infas

Methodenbericht

Deutscher Viktimisierungssurvey 2017

Autoren: Stefan Schiel, Petra Knerr, Dennis Schumacher

infas Institut für angewandte Sozialwissenschaft GmbH

Friedrich-Wilhelm-Straße 18
D-53113 Bonn
Tel. +49 (0)228/38 22-0
Fax +49 (0)228/31 00 71
info@infas.de
www.infas.de

infas

Bericht an

Bundeskriminalamt
Äppelallee 45
65173 Wiesbaden

Projekt

5775
Bonn, Juli 2018
Sl, Kp

Vorgelegt von

infas Institut für angewandte
Sozialwissenschaft GmbH
Friedrich-Wilhelm-Straße 18
53113 Bonn

Kontakt

Stefan Schiel
Senior-Projektleiter Sozialforschung

Tel. +49 (0)228/38 22-424
Fax +49 (0)228/310071
E-Mail s.schiel@infas.de

Autoren

Stefan Schiel, Petra Knerr, Dennis Schumacher

infas ist zertifiziert

nach ISO 20252 für die Markt-, Meinungs- und Sozialforschung

ISO 20252

CERTIFIED BY AUSTRIAN STANDARDS

infas ist Mitglied im

Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. (ADM) und ESOMAR

adm

ESOMAR member

© infas Institut für angewandte Sozialwissenschaft GmbH
Der Inhalt dieses Berichts darf ganz oder teilweise nur mit unserer schriftlichen Genehmigung veröffentlicht, vervielfältigt, gedruckt oder in Informations- und Dokumentationssystemen (information storage and retrieval systems) gespeichert, verarbeitet oder ausgegeben werden.

Vorbemerkung

Im Jahr 2012 hat das infas Institut für angewandte Sozialwissenschaft eine erste bundesweite Telefonbefragung bei rund 35.000 Männern und Frauen im Auftrag des Bundeskriminalamts (BKA) durchgeführt. Ziel der Befragung war es, umfassende Informationen über die Themen Kriminalitätsfurcht, Opfererfahrungen, Anzeigeverhalten, Betrugsdelikte mit EC- und Kreditkarte oder im Internet sowie Gewalterfahrungen zu erheben. Die erste Erhebung des Deutschen Viktimisierungssurveys 2012 war als repräsentative CATI-Befragung konzipiert. Mittels Telefonbefragung wurden erstmals bei einer großen Stichprobe in allen Bundesländern mit demselben Erhebungsinstrument zeitgleich Prävalenzraten zu Opfererfahrungen ermittelt.

Mit dieser bundesweiten Dunkelfelderhebung wurde eine Forschungslücke geschlossen und an internationale Standards aufgeschlossen. Die Befragung war als Querschnitterhebung angelegt.

Im Jahr 2017 – also fünf Jahre nach der ersten bundesweiten Telefonbefragung – wurde erneut die bundesdeutsche Bevölkerung ab 16 Jahren per Telefon zu den genannten Themen befragt. Das BKA hat infas beauftragt, mit einem im Vergleich zur Erhebung 2012 nur leicht modifizierten Instrument rund 31.000 Frauen und Männer zu befragen.

Die Datenerhebung des Deutschen Viktimisierungssurveys 2017 erfolgte erneut mittels computerunterstützter Telefoninterviews (CATI). Mit dieser zweiten umfangreichen Querschnitterhebung sollen Veränderungen zwischen 2012 und 2017 gemessen werden.

Die Interviews wurden analog zu 2012 in den Sprachen Deutsch, Türkisch und Russisch durchgeführt. Im Vorfeld der Haupterhebung wurde das Fragenprogramm in einem quantitativen Pretest auf Feldfähigkeit getestet. Die Durchführung des Tests ist in einem Pretestbericht[1] dokumentiert.

Der vorliegende Methodenbericht dokumentiert das Erhebungsdesign, die Stichprobenziehung, die Feldarbeit der Hauptstudie sowie die Prüfung und Gewichtung der 2017 erhobenen Daten.

infas Institut für angewandte Sozialwissenschaft GmbH

[1] Schiel, S., Knerr, P. (2017): „Pretestbericht, Deutscher Viktimisierungssurvey 2017 – Repräsentative Bevölkerungsbefragung“, Bonn.

Inhaltsverzeichnis

1	**Erhebungsdesign**		**7**
	1.1	Grundgesamtheit	8
	1.2	Stichprobenziehung	8
	1.2.1	Basisstichprobe	8
	1.2.2	Zusatzstichprobe nach onomastischem Verfahren	10
	1.3	Das Erhebungsinstrument	11
2	**Durchführung der Hauptstudie**		**17**
	2.1	Feldzeit	17
	2.2	Interviewerschulung	17
	2.3	Interviewereinsatz	18
	2.4	Durchführung fremdsprachiger Interviews	19
	2.5	Feldbegleitende Prüfung möglicher Interviewereffekte	20
	2.6	Supervision zur Qualitätssicherung	21
	2.7	Feldsteuerung	21
	2.8	Adressnachbearbeitung zur Erhöhung der Ausschöpfung	22
	2.9	Stichprobenausschöpfung	23
	2.9.1	Basisstichprobe	23
	2.9.2	Onomastische Stichprobe	26
	2.10	Interviewdauer	29
	2.11	Interviewsprache	29
	2.12	Kontakthäufigkeit	30
	2.13	Fallzahl pro Fragenmodul	31
	2.14	Einsatz der Vignettensets	32
	2.15	Verhältnis auswertbare Interviews aus Festnetz- und Mobilfunkstichprobe	33
	2.16	Interviewabbrüche und Abbruchgründe	34
3	**Datenprüfung und –gewichtung**		**39**
	3.1	Datenprüfung und –aufbereitung	39
	3.2	Gewichtung und Hochrechnung	42
	3.2.1	Bestimmung der Auswahlwahrscheinlichkeit für die Dual-Frame-Stichproben	43
	3.2.2	Neuerer Ansatz: Erweiterung der Gewichtung um einen Erreichbarkeitsterm	45
	3.2.3	Neuerer Ansatz: iterative Designgewichtung	45
	3.2.4	Proportionalisierung des Dual-Use-Anteils	46
	3.2.5	Auswahlwahrscheinlichkeit für die Onomastikstichprobe	46
	3.2.6	Integration der Onomastikstichprobe	47
	3.2.7	Unterschiedliche Varianten der Designgewichtung	48
	3.2.8	Trimmen der Designgewichte	48
	3.2.9	Kalibrierung und Hochrechnung	49
4	**Literaturverzeichnis**		**60**
Anhang			**61**

Abbildungsverzeichnis

Abbildung 1 Struktur des Fragebogens 14

Tabellenverzeichnis

Tabelle 1	Überblick über das Studiendesign	7
Tabelle 2	Regionale Verteilung von Personen mit türkischem Migrationshintergrund nach Bundesländern in der Grundgesamtheit und der Bruttostichprobe der onomastischen Stichprobe	11
Tabelle 3	Angezielte Fallzahl pro Modul in der Basisstichprobe	15
Tabelle 4	Fragebogeninhalte	16
Tabelle 5	Telefoninterviewer nach Geschlecht	18
Tabelle 6	Telefoninterviewer nach Alter (gruppiert)	19
Tabelle 7	Realisierte Interviews pro Interviewer	19
Tabelle 8	Anzahl realisierte Interviews pro Interviewer (gruppiert)	19
Tabelle 9	Feldergebnis der Basisstichprobe nach AAPOR-Klassifikation	25
Tabelle 10	Final Outcome nach AAPOR, differenziert nach Teilstichproben Festnetz und Mobilfunk	26
Tabelle 11	Feldergebnis der onomastischen Stichprobe nach AAPOR-Klassifikation	28
Tabelle 12	Interviewdauer	29
Tabelle 13	Basisstichprobe: Interviewsprache	30
Tabelle 14	Onomastische Stichprobe: Interviewsprache	30
Tabelle 15	Kontakthäufigkeit der auswertbaren Interviews	31
Tabelle 16	Fallzahl pro Modul in der Basisstichprobe	32
Tabelle 17	Fallzahl pro Modul insgesamt	32
Tabelle 18	Vignettensetverteilung für die Basisstichprobe	33
Tabelle 19	Festnetz- und Mobilfunkstichprobe in der Basisstichprobe	34
Tabelle 20	Festnetz- und Mobilfunkstichprobe insgesamt	34
Tabelle 21	Abbruchstellen im Fragebogen	38
Tabelle 22	Vercodung der offenen Ortsangaben	41
Tabelle 23	Trimmungsgrenzen der Designgewichte	49
Tabelle 24	Prozentualer Verteilungsvergleich (Personen)	52
Tabelle 25	Effektivitätsmaße und effektive Fallzahl (Personen)	55
Tabelle 26	Kennwerte der kalibrierten Gewichte und Hochrechnungsfaktoren Personen (Gewichtungsvariante c1 – klassische Designgewichtung mit Onomastik)	56
Tabelle 27	Prozentualer Verteilungsvergleich (Haushalte)	58
Tabelle 28	Effektivitätsmaß und effektive Fallzahl im Vergleich (Haushalte)	59
Tabelle 29	Kennwerte der kalibrierten Gewichte und Hochrechnungsfaktoren Haushalte (Gewichtungsvariante c1 – klassische Designgewichtung mit Onomastik)	59
Tabelle 30	Änderungen am Fragebogen im Vergleich zur Erhebung 2012	61

1 Erhebungsdesign

Bei der Datenerhebung des Deutschen Viktimisierungssurveys 2017 wurde weitgehend analog zur Erhebung des Deutschen Viktimisierungssurveys 2012[2] vorgegangen. Ein zentrales Ziel der zweiten bundesweiten Querschnitterhebung war, Vergleichszahlen zur ersten Erhebung 2012 zu ermitteln, um Veränderungen messen zu können. Wie 2012 erfolgte die Datenerhebung mit computerunterstützten Telefoninterviews (CATI). Schwerpunkte der Befragung waren unverändert die Themen Kriminalitätsfurcht, Opfererfahrungen, Anzeigeverhalten, Betrugsdelikte mit EC- und Kreditkarte oder im Internet sowie Gewalterfahrungen. Zudem wurden Fragen zur Soziodemografie gestellt. Es wurden Personen in Privathaushalten im Alter ab 16 Jahren befragt. Die nachfolgende Synopse gibt einen Überblick über das Studiendesign.

Tabelle 1 Überblick über das Studiendesign

Studientitel	Lebenssituation und Sicherheit in Deutschland
Grundgesamtheit	Bevölkerung in Privathaushalten ab 16 Jahren in Deutschland
Erhebungsmethode	Computergestützte Telefoninterviews (CATI)
Erhebungsinstrument	CATI-Fragebogen
Erhebungssprachen	Deutsch, Türkisch und Russisch
Pretest	– Deutsch n=150 – Türkisch n=19 – Russisch n=18
Stichprobenansatz	Basisstichprobe – Dual-Frame-Design, angezieltes Verhältnis Festnetz- und Mobilfunkstichprobe 22.500 (75%) zu 7.500 (25%) Interviews; je zufällig generierte Rufnummern – Ziel n=30.000 Interviews Onomastische Stichprobe – Listenauswahl unter Anwendung onomastischer Verfahren – Türkisch, Festnetz- und Mobilfunknummern – Ziel n=1.000 Interviews
Feldzeit Hauptfeld	– Basisstichprobe: 10.07.2017 bis 05.01.2018 – Onomastische Stichprobe: 16.08.2017 bis 02.01.2018
Auswahl Zielperson	Festnetz: Last-Birthday-Schlüssel, Mobilfunk: Hauptnutzer des Telefons
Interviewerschulung	Mündliche Schulung und Einsatz schriftliches Interviewerhandbuch
Auswertbare Interviews	– Basisstichprobe: n=30.180 – Onomastische Stichprobe: n=1.012
Gesamtdauer CATI-Interviews	Durchschnittlich 22,1 Minuten
Rückversicherungs-möglichkeit	– Telefonische Ansprechpartner bei infas und BKA – infas-Homepage – BKA-Homepage
Datensätze und Dokumentation	– Gelabelter Gesamtdatensatz – Gewichte u. Hochrechnungsfaktoren (insgesamt 8 Gewichtungsvarianten, jeweils Gewichtungs- u. Hochrechnungsfaktoren auf Personen und Haushalt) – Tabellenbände (2 Tabellenbände: gewichtete und ungewichtete Häufigkeiten) – Methodenbericht

[2] Vgl. Schiel, S., Dickmann, C., Gilberg, R. und Malina, A. (2013): Repräsentative Bevölkerungsbefragung im Rahmen des BaSiD-Teilvorhabens „Sicherheitsgefährdungen durch Kriminalität" – Methodenbericht. Bonn.

1.1 Grundgesamtheit

Grundgesamtheit der Erhebung ist die deutschsprachige Wohnbevölkerung in Privathaushalten ab 16 Jahren sowie die nicht über ausreichend Deutschkenntnisse verfügende Wohnbevölkerung der genannten Altersgruppe, soweit sie auf Türkisch oder Russisch befragt werden kann.

1.2 Stichprobenziehung

Die Stichprobe für die Erhebung war von zentraler Bedeutung, da Prävalenzschätzungen eines Untersuchungsgegenstands mit geringer Prävalenz (Opfererfahrungen) vorgenommen werden sollten. Dies setzte voraus, dass der Auswahlrahmen die gesamte Zielpopulation abdeckt. Jede Einheit der Zielpopulation musste also eine berechenbare Wahrscheinlichkeit größer 0 haben, in die Stichprobe zu gelangen (kein Coverage-Problem). In den folgenden Kapiteln wird beschrieben wie dies auf Basis der eingesetzten Stichproben erreicht werden konnte.

Es kamen zwei Teilstichproben zum Einsatz: eine bundesweite Festnetz- und Mobilfunkstichprobe (Basisstichprobe) sowie eine zusätzliche Stichprobe, um den Anteil Personen mit türkischem Migrationshintergrund zu erhöhen.

