



**STIFTUNG  
GESUNDHEITS  
WISSEN**

---

**Health Information National Trends  
Survey (HINTS) Germany**

– Welle 1 –

**Methodenreport**

## Zitationsvorgaben

Um auf diesen Methodenreport zu verweisen, geben Sie bitte alle folgenden Metadaten an. Sie können diese entsprechend Ihres gewählten Zitationsstils anpassen.

### Metadaten

Urheber	Stiftung Gesundheitswissen
Veröffentlichungsdatum	2020
Titel	Health Information National Trends Survey (HINTS) Germany, Welle 1
Ressourcentyp	Methodenreport
Identifikator	<a href="http://hints-germany.de/">http://hints-germany.de/</a>

### Beispielzitationen

#### APA Reference Style:

Stiftung Gesundheitswissen. (2020). Health Information National Trends Survey (HINTS) Germany, Welle 1 [Methodenreport]. Abgerufen von <http://hints-germany.de/>

#### Vancouver Reference Style:

Stiftung Gesundheitswissen. Health Information National Trends Survey (HINTS) Germany, Welle 1 [Methodenreport]. 2020 [Stand 14.07.2020]. Verfügbar unter: <http://hints-germany.de/>

## Projektpartner

Stiftung Gesundheitswissen  
Friedrichstraße 134, 10117 Berlin  
[info@hints-germany.de](mailto:info@hints-germany.de)  
[www.stiftung-gesundheitswissen.de](http://www.stiftung-gesundheitswissen.de)

in Zusammenarbeit mit

Hochschule für Musik, Theater und Medien Hannover  
Institut für Journalistik und Kommunikationsforschung  
Hanover Center for Health Communication [HC]<sup>2</sup>  
Expo Plaza 12, 30539 Hannover  
[www.ijk.hmtm-hannover.de/de/hc-quadrat/start](http://www.ijk.hmtm-hannover.de/de/hc-quadrat/start)

## Befragungsinstitut

USUMA GmbH  
Institut für Marktforschung und Sozialforschung  
Berliner Allee 96, 13088 Berlin  
[www.usuma.com](http://www.usuma.com)

## Inhalt

1. Projektübersicht.....	04
2. Adaption des Fragebogens .....	04
3. Pretests.....	05
4. Stichprobe.....	05
5. Interviewerschulung .....	07
6. Interviewerkontrolle.....	07
7. Feldarbeit .....	07
8. Stichprobenprotokoll .....	08
9. Datenaufbereitung .....	09
10. Gewichtung .....	09
11. Datenübergabe .....	12
Quellenangaben .....	13

# 1. Projektübersicht

Mit dem Projekt *Health Information National Trends Survey (HINTS) Germany* möchten die Stiftung Gesundheitswissen und das Hanover Center for Health Communication eine nationale Datenerhebung zum Gesundheitsinformationsverhalten in Deutschland etablieren. In mehreren Erhebungswellen soll eine einheitliche und wissenschaftlich fundierte Datenbasis entstehen, die Trends im Gesundheitsinformationsverhalten erkennen lässt, eine Planungsgrundlage für die zielgruppengerechte Gestaltung effektiver Gesundheitskommunikation schafft und darüber hinaus Hinweise auf die Wirksamkeit präventiver, gesundheitspolitischer Maßnahmen sowie versorgungspraktische Erkenntnisse liefert.

Für die Trendstudie wurde das Frageninstrumentarium des seit 2003 regelmäßig vom National Cancer Institute (NCI) durchgeführten US-amerikanischen Gesundheitssurvey *HINTS* adaptiert, um international vergleichbare Messwerte für Deutschland zu erhalten.

Die 1. Welle von *HINTS Germany* wurde 2018/2019 als Querschnittserhebung durchgeführt. Dabei wurden in einer repräsentativen telefonischen Umfrage (CATI) insgesamt 2.902 zufällig ausgewählte Personen zu ihrem Gesundheitsinformationsverhalten befragt. Die Grundgesamtheit der Umfrage entspricht der deutschsprachigen Wohnbevölkerung in Privathaushalten zwischen 18 und 79 Jahren, die dauerhaft in der Bundesrepublik Deutschland lebt.

In diesem Methodenreport wird das Vorgehen bei der Durchführung der 1. Welle beschrieben.

