer und alphabetischer Teil mit Erläuterungen. Nürnberg.
- Bundesagentur für Arbeit (2021): Statistik der Bundesagentur für Arbeit, Grundlagen: Hintergrundinfo – Geänderte Zuordnungen von Berufen in der Klassifikation der Berufe – Auswirkungen auf die Arbeitsmarktstatistiken. Nürnberg.
- Bundesagentur für Arbeit (2024): Klassifikation der Berufe 2010 – Überarbeitete Fassung 2020 – alphabetisches Verzeichnis der Berufsbenennungen. Nürnberg
- Carmines, E.G., Zeller, R.A. (1979): Reliability and Validity Assessment. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-017. Beverly Hills: Sage.
- Cassel, C.M., Särndal, C.E., Wretman, J.H. (1977): Foundations of Inference in Survey Sampling. New York: Wiley.
- Cochran, W.G. (1968): The effectiveness of adjustment by subclassification in removing bias in observational studies. Biometrics. 24. S. 295-313.
- Cox, L. H. (1987). A Constructive Procedure for Unbiased Controlled Rounding. Journal of the American Statistical Association, 82(398), 520-524.
- Cox, L. H. (1987): A constructive procedure for unbiased controlled rounding; in: Journal of the American Statistical Association 82, S.520 – 524, 1987.
- Deming, W.E., Stephan, F.F. (1940): On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Annals of Mathematical Statistics. 11. S. 427-444.
- Deville J.-C., Särndal, C.-E., Sautory, O. (1993): Generalized Raking Procedures in Survey Sampling. Journal of the American Statistical Association, Vol. 88, No. 423, (Sep., 1993), S. 1013–1020.
- Deville, J.-C., Särndal, C.-E. (1992): Calibration Estimators in Survey Sampling. In: Journal of the American Statistical Association. Vol. 87, No. 418, S. 376–382.
- Dillman, D.A., Smyth, J.D., Christian, L.M. (2014): Internet, Phone, Mail, and Mixed-Mode Surveys. The Tailored Design Method. Hoboken, New Jersey: John Wiley & Sons.
- Eurostat (2015): ESS Handbook for Quality Reports. 2014 edition. Luxembourg: Publications Office of the European Union. URL: <http://ec.europa.eu/eurostat/documents/3859598/6651706/KS-GQ-15-003-EN-N.pdf/18dd4bf0-8de6-4f3f-9adb-fab92db1a568>, letzter Zugriff: 12.01.2023.
- Faulbaum, F. (2014): „Total Survey Error“. In: Handbuch Methoden der empirischen Sozialforschung, S. 439-453.
- Frankel, M. (2010): Sampling Theory. In: Marsden P.V., and Wright, J.D. (eds.): Handbook of Survey Research. Second edition. Bingley: Emerald Group Publishing, pp. 83–137.
- Gabler, S., Ayhan, Ö. (2007): Gewichtung bei Erhebungen im Festnetz und über Mobilfunk: Ein Dual-Frame-Ansatz, in: Mobilfunktelefonie – Eine Herausforderung für die Umfrageforschung, Gabler und Häder (Hrsg.), GESIS - ZUMA Nachrichten, Spezial-Band 13.
- Gabler, S., Häder, S. (1998): Ein neues Stichprobendesign für telefonische Umfragen in Deutschland. S. 69 ff. In: S. Gabler, S., Häder, J.H.P. Hoffmeyer-Zlotnik, (Hrsg.): Telefonstichproben in Deutschland. Opladen: Westdeutscher Verlag.