1.2.1 Basisstichprobe

Bei der Ziehung der Basisstichprobe ist analog zum Deutschen Viktimisierungssurvey 2012[3] vorgegangen worden. Der Erhebung liegt eine Telefonstichprobe nach dem Häder-Gabler-Verfahren, dem Standardverfahren der ADM[4]-Institute, zugrunde.[5] Bei diesem Verfahren werden alle im Feld eingesetzten Telefonnummern synthetisch generiert. Dieses Vorgehen ist notwendig, da allgemein zugängliche Verzeichnisse wie Telefonbücher nur einen eingeschränkten Auswahlrahmen bieten. Aufgrund der Ergebnisse verschiedener Studien ist davon auszugehen, dass nicht alle Personen in allgemein zugänglichen Registern gelistet sind. Insbesondere jüngere Leute, Single-Haushalte und Personen mit niedrigem Bildungsniveau sind häufiger nicht im Telefonbuch eingetragen als ältere Personen oder Mehrpersonenhaushalte. Neben gelisteten (in Telefonregistern eingetragenen) Nummern sind daher zufällig generierte Nummern bei der Stichprobenziehung zu berücksichtigen bzw. mit entsprechenden Inklusionswahrscheinlichkeiten zu versehen. Das weiter oben genannte Häder-Gabler-Verfahren ist in der Praxis in großem Maßstab erprobt und gehört mittlerweile zur gängigen Praxis bei telefonischen Bevölkerungsbefragungen.

[3] Vgl. Schiel, S., Dickmann, C., Gilberg, R. und Malina, A. (2013): Repräsentative Bevölkerungsbefragung im Rahmen des BaSiD-Teilvorhabens „Sicherheitsgefährdungen durch Kriminalität". Methodenbericht. Bonn. S.9ff.

[4] ADM = Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V.

[5] Häder, S., Gabler, S. (1998): Ein neues Stichprobendesign für telefonische Umfragen in Deutschland. In Gabler S. , S. Häder und J.H.P. Hoffmeyer-Zlotnik (Hrsg.) Telefonstichproben in Deutschland. Opladen, Westdeutscher Verlag.

Eine zusätzliche Herausforderung stellen Haushalte dar, die nur noch einen Mobilfunkanschluss haben und über keinen Festnetzanschluss mehr verfügen (sogenannte „Mobile onlys"). Sie haben nach der Logik der Stichprobenziehung aus der Menge aller Festnetznummern keine Inklusionswahrscheinlichkeit. Das bedeutet, sie wären also aus der Stichprobenziehung ausgeschlossen (Coverage-Problem). Im Jahr 2011 waren nach einer Studie des Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute (ADM) etwa 13 Prozent der Bevölkerung (vermutlich etwa 18 Prozent der Haushalte[6]) nur noch über Mobilfunknummern erreichbar; der Anteil dürfte inzwischen sogar noch gestiegen sein. Diese Haushalte unterscheiden sich zudem in ihrer Zusammensetzung stark von den Haushalten, die nach wie vor über einen Festnetzanschluss erreichbar sind. Es handelt sich demnach um ein systematisches Coverage-Problem.

Bereits für die Datenerhebung im Deutschen Viktimisierungssurveys 2012 wurde deshalb zusätzlich eine Mobilfunkstichprobe aus synthetisch generierten Mobilfunknummern verwendet. Diese Vorgehensweise – die Ziehung einer Stichprobe aus zwei unterschiedlichen Auswahlrahmen – wird als „Dual-Frame-Ansatz" bezeichnet und bedeutet, dass zwei unterschiedliche Stichprobenansätze (einer für Nummern mit Festnetzanschluss, ein anderer für die Mobilnummern) miteinander verbunden werden. Eine solche Dual-Frame-Stichprobe ist beim Viktimisierungssurvey 2017 analog zu 2012[7] eingesetzt worden.

Auch bei den Mobilfunknummern handelt es sich um generierte Nummern, da nur ein sehr kleiner Teil der Mobilfunknummern in einem Verzeichnis gelistet ist. Im Mobilfunkbereich ist allerdings keinerlei regionale Verortung über Vorwahl und Rufnummer möglich, so dass keine regionale bzw. regionalstrukturelle Schichtung und auch keine regionale Selektion möglich sind.

Die Auswahlrahmen für Festnetz- und Mobilfunkstichproben wurden von der Arbeitsgemeinschaft ADM-Telefonstichproben[8] zur Verfügung gestellt. Aus diesen beiden beschriebenen Auswahlrahmen wurde die Bruttostichprobe der Telefonnummern gezogen.

Das beschriebene Auswahlverfahren führt dazu, dass sowohl in der Festnetz- als auch in der Mobilfunkstichprobe ein erheblicher Teil der Telefonnummern der Bruttostichprobe technisch nicht geschaltet ist. Für die Berechnung der Bruttostichprobengröße wurde anhand der vorliegenden Erfahrung mit den ADM-Auswahlgrundlagen ein Anteil von 60 Prozent nicht geschalteter Nummern in der Festnetzstichprobe und ein Anteil von 67 Prozent nicht geschalteter Nummern in der Mobilfunkstichprobe unterstellt.

[6] ADM (2012), ADM-Forschungsprojekt „Dual-Frame-Ansätze" 2011/2012, Forschungsbericht.

[7] Vgl. Schiel, S., Dickmann, C., Gilberg, R. und Malina, A. (2013): Repräsentative Bevölkerungsbefragung im Rahmen des BaSiD-Teilvorhabens „Sicherheitsgefährdungen durch Kriminalität". Methodenbericht. Bonn. S.9ff.

[8] Die kontinuierliche Aktualisierung des Frames wird von der Arbeitsgemeinschaft „Stichproben" des Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute (ADM) e.V. organisiert. Zugriff haben nur Institute, die Mitglied dieser AG sind. infas ist Mitglied und hat entsprechend Zugriff auf die aktualisierte Grundgesamtheit.

Die Bruttoeinsatzstichprobe für die Basisstichprobe bestand aus insgesamt 499.140 Mobilfunk- und 1.183.165 Festnetznummern. Das Mischungsverhältnis von Festnetznummern zu Mobilfunknummern in der Basisstichprobe lag somit bei 70 Prozent Festnetznummern zu 30 Prozent Mobilfunknummern. Es wurde ein Verhältnis von 75 Prozent Festnetznummern und 25 Prozent Mobilfunknummern in der realisierten Stichprobe angestrebt. Die Bruttoeinsatzstichprobe wurde in mehrere Teilstichproben aufgeteilt und abhängig vom Feldverlauf in einzelnen Tranchen ins CATI-Feld gegeben.

1.2.2 Zusatzstichprobe nach onomastischem Verfahren

Die Datenerhebung zum Deutschen Viktimisierungssurvey 2012 hatte gezeigt, dass in der realisierten Basisstichprobe der Anteil Personen mit einem türkischem Migrationshintergrund niedriger war als in der Bevölkerung. Um den Anteil dieser Personengruppe zu erhöhen wurde damals eine zusätzliche sogenannte Onomastikstichprobe[9] eingesetzt. Das Verfahren wurde für die Erhebung 2017 repliziert.

Die Onomastikstichprobe wurde von „Dr. Andreas Humpert und Klaus Schneiderheinze GbR“ gezogen.[10] Der Auswahlrahmen wurde dabei so gebildet, dass mithilfe der Onomastik den im aktuellen Telefonverzeichnis eingetragenen Privathaushalten (17,2 Mio. Eintragungen, Januar 2017) eine regionale Herkunft bzw. ein spezifischer Migrationshintergrund zugeordnet wurde. Die Kennzeichnung basiert auf 3,4 Mio. Namen (Vor- und Nachnamen) aus 581 Quellen der Onomastik. Die zentrale Zuordnungsdatei enthält 30,1 Mio. verschiedene Personennamen (Nachname, Vorname) mit einer Zuordnung der Herkunftsregion. Daraus ergaben sich 145.017 Telefonnummern (davon 26.489 nur Mobilfunkanschlüsse) von eingetragenen Personen, die mit hoher Wahrscheinlichkeit über einen türkischen Migrationshintergrund verfügen. Sie bildeten den Auswahlrahmen für die Stichprobenziehung.

Aus diesem Auswahlrahmen wurde eine regional geschichtete Zufallsstichprobe gezogen. Die regionale Schichtung nach Bundesländern erfolgte anhand der Verteilung von Personen mit türkischem Migrationshintergrund aus dem Mikrozensus 2015.

[9] Onomastik = Namenforschung, Namenkunde

[10] Humpert, A. und K. Schneiderheinze (2002), Stichprobenziehung für telefonische Zuwandererumfragen – Praktische Erfahrungen und Erweiterung der Auswahlgrundlage, in: Gabler, S. und S. Häder (Hg.), Telefonstichproben – Methodische Innovationen und Anwendungen in Deutschland, Münster: Waxmann, S. 187-208. Humpert, A. und K. Schneiderheinze (2000), Stichprobenziehung für telefonische Zuwandererumfragen – Einsatzmöglichkeiten der Namenforschung (Onomastik), in: ZUMA-Nachrichten 47, S36-63.

Tabelle 2 Regionale Verteilung von Personen mit türkischem Migrationshintergrund nach Bundesländern in der Grundgesamtheit und der Bruttostichprobe der onomastischen Stichprobe

Bundesland	Grundgesamtheit	Bruttostichprobe
Spalten%	*%*	*%*
Schleswig-Holstein	2,0	1,9
Hamburg	3,2	2,2
Niedersachsen	6,8	7,3
Bremen	1,7	1,4
Nordrhein-Westfalen	33,1	31,4
Hessen	10,6	11,0
Rheinland-Pfalz	4,2	4,8
Baden-Württemberg	17,2	19,9
Bayern	11,8	14,4
Saarland	0,8	0,8
Berlin	8,2	4,2
Neue Länder zusammen	0,7	0,7

Quelle: Mikrozensus 2015, eigene Berechnung

Vor dem Feldeinsatz wurden die Telefonnummern der onomastischen und der Basisstichprobe abgeglichen, um sicherzustellen, dass es keine Überschneidungen zwischen den Stichproben gab. Nach dieser Prüfung wurden 20.791 Fälle der onomastischen Stichprobe ins CATI-Feld gegeben.

Für alle eingesetzten Stichproben standen lediglich Telefonnummern für den Feldeinsatz zur Verfügung, d.h. es lagen weder Namen noch Adressen vor. Die Zielpersonen wurden in den Festnetzstichproben anhand des Last-Birthday-Schlüssels und in der Mobilfunkstichprobe der Hauptnutzer des Telefons ausgewählt. Es bestand somit keine Möglichkeit, mithilfe eines Anschreibens über das Projekt zu informieren und die Betroffenen auf den Anruf vorzubereiten. Der Anruf durch die Interviewer erfolgte also „kalt“.

1.3 Das Erhebungsinstrument

Der Fragebogen für den Deutschen Viktimisierungssurvey 2017 basiert auf dem im Jahr 2012 eingesetzten Erhebungsinstrument; die grundsätzliche Struktur des Fragebogens blieb unverändert. Es wurden punktuell Änderungen im Fragebogen und im Kontaktgespräch vorgenommen. Das Kontaktgespräch wurde aufgrund der Pretesterfahrungen leicht modifiziert.

Die größten inhaltlichen Änderungen im Fragenprogramm umfassen[11]:

- Reduzierung der Fragebogenmodule von 7 auf 5 (gestrichen wurden die bisherigen Module 1 „Kriminalitätsfurcht letzte 7 Tage“ und 2 „Mediennutzung“), Neunummerierung der Module für die Erhebung 2017,
- Aufnahme von zusätzlichen Vignetten im Modul 2 (eine zusätzliche Vignette zu Raub, neue Vignetten zu Wohnungseinbruchdiebstahl),
- Aufnahme von Folgefragen zur Viktimisierung im Kontext der Internetnutzung (Schädigungen durch Schadware; Phishing; Pharming),
- Ergänzung von Fragen zur telefonischen Erreichbarkeit der Befragungsperson in den letzten 3 Tagen für die Gewichtung der Daten.

Darüber hinaus wurde nach dem Pretest noch ein Set mit 8 neuen Fragen in das Fragenprogramm aufgenommen. Diese Fragen betrafen die Arbeit der Justiz in Deutschland.

Nachdem der Fragebogen in deutscher Sprache für das Hauptfeld vollständig mit dem BKA abgestimmt war und vollständig programmiert vorlag, wurden alle Änderungen im Vergleich zum Fragebogen aus dem Jahr 2012 ins Türkische und Russische übersetzt. Die neu übersetzten Passagen wurden anschließend in den fremdsprachigen Fragebogen aus der Vorgängererhebung integriert. Damit wurde gewährleistet, dass die Übersetzung des Fragebogens – bis auf die wenigen Änderungen – identisch war mit der in 2012 verwendeten Übersetzung. Nach Abschluss der Fragebogenprogrammierung fanden auch für die beiden fremdsprachigen Instrumente ein Pretest[12] statt.

Der Fragebogen gliedert sich in 8 Elemente (vgl. Abbildung 1, Elemente A bis H). Im Kontaktgespräch (Element A) wurde in der Festnetzstichprobe bei Mehrpersonenhaushalten mittels Last-Birthday-Schlüssel die Zielperson zufällig ausgewählt. In der Mobilfunkstichprobe wurde der Hauptnutzer des Handys befragt. Da bei der Festnetzstichprobe nur Privatanschlüsse zur Grundgesamtheit zählen, wurde auch bei der Mobilfunkstichprobe analog vorgegangen, indem Handyanschlüsse, die ausschließlich beruflich genutzt werden, zum Ausschluss des Hauptnutzers aus der Befragung führten.

Im Zuge der Zielpersonenauswahl in der Festnetzstichprobe wurde bereits im Kontaktgespräch die reduzierte Haushaltsgröße (Anzahl Personen ab 16 Jahren im Haushalt) erfragt. In der Mobilfunkstichprobe wurde die reduzierte Haushaltsgröße erst im Verlauf des Interviews ermittelt, da für die Auswahl der Zielperson die Haushaltsgröße irrelevant war.

[11] Eine vollständige Übersicht der Änderungen am Fragebogen im Vergleich zur Erhebung 2012 befindet sich im Anhang.

[12] Vgl. Knerr, P., Schiel, S. und Quandt, S. (2017): Sachstandsbericht. Deutscher Viktimisierungssurvey 2017. Bonn.

Nachdem im Kontaktgespräch (Element A) die Zielperson ausgewählt wurde und sie ihre Bereitschaft erklärt hatte, am Interview teilzunehmen, begann das inhaltliche Interview (ab Element B).