# 2. Adaption des Fragebogens

Die Entwicklung des Fragebogens für *HINTS Germany* erfolgte auf Grundlage des Original-Fragebogens des US-Referenzprojektes (*HINTS 5, Cycle 1, 2017*) (1). Der für eine postalische Befragung entwickelte Original-Fragebogen wurde zur Anpassung an die telefonische Befragung gekürzt. Die Übersetzung wurde dabei nicht mit dem traditionellen Ansatz der „back translation“ durchgeführt, sondern mit dem moderneren und inzwischen als Gold-Standard geltenden Ansatz der „team translation“ (2). Die erste Übersetzung wurde von einer professionellen Übersetzerin mit Erfahrungen in medizinischen Studien vorgenommen, die zweite Übersetzung entstand im *HINTS Germany*-Team (insgesamt sechs Personen mit gesundheits- bzw. kommunikationswissenschaftlicher Expertise). Die beiden Versionen wurden anschließend im Rahmen eines Review-Verfahrens unter Moderation einer erfahrenen Expertin für internationale Survey-Forschung des GESIS-Instituts (Leibniz-Institut für Sozialwissenschaften) zusammengeführt, wobei Abweichungen zwischen den Übersetzungen ausführlich diskutiert und in eine konsenterte Fassung überführt wurden.

Ergänzend zum adaptierten *HINTS*-US-Fragebogen wurden eine Kurzskala zur Informationsvermeidung (3), eine Skala zum Sense of Control (4) sowie eine vom Projektteam selbst entwickelte Skala zum Interesse an Angeboten eingesetzt.

Der Fragebogen besteht aus den folgenden Blöcken:

- Block A: Suche nach Gesundheitsinformationen
- Block B: Nutzung des Internets zur Suche nach Informationen
- Block C: Gesundheitsversorgung
- Block G: Ihre Gesundheit
- Block I: Körperliche Aktivität
- Block J: Tabak- und Alkoholkonsum
- Block K: Krebs-Früherkennung
- Block M: Ihre Krebsgeschichte
- Block N: Ansichten und Überzeugungen zu Krebs
- Block O: Soziodemografie

### 3. Pretests

Es wurden zwei verschiedene Pretests durchgeführt: Im Juli 2018 wurden zehn Fragen einem tiefergehenden kognitiven Pretest unterzogen, um etwaige Probleme der Befragten bei der sinnhaften Erfassung der Fragenformulierung und der Antwortskalen zu erkennen. Dazu wurden persönliche Interviews mit 13 Personen durchgeführt, aufgezeichnet, transkribiert und ausgewertet. Die Ergebnisse zeigten bei einer Frage große Schwierigkeiten mit der Antwortskala, die daraufhin umformuliert wurde.

Nach der anschließenden CATI-Programmierung wurden im Rahmen eines Feld-Pretests 50 vollständige Interviews geführt, um das Erhebungsinstrument zu überprüfen. Nach Auswertung der Pretest-Ergebnisse wurden die Fragebogendokumente redaktionell überarbeitet. Dabei wurden verschiedene methodenrelevante Aspekte geprüft, wie beispielsweise die Verständlichkeit einzelner Formulierungen oder gewählter Skalierungen, die Vollständigkeit der Antwortkategorien, die Einfachheit und Strukturiertheit der Fragestellung sowie die Positionierung von Fragen im Fragebogenverlauf und die Interviewer-Hinweise. Da die im Pretest gemessene Interviewlänge über der geplanten durchschnittlichen Fragebogendauer lag, wurden auch Kürzungen im Fragebogendesign durchgeführt.

Am 25.09.2018 wurde die finale Version des Fragebogens nach nochmaligem Test der CATI-Programmierung für den Start der Haupterhebung verabschiedet.

### 4. Stichprobe

Im Rahmen der computergestützten telefonischen Umfrage war geplant, bundesweit insgesamt 3.000 Interviews durchzuführen.

Die Stichprobe für die Haupterhebung wurde zum einen über die ADM-Festnetz-Telefonstichprobe „eASYSAMPLE“ ausgewählt, in der auch nicht ins Telefonbuch eingetragene Haushalte nach dem Gabler/Häder-Verfahren berücksichtigt sind. Gezogen wurden hier eingetragene und generierte Telefonnummern nach dem Flächennetzsystem der Bundesnetzagentur in ganz Deutschland, proportional zur regionalen Einwohnerstruktur auf Bundesland-Ebene, geschichtet nach den bekannten Ortsgrößenklassen der einzelnen Kreise bzw. Gemeinden.

Zum anderen wurden Mobilfunknummern aus der ADM-Mobilfunkstichprobe „eASYMOBILE“ eingesetzt, da eine Kombination aus Festnetznummern und Mobilfunknummern besonders auf Grund eher schwierig zu erreichender Zielgruppen sowie den sogenannten „Mobile-Only“-Nutzern als methodische Notwendigkeit erachtet wurde.