- Ganninger, M., Häder, S., & Gabler, S. (2007): Design Effects and Interviewer Effects in the European Social Survey: Where are we now and where do we want to go tomorrow? Mannheim: ZUMA
- Ganninger, M. (2010): Design effects, model-based versus design-based approach, (GESIS - Leibniz-Institut für Sozialwissenschaften: GESIS-Schriftenreihe; Bd. 3, GESIS: Mannheim.
- Groves, R.M. (2004): Survey Errors and Survey Costs. Hoboken, New Jersey: John Wiley & Sons.
- Groves, R.M., Fowler, F.J. Jr., Couper, M.P., Lepkowski J.M., Singer, E., Tourangeau, R. (2009): Survey Methodology. Hoboken, New Jersey: John Wiley & Sons.
- Groves, R.M., Lyberg, L. (2010): Total Survey Error – Past, Present, and Future. In: Public Opinion Quarterly, 74 (5), S. 849–879.
- Häder, S. & Sand, M. (2019): Telefonstichproben. In Häder, S., Häder, M. & Schmich, P. (Hrsg.): Telefonumfragen in Deutschland. Wiesbaden [Springer VS] 2019, S. 113–151.
- Häder, S., Ganninger, M., Gabler, S. (2009): Die Stichprobenziehung für den European Social Survey: Prinzipien und Ergebnisse. In: Weichbold M., Bacher J., Wolf C. (Hrsg.) Umfrageforschung. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 181-193.
- Hanefeld, U. (1987): Das sozio-ökonomisch Panel. Grundlagen und Konzeption. Frankfurt: Campus.
- Häring, A., Schütz, H., Kleuden, M., Nold, J., Vieten, L., Entgelmeier, I., Backhaus, N. (2022): Methodenbericht und Fragebogen zur BAuA-Arbeitszeitbefragung 2021. Dortmund/Berlin/Dresden.
- Heyde, C. von der (2014): Techniken und Möglichkeiten der Realisierung von Random-Stichproben. S. 37 – 65 in: ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V.: Stichproben-Verfahren in der Umfrageforschung. Eine Darstellung für die Praxis. 2. Auflage. Wiesbaden: Springer VS.
- International Labour Organisation (2009): International Standard Classification of Occupations 2008 (ISCO-08): Structure, group definitions and correspondence tables, Chapter 5 “Adapting ISCO-08 for National Use and for the Production of Internationally Comparable Data.
- Jonas, K., Stroebe, W., Hewstone, M. (2014): Sozialpsychologie. 6. Aufl. Berlin, Heidelberg: Springer.
- Kish, L. (1992): Weighting for unequal Pi. Journal of Official Statistics (1992), 8, S. 183–200.
- Kish, L. (1965): Survey Sampling. New York: John Wiley & Sons.
- Kolenikov, S. (2016): Is It a Post-Stratification or a Non-Response Adjustment. Survey Practice Vol. 9, no 2, S. 1–12.
- Lenzner, T., Neuert, C. Otto, W. (2015): Kognitives Pretesting. Mannheim, GESIS Leibniz-Institut für Sozialwissenschaften (GESIS Survey Guidelines).
- Lohr, S. L. (2010): Sampling: Design and Analysis. Second Edition. Boston: Brooks/Cole, S. 28.
- Nunnally J.C. (1967): Psychometric Theory. New York: McGraw Hill.
- Särndal, C.-E., Swensson, B., Wretman, J. (1992): Model Assisted Survey Sampling. New York: Springer.
- Smith, E.R., Mackie, D.M., Claypool, H.M. (2015): Social Psychology. Fourth edition. New York, London: Psychology Press.
- Statistisches Bundesamt (2008): Klassifikation der Wirtschaftszweige. Mit Erläuterungen. Wiesbaden: Statistisches Bundes (URL: <https://www.destatis.de/static/DE/dokumente/klassifikation-wz-2008-3100100089004.pdf>; letzter Zugriff: 13.01.2023).
- Statistisches Bundesamt (2021): Glossar Mikrozensus 2020.
- Statistisches Bundesamt (2022): Fachserie 1 Reihe 3: Bevölkerung und Erwerbstätigkeit. Haushalte und Familien. Ergebnisse des Mikrozensus 2020 (Endergebnisse).
- Statistisches Bundesamt (2023): Statistischer Bericht. Mikrozensus - Arbeitsmarkt 2022 (Erstergebnis). EVAS-Nummern 12211 und 12251. Ergänzung zur Datenbank GENESIS-Online. Erschienen am 31.03.2023.
- Statistisches Bundesamt (2023a): Mikrozensus – Arbeitsmarkt. 2021 (Endgültiges Ergebnis). Ergänzung zur Datenbank GENESIS-Online. Erschienen am 31.01.2023.
<https://www.destatis.de/DE/Themen/Arbeit/Arbeitsmarkt/Erwerbstaetigkeit/Publikationen/Downloads-Erwerbstaetigkeit/statistischer-bericht-mikrozensus-arbeitsmarkt->