Die Fragen zu den Themen Lebenszufriedenheit und Vertrauen wurden allen Zielpersonen gestellt. Im Anschluss daran wurden einem Teil der Zielpersonen Modulfragen gestellt. Es waren 5 Fragenmodule im Fragebogen implementiert (Element C, vgl. Abbildung 1). Jedes Fragenmodul umfasste mehrere Fragen zu einem spezifischen Thema. Diese Fragen wurden jeweils nur einer Teilgruppe gestellt. Die Zuordnung einer Zielperson zu den Teilgruppen erfolgte zufallsgesteuert. Die zufällige Zuordnung, ob ein Modul und welches Modul für diesen Fall vorgelegt wurde, erfolgte über eine Kennung der Fallnummern in der Bruttostichprobe. Weder das Antwortverhalten der Zielperson noch die Interviewer hatten einen Einfluss auf die Auswahl der Modulfragen.

In Modul 2 wurde wieder die Messung von Strafeinstellungen anhand unterschiedlicher Fallbeispiele (sogenannte Vignetten[13]) vorgesehen, die die Zielpersonen bewerten sollten. Die zufällige Zuweisung der Vignettensets zu einzelnen Fallnummern (Case-IDs) erfolgte analog zur Erhebung aus dem Jahr 2012 über die Bruttostichprobe. Weder das Antwortverhalten der Zielperson noch der Interviewer hatten Einfluss darauf, welches Vignettenset im Interview vorgelegt wurde.

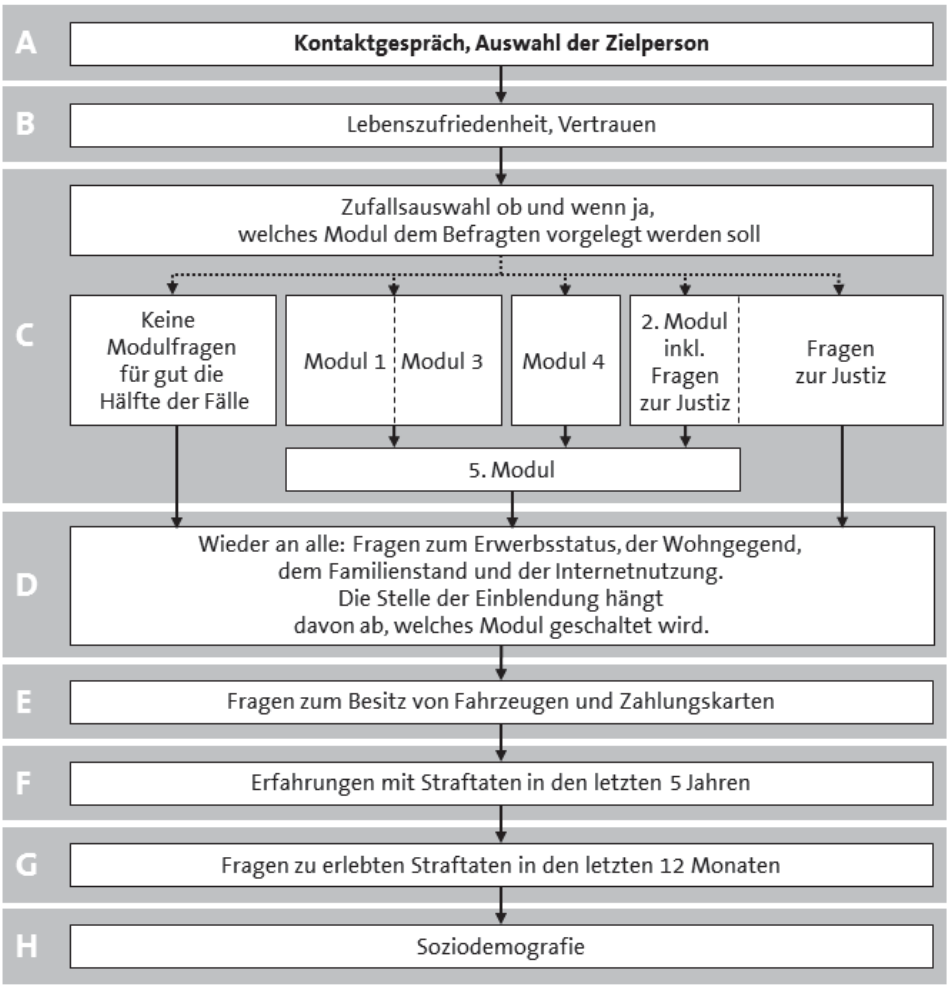
Die Fragen in Modul 5 wurden stets in Kombination mit den Fragen aus einem der anderen 4 Module gestellt. Außerdem wurden die Module 1 und 3 immer gemeinsam geschaltet, d.h. eine Person, die im Interview die Fragen des Moduls 1 erhielt, durchlief immer auch Modul 3. Die übrigen Module (2 und 4) wurden alternativ eingesetzt, so dass es hier zu keinen Überschneidungen untereinander oder mit Modul 1 und 3 kam. In der Zusatzstichprobe (onomastische Stichprobe) wurde analog zur Basisstichprobe verfahren.

Die Zusatzfragen zur Justiz folgen nicht der Modullogik. Zum einen wurden die 8 Justizfragen immer zusammen mit den Fragen aus Modul 2 gestellt. Zum anderen wurden die Justizfragen in weiteren 8.000 Interviews[14] gestellt, in denen keines Module zum Einsatz kam. Infolgedessen können die Justizfragen lediglich in Verbindung mit den Fragen aus Modul 2 und Modul 5 ausgewertet werden, aber mit keinem der Module 1, 3 oder 4.

[13] Zu Möglichkeiten und Grenzen von Vignetten in der sozialwissenschaftlichen Forschung siehe Auspurg, K., Hinz, T. und Liebig, S. (2009): „Komplexität von Vignetten, Lerneffekte und Plausibilität im Faktoriellen Survey“, in Methoden – Daten – Analysen, Jg. 3, Heft 1, S. 59-96.

[14] Vgl. angezielte Fallzahlen pro Modul in Tabelle 3.

Abbildung 1 Struktur des Fragebogens



Das beschrieben Vorgehen garantierte, dass für jedes Modul nach Abschluss der Feldarbeit wie geplant (mindestens) 1.500 bis 5.700 Interviews realisiert wurden und damit ausreichend viele Fälle für spätere Analysen vorlagen. Die Justizfragen sollten in (mindestens) 10.000 Interviews gestellt werden. Die angestrebten Fallzahlen pro Modul in der Basisstichprobe sind in Tabelle 3 aufgeführt.[15] Die Beschränkung des Einsatzes der Module auf Fallzahlen, die deutlich unter dem Gesamtumfang der angezielten 30.000 Interviews lagen, garantierte, dass die geplante durchschnittliche Interviewdauer von rund 20 Minuten insgesamt eingehalten werden konnte.

Die Themenfelder, die in Abbildung 1 unter Element D zusammengefasst sind, wurden an unterschiedlichen Stellen im Fragenkomplex Bereich C eingestreut (s.

[15] In der onomastischen Stichprobe war die prozentuale Verteilung der einzelnen Module genauso wie in der Basisstichprobe.

dazu auch Tabelle 4). Es wurde darauf geachtet, dass die Fragen in den Kontext des Interviews passten. Fragen zu Erwerbsstatus oder Familienstand sind beispielsweise so platziert, dass die Antworten für eine spätere Steuerung von entsprechenden Einblendungen genutzt werden können. Bei Folgefragen war es in manchen Fällen notwendig zu wissen, ob die Zielperson verheiratet oder ledig, berufstätig oder noch Schüler war. Dieses Vorgehen entspricht dem Verfahren aus der Befragung 2012.

Unverändert im Vergleich zur Erhebung 2012 wurde auch bei der Befragung 2017 im Screener-Teil I (Fragebogenelement E) ermittelt, ob die Zielperson (oder ein anderes Haushaltsmitglied) ausgewählte Fahrzeuge besitzt. Im zweiten Teil der Screenerfragen (Fragebogenelement F) wurden die Zielpersonen zu Beginn gefragt, ob sie persönlich in den vergangenen rund 5 Jahren (d.h. seit Anfang 2012) Erfahrungen mit ausgewählten Straftaten gemacht hatten. Sollten sie Opfer einer der genannten Straftaten geworden sein, dann wurde in Element G gezielter nach Opfererfahrungen in den letzten 12 Monaten gefragt. Das Interview endete mit ausgewählten Fragen zur Soziodemografie.

Den Screenerfragen unter E und F kam eine zentrale Bedeutung bei der Ermittlung von Prävalenzen zu. Sie sind im Vorfeld der Erhebung 2012 intensiv getestet und im Vergleich zur Erhebung 2012 unverändert eingesetzt worden. In Tabelle 4 sind die Themen der einzelnen Frageblöcke bzw. -module aufgeführt. Darüber hinaus kann der Tabelle entnommen werden, welcher Zielgruppe die Fragen gestellt worden sind. Der Fragebogen wurde in 3 Sprachen (Deutsch, Russisch und Türkisch) eingesetzt.

Tabelle 3 Angezielte Fallzahl pro Modul in der Basisstichprobe

Modulnummer	Angestrebte Fallzahl (absolut)	Anteil an Basisstichprobe (in %)
1*	1.500	5,0
2	2.000	6,7
3*	1.500	5,0
4	2.200	7,3
5	5.700	19,0
Zusatzfragen zur Justiz	10.000**	3,3

Quelle: Stichprobenplan
*Anmerkung: Modul 1 und Modul 3 werden immer in denselben Interviews geschaltet.
** Davon 2.000 im Modul 2 und 8.000 weitere, die keines der Module erhalten

Tabelle 4 Fragebogeninhalte

Zielgruppe	Inhalt
Alle	Kontaktgespräch und Auswahl der Zielperson (Last-Birthday-Schlüssel, Hauptnutzer des Mobilfunktelefons), Ermittlung der reduzierten Haushaltsgröße bei Festnetz
Alle	Lebenszufriedenheit, Vertrauen
Zufallsstichprobe	Modul 2: Kriminalitätsbezogene Einstellungen, inkl. Fragen zur Justiz
Zufallsstichprobe	Modul 1: Einstellungen zur Zuwanderung
Alle	Allgemeiner und beruflicher Bildungsabschluss, Erwerbsstatus, Fragen zur Wohngegend
Zufallsstichprobe, immer gemeinsam mit Modul 1	Modul 3: Wohngegend, Disorder/Incivilities
Alle	Ausgehverhalten, Familienstand, Internetnutzung
Zufallsstichprobe	Modul 4: Sicherheitsgefühl bei Internetnutzung
Alle	Standardfrage zur Kriminalitätsfurcht
Nur Befragte, denen auch Fragen zu einem anderen Modul gestellt wurden	Modul 5: Deliktspezifische Kriminalitätsfurcht
Alle	Screenerfragen zu den Themen – Besitz und Diebstahl von Kfz, Krad, Fahrrädern sowie Kreditkartenbesitz – Diebstahl aus Wohnungen, Wohnungseinbruchsdiebstahl – Versuchter Diebstahl aus Wohnungen, Wohnungseinbruchs-diebstahl – Fahrzeugdiebstahl – Konsumentenbetrug – Betrug mittels rechtswidrig erlangter Zahlungskarten – Viktimisierung bei Internetnutzung – Raub – Sonstiger Diebstahl – Körperverletzung
Befragte, die positiv auf Screener-frage(n) geantwortet haben	Bündel Folgefragen zur Inzidenz, zum einzelnen Ereignis (einfache/mehrfache Viktimisierung)
Alle	Soziodemografie
Alle	Ausstattung mit Telekommunikationsmedien und telefonische Erreichbarkeit in den letzten 3 Tagen
Alle	Postleitzahl
Alle	Verabschiedung

2 Durchführung der Hauptstudie

2.1 Feldzeit

Die Hauptstudie startete am 10.07.2017. Ab dem 16.08.2017 kamen auch die fremdsprachigen Instrumente zum Einsatz. In der Basisstichprobe wurden Anfang Dezember die letzten definitiven Termine abgearbeitet. Parallel dazu startete Anfang Dezember die Konvertierungsstudie.

Kurz darauf wurde ab dem 12.12.2017 auch in der onomastischen Stichprobe eine Konvertierungsstudie aufgesetzt. Am 05.01.2018 wurden die letzten Interviews erhoben, da die angestrebten Fallzahlen sowohl in der Basisstichprobe als auch in der onomastischen Stichprobe erreicht waren.

2.2 Interviewerschulung

Grundsätzlich haben bei infas alle eingesetzten Interviewer vor ihrem ersten Interview an einer Grundschulung teilgenommen, in der u.a. auch die Faktoren einer erfolgreichen Felddurchführung und Kontaktierung vertieft werden. Insbesondere der Umgang mit nicht teilnahmebereiten Zielpersonen wird dort eingeübt. Dazu besteht ein Schwerpunkt der Grundschulung aus einem speziellen Training zur Einwandbehandlung (RAT = Refusal Avoidance Training[16]). Die Grundschulungen erfolgen vor dem Ersteinsatz der Interviewer und werden bei Bedarf im Einzelfall zur Vertiefung wiederholt.

Um eine hohe Qualität der Datenerhebung im Feld sicherzustellen und positiv auf die Ausschöpfung der Befragung einzuwirken, haben alle für die Erhebung vorgesehenen Interviewerinnen und Interviewer an einer studienspezifischen Schulung teilgenommen. Sie ist obligatorisch für jede CATI-Befragung bei infas. Die Inhalte der mündlichen Schulung orientierten sich an der Schulung des Viktimisierungssurveys 2012. Die Schulungsinhalte wurden mit Blick auf die Neuerungen im Fragebogen und insbesondere aufgrund der Erfahrungen aus dem Pretest mit unterschiedlichen Kontaktierungsstrategien aktualisiert. Ebenfalls wurde das Interviewerhandbuch aus dem Jahr 2012 überarbeitet und aktualisiert. Das Handbuch umfasste alle wichtigen Hinweise zum Hintergrund der Studie, zur Stichprobe, zu Besonderheiten im Fragebogen sowie Hilfestellungen beim Umgang mit Zielpersonen, die aufgrund der Fragen ggf. emotional aufgewühlt reagieren.

Die Unterlagen wurden vor dem Einsatz mit dem Auftraggeber abgestimmt. Ziel der projektspezifischen Schulung war eine auf gleich gutem Niveau wie 2012 stattfindende umfassende Vorbereitung der Interviewer.

[16] Verweigerungsvermeidungstraining: Besondere Aufmerksamkeit wird bei diesem Training den Maßnahmen und Interviewerstrategien zur Vermeidung von Teilnahmeverweigerung zuteil. Die Bemühungen zielen darauf ab, schwer motivierbare Personen zur Teilnahme an der Befragung zu bewegen, um die Stichprobe maximal auszuschöpfen. Es gibt kein allgemeingültiges Argument, mit dem garantiert jede Zielperson überzeugt werden kann, an der Befragung teilzunehmen. Jedes Interview ist individuell und einzeln auszuhandeln.