Das Studiendesign folgte dabei dem „Dual-Frame Ansatz“ und enthielt in der Bruttostichprobe einen Mobilfunknummern-Anteil von etwa 40 %.

#### **Kurzbeschreibung der ADM-Stichprobenziehung (Auszug aus ADM-Webseite):**

„In der Bundesrepublik gibt es weder ein allgemein zugängliches Verzeichnis aller Privathaushalte oder Personen noch ein vollständiges Verzeichnis aller privat genutzten Telefonanschlüsse. Die mit telefonischer Befragung erreichbare Bevölkerung ist in keinem amtlichen Telefonverzeichnis zu finden, so dass weder komplette individuelle noch statistische Angaben dazu existieren.“

Die Arbeitsgemeinschaft ADM-Telefonstichproben füllt diese Lücke, indem sie den beteiligten Instituten einen Auswahlrahmen zur Verfügung stellt, der es erlaubt, repräsentative Stichproben der Bevölkerung in Privathaushalten – soweit diese telefonisch erreichbar ist – zu erstellen. Parallel dazu werden die wichtigsten statistischen Daten für diese Grundgesamtheiten zur Verfügung gestellt.

Der Auswahlrahmen, das sogenannte ADM-Telefonstichproben-System, basiert auf den im Internet veröffentlichten Eckdaten des Nummernraums, den die Bundesnetzagentur den Telefonanbietern zur Verfügung stellt und verwaltet. Dieser Nummernraum umfasst prinzipiell alle in der Bundesrepublik Deutschland nutzbaren Telefonnummern.

Unter adaptiver Nutzung des von Sabine Häder und Siegfried Gabler entwickelten Generierungsverfahrens (5) werden die relevanten Eckdaten der Bundesnetzagentur so aufbereitet,

- dass sie die Basis für die Generierung von 10er Blöcken bilden, die alle in der BRD nutzbaren Festnetznummern umfassen (sogenannte Festnetzdatei),
- dass alle in der BRD nutzbaren Mobilfunknummern auf Basis von 10.000er Blöcken generiert werden (sogenannte Mobilfunkdatei).

Da rund 99 % der Bevölkerung mit mindestens einem Telefonanschluss erreichbar sind und davon rund 12 % ausschließlich über Mobilfunk (Stand 2011), sind repräsentative Abbildungen der Bevölkerung auf Basis beider Dateien jederzeit möglich. Das vom ADM initiierte und 2011/2012 durchgeführte Forschungsprojekt „Dual-Frame-Ansätze“ hat darüber hinaus die Standards erarbeitet, die die optimale Nutzung beider Auswahlgrundlagen zur Erzielung repräsentativer Abbildungen ermöglichen (6).

Stichprobenziehungen aus der Festnetzdatei haben einen weiteren Vorteil: Die Festnetznummern sind mit regionalen Zuordnungen versehen, die es erlauben, bis auf Ebene des amtlichen Gemeindegemeinschaftsschlüssels (AGS) – und bei einer Reihe von Großstädten bis auf Stadtteilebene – regional zu schichten, und so das aus der f2f-Stichprobe bekannte Schichtungssystem auch für die Telefonstichproben zu realisieren. Wegen der geringen Eintragsdichte bei den Mobilfunknummern ist eine Regionalschichtung dort nicht möglich. Stattdessen können dort die Vorwahlnummern oder die Zuordnung zu Providern als externe Schichtkriterien zur Verbesserung des Stichprobendesigns genutzt werden.

Um diese Schichtungseffekte voll nutzen zu können, werden die Stichproben aus der jeweiligen Auswahlgrundlage in der Regel nach dem von L. H. Cox (7) entwickelten Verfahren der Zufalls-Allokation mit Proportion zur Größe der Nachkommastellen realisiert.

Da nicht nur die Ziehung der Nummernstichproben (und damit der Zielpersonen bei den Mobilfunknummern) sondern auch die Auswahl der Zielperson im jeweils kontaktierten Privathaushalt bei den Festnetzstichproben als Zufallsauswahl realisiert werden, sind die ADM-Telefonstichproben als repräsentative Zufallsstichproben zu bewerten, die den in der mathematischen Statistik entwickelten Zufallsmodellen gehorchen.“

(Quelle: [www.adm-ev.de/leistungen/arbeitsgemeinschaft-adm-stichproben](http://www.adm-ev.de/leistungen/arbeitsgemeinschaft-adm-stichproben))

Aus stichprobentheoretischer Sicht ist beim Dual-Frame Ansatz wesentlich, dass die Inklusionswahrscheinlichkeiten für jede Person genau festgestellt und so später in einer Gewichtung berücksichtigt werden können. Daraus ergibt sich, dass die Interviews immer mit dem Gerät (Festnetzanschluss oder Mobiltelefon) durchgeführt werden, mit dem der Kontakt zustande kam.