2010410217005-endergebnisse.xlsx?__blob=publicationFile (zuletzt aufgerufen am 16. Februar 2023)

Statistisches Bundesamt (2023b): GENESIS-ONLINE Ergebnis 12211-0005. Erwerbstätige aus Hauptwohnsitzhaushalten: Deutschland, Jahre, Geschlecht, Stellung im Beruf, Art der geleisteten Arbeitsstunden, Größenklasse der geleisteten Arbeitsstunden je Woche

<https://www-genesis.destatis.de/genesis/online?operation=abruftabelleBearbeiten&levelindex=1&levelid=1681832486360&auswahloperation=abruftabelleAuspraegungAuswaehlen&auswahlverzeichnis=ordnungsstruktur&auswahlziel=werteabruf&code=12211-0005&auswahltext=&werteabruf=Werteabruf#abreadcrumb> (zuletzt aufgerufen am 18. April 2023)

Statistisches Bundesamt (2023b): Mikrozensus – Arbeitsmarkt. 2022 (Erstergebnis). Ergänzung zur Datenbank GENESIS-Online. Erschienen am 31.03.2023.

https://www.destatis.de/DE/Themen/Arbeit/Arbeitsmarkt/Erwerbstaetigkeit/Publikationen/Downloads-Erwerbstaetigkeit/statistischer-bericht-mikrozensus-arbeitsmarkt-2010410227005-erstergebnisse.xlsx?__blob=publicationFile (zuletzt aufgerufen am 18. April 2023)

STATISTIK AUSTRIA – Die Informationsmanager (2010): ISCO 08 – gemeinsame deutschsprachige Titel und Erläuterungen auf Basis der englischsprachigen Version von Juni 2009.

Strippel, C. & Emmerich, M (2021): Stichproben für Telefonbefragungen in Deutschland: Ein Werkstattbericht. In: Jünger, J., Gochermann, U., Peter, C., Bachl, M. (Hrsg.) (2021). Grenzen, Probleme und Lösungen bei der Stichprobenziehung (=Methoden und Forschungslogik der Kommunikationswissenschaft, Band 17). Köln: Halem. Preprint; URL: <https://osf.io/5j974/download/?format=pdf>, letzter Zugriff: 07.04.2023.

The American Association for Public Opinion Research (2023): Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 10th edition. AAPOR.

Tourangeau, R., Rips, L.J., Rasinski, K. (2000): The Psychology of Survey Response. Cambridge: Cambridge University Press.

Weisberg, H.F. (2005): The Total Survey Error Approach. A Guide to the New Science of Survey Research. Chicago: University of Chicago Press.

Willis, G., Lessler, H. (1999): Question Appraisal System QAS-99. Research Triangle Institute: Rockville.

(URL: https://www.researchgate.net/profile/Gordon-Willis/publication/267938670_Question_Appraisal_System_QAS-99_By/links/54b7b26a0cf2e68eb2803f6a/Question-Appraisal-System-QAS-99-By.pdf; letzter Zugriff: 18.01.2023)

Zabal, A., Martin, S., Massing, N., Ackermann, D., Helmschrott, S., Barkow, I. & Rammstedt, B. (2014): PIAAC Germany 2012: Technical Report. Münster, New York: Waxmann.

Zeller, R.A., Carmines, E.G. (1980): Measurement in the Social Sciences. The Link between Theory and Data. Cambridge: Cambridge University Press.