Um Interviewereffekte möglichst gering zu halten, sollte jeder einzelne Interviewer nur einen geringen Anteil der 31.000 Interviews durchführen. Deshalb wurde eine Vielzahl an Interviewern projektspezifisch geschult und im Telefonfeld eingesetzt. Insgesamt wurden 843 Interviewer für die Studie in einer mündlichen Schulung auf ihren Einsatz vorbereitet. Es fanden insgesamt 25 Schulungen statt. Bevor die geschulten Interviewer im Telefonstudio eingesetzt wurden, waren sie gehalten, sich intensiv mit der Testfassung des Fragebogens vertraut zu machen.

2.3 Interviewereinsatz

843 Interviewer haben mindestens ein Interview vollständig realisiert. Der Anteil Männer ist mit 41,6 Prozent etwas niedriger als der Frauenanteil (54,0 Prozent, Tabelle 5). Es kamen Interviewer aller Altersklassen ab 16 Jahren zum Einsatz (vgl. 0).

Die hohe Anzahl an eingesetzten Interviewerinnen und Interviewern garantierte, dass jeder einzelne nur einen geringen Anteil an der Gesamtfallzahl von mehr als 31.000 Interviews durchführte. Nur in sehr wenigen Fällen (3 Mal) haben Interviewer mehr als 300 Interviews über beide Teilstudien und die Konvertierungsstudien[17] hinweg realisiert. Und lediglich 6 Interviewer haben zwischen 201 und 250 Interviews geführt. Hingegen haben 74,1 Prozent der Interviewer bis zu 50 Interviews realisiert und weitere 17,1 Prozent haben 51 bis 100 Interviews durchgeführt (vgl. Tabelle 8). Im Durchschnitt führten die Interviewer rund 37 Interviews durch (vgl. Tabelle 7).

Tabelle 5 Telefoninterviewer nach Geschlecht

Geschlecht		
Spalten%	*abs.*	*%*
Männlich	351	41,6
Weiblich	455	54,0
Keine Angabe	37	4,4
Gesamt	843	100,0

Quelle: Einsatzdatenbank

[17] Vgl. dazu Kapitel 2.8.

Tabelle 6 Telefoninterviewer nach Alter (gruppiert)

Altersgruppen		
Spalten%	*abs.*	*%*
Zwischen 16 und 24 Jahre	432	51,2
Zwischen 25 und 34 Jahre	194	23,0
Zwischen 35 und 44 Jahre	44	5,2
Zwischen 45 und 54 Jahre	55	6,5
Zwischen 55 und 64 Jahre	72	8,5
Zwischen 65 und 74 Jahre	37	4,4
75 Jahre und älter	5	0,6
Keine Angabe	4	0,5
Gesamt	843	100,0

Quelle: Einsatzdatenbank

Tabelle 7 Realisierte Interviews pro Interviewer

	abs.
Durchschnittliche Anzahl	37,1
Maximum	409
Minimum	1
Standardabweichung	43,9
Anzahl Interviewer mit mindestens einem realisierten Interview	843

Quelle: Einsatzdatenbank

Tabelle 8 Anzahl realisierte Interviews pro Interviewer (gruppiert)

Anzahl realisierte Interviews pro Interviewer		
Spalten%	*abs.*	*%*
1 bis 50	625	74,1
51 bis 100	144	17,1
101 bis 150	55	6,5
151 bis 200	10	1,2
201 bis 250	6	0,7
251 bis 300	-	-
Über 300	3	0,4
Gesamt	843	

Quelle: Einsatzdatenbank

2.4 Durchführung fremdsprachiger Interviews

In der Basisstichprobe sprachen die Interviewer die Kontaktperson im Haushalt immer auf Deutsch an. Wenn sich im Gespräch herausstellte, dass die Kontaktperson (bzw. Zielperson) nicht ausreichend Deutsch sprach, wurde das Gespräch an dieser Stelle unterbrochen. Die Interviewer legten den Fall ab mit einem Hinweis auf die Fremdsprache, die der Gesprächspartner bzw. die Zielperson (wahrscheinlich) spricht. Wenn es sich bei dieser Sprache um Russisch oder Türkisch handelte, dann wurden diese Fälle in eine separate Stichprobendatei überführt. Diese Substichprobe wurde in den folgenden Tagen und Wochen aus-

schließlich von zweisprachigen Interviewern bearbeitet. Wenn der Gesprächspartner bzw. die Zielperson keine der 3 Interviewsprachen (Deutsch, Russisch oder Türkisch) sprach, dann wurde der Endstatus „keine Verständigung in einer der Zielsprachen möglich“ vergeben. Das Verfahren wurde analog zur Vorgehensweise in der Erhebung 2012 angewendet.

In der onomastischen Stichprobe wurden ebenfalls analog zur Erhebung 2012 verfahren. Es wurden ausschließlich zweisprachige Interviewer eingesetzt. Ihnen lag der Fragebogen in den beiden Sprachen Türkisch und Deutsch vor. Sie hatten so die Möglichkeit, bei Bedarf das Kontaktgespräch direkt auf Türkisch oder Deutsch zu führen. Sie konnten auch im späteren Verlauf des Gesprächs auf die jeweils andere Sprache wechseln, wenn sich herausstellte, dass die Zielperson nicht ausreichend Deutsch versteht und/oder spricht.

2.5 Feldbegleitende Prüfung möglicher Interviewereffekte

Während der gesamten Feldzeit wurde die Güte der realisierten Stichprobe anhand mehrerer Indikatoren kontinuierlich geprüft. Kern der feldbegleitenden Prüfung der realisierten Fälle war die Entwicklung der Prävalenzraten für 8 ausgewählte Delikte („Konsumentenbetrug“, „Raub“, „persönlicher Diebstahl“, „Körperverletzung“ und „Fahrraddiebstahl“, „Schädigung von Computern durch Schadsoftware“, „Phishing“[18], „Pharming“[19]). Die Auswahl dieser 8 Delikte orientierte sich an den Prüfungen, die bereits 2012 durchgeführt worden waren. Zusätzlich sind die 2017 in dieser Tiefe neu berücksichtigten Delikte „Schädigung durch Schadsoftware“, „Phishing“ und „Pharming“ in die Liste der beobachteten Prävalenzen aufgenommen worden.

Des Weiteren wurde geprüft, ob die Erfassung der Prävalenzraten von Interviewereffekten beeinflusst wird. Dazu wurden die designgewichteten Prävalenzraten der 8 genannten Delikte wöchentlich auf mögliche Interviewereffekte geprüft. Für diese Prüfung wurde in linearen Modellen ein Interviewerklumpungseffekt berechnet. Als Maß wurde der Intraclusterkorrelationskoeffizient (ICC) verwendet, wobei die eingesetzten Interviewer als Cluster definiert wurden. Anhand dieser Analysen wurde geprüft, ob die aufgenommenen Angaben eines einzelnen Interviewers ein charakteristisches Muster zeigen, das sich von den durchschnittlichen aufgenommenen Angaben aller Interviewer signifikant unterschied. Die Analysen möglicher Interviewereffekte haben im Feldverlauf keine Auffälligkeiten erkennen lassen.

[18] Gemeint ist die Preisgabe vertraulicher Daten, wie etwa Kontonummer oder Geheimzahl, nachdem in einer E-Mail mit gefälschter Absenderadresse dazu aufgefordert worden waren.

[19] Hierbei handelt es sich um die Preisgabe von Zugangsdaten (z.B. für das Online-Banking) nach Umleitung auf eine gefälschte Webseite.

2.6 Supervision zur Qualitätssicherung

Während der gesamten Feldphase fand eine Qualitätssicherung der Interviewdurchführung statt. Die Qualitätssicherung der Interviewführung wurde durch geschulte Supervisoren verantwortet, die mit methodischen Anforderungen der Erhebung und der CATI-Befragungstechnik eng vertraut sind. Diese Supervisoren wurden gezielt inhaltlich geschult, um in enger Rückbindung mit der Projektleitung studienspezifische Rückfragen der Interviewer beantworten zu können. Zielsetzung der Supervision war es, durch kontinuierliche Betreuung und persönliche Ansprechbarkeit eine hohe Qualität der Befragungsdaten zu erhalten und die Ausschöpfung der Studie durch adäquate Kontaktaufnahme zu maximieren. Zu den Aufgaben der Supervision gehörten im Einzelnen:

- Die Kontrolle durch Mithören der Interviews im Telefonstudio und Überprüfung der Eingaben durch Beobachtung auf Supervisionsmaske (Übernahme der Interviewermonitoranzeigen auf den Supervisionsarbeitsplatz).
- Bei Bedarf die Durchführung von Einzelgesprächen oder Nachschulungen mit allen Interviewern hinsichtlich studienspezifischer Probleme.
- Bei Bedarf die Durchführung von Einzel- und Gruppengesprächen oder auch Nachschulungen direkt im Anschluss an das Interview bei Rückmeldungen von Problemen. Dabei werden bei Bedarf auch schriftliche Rückmeldungen an die Interviewer ausgeteilt.
- Direktes Eingreifen in das Interview bei beobachteten Fehlern bzw. Zuordnungsschwierigkeiten, die zu Fehlsteuerungen und Lücken zu führen drohen. Wobei dies entweder über kurze mündliche oder schriftliche Hinweise während des Interviews geschieht oder bei größeren Problemen und in Überforderungssituationen der Interviewer (meist infolge sehr komplizierter Verläufe oder widersprüchlicher Angaben der Zielperson) auch Eingriffe der Supervision in das Gespräch zur Folge haben kann.

Die Aufgaben der inhaltlichen Supervision waren sehr vielfältig und dauerten während der gesamten Feldphase an. Es bestand ein enger Informationsfluss zur Projektleitung, so dass inhaltliche Probleme möglichst rasch und auf Basis umfassender Kommunikation aller Beteiligten gelöst werden konnten. Auch eine sofortige Rückmeldung der gemeinsam erarbeiteten Problemlösungen ins Feld hinein war damit gewährleistet.

2.7 Feldsteuerung

Den Verlauf der Erhebung beobachtete die Projektleitung anhand von täglich aktualisierten Feldreportings.

Der Feldeinsatz stand unter dem Ziel einer maximal möglichen Ausschöpfung der Stichprobe in der vorgesehenen Feldzeit. Dazu gehört, dass einerseits schwer erreichbare Personen zu unterschiedlichen Tageszeiten angerufen werden, um die Wahrscheinlichkeit zu erhöhen, dass ein Kontakt zustande kommt. Eine zweite wichtige Feldstrategie ist das Motivieren von Befragten und das Vermeiden von Verweigerungen.

Darüber hinaus ermöglichten Terminabsprachen eine flexible und individuell ausgerichtete Bearbeitung der Stichproben. Die Stichprobe wurde in der gesamten Feldzeit an 6 Tagen in der Woche zwischen 9:00 Uhr und 21:00 Uhr (Montag bis Freitag) bzw. 10:00 bis 18:00 Uhr (Samstag), und in Ausnahmefällen auch sonntags, bearbeitet. Für jede Rufnummer wurde die Tageszeit für den nächsten Kontakt kontinuierlich verändert, um eine möglichst große Variation an unterschiedlichen Wochentagen und Tageszeiten für einen Kontaktversuch zu gewährleisten.

Die Kontaktzeiten wurden über die Feldzeit hin gestreckt, um Zielpersonen auch nach möglichen Urlaubs- oder längeren anderen Abwesenheitszeiten erreichen zu können. Die kontinuierliche Kontaktierung der Stichprobe über die gesamte Feldphase machte es möglich, auch Personen zu erreichen, die zum Beispiel in den Sommerferien längere Zeit verreist waren. Es bestand beispielsweise die Möglichkeit, vor Reisebeginn einen Kontakttermin für die Zeit nach dem Urlaub zu vereinbaren.

2.8 Adressnachbearbeitung zur Erhöhung der Ausschöpfung

Alle Interviewer sind in Schulungen mit Zugangsstrategien und der Behandlung von Einwänden vertraut gemacht worden. Ein fester Bestandteil der studienspezifischen Schulung war zudem das Pretestergebnis für ein erfolgreiches Kontaktgespräch. Zusätzlich enthielt das Interviewerhandbuch ausführliche Hinweise für ein zielführendes und erfolgreiches Kontaktgespräch.

Kam es trotzdem zu einer Teilnahmeverweigerung, so erlaubten die differenziert dokumentierten Ausfallgründe eine Analyse, weshalb es nicht zu einem Interview gekommen war. Es wurden „harte“ Verweigerungen – strikte Ablehnung einer Teilnahme – von „weichen“ Verweigerungen, die in der jeweiligen Situation begründet sind – z.B. „keine Zeit“ – unterschieden.

Zielpersonen, die während des Kontaktgesprächs situativ das Interview verweigerten[20], wurden nach einer längeren Karenzzeit im Dezember 2017 bis Anfang Januar 2018 erneut angerufen und noch einmal um Mitwirkung gebeten (sog. Konvertierungsstudien). Eine solche Nachbearbeitung wurde für die Basisstichprobe Anfang Dezember und für die onomastische Stichprobe Mitte Dezember 2017 gestartet. Für die Nachbearbeitung wurden die entsprechenden Fälle in eine Teilstichprobe separiert. Diese Konvertierungsstichproben wurden von speziell geschulten Interviewern bearbeitet und dauerten bis Anfang Januar 2018 an.

Die erneute Kontaktierung dieser ausgewählten Fälle erwies sich als erfolgreich. Insgesamt konnten durch die intensive Nachbearbeitung bestimmter Adressen zusätzlich 2.122 Interviews realisiert werden. Davon zählen 2.021 zur Basisstichprobe und 101 wurden in der onomastischen Stichprobe mittels Nachbearbei-

[20] Berücksichtigt wurden die Codes „keine Zeit“ und „kein Interesse“.

tung realisiert. Die Konvertierungsstudien wurden nach dem Erreichen der angezielten Gesamtfallzahlen beendet.

2.9 Stichprobenausschöpfung

Das Feldergebnis und somit die Stichprobenausschöpfung werden für beide Teilstichproben (Basis und Onomastik) im Folgenden differenziert dokumentiert.