Damit sollte zunächst eine zufällige Auswahl der kontaktierten Haushalte innerhalb des Nummernraumes der Bundesrepublik Deutschland sichergestellt werden. Bei Mobilfunknutzern war die Zielperson automatisch der Besitzer bzw. Hauptnutzer des Telefons (Handy, Smartphone). War diese Person jünger als 18 oder älter als 79 Jahre, wurde der Anschluss entsprechend als neutraler Ausfall abgelegt. Bei einem Festnetzanschluss wurde die Zielperson innerhalb des Haushaltes anhand der Schwedenschlüssel-Methode zufällig ermittelt. Bei diesem Verfahren wurden zunächst über eine Kontaktperson alle Personen im Haushalt ermittelt, die in die Zielgruppe fielen (zwischen 18 und 79 Jahre alt). Der Schwedenschlüssel selektierte zufällig und damit unabhängig von den Interviewern die Person im Haushalt, mit der das Interview geführt werden sollte.

## 5. Interviewerschulung

Insgesamt waren 92 Interviewer an der Studie beteiligt. Am 08.10.18, am 17.10.18, am 05.11.18 und am 09.01.19 fanden umfassende Schulungen der beteiligten Interviewer statt. In der Schulung wurden die Interviewer in die speziellen Anforderungen der telefonischen Umfrage eingewiesen und mit den Hintergründen der Befragungsmethode sowie den verwendeten Begriffsdefinitionen vertraut gemacht. Zusätzlich wurde den Interviewern erklärendes Begleitmaterial präsentiert, um die Studie im richtigen Kontext zu erfassen. Dazu gehörte neben der genauen Umsetzung der Interviewer-Hinweise und erklärenden Textvorgaben vor allem die Erfassung des inhaltlichen und methodischen Hintergrundes der standardisierten Fragen und das exakte Vorlesen der Texte und Skalen. Jeder Interviewer erhielt zusätzlich eine schriftliche Projektinformation, in der die wichtigsten Hinweise zur Durchführung der Befragung enthalten waren.

## 6. Interviewerkontrolle

Während der Feldzeit wurden die Interviewer zum Zweck der Qualitätssicherung kontinuierlich durch erfahrene und geschulte Betreuer begleitet und in unregelmäßigen Abständen stichprobenartig durch die verantwortlichen Supervisoren durch paralleles „Monitoring“ (begleitendes Mithören) beobachtet. Wurden über diese Kontrollen Schwierigkeiten oder Mängel in der Umsetzung entdeckt, kam es so unmittelbar zu gezielten Nachschulungen der Interviewer.

Die Kontrollen erfolgten in 3 Schritten:

- Kontrolle der Arbeit der Interviewer im Studio direkt durch die Projektbetreuer,
- Durchführung eines Monitorings, bei dem die Beantwortung der Fragen auf dem Bildschirm mitverfolgt wird, sowie
- Kontrolle der eingehenden Daten auf Logik und Vollständigkeit.

## 7. Feldarbeit

Der gesamte Erhebungszeitraum der Studie dauerte mit Unterbrechungen vom 04.09.2018 bis zum 24.01.2019. Während der Erhebung fanden wöchentliche Feldstand-Reports und Rücksprachen zwischen dem Erhebungsinstitut und der wissenschaftlichen Projektleitung statt.

Auf Grund der ermittelten durchschnittlichen Interviewdauer von 28 statt 25 Minuten wurden insgesamt 2.902 vollständige Interviews statt der geplanten 3.000 Interviews durchgeführt. Der Median beträgt 34 Minuten. Der Anteil der eingesetzten Mobilfunknummern konnte bei 40 % in der Stichprobe beibehalten werden.

Als Anreiz zur weiterführenden Beschäftigung mit den Inhalten der Studie wurde den Befragten auf Nachfrage der Link zur Projektwebseite genannt (<http://hints-germany.de/>), wo sie eine Übersicht mit Informationen zum Projekt *HINTS Germany* und den beteiligten wissenschaftlichen Einrichtungen fanden.

## 8. Stichprobenprotokoll

In Tabelle 1 sind die Ausfälle im Verhältnis zu den geführten Interviews anhand der internationalen Standards der [American Association for Public Opinion Research](#) (AAPOR) dargestellt.

**Tabelle 1: Stichprobenprotokoll**

Bez.	Stichprobenverteilung	Festnetz	Mobilfunk	Gesamt
I	Vollständige Interviews	1.698	1.204	2.902
P	Partielle Interviews	64	49	113
R	Verweigerung (oder Abbruch mit qualifizierter Person)	4.578	876	5.454
NC	Nicht-Kontakt mit qualifiziertem Haushalt (Anrufbeantworter)	1.236	0	1.236
O	Andere Ausfälle mit qualifiziertem Haushalt (Sprachprobleme, Verständnisprobleme)	724	976	1.699
UH	Unbekannt, ob Zielhaushalt (Freizeichen, Besetzt)	4.573	5.204	9.777
UO	Unbekannt aus anderen Gründen (Haushalt, aber keine Zielperson ermittelt)	1.162	1.810	2.972
INNR	Ungültig: kein Haushalt (ungültige Telefonnummern, kein Privathaushalt)	49.276	32.696	81.972
INR	Ungültig: Haushalt, aber nicht für Studie qualifiziert (in Feldzeit nicht anwesend, krank, verstorben)	253	85	338
<b>Total</b>		<b>63.564</b>	<b>42.900</b>	<b>106.463</b>

In Tabelle 2 sind die Beteiligungsraten aufgeführt. Wählt man den in der Sozialwissenschaft üblichen Berechnungsschlüssel 3 ergibt sich bei den brutto geführten Interviews eine Beteiligungsrate von ca. 19 %.

**Tabelle 2: Ermittlung der Beteiligungsrate nach AAPOR**

Berechnung der Fälle mit unbestimmter Qualifizierung	Festnetz	Mobilfunk
e1 = geschätzter % Anteil der als Haushalte identifizierte Fälle, die für das Interview qualifiziert wären	97,0 %	97,3 %
e2 = geschätzter % Anteil der Fälle ohne Kontakt, die als Haushalte gezählt werden	16,5 %	13,3 %
Response Rate 1 $I / (I+P+R+NC+O+UH+UO)$	12,1 %	11,9 %
Response Rate 2 $(I+P) / (I+P+R+NC+O+UH+UO)$	12,6 %	12,4 %
Response Rate 3 $I / (I+P+R+NC+O+[e1 * e2 * UH]+[e1 * UO])$	16,7 %	21,7 %
Response Rate 4 $(I+P) / (I+P+R+NC+O+[e1 * e2 * UH]+[e1 * UO])$	17,3 %	22,6 %
KLL = % Anteil aller Interviews aus der Festnetzstichprobe	61,3 %	
Combined Response Rate 1 $(RR1LL * KLL) + (RR1CP * (1-KLL))$		12,0 %
Combined Response Rate 2 $(RR2LL * KLL) + (RR2CP * (1-KLL))$		12,5 %
<b>Combined Response Rate 3 <math>(RR3LL * KLL) + (RR3CP * (1-KLL))</math></b>		<b>18,7 %</b>
Combined Response Rate 4 $(RR4LL * KLL) + (RR4CP * (1-KLL))$		19,4 %

## 9. Datenaufbereitung

Die Daten wurden auf Plausibilität und formale Integrität geprüft und bereinigt. Fehlende Werte wurden durch „99“ und Ausfälle wegen Selektions- oder Filtersprüngen durch „-2“ verschlüsselt. Zu allen Mehrfachnennungsfragen gibt es in der Datei zugehörige dichotomisierte Variablen. Alle offen erfassten Texte wurden im SPSS-Datensatz vollständig integriert.

Nach Prüfung der Daten auf Logik, Konsistenz und Filterführung musste kein Fall von der Analyse ausgeschlossen werden.

## 10. Gewichtung

Bei der Auswertung der Befragungsergebnisse sollen Aussagen getroffen werden, die als repräsentativ für die Grundgesamtheit der deutschsprachigen Wohnbevölkerung angesehen werden können. Daher wurden die Daten anhand der folgenden Schritte gewichtet.

### 10.1. Anpassung der Haushaltsgrößen

Vor der Designgewichtung wird die Verteilung der Haushaltsgrößen in der realisierten Stichprobe an die bekannte Verteilung in der Grundgesamtheit angepasst. Die auftretende Verzerrung ist hier auf die erhöhte Nonresponse-Rate in Ein-Personen-Haushalten wegen der geringeren Antreffwahrscheinlichkeit zurückzuführen.