Das Ergebnis der einzelnen Kontaktversuche wurde anhand differenzierter Response-Codes erfasst. Die differenzierten Response-Codes orientieren sich an den „Standard Definitions" der American Association for Public Opinion Research (AAPOR 2016[21]). Zu Gruppen von Bearbeitungsstatus zusammengefasst, bilden sie die Grundlage für die Berechnung von Kennwerten zur Stichprobenausschöpfung nach AAPOR-Standard. Nach AAPOR werden die Bearbeitungsergebnisse folgenden Gruppen zugeordnet:

- nicht Zielgruppe (NE),
- unbekannte Auswahl (Unknown Eligibility, UE),
- Nonresponse – nicht befragbar (NR-NA),
- Nonresponse – nicht erreicht (NR-NC),
- Nonresponse – Verweigerung (NR-R),
- realisierte Interviews (vollständig) (I).

Die im Folgenden ausgewiesenen Bearbeitungscodes bilden den endgültigen Bearbeitungsstatus einer Telefonnummer ab (Final Outcome). Zur Ermittlung des Final Outcomes wird zwischen allen Feldcodes innerhalb eines Kontaktverlaufs eine Priorisierung vorgenommen. So gilt grundsätzlich der Status aus dem letzten Sprachkontakt zum Haushalt, auch wenn danach noch vergebliche Kontaktversuche unternommen worden sein sollten. Eine Telefonnummer wird z.B. nur dann als „nicht erreicht" ausgewiesen, wenn im gesamten Kontaktverlauf kein Sprachkontakt zur Zielperson oder ihrem Haushalt hergestellt werden konnte. Telefonische Terminvereinbarungen oder telefonisch erteilte Teilnahmeverweigerungen werden bei der Dokumentation des Feldstatus ausgewiesen, auch wenn danach ein Interviewer noch vergeblich versucht haben sollte, die Zielperson zu erreichen.

2.9.1 Basisstichprobe

Die Bruttoeinsatzstichprobe der Basisstichprobe umfasste 1.682.305 zufällig generierte Telefonnummern. Rund 72 Prozent der generierten Rufnummern führten entweder zu keinem geschalteten Anschluss eines Privathaushalts, es war keine Verständigung in einer der drei Sprachen Deutsch, Türkisch oder Russisch möglich oder im Haushalt lebte keine Person aus der Zielgruppe bzw. der Hauptnutzer des Mobiltelefons war jünger als 16 Jahre und gehörte infolgedes-

[21] The American Association for Public Opinion Research (2016). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th edition. AAPOR.

sen nicht zur Zielgruppe (vgl. Tabelle 9, Spalte II). Bei 16,5 Prozent konnte die Zielgruppenzugehörigkeit nicht geklärt werden.

Abzüglich der genannten Rufnummern verbleiben noch 196.321 Telefonnummern (rund 12 Prozent der Bruttostichprobe), für die bekannt ist, dass eine Zielperson vorhanden ist. In 381 Fällen (0,2 Prozent der 196.321 Telefonnummern, vgl. Tabelle 9, Spalte IV) konnte mit der Zielperson kein Interview durchgeführt werden, da sie aus gesundheitlichen Gründen nicht befragbar war. Weitere 23 Prozent der 196.321 kontaktierten Haushalte bzw. Personen konnten trotz mehrfacher Kontaktversuche in der Feldzeit nicht für ein Interview erreicht werden. Für einen Teil dieser Gruppe war es bereits zu einem vereinbarten Interviewtermin gekommen. Dieser Termin wurde von der Zielperson allerdings nicht wahrgenommen.

In den meisten Fällen, in denen die Zielperson bis Feldende nicht erreicht werden konnte, handelte es sich um „vage Termine“ (rund 21 Prozent, Tabelle 9). In diesen Fällen ist es zwar zu Sprachkontakten gekommen. In der Feldzeit war es jedoch nicht möglich, mit der Zielperson einen (definitiven) Interviewtermin zu vereinbaren.

Abzüglich der nicht erreichten und nicht befragbaren Zielpersonen verbleiben noch 150.689 Fälle, in denen geklärt werden konnte, dass eine Zielperson vorhanden und befragbar war. In knapp 22 Prozent der Fälle (n=32.837) ist es auch zu einem Interview gekommen. Von den 32.837 Interviews sind 2.616 Interviews vorzeitig von den Zielpersonen beendet worden. Nach der Datenprüfung sind von den übrigen 30.221 vollständig realisierten Interviews 41 als nicht auswertbar deklariert worden (vgl. Abschnitt 3.1). Die verbleibenden 30.180 Interviews sind für die weiteren Analysen an das BKA-Forscherteam übermittelt worden.

Die Hauptgründe, nicht an einem Interview teilzunehmen, waren „kein Interesse, wegen des Themas“ und „grundsätzliche Teilnahmeverweigerung“. Diese beiden Verweigerungsgründe decken rund 56 Prozent der Verweigerungsfälle ab (Tabelle 9).[22]

[22] Zu Vergleichszwecken mit anderen Erhebungen wird an dieser Stelle zusätzlich die Response-Rate ausgewiesen. Analog zur Vorgehensweise 2012 wird die Response-Rate 4 nach AAPOR ausgewiesen. In diese Rate fließt eine Schätzung über den Anteil (e) der Telefonnummern mit Status „Zielgruppenzugehörigkeit unbekannt“ ein, der zur Zielgruppe gehört. Die Schätzung unterstellt, dass dieser Anteil genauso hoch ist wie bei den in der Befragung aufgeklärten Fällen. Der Zähler der Rate umfasst vollständig und unvollständig realisierte Interviews. Für den Nenner der Rate wird die Bruttostichprobe bereinigt um Adressen außerhalb der Zielgruppe und um einen geschätzten Anteil der ebenfalls nicht zur Zielgruppe gehörenden Personen aus der Kategorie „Zielgruppenzugehörigkeit unbekannt“. Unter dieser Annahme ergibt sich eine Response-Rate 4 in Höhe von 14,0 Prozent. Zur Berechnung der Response-Raten vgl.: The American Association for Public Opinion Research (AAPOR) (2016): Standard Definitions. Final Dispositions of Case Codes and Outcome Rate for Surveys. Revised Version 2016, S. 62.

Tabelle 9 Feldergebnis der Basisstichprobe nach AAPOR-Klassifikation

Spalten%	*abs.*	*%*	*abs.*	*%*	*abs.*	*%*
	I	**II**	**III**	**IV**	**V**	**VI**
Bruttostichprobe	**1.682.305**	**100,0**				
Nicht Zielgruppe (NE)	**1.208.233**	**71,8**				
Kein Anschluss	1.153.590	68,6				
Anschluss nur Fax/Modem	20.210	1,2				
Kein Privathaushalt unter diesem Anschluss	22.507	1,3				
Nicht in Zielgruppe lt. Angabe im Kontaktgespräch	5.488	0,3				
Bereits befragt	827	0,0				
Keine Verständigung in einer der 3 Zielsprachen möglich	5.611	0,3				
Unbekannte Auswahl (UE)	**277.751**	**16,5**				
Nicht abgehoben	173.492	10,3				
Anrufbeantworter	36.778	2,2				
Besetzt	2.799	0,2				
Legt sofort auf	64.682	3,8				
Zugehörigkeit zur Zielgruppe bekannt	**196.321**	**11,7**	**196.321**	**100,0**		
Nonresponse – nicht befragbar (NR-NA)	**381**	**0,0**	**381**	**0,2**		
Zielperson (lt. Auskunft) nicht befragbar/ dauerhaft krank/behindert	381	0,0	381	0,2		
Nonresponse – nicht erreicht (NR-NC)	**45.251**	**2,7**	**45.251**	**23,0**		
Definitiver Termin	3.138	0,2	3.138	1,6		
Vager Termin	41.947	2,5	41.947	21,4		
Zielperson in Feldzeit nicht zu erreichen	166	0,0	166	0,1		
Zielperson/Haushalt erreicht und befragbar					**150.689**	**100,0**
Nonresponse – Verweigerung (NR-R)	**117.852**	**7,0**	**117.852**	**60,0**	**117.852**	**78,2**
Zielperson verweigert grundsätzlich	32.896	2,0	32.896	16,8	32.896	21,8
Zielperson verweigert: keine Zeit/dauert zu lange	8.200	0,5	8.200	4,2	8.200	5,4
Zielperson verweigert: will nur persönlich befragt werden	5.454	0,3	5.454	2,8	5.454	3,6
Zielperson verweigert krank	2.165	0,1	2.165	1,1	2.165	1,4
Kein Zugang zu Zielperson / Teilnahme untersagt/lt. Auskunft nicht bereit	1.017	0,1	1.017	0,5	1.017	0,7
Kontaktperson verweigert jegliche Auskunft	15.591	0,9	15.591	7,9	15.591	10,3
Zielperson verweigert: kein Interesse, wegen des Themas	33.565	2,0	33.565	17,1	33.565	22,3
Zielperson verweigert: Datenschutzgründe/ u persönlich	2.343	0,1	2.343	1,2	2.343	1,6
Zielperson verweigert: sonstige Gründe	16.621	1,0	16.621	8,5	16.621	11,0
Realisierte Interviews (I)	**32.837**	**2,0**	**32.837**	**16,7**	**32.837**	**21,8**
Interviews vollständig realisiert, ausgeliefert	30.180	1,8	30.180	15,4	30.180	20,0
Interviews vollständig realisiert, nicht auswertbar	41	0,0	41	0,0	41	0,0
Interviews unvollständig realisiert	2.616	0,2	2.616	1,3	2.616	1,7

Quelle: Rücklaufdatenbank, eigene Berechnung

Die Interviews wurden mittels Dual-Frame-Stichprobe realisiert, das heißt es kamen sowohl Festnetz- als auch Mobilfunknummern zum Einsatz. In Tabelle 10 ist das Feldergebnis nach AAPOR-Klassifikation differenziert für die beiden Teilstichproben Festnetz und Mobilfunk der Basisstichprobe dargestellt. In der Mobilfunkstichprobe liegt der Anteil der Anschlüsse, der nicht zur Zielgruppe führte, 4,5 Prozentpunkte niedriger als in der Festnetzstichprobe. Der Anteil „unbekannte Auswahl" liegt dagegen in der Mobilfunkstichprobe deutlich höher als bei den Festnetznummern.

Bezogen auf die jeweiligen Einsatzbruttostichproben liegen die Ausschöpfungen zwar auf nahezu gleichem Niveau. Das Verhältnis von Interviews zu Verweigerungen unterscheidet sich jedoch erkennbar. Unter den Rufnummern, unter denen ein Haushalt bzw. eine Zielperson erreicht wurde und sie auch befragbar waren, haben in der Festnetzstichprobe 20,4 Prozent am Interview teilgenommen, bei der Mobilfunkstichprobe waren es mit 26,5 Prozent deutlich mehr.

Tabelle 10 Final Outcome nach AAPOR, differenziert nach Teilstichproben Festnetz und Mobilfunk

Final Outcome	**Festnetz**		**Mobilfunk**		**Insgesamt**	
Spalten%	*abs.*	*%*	*abs.*	*%*	*abs.*	*%*
Bruttostichprobe	1.183.165	100,0	499.140	100,0	1.682.305	100,0
Nicht Zielgruppe (NE)	865.323	73,1	342.910	68,7	1.208.233	71,8
Unbekannte Auswahl (UE)	172.067	14,5	105.684	21,2	277.751	16,5
Nonresponse – nicht befragbar (NR-NA)	373	0,0	8	0,0	381	0,0
Nonresponse – nicht erreicht (NR-NC)	28.516	2,4	16.735	3,4	45.251	2,7
Nonresponse – Verweigerung (NR-R)	93.030	7,9	24.822	5,0	117.852	7,0
Realisierte Interviews (I)	23.856	2,0	8.981	1,8	32.837	2,0

Quelle: Rücklaufdatenbank, eigene Berechnung

2.9.2 Onomastische Stichprobe

Wie bereits in Abschnitt 1.2.2 beschrieben, handelt es sich bei der onomastischen Stichprobe ausschließlich um eine Listenauswahl. Die Wahrscheinlichkeit, dass sich hinter einer Telefonnummer ein geschalteter Anschluss befindet, ist daher weitaus höher als bei vollständig generierten Rufnummern (vgl. Ausführungen zur Basisstichprobe weiter oben). Insgesamt sind 20.791 Rufnummern eingesetzt worden. Der Anteil Rufnummern, die zu keinem Anschluss führten, liegt mit rund 32 Prozent der Bruttoeinsatzstichprobe erwartungsgemäß deutlich unter dem Anteil in der Basisstichprobe (s. Tabelle 9 oben). Insgesamt führten rund 40 Prozent der Telefonnummern nicht zu einem Zielgruppenhaushalt. Der Anteil Rufnummern, bei denen nicht geklärt werden konnte, ob sich dahinter ein Zielhaushalt verbirgt, liegt bei rund 10 Prozent. In fast allen dieser 2.128 Fälle konnte die Zielgruppenzugehörigkeit nicht geklärt werden, weil der Telefonhörer umgehend wieder aufgelegt wurde, als sich der Interviewer vorstellte.

Bei 10.232 Rufnummern, und damit in rund der Hälfte der Fälle (rund 49 Prozent), konnte die Zielgruppenzugehörigkeit geklärt werden. In rund 22 Prozent dieser Fälle konnte die Zielperson bis Feldende nicht für ein Interview erreicht werden. In einem sehr geringen Anteil betraf dies auch definitiv abgestimmte Interviewtermine (Tabelle 9). Für 7.970 Rufnummern konnte geklärt werden, dass eine Zielperson im Haushalt lebt und befragbar ist. In 1.226 Fällen (rund 15 Prozent) ist es zu einem Interview gekommen. 1.012 Interviews sind vollständig und auswertbar, 214 sind vorzeitig beendet worden. Lediglich ein vollständiges Interview ist von der weiteren Auswertung aufgrund des Datenprüfungsergebnisses ausgeschlossen worden. Die Gründe, weshalb es nicht zu einem Interview kam, sind in der onomastischen Stichprobe ähnlich verteilt wie in der Basisstichprobe. Als am häufigsten genannte Verweigerungsgründe notierten die Interviewer „kein Interesse, wegen des Themas“ und eine grundsätzliche Teilnahmeverweigerung der Zielperson.[23].

[23] Auch für die onomastische Stichprobe wurde die Response-Rate 4 gemäß AAPOR-Standarddefinition berechnet. Der Anteil liegt bei 10,8 Prozent.

Tabelle 11 Feldergebnis der onomastischen Stichprobe nach AAPOR-Klassifikation

Spalten%	*abs.*	*%*	*abs.*	*%*	*abs.*	*%*
	I	**II**	**III**	**IV**	**V**	**VI**
Bruttostichprobe	**20.791**	**100,0**				
Nicht Zielgruppe (NE)	**8.431**	**40,6**				
Kein Anschluss	6.605	31,8				
Anschluss nur Fax/Modem	225	1,1				
Kein Privathaushalt unter diesem Anschluss	1.196	5,8				
Nicht in Zielgruppe lt. Angabe im Kontaktgespräch	121	0,6				
Bereits befragt	75	0,4				
Keine Verständigung in einer der beiden Zielsprachen möglich	209	1,0				
Unbekannte Auswahl (UE)	**2.128**	**10,2**				
Nicht abgehoben	4	0,0				
Anrufbeantworter	1	0,0				
Legt sofort auf	2.123	10,2				
Zugehörigkeit zur Zielgruppe bekannt	**10.232**	**49,2**	**10.232**	**100,0**		
Nonresponse – nicht befragbar (NR-NA)	**19**	**0,1**	**19**	**0,2**		
Zielperson (lt. Auskunft) nicht befragbar/ dauerhaft krank/behindert	19	0,1	19	0,2		
Nonresponse – nicht erreicht (NR-NC)	**2.243**	**10,8**	**2.243**	**21,9**		
Definitiver Termin	39	0,2	39	0,4		
Vager Termin	2.199	10,6	2.199	21,5		
Zielperson in Feldzeit nicht zu erreichen	5	0,0	5	0,0		
Zielperson/Haushalt erreicht und befragbar					**7.970**	**100,0**
Nonresponse – Verweigerung (NR-R)	**6.744**	**32,4**	**6.744**	**65,9**	**6.744**	**84,6**
Zielperson verweigert grundsätzlich	1.782	8,6	1.782	17,4	1.782	22,4
Zielperson verweigert: keine Zeit/dauert zu lange	682	3,3	682	6,7	682	8,6
Zielperson verweigert: will nur persönlich befragt werden	183	0,9	183	1,8	183	2,3
Zielperson verweigert krank	215	1,0	215	2,1	215	2,7
Kein Zugang zu Zielperson/Teilnahme untersagt/ lt. Auskunft nicht bereit	74	0,4	74	0,7	74	0,9
Kontaktperson verweigert jegliche Auskunft	1.028	4,9	1.028	10,0	1.028	12,9
Zielperson verweigert: kein Interesse, wegen des Thema	1.891	9,1	1.891	18,5	1.891	23,7
Zielperson verweigert: Datenschutzgründe/zu persönlich	98	0,5	98	1,0	98	1,2
Zielperson verweigert: sonstige Gründe	791	3,8	791	7,7	791	9,9
Realisierte Interviews(I)	**1.226**	**5,9**	**1.226**	**12,0**	**1.226**	**15,4**
Interviews vollständig realisiert, ausgeliefert	1.012	4,9	1.012	9,9	1.012	12,7
Interviews vollständig realisiert, nicht auswertbar	1	0,0	1	0,0	1	0,0
Interviews unvollständig realisiert	213	1,0	213	2,1	213	2,7

Quelle: Rücklaufdatenbank, eigene Berechnung

2.10 Interviewdauer

Die durchschnittliche Dauer der CATI-Interviews beträgt für die n=31.048 auswertbaren Interviews mit gültigen Zeitstempeln 22,1 Minuten.[24] Angezielt wurde eine durchschnittliche Dauer von 21 Minuten, allerdings ohne Berücksichtigung der Dauer von in 2017 neu aufgenommenen Fragen zur Erreichbarkeit der Befragungsperson.[25] Die durchschnittliche Dauer dieser Erreichbarkeitsfragen betrug rund 1 Minute. Damit wurde die angezielte Durchschnittsdauer des übrigen Fragebogens also sehr gut erreicht.

Für die Basisstichprobe liegt der Wert bei 22,0 Minuten (vgl. Tabelle 12). Die fremdsprachigen Interviews dauerten meist länger als die deutschsprachigen. Dies entspricht den Erfahrungen aus dem Jahr 2012 sowie den Pretests 2012 und 2017. Für die türkischsprachigen Interviews der Basisstichprobe wurden im Durchschnitt 28,4 Minuten benötigt, bei den russischsprachigen Interviews liegt der Wert bei 27,5 Minuten.

Die durchschnittliche Interviewdauer lag bei der onomastischen Stichprobe über alle Interviews mit gültiger Zeitmessung bei 26,1 Minuten (vgl. Tabelle 12).

Tabelle 12 Interviewdauer

	Interviews	**Durchschnittsdauer**
	abs.	*in Minuten*
Basis	30.048	22,0
– Davon auf Türkisch	86	28,4
– Davon auf Russisch	267	27,5
Onomastik	1.000	26,1
Gesamt	31.048	22,1

Quelle: Zeitstempel (time stamps) im Fragebogen, Fälle mit gültigen Angaben, eigene Berechnung

2.11 Interviewsprache

Analog zu 2012 wurde auch in der zweiten Erhebung des Deutschen Viktimisierungssurveys Personen, die nicht ausreichend Deutsch verstehen und/oder sprechen, angeboten, das Interview in Türkisch oder Russisch zu führen. Insgesamt wurden 88 Interviews in der Basisstichprobe auf Türkisch und 271 Interviews auf Russisch geführt (vgl. Tabelle 13). Ohne die Vorhaltung der beiden Fremdsprachen wären diese 359 Interviews mit Personen, die der deutschen Sprache nicht (ausreichend) mächtig sind, nicht möglich gewesen. In der Summe wurden 1,2 Prozent der 30.180 Interviews der Basisstichprobe nicht in deutscher Sprache geführt.

[24] Inklusive Ermittlung der reduzierten Haushaltsgröße im Kontaktgespräch und Gewinnung der Zielperson.

[25] Diese Fragen zur Erreichbarkeit wurden in einem neueren Gewichtungsansatz verwendet, den infas in Rücksprache mit dem Auftraggeber mit den Daten des DVS 2017 zusätzlich berechnet hat (vgl. dazu Kapitel 3.3.2). Die Dauer dieser Fragen wurde bei der angezielten Durchschnittsdauer daher nicht berücksichtigt.

Tabelle 13 Basisstichprobe: Interviewsprache

Interviewsprache		
Spalten%	*abs.*	*%*
Deutsch	29.821	98,8
Türkisch	88	0,3
Russisch	271	0,9
Interviews gesamt	30.180	100,0

Quelle: Stichprobeninformation

In der onomastische Stichprobe wurde das Interview deutlich häufiger auf Türkisch geführt. Dies verwundert nicht, denn diese Zusatzstichprobe umfasste mit hoher Wahrscheinlichkeit Haushalte, in denen Personen mit türkischem Migrationshintergrund leben. Das Interview sollte unabhängig von den Deutschkenntnissen im Haushalt (insbesondere der Zielperson) durchgeführt werden. Rund 48 Prozent der Interviews wurden vollständig und rund 10 Prozent zum Teil auf Türkisch geführt (vgl. Tabelle 14).

Tabelle 14 Onomastische Stichprobe: Interviewsprache

Interviewsprache		
Spalten%	*abs.*	*%*
Deutsch	381	41,8
Türkisch	438	48,1
Teils/teils	92	10,1
Interviews gesamt	911	100,0

Quelle: Angaben der Interviewer, Fälle mit gültigen Angaben zur Interviewsprache

2.12 Kontakthäufigkeit

Die beiden Stichproben wurden über die gesamte Feldzeit kontinuierlich kontaktiert, um auch schwer erreichbaren Zielpersonen die Möglichkeit zu geben, an einem Interview teilzunehmen. Für Folgekontakte wurden in allen Fällen, bei denen niemand erreicht wurde, die Uhrzeit und der Wochentag variiert. Mit diesem Vorgehen sollten möglichst viele Personen erreicht und um Teilnahme geworben werden. Das Feldergebnis konnte nur durch diese konsequente Kontaktierungsstrategie erreicht werden.

Um ein Interview zu realisieren, waren im Durchschnitt knapp 4,5 Kontakte notwendig (Tabelle 15). Es kam durchaus vor, dass das Interview bereits beim Erstkontakt geführt werden konnte. Im Maximum benötigten die Interviewer jedoch bis zu 27 Kontakte, um ein Interview mit der Zielperson zu führen (vgl. Tabelle 15).

Lässt man bei der Betrachtung der Kontaktanzahl für jene Fälle, bei denen es zu keinem Interview kam, die nicht geschalteten Rufnummern unberücksichtigt, dann wird der Umfang der Bemühungen um ein Interview deutlich. Bei den nicht realisierten Interviews (ohne nicht geschaltete Nummern) liegt die durchschnittliche Kontakthäufigkeit bei gut 5 Kontakten und einem Maximalwert von 33 Kontakten spürbar höher als bei den realisierten Interviews.

Tabelle 15 Kontakthäufigkeit der auswertbaren Interviews

	abs.
Durchschnittliche Kontakthäufigkeit	4,4
Minimum Kontakthäufigkeit	1
Maximum Kontakthäufigkeit	27
Standardabweichung	3,32
Fallzahl	31.192

Quelle: Kontaktdatei, eigene Berechnungen

2.13 Fallzahl pro Fragenmodul

Im gesamten Fragebogen kamen 5 inhaltlich verschiedene Fragenmodule zum Einsatz. Sie sollten nur einer zufällig ausgewählten Teilstichprobe vorgelegt werden, damit die durchschnittliche Dauer der Befragung im vereinbarten Rahmen blieb. Für die spätere Auswertung der Antworten wurde eine Fallzahl pro Modul zwischen 1.500 und 2.200 als ausreichend erachtet (siehe Spalte „angezielte Fallzahl“ in Tabelle 16).

Zu beachten ist dabei, dass die Module 1 und 3 immer in denselben Interviews geschaltet wurden; d.h. eine Person, die im Interview die Fragen des Moduls 1 erhielt, durchlief immer auch Modul 3. Die übrigen Module 2 und 4 wurden alternativ eingesetzt, so dass es hier zu keinen Überschneidungen untereinander oder mit Modul 1 und 3 kam. Die Fragen von Modul 5 wurden in allen Interviews gestellt, in denen eines der übrigen Module 1 bis 4 geschalten war. Die Fragen der Module 1 bis 4 können also jeweils zusammen mit den Fragen des Moduls 5 ausgewertet werden.

Außerdem enthält der Fragebogen einige Fragen zur Arbeit der Justiz in Deutschland. Diese Fragen sollten an insgesamt 10.000 Personen gestellt werden, und zwar

- an alle 2.000 Befragten, die das Modul 2 durchliefen, sowie
- an weitere 8.000 Befragte, die zufällig aus denjenigen Befragten ausgewählt wurden, die keines der 5 Module erhalten.

In der nachfolgenden Tabelle sind die angezielten und die tatsächlich erreichten Fallzahlen pro Modul und für die Zusatzfragen zur Justiz ausgewiesen. Wie dort zu sehen ist, konnten die angezielten Fallzahlen in jedem Modul erreicht bzw. übertroffen werden.

Tabelle 16 Fallzahl pro Modul in der Basisstichprobe

Modulnummer	Realisierte Fallzahl	Anteil an Basisstichprobe	Angezielte Fallzahl
Spalten%	*abs.*	*%*	*abs.*
1*	1.625	5,4	1.500
2	2.061	6,8	2.000
3*	1.625	5,4	1.500
4**	2.204 1.861	7,3 6,2	2.200
5	5.890 5.547***	19,5 18,4	5.700
Zusatzfragen Justiz	10.322	34,2	10.000****

Quelle: auswertbare Interviews, eigene Berechnung

* Modul 1 und 3 wurden immer in denselben Interviews geschaltet.
** Modul 4 startet mit einer Filterfrage (Frage 607, Internetnutzung). In der ersten Zeile werden die Fälle ausgewiesen, die die Filterfrage bekommen haben. In der zweiten Zeile sind jene Fälle ausgewiesen, die auch die Folgefragen erhalten haben (d.h. nicht überfiltert wurden).
*** Modul 5 wurde immer dann geschaltet, wenn eines der Module 1 bis 4 geschaltet war. Für 5.548 Fälle kann Modul 5 zusammen mit einem der Module 1 bis 4 ausgewertet werden. Die Differenz zu den angegebenen 5.890 Fällen ergibt sich aufgrund der Filterung zu Beginn von Modul 4. Vergleiche dazu die Anmerkungen zu Modul 4.
**** Davon 2.000 im Modul 2 und 8.000 weitere, die keines der Modul erhalten.

Die Gesamtfallzahl pro Modul sowie die Verteilung auf Basis- und onomastische Stichprobe sind in Tabelle 17 ausgewiesen. Da der Umfang der Basisstichprobe deutlich größer ist als der der onomastischen Stichprobe, entfallen pro Modul rund 97 Prozent der Fälle auf die Basisstichprobe und rund 3 Prozent auf die onomastische Stichprobe.

Tabelle 17 Fallzahl pro Modul insgesamt

Modulnummer	Realisierte Fallzahl in der Basisstichprobe		Realisierte Fallzahl in der onomastischen Stichprobe		Realisierte Fallzahl insgesamt	
Zeilen%	*abs.*	*%*	*abs.*	*%*	*abs.*	*%*
1	1.625	96,9	52	3,1	1.677	100,0
2	2.061	97,1	61	2,9	2.122	100,0
3	1.625	96,9	52	3,2	1.677	100,0
4	2.204	96,7	76	3,3	2.280	100,0
5	5.890	96,9	189	3,1	6.079	100,0
Zusatzfragen Justiz	10.322	97,0	321	3,0	10.643	100,0

Quelle: auswertbare Interviews, eigene Berechnung

2.14 Einsatz der Vignettensets

In Modul 2 wurden unterschiedliche Vignettensets eingesetzt (vgl. Abschnitt 1.3). Die Zuordnung der Vignettensets zu den Interviews wurde über die Bruttostichprobe gesteuert, so dass eine zufällige Einspielung der Vignettensets gewährleistet war. Die Interviewer hatten weder einen Einfluss darauf, ob ein Set überhaupt zum Einsatz kam, noch welches Vignettenset eingespielt wurde. In den auswertbaren Interviews der Basisstichprobe[26] liegen zwischen 275 und 317 eingesetzte Sets vor (vgl. Tabelle 18). Insgesamt summieren sich die beantworteten Vignettensets auf 2.061 – damit 61 mehr, als ursprünglich für die Basisstichprobe geplant. Der Anteil der einzelnen eingesetzten Sets an allen Vignettensets variierte zwischen 13,3 Prozent bei Vignettenset 2 und 15,4 Prozent für Vignettenset 6. Das Ziel einer annähernd gleichen Verteilung über die 7 Vignettensets ist erreicht worden.