**Tabelle 3: Prozentuale Verteilung der Haushaltsgrößen**

Haushaltsgröße	Ist	Soll	Faktor
1 Person	33,2	41,1	1,17
2 Personen	37	34	0,93
3 Personen	13,2	12,3	0,98
4 und mehr Personen	16,5	12,7	0,79
<b>Gesamt</b>	<b>100</b>	<b>100</b>	

Nach dieser Anpassungsgewichtung auf Basis der Haushalte kann man davon ausgehen, dass eine repräsentative Haushaltsstichprobe vorliegt.

## 10.2. Designgewichtung

Die Designgewichtung korrigiert die durch das Stichprobendesign zwangsläufig entstehenden verschiedenen Auswahlwahrscheinlichkeiten der Zielpersonen. Es kommt ein Standard-Berechnungsverfahren für das hier vorliegende Dual-Frame-Design zum Einsatz, das vom ADM empfohlen wird. Dabei berechnet sich das Designgewicht wie folgt:

$$\text{Designgewicht} = 1 / \left[ K_{\text{Festnetz}} * \frac{1}{Z_{\text{HH}}} + \frac{E_{\text{Festnetz}}}{\epsilon_{\text{Festnetz}}} * \frac{\epsilon_{\text{Mobil}}}{E_{\text{Mobil}}} * K_{\text{Mobil}} * \frac{1}{Z_{\text{Mobil}}} \right]$$

mit den Parametern:

$Z_{\text{HH}}$  = Anzahl Zielpersonen im Haushalt (max. 4)

$Z_{\text{Mobil}}$  = Anzahl Zielpersonen, die das angerufene Handy nutzen (max. 3)

$K_{\text{Festnetz}}$  = Anzahl Festnetzrufnummern, über die der Haushalt erreichbar ist (max. 4)

$K_{\text{Mobil}}$  = Anzahl Handynummern, über die der Befragte persönlich erreichbar ist (max. 4)

$E_{\text{Festnetz}}$  = Anzahl der Festnetznummern im Auswahlrahmen (bereinigt um ungültige Nummern, entspricht 46.659.600)

$E_{\text{Mobil}}$  = Anzahl der Handynummern im Auswahlrahmen (bereinigt um ungültige Nummern, entspricht 69.820.854)

$\epsilon_{\text{Festnetz}}$  = realisierte Stichprobe im Festnetz (hier 1.698)

$\epsilon_{\text{Mobil}}$  = realisierte Stichprobe im Mobilnetz (hier 1.204)

## 10.3. Anpassungsgewichtung

Bei gleicher Teilnahmebereitschaft über alle Bevölkerungsgruppen hinweg, würde nach der Designgewichtung eine repräsentative Personenstichprobe vorliegen. In der Umfragepraxis ergeben sich jedoch durch schwer messbare Einflussfaktoren (z. B. höhere Nonresponse-Raten bestimmter Bevölkerungsgruppen) Verzerrungen bei der Verteilung verschiedener Merkmale gegenüber der Grundgesamtheit. Diese gleicht man durch eine weitere Anpassung dieser ausgewählten Merkmale an deren Verteilung in der Grundgesamtheit aus. Bei der Auswahl der interessierenden Merkmale ist es zwingend erforderlich, dass entsprechende Informationen über deren Verteilung in der Grundgesamtheit vorliegen. Für die Merkmale Alter, Geschlecht, Bildung sowie Wohnort nach Bundesland liegen diese Werte auf Basis des Mikrozensus in aktuellen Veröffentlichungen des Statistischen Bundesamtes vor.

Bei der Berechnung der Gewichte wird iterativ vorgegangen, sodass Informationen zu den Randverteilungen (keine kombinierten Verteilungen) der einzelnen Merkmale ausreichen. Anschließend werden die Gewichte schrittweise jeweils so angepasst, dass die gewichtete Stichprobe proportional hinsichtlich eines ausgewählten Merkmals in der Grundgesamtheit ist. Dieses Verfahren wird für alle relevanten sozio-demografischen Merkmale nacheinander durchgeführt und so lange wiederholt, bis sich die Gewichte nicht mehr ändern (Iterative Proportional Fitting).

In der vorliegenden Studie wurden die Randverteilungen der Merkmale Alter x Geschlecht (in 7 x 2 Stufen), Bildung (in 4 Stufen) und Wohnort (in 16 Stufen) durch dieses Verfahren an die tatsächlichen Verteilungen dieser Merkmale in der Grundgesamtheit angeglichen.