Tabelle 18 Vignettensetverteilung für die Basisstichprobe

Vignettenset	Realisierte Fallzahl pro Vignettenset	Anteil an allen Vignettensets
Spalten%	*abs.*	*%*
Set 1	282	13,7
Set 2	275	13,3
Set 3	302	14,7
Set 4	289	14,0
Set 5	302	14,7
Set 6	317	15,4
Set 7	294	14,3
Gesamt	2.061	100,0

Quelle: auswertbare Interviews

2.15 Verhältnis auswertbare Interviews aus Festnetz- und Mobilfunkstichprobe

Für die Basisstichprobe war ein Verhältnis der Interviews aus Festnetzstichprobe und Mobilfunkstichprobe von 75:25 vorgesehen. Die Einsatzstichproben der Basisstichprobe wurden in einem Verhältnis von 70:30 ins Feld gegeben (vgl. Abschnitt 1.2.1). Mit 7.958 Interviews konnten 458 Interviews mehr als geplant in der Mobilfunkstichprobe realisiert werden (vgl. Tabelle 19). Der Anteil liegt bei 26,4 Prozent der realisierten Basisstichprobe. Damit konnte das angezielte Verhältnis von 75:25 in den realisierten Interviews also nahezu erreicht werden.

[26] Die nachfolgende Betrachtung der Verteilung der Vignettensets beschränkt sich auf die Basisstichprobe, da für diese Teilstichprobe in 2.000 Fällen Vignetten eingesetzt werden sollten. Da in der onomastischen Stichprobe weitere 61 Vignettensets zum Einsatz kamen, erhöht sich die Anzahl der insgesamt vorliegenden Vignettensets entsprechend.

Tabelle 19 Festnetz- und Mobilfunkstichprobe in der Basisstichprobe

Festnetz- und Mobilfunkstichprobe		
Spalten%	*abs.*	*%*
Festnetz	22.222	73,6
Mobilfunk	7.958	26,4
Gesamt	30.180	100,0

Quelle: auswertbare Interviews

Tabelle 20 Festnetz- und Mobilfunkstichprobe insgesamt

Festnetz- und Mobilfunkstichprobe		
Spalten%	*abs.*	*%*
Festnetz	23.120	74,1
Mobilfunk	8.072	25,9
Gesamt	31.192	100,0

Quelle: auswertbare Interviews

Die Listenauswahl für die Onomastikstichprobe umfasste mehrheitlich Festnetznummern, daneben aber auch einige Mobilfunknummern. In Tabelle 20 sind die 898 Festnetzinterviews sowie die 114 Fälle mit Mobilfunknummer der Onomastikstichprobe ergänzt. Die Gesamtstichprobe enthält somit rund 26 Prozent Mobilfunknummern (Tabelle 20).

2.16 Interviewabbrüche und Abbruchgründe

Über die beiden Teilstichproben hinweg kam es in 2.829 Fällen zu einem Interviewabbruch.[27] Ein Teil der Abbrüche fand bereits im Kontaktgespräch statt, während der Interviewer versuchte, die richtige Zielperson im Haushalt zu ermitteln und sie für eine Teilnahme an der Befragung zu gewinnen. Die restlichen 2.110 Abbrüche ereigneten sich während des Interviews. 184 Abbrüche erfolgten in der onomastischen Stichprobe und 1.926 in der Basisstichprobe.

[27] Darunter sind sowohl Interviews gefasst, die von der Befragungsperson explizit abgebrochen wurden, als auch Interviews, die während der Feldzeit zunächst als unterbrochenes Interview abgelegt wurden, bis zur Ende der Feldzeit aber nicht beendet werden konnten.

Im Folgenden wird dargelegt,

a. inwieweit sich die Interviewabbrüche an bestimmten Stellen im Fragebogen häufen oder sie sich eher unspezifisch über den gesamten Bogen verteilen,

b. inwieweit es Abbruchgründe gibt, die besonders häufig genannt wurden und/oder möglicherweise mit Besonderheiten des Fragebogens zusammenhängen,

c. ob in diesem Zusammenhang Unterschiede zwischen der Basisstichprobe auf der einen Seite und der onomastischen Stichprobe auf der anderen Seite zu finden sind.

Die Abbruchquoten (Anteil der abgebrochenen Interviews an der Gesamtzahl vollständig realisierter Interviews plus abgebrochene Interviews) unterscheiden sich zwischen der Basisstichprobe und der onomastischen Stichprobe spürbar. Während die Abbruchquote in der Basisstichprobe bei rund 8 Prozent liegt[28], bewegt sie sich in der onomastischen Stichprobe mit gut 17 Prozent deutlich darüber.

Die mit Abstand meisten Interviewabbrüche ereignen sich zu Beginn des Interviews bei den Fragen zum Alter sowie zu Lebenszufriedenheit und Vertrauen (Fragenblock 100, vgl. Tabelle 21). In der Basisstichprobe entfallen gut 30 Prozent aller Abbrüche auf die ersten Fragen nach dem Interviewstart inklusive Block 100. Bei der onomastischen Stichprobe liegt der Anteil für die ersten Fragen bis einschließlich Block 100 mit rund 42 Prozent noch etwas höher (vgl. Tabelle 21). Das ist gleichzeitig auch der größte Unterschied in der Verteilung der Abbrüche über das Interview zwischen den beiden Teilstichproben. Im weiteren Verlauf gibt es keine nennenswerten Unterschiede in der Verteilung der Abbruchstellen über das Interview.

[28] Die Abbruchquoten unterscheiden sich auch in den beiden Teilstichproben der Basisstichprobe. In der Festnetzstichprobe liegt sie bei knapp 7 Prozent, in der Mobilfunkstichprobe bei gut 11 Prozent.

Die im Vergleich zu anderen Stellen im Fragebogen hohe Abbruchquote zu Beginn des inhaltlichen Interviews lässt sich auf unterschiedliche Faktoren zurückführen. Dies zeigen die von den Interviewern notierten Erläuterungen zum Abbruchgrund. Sie geben wichtige Hinweise auf die Gesprächssituation und eine Begründung für das vorzeitige Ende der Befragung.

Der Start der Befragung spielt eine sehr entscheidende Rolle für die Realisierungschancen sowie den gesamten Verlauf eines Interviews. Zu Beginn muss der Interviewer ein gewisses Maß an Vertrauen aufbauen, damit die Zielperson bereit ist, auf die vielfach persönlichen Fragen zu antworten. Insbesondere bei zweifelnden Personen, die zur Nichtteilnahme tendieren, schlägt der Interviewer oftmals vor, erst einmal mit dem Interview zu beginnen, damit sich die Zielperson ein besseres Bild über die Inhalte der Befragung machen kann. Zum Teil gelingt es dann, das Interview bis zum Ende durchzuführen. In anderen Fällen entscheiden sich die Zielpersonen nach den ersten Fragen doch gegen eine Fortführung des Interviews. Zudem kann in einigen Fällen erst nach den ersten inhaltlichen Fragen festgestellt werden, ob die Zielperson dem Interview geistig, sprachlich oder trotz eingeschränkter Gesundheit überhaupt folgen kann. So stehen am Anfang eines jeden Interviews die Fragen nach der Befragbarkeit der Zielperson und ihrem Interesse, an der Befragung teilzunehmen bzw. die Akzeptanz der Fragen auf Seiten der Zielperson im Raum.

Neben den vermehrten Abbrüchen zu Beginn der Befragung verteilen sich die übrigen Abbruchstellen über den gesamten Fragebogen. Sie schwanken meist zwischen 0,1 und 5 Prozent der Abbrüche und erreichen nur selten einen Anteil von mehr als 7 Prozent. Erhöhte Abbruchquoten finden sich in beiden Teilstichproben bei den Fragen zur Wohngegend (rund 8 Prozent in beiden Teilstichproben). Etwas höhere Abbruchquoten zeigen sich auch in Modul 2 und hier insbesondere bei den Zusatzfragen zur Justiz (rund 5 Prozent in beiden Teilstichproben). Beide Fragenblöcke liegen auch recht weit vorn im Interview.

Jeweils 5 bis 6 Prozent der Abbrüche entfallen für die Basisstichprobe auf das Modul 5 zur deliktspezifischen Kriminalitätsfurcht und in beiden Stichproben auf den Block 800 zum Besitz von Zahlungskarten und auf die Detailfragen zur Viktimisierung.

Die Durchsicht der Interviewerangaben zu den Abbruchgründen ergab eine Vielschichtigkeit von Gründen für ein vorzeitiges Ende des Interviews. Insbesondere wurde zu Beginn des Interviews deutlich, dass die Zielperson aus alters- oder gesundheitlichen Gründen dem Interview nicht folgen konnte und es dadurch zu einem Abbruch kam. Im gesamten Interviewverlauf kam es auch mehrfach vor, dass die Zielperson einfach auflegte, ohne dem Interviewer einen Grund für das abrupte Ende des Gesprächs zu nennen. Der Versuch, das Telefonat erneut aufzunehmen, scheiterte in diesen Fällen. Zudem entschieden sich Zielpersonen nach den ersten inhaltlichen Fragen gegen die Fortführung des Interviews, weil ihnen das Thema nicht zusagte, ihnen das Interview zu lange dauerte, sie keine Zeit oder allgemein kein besonderes Interesse an der Interviewteilnahme hatten. Das waren sowohl in der Basisstichprobe als auch in der onomastischen Stichprobe die weitaus häufigsten Gründe für eine vorzeitige Beendigung des Interviews.

Zusammenfassend lässt sich feststellen, dass gehäuft Abbrüche zu Beginn des Interviews stattgefunden haben. An dieser Stelle des Interviews musste oftmals erst geklärt werden, ob die Zielperson tatsächlich befragbar ist bzw. ob die Befragung auf Akzeptanz bei der Zielperson stößt. Die Anzahl an Abbrüchen variiert auch im weiteren Interview zwischen den Fragenblöcken. Die breite Streuung der Abbrüche über den gesamten Fragebogen und die von den Interviewern notierten Gründe für das vorzeitige Ende der Befragung deuten darauf hin, dass es keine Stelle im Fragebogen gibt, die aufgrund der Fragestellung zu einer deutlich höheren Abbruchquote geführt hat.

Tabelle 21 Abbruchstellen im Fragebogen

Fragebogenabschnitt	Basisstichprobe		Onomastikstichprobe	
Spalten%	*abs.*	*%*	*abs.*	*%*
Frage 1-1a (Alter)	253	13,1	28	15,2
Block 100: Lebenszufriedenheit, Vertrauen	394	20,5	49	26,6
Modul 2: Kriminalitätsbezogene Einstellungen	29	1,5	3	1,6
Modul 2: Zusatzfragen zur Justiz	102	5,3	9	4,9
Modul 2: Vignetten *(Frage 308-320)*	66	3,4	7	3,8
Modul 1: Einstellungen zur Zuwanderung	11	0,6	-	-
Allgemeiner u. beruflicher Bildungsabschluss/Erwerbsstatus	93	4,8	8	4,3
Wohngegend	160	8,3	14	7,6
Modul 3: Wohngegend, Disorder/Incivilities	11	0,6	1	0,5
Ausgehverhalten	22	1,1	1	0,5
Familienstand	67	3,5	6	3,3
Internetnutzung	20	1,0	2	1,1
Modul 4: Sicherheitsgefühl bei Internetnutzung	6	0,3	4	2,2
Kriminalitätsfurcht	22	1,1	5	2,7
Modul 5: Deliktspezifische Kriminalitätsfurcht	96	5,0	4	2,2
Block 800: Besitz von Fahrzeugen	-	-	-	-
Block 800 KB: Besitz von Zahlungskarten	129	6,7	11	6,0
Screener – Block 900 Wohnungseinbruch	17	0,9	1	0,5
Screener – Block 900 Versuchter Wohnungseinbruch	14	0,7	1	0,5
Screener – Block 900 Fahrzeugdiebstahl	-	-	-	-
Screener – Block 900 Konsumentenbetrug	17	0,9	1	0,5
Screener – Block 900 Kreditkartenbetrug	14	0,7	2	1,1
Screener – Block 900 Viktimisierung bei Internetnutzung	14	0,7	1	0,5
Screener – Block 900 Raub	45	2,3	3	1,6
Screener – Block 900 Sonstiger Diebstahl	26	1,3	4	2,2
Screener – Block 900 Körperverletzung	72	3,7	3	1,6
Viktimisierungsblöcke (alle Delikte)	118	6,1	11	6,0
Soziodemografie	66	3,4	2	1,1
Ausstattung Telekommunikationsmedien	42	2,2	3	1,6
Erreichbarkeit	-	-	-	-
Postleitzahl	-	-	-	-
Gesamt	1.926	100,0	184	100,0

Quelle: abgebrochene Interviews, eigene Berechnungen

3 Datenprüfung und –gewichtung

3.1 Datenprüfung und –aufbereitung

Die ersten Schritte der Datenprüfung begannen bereits vor der eigentlichen Feldphase. Nachdem der abgestimmte Fragebogen programmiert war, wurde die korrekte Ablage der Daten geprüft. Dabei wurde der gesamte Filterverlauf im programmierten Instrument nachvollzogen. Für alle Fragen wurde zudem auf die Einhaltung des vorgesehenen Wertebereichs geachtet, damit bei geschlossenen Skalen lediglich die in der Fragebogenvorlage vorgesehen Codes auftreten können. Bei offenen Zahlenangaben durften die erhobenen Werte nur im vorgegebenen Wertebereich auftreten.

Auch die Richtigkeit der Steuerung einzelner Fragenmodule, die kombinierte Schaltung der Module (sofern vorgesehen) und die Zuweisung der Vignetten in Modul 2 wurden in der Datenprüfung nachvollzogen und auf ihre Regelkonformität hin geprüft. Diese intensiven Prüfungen der korrekten Datenablage gewährleisten eine fehlerfreie technische Erhebung der Daten im Feld. Diese Prüfungen erfolgten vor dem Pretest sowie vor dem Feldstart der Haupterhebung anhand von speziell dafür eingegebenen Testfällen. Die Prüfungen wurden sowohl nach dem Pretest als auch zu Beginn des Hauptfelds anhand der realisierten Fälle fortgesetzt.