**Tabelle 4: Prozentuale Verteilung des Merkmals Bildungsabschluss**

Haushaltsgröße	Ist	Soll
Niedrig (bis Hauptschulabschluss)	16,0	31,5
Mittel (Mittlere Reife)	25,9	29,8
Hoch (Abitur, Fachabitur)	57,5	33,8
Sonstiges (Schüler, k. A.)	0,5	0,2
<b>Gesamt</b>	<b>100</b>	<b>100</b>

**Tabelle 5: Prozentuale Verteilung der Merkmalskombination „Alter und Geschlecht“**

Altersgruppe	Geschlecht	Ist	Soll
18 – 24 Jahre	Männlich	4,4	5,1
	Weiblich	3,8	4,7
25 – 34 Jahre	Männlich	6,1	8,6
	Weiblich	6,2	8,0
35 – 44 Jahre	Männlich	7,1	8,0
	Weiblich	7,1	7,7
45 – 54 Jahre	Männlich	9,3	10,4
	Weiblich	9,9	10,3
55 – 64 Jahre	Männlich	10,8	8,8
	Weiblich	11,2	9,0
65 – 74 Jahre	Männlich	7,4	6,1
	Weiblich	8,7	6,8
75+ Jahre	Männlich	3,4	3,0
	Weiblich	4,5	3,7
<b>Gesamt</b>		<b>100</b>	<b>100</b>

**Tabelle 6: Prozentuale Verteilung des Merkmals Bundesland**

Bundesland	Ist	Soll
Schleswig-Holstein	3,4	3,5
Hamburg	2,6	2,2
Niedersachsen	9,6	9,6
Bremen	0,5	0,8
Nordrhein-Westfalen	18,5	21,5
Hessen	7,2	7,5
Rheinland-Pfalz	4,1	4,9
Baden-Württemberg	12,3	13,2
Bayern	17,1	15,9
Saarland	1,2	1,2
Berlin	5,6	4,3
Brandenburg	3,5	3,0
Mecklenburg-Vorpommern	2,0	2,0
Sachsen	5,2	4,9
Sachsen-Anhalt	3,0	2,7
Thüringen	2,7	2,6
<b>Gesamt</b>	<b>100</b>	<b>100</b>

#### 10.4. Schätzung der Gewichtung-Varianz (Jackknife-Methode)

Zusätzlich zur Gewichtungsvariable wurde im Datensatz für jeden Befragten ein Satz von 50 Wiederholungsgewichten bereitgestellt. Diese Replikationsgewichte werden verwendet, um Standardfehler von Schätzungen innerhalb der *HINTS Germany*-Daten zu berechnen. Verwendet wurde dafür die Replikationsmethode „Delete One Jackknife“ (JK1).

Die JK1-Methode ist kompatibel zum verwendeten Stichprobendesign und den Gewichtungsverfahren für *HINTS Germany*. Bei dieser Technik der Varianzschätzung werden 50 gleichgroße und disjunkte Teilmengen („Replikate“) der Daten selektiert.

Für jedes dieser 50 Replikate und für jeden Befragten in der Replikat-Untermenge wird ein Stichprobengewicht bestimmt, so als ob die Replikat-Untermenge tatsächlich der Gesamtdatensatz wäre. Die Replikat-Teilmenge entspricht jeweils 98 % der gesamten Stichprobe, mit Ausnahme der Gruppe von Befragten, die für diese Replikation „gelöscht“ werden, also jeweils 1/50. Die resultierenden Gewichte werden als Replikatgewichte bezeichnet.

Der Schätzer für die Varianz der Jackknife-Methode erfordert die Verwendung von replizierten Gewichten. Für den *HINTS Germany*-Datensatz wurde daher jeder Befragungsperson ein Satz von 50 replizierten Gewichten zugeordnet.

Zur Veranschaulichung der Verwendung der replizierten Varianzschätzung nehmen wir an, dass P ein Prozentsatz der Befragten mit einem bestimmten Merkmal ist (z. B. „ja, im gesamten Leben mehr als 100 Zigaretten geraucht“). Ein national repräsentativer Schätzer p kann berechnet werden, indem die Stichprobengewichte aller antwortenden Befragten mit diesem Merkmal (z. B. „ja, im gesamten Leben mehr als 100 Zigaretten geraucht“) aggregiert werden. Ein JK1-Varianzschätzer der Stichprobenvarianz von p kann in zwei Schritten berechnet werden:

**Schritt 1.** Neuberechnung der Schätzer  $p(r)$ ,  $r = 1, \dots, 50$ , durch Aggregation der replizierten Stichprobengewichte, die dem replizierten  $r$  entsprechen, für alle antwortenden Befragten mit dem Merkmal.

**Schritt 2.** Berechnung des JK1-Varianzschätzers

$$v(p) = \frac{R-1}{R} + \sum_{(r=1)}^{50} (p(r)-p)^2$$

Die Replikationsgewichte ergeben sich durch systematisches Löschen eines Teiles (1/50) der Originalstichprobe (sogenannte „deletion groups“:  $D(r)$ ,  $r = 1, \dots, 50$ ) und sukzessive Neuberechnung der einzelnen Replika-Gewichtungen.

Der jeweils verbleibende Stichprobenanteil (98 %), bei dem die „deletion group“ entfernt wurde, wird Replikat-Subset genannt.

Die replizierten Stichprobengewichte wurden in einer Reihe von Schritten erzeugt, die parallel zu den Schritten zur Berechnung der Gesamtstichprobengewichtung verlaufen. Das replizierte Basisgewicht für jeden Befragten im Datensatz und jede replizierte Kopie ist entweder gleich dem  $R/(R-1)$ -Faktor des vollen Basisgewichts der Stichprobe (wenn die Person in der replizierten Teilmenge enthalten ist) oder gleich 0 (wenn die Befragungsperson nicht in der replizierten Teilmenge enthalten ist, sondern im „gelöschten“ Satz für diese replizierte Kopie).

## 11. Datenübergabe

Die Ergebnisse der Befragung liegen als kompletter, analysefähiger SPSS-Datensatz mit Variablen- und Wertelabels und Gewichtungsfaktoren (Dual-Frame-Gewichtung inklusive der „Jackknife-Methode“ zur Varianzschätzung der Gewichtungsfaktoren) vor.

Die Variable *gewicht* enthält für jeden Fall den ermittelten Faktor auf Basis der Gesamtgewichtung. Die Variablen *gewicht\_1* bis *gewicht\_50* enthalten den Gewichtungsfaktor der replizierten Stichproben.

Im Falle von nicht beantworteten Fragen mit Einfachnennung wurde der Wert der entsprechenden Variablen auf „99“ gesetzt. Durch Filterführung übersprungene Fragen mit Einfachnennung enthalten in der zugehörigen Variablen den Wert „-2“.

Zu allen Mehrfachnennungsfragen gibt es in der SPSS-Datei entsprechende dichotomisierte Variablen. Variablen mit der Endung „\_00“ im Variablennamen enthalten bei den Fragen mit Mehrfachnennungen den Hinweis auf einen Filter. Der Wert „1“ in dieser Variablen bedeutet hierbei, dass der Befragte die Frage durch Filterführung übersprungen hat.

Die Variable *O8\_PLZ* wurde im Zuge der Anonymisierung der Daten für die Open-Access-Bereitstellung aus dem SPSS-Datensatz entfernt.

## Quellenangaben

1. National Cancer Institute. Health Information National Trends Survey (HINTS) 5, Cycle 1 [Fragebogen]. 2017 [Stand: 14.07.2020]. Verfügbar unter: <https://hints.cancer.gov/data/survey-instruments.aspx>.
2. Survey Research Center. Guidelines for Best Practice in Cross-Cultural Surveys. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan; 2016 [Stand: 14.07.2020]. Verfügbar unter: <http://www.ccsq.isr.umich.edu/>.
3. Howell JL, Shepperd JA. Establishing an Information Avoidance Scale. *Psychol Assess* 2016; 28(12):1695–708. doi: 10.1037/pas0000315.
4. Oh HJ, Lee B. The effect of computer-mediated social support in online communities on patient empowerment and doctor-patient communication. *Health Commun* 2012; 27(1):30–41. doi: 10.1080/10410236.2011.567449.
5. Häder S, Gabler S. Ein neues Stichprobendesign für telefonische Umfragen in Deutschland. In: Gabler S, Häder S, Hoffmeyer-Zlotnik JHP, Hrsg. *Telefonstichproben in Deutschland*. Wiesbaden: VS Verlag für Sozialwissenschaften; 1998. S. 69–88.
6. ADM - Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V. Forschungsbericht zum ADM-Forschungsprojekt „Dual-Frame-Ansätze“ 2011/ 2012 [Stand: 14.07.2020]. Verfügbar unter: <https://www.adm-ev.de/leistungen/forschungsprojekte/>.
7. Cox LH. A Constructive Procedure for Unbiased Controlled Rounding. *Journal of the American Statistical Association* 1987; 82(398):520–4. doi: 10.1080/01621459.1987.10478456.