Nach Feldende wurden die 31.234 vollständig realisierten Interviews auf ihre Gültigkeit geprüft. Prüfkriterien waren die Dauer des Interviews, der Anteil an Missing-Werten (weiß nicht/verweigert), der Anteil an Missing-Werten in den Screening-Fragen zu erlebten Delikten sowie eine Sichtung der offenen Nennungen. Als Ergebnis dieser Prüfungen mussten insgesamt nur 42 Interviews als nicht auswertbar deklariert werden, davon 41 Interview der Basis- und ein einziger Fall aus der onomastischen Stichprobe.

In einem weiteren Schritt wurden die Häufigkeitsauszählungen durchgesehen. Bei den Fragen nach offenen Zahlenangaben (z.B. Anzahl Haushaltsmitglieder, Häufigkeit der einzelnen Opfererfahrungen, Anzahl Telefonnummern im Haushalt) kam es in sehr wenigen Fällen zu vergleichsweise hohen Angaben. Nach Rücksprache mit dem BKA werden diese Angaben im Rahmen der Auswertung im BKA auf ihre Plausibilität hin geprüft und bei Bedarf bereinigt.

Bei den Datenbereinigungen wurde – soweit möglich – analog zum Viktimisierungssurvey 2012 vorgegangen. An den im Folgenden beschriebenen Variablen sind in enger Abstimmung mit dem BKA Recodierung vorgenommen worden. Die Ursprungsangabe der Zielperson ist in allen Fällen in der Originalvariable enthalten geblieben, so dass es zu keinem Informationsverlust gekommen ist.

Angabe zum Geburtsland

In den Fragen zum Geburtsland der Zielperson (SD12/SD13), zum Geburtsland der Mutter (SD15/SD16) und zum Geburtsland des Vaters der Zielperson (SD18/SD19) haben die Zielpersonen in einigen wenigen Fällen erst bei der offenen Nachfrage zum Geburtsland angegeben, dass ein Elternteil in Deutschland geboren wurde. Für diese Fälle wurden die Korrekturvariablen SD12k/SD13k,

SD15k/SD16k sowie SD18k/SD19k angelegt. In der Eingangsfrage SD12k (bzw. SD15k oder SD18k) wurden diese Fälle auf Code 1 „Bundesrepublik Deutschland oder DDR“ umgesetzt. Die Angabe aus der Nachfrage SD13k (bzw. SD16k oder SD19k) wurde gelöscht. Die Originalvariablen SD12, SD13, SD15, SD16, SD18 und SD19 blieben hingegen unverändert. Die betroffenen Fälle können also durch den Abgleich von Original- und Korrekturvariable nachvollzogen werden. Die kategorisierten Variablen zum Geburtsland SD12t, SD15t und SD18t basieren auf den Korrekturvariablen.

Recodierung „Tatort“

Für die Fälle, in denen die Zielperson nicht angeben konnte, in welchem Bundesland sie Opfer einer Straftat geworden war, wurde mit einer offenen Frage die Ortschaft oder die Stadt ermittelt. Diese Ortsangaben konnten im Rahmen der Datenaufbereitung eindeutig einzelnen Bundesländern zugeordnet werden (vgl. Tabelle 22). Nach Rücksprache mit dem BKA wurden die Recodierungen im Datensatz vorgenommen. Auch diese Umsetzungen können in den Daten nachvollzogen werden. Das nachfolgende Beispiel illustriert die Vorgehensweise, die bei allen entsprechenden Fragen analog angewandt wurde:

Wenn einer Zielperson lediglich einmal das Fahrrad gestohlen wurde und sie in Frage FD6 das Bundesland nicht angegeben hatte (Code 97 „verweigert“/ 98 „weiß nicht“), dann wurde ihr die Nachfrage zur Stadt bzw. Ortschaft des Vorfalls (hier FD7) gestellt. Sofern die Zielperson hier einen Ort nennen konnte, ist in FD7 der Code 1 „offene Nennung erfassen“ abgelegt. Es gilt also allgemein, dass die Nachfrage zur Stadt bzw. Ortschaft nur gestellt wird, wenn in der vorherigen Abfrage des Bundeslands einer der Codes 97 „verweigert“ oder 98 „weiß nicht“ vorlag. Bei den nachcodierten Fällen wird hiervon abgewichen. Bei der Nachcodierung wurde in der Frage nach dem Bundesland von den Codes 97 oder 98 auf einen der Codes 1 bis 16 recodiert. Die Nachfrage nach der Stadt bzw. Ortschaft wurde in diesen Fällen nicht gelöscht. Nachcodierte Angaben zum Bundesland, in dem sich ein Vorfall ereignet hat, lassen sich also allgemein daran erkennen, dass in der Frage zum Bundesland einer der Antwortcodes 1 bis 16 und in der Nachfrage zur Stadt bzw. Ortschaft Code 1 „offene Nennung erfassen“ vorliegt.

Bei der Nachfrage zur Stadt bzw. Ortschaft beim Delikt „Waren- oder Leistungsbetrug“ (nur ein berichteter Vorfall, Variable B9) wurde in mehreren Fällen angegeben, dass diese Frage so nicht beantwortet werden könne, weil sich der Vorfall über das Internet ereignet habe. Diese Fälle wurden in Variable B9 mit dem Sondercode 2 „offene Nennung: über das Internet“ gekennzeichnet.

Tabelle 22 Vercodung der offenen Ortsangaben

Ortsangabe aus dem Interview	Vercodetes Bundesland
Hamburg	Hamburg
Frankfurt	Hessen
Itzingen	Baden-Württemberg
Landsberg am Lech	Bayern
Hannover	Niedersachsen
Halle	Sachsen-Anhalt
Voerde	Nordrhein-Westfalen
Nürnberg	Bayern
Düsseldorf	Nordrhein-Westfalen
Berlin	Berlin
Bad Staffelstein	Bayern

Quelle: Rohdaten, Datenaufbereitung.

Negative Missing-Codes

In allen Fragen, bei denen Länderlisten hinterlegt waren, um eine schnelle und fehlerfreie Erfassung der Nennungen zu gewährleisten, wurden die Missing-Codes für „weiß nicht“ (98) und „verweigert“ (97) mit einem negativen Vorzeichen versehen. Dies war notwendig, da die Codes 97 und 98 zu den gültigen Codes der Staaten und Territorien in der Länderliste zählen. Die Codierung mit „-97“ bzw. „-98“ betrifft die Variablen SD11, SD11b, SD13, SD16 und SD19.

Darüber hinaus wird bei der Fragen zur Staatsangehörigkeit SD11 ein weiterer Sondercode verwendet. Es handelt sich dabei um die Codes „-96“ für „ohne Staatsangehörigkeit“.

Abgleich Zuzugsjahr mit Geburtsjahr

Am Ende des Fragebogens wurden die Zielpersonen, die nicht in Deutschland geboren wurden, gefragt, in welchem Jahr sie nach Deutschland (heutige Gebiet der Bundesrepublik) gezogen sind (Frage SD14). Es wurde eine Plausibilitätsprüfung vollzogen, indem geprüft wurde, ob das Zuzugsjahr vor oder nach dem Geburtsjahr der Zielperson liegt. Da für die Zielpersonen nur eine Angabe zum Alter und nicht zum Geburtsjahr vorliegt, wurde eine entsprechende Umrechnung aufgrund der Angabe zum Alter vorgenommen. Der Einfachheit halber wurde bei allen Zielpersonen unterstellt, dass diese im Befragungsjahr noch nicht Geburtstag hatten. Für eine Zielperson, die ihr Alter mit 20 Jahren angegeben hatte, wurde beispielsweise das Geburtsjahr 1997 angenommen. Sofern nur eine Angabe der Altersgruppe aus Frage 1a vorliegt, wurde ein Alter am oberen Ende der Skala für die Prüfung berücksichtigt.[29] Für Personen mit der Altersgruppenanga-

[29] Für eine Zielperson, die ihr Alter beispielsweise mit „zwischen 25 und 34 Jahre“ angegeben hat, wurde ein Alter von 34 Jahren und somit das Geburtsjahr 1982 angenommen.

be „75 Jahre und älter“ wurde auf einen Abgleich der Altersangabe mit der Antwort aus Frage SD14 verzichtet. Unter den genannten Rahmenbedingungen ergab sich lediglich für 18 Fälle eine Situation, bei der das Zuzugsjahr *vor* dem errechneten Geburtsjahr lag. Die Entscheidung über den Umgang mit diesen wenigen Fälle obliegt dem BKA. Daher wurden keine Änderungen im Datensatz vorgenommen.

Zuspielung der Kreiskennziffer (KKZ)

Um regionale Einflüsse bei der Datenanalyse berücksichtigen zu können, wurde den Befragungsdaten die Kreiskennziffer (KKZ) – sofern möglich – zugespielt. Das im Folgenden beschriebene Verfahren war sehr erfolgreich. Es konnte für 96,7 Prozent der Interviews eine KKZ und zugleich auch eine Bundeslandkennung zugespielt werden. Die KKZ wurde für alle Fälle aus den Festnetzstichproben über die Stichprobe anhand der Vorwahl zugespielt. Für die Fälle der Mobilfunkstichprobe hingegen liegt in der Stichprobe keine Information zur KKZ vor. Für die realisierten Fälle der Mobilfunkstichprobe wurde die KKZ aus der am Ende der Befragung von der Zielperson genannten Postleitzahl ihres Wohnorts generiert. Voraussetzung hierfür war, dass die Zielperson alle 5 Stellen der Postleitzahl angegeben hatte. Erfahrungsgemäß sind nicht alle genannten Postleitzahlen korrekt. Ein Grund für die Nennung von nicht existenten Postleitzahlen können Erinnerungsfehler der Zielperson sein. Es kann sich aber auch um verdeckte Verweigerungen der Angabe handeln, weshalb fiktive Postleitzahlen genannt werden.

3.2 Gewichtung und Hochrechnung

Die Berechnung der Gewichte und Hochrechnungsfaktoren erfolgte in einem mehrstufigen Prozess. Bei der Berechnung der Gewichte mussten unterschiedliche Auswahlrahmen (Frames) berücksichtigt und in einen gemeinsamen Gewichtungsfaktor integriert werden. Grundlegend war die Entscheidung, die ADM-Stichprobe, bestehend aus einer Festnetzstichprobe und einer Mobilfunkstichprobe, sowie die onomastische Stichprobe von (mehrheitlich) Personen mit türkischem Migrationshintergrund in eine Gesamtstichprobe zu integrieren. Es wurden unterschiedliche Gewichtungsvarianten berechnet.

Die Auswahlrahmen waren:

- ADM-Telefonnummerndatei Festnetz gesamt,
- ADM-Telefonnummerndatei Mobilfunk,
- Telefonbucheinträge (Festnetz und Mobilfunkeinträge) nach onomastischen Gesichtspunkten für Personen mit türkischem Migrationshintergrund.[30]

[30] Die Stichprobe wurde gezogen von Humpert & Schneiderheinze GbR.

Zudem erfolgte eine Umgewichtung der Stichproben für Personen und Haushalte. Während es sich bei einer Festnetzstichprobe um eine Haushaltsstichprobe handelt, lässt sich eine Mobilfunkstichprobe als Personenstichprobe auffassen. Die meisten Fragen im Fragebogen bezogen sich auf die befragte Zielperson. Es wurden darüber hinaus aber auch haushaltsbezogene Fragen gestellt. Die personenbezogenen Fragen sollten mit dem Personengewicht ausgewertet und mit dem Personenhochrechnungsfaktor auf alle Personen in der Grundgesamtheit hochgerechnet werden. Die (wenigen) haushaltsbezogenen Fragen (z.B. Fahrrad- oder Kfz-Diebstahl) sollten mit dem Haushaltsgewicht ausgewertet und dem Haushaltshochrechnungsfaktor auf die Grundgesamtheit aller Haushalte (mit mindestens einem 16-jährigen oder älteren Haushaltsmitglied) hochgerechnet werden. Alle Gewichtungsschritte und die Berechnung der Hochrechnungsfaktoren erfolgten daher getrennt für Personen und Haushalte.

Folgende Arbeitsschritte wurden bei der Gewichteberechnung vorgenommen:

1. Bestimmung der Auswahlwahrscheinlichkeit für die Dual-Frame-Stichprobe.
2. Bestimmung der Auswahlwahrscheinlichkeit für die Onomastikstichprobe.
3. Integration der Onomastikstichprobe über eine Konvexverbindung mit türkischen Migranten in der ADM-Stichprobe.
4. Trimmen der integrierten adjustierten Designgewichte.
5. Kalibrierung der integrierten Stichprobe via GREG (generalized regression estimator).
6. Fallzahlnormierung der Gewichte und Berechnung der Hochrechnungsfaktoren für die Grundgesamtheit.

Zu beachten ist, dass der erste Gewichtungsschritt, also die Bestimmung der Auswahlwahrscheinlichkeit der beiden Teilstichproben der Dual-Frame-Stichprobe, in unterschiedlichen Varianten durchgeführt wurde. Die Erläuterung dieser Variationen der Berechnung folgt in den nächsten Kapiteln.

3.2.1 Bestimmung der Auswahlwahrscheinlichkeit für die Dual-Frame-Stichproben

Für die ADM-Stichprobe wurden Telefonnummern aus zwei Auswahlrahmen gezogen. Zum einen wurde eine Stichprobe gezogen aus der „ADM-Auswahlgrundlage für Telefonstichproben“ (Festnetz) gezogen und zum anderen aus der „ADM-Auswahlgrundlage für Mobilfunkstichproben“ (Mobilfunk). Die beiden Stichproben lassen sich in Anlehnung an das von Gabler/Häder[31] vorgeschlagene Verfahren grundsätzlich so zusammenführen, dass die beiden Auswahlwahrscheinlichkeiten (Auswahlwahrscheinlichkeit in Festnetzstichprobe und Aus-

[31] Hoffmann, H. (2007): Kombinierte Stichproben für Telefonumfragen – Ansätze in Europa. In: Gabler, S., Häder, S. (Hg.): Mobilfunktelefonie – Eine Herausforderung für die Umfrageforschung. ZUMA-Nachrichten Spezial Band 13. Mannheim, S. 47-57; Häder, S., Häder, M. (Hrsg.): Telefonbefragungen über das Mobilfunknetz: Konzept, Design und Umsetzung einer Strategie zur Datenerhebung. Wiesbaden: VS Verlag für Sozialwissenschaften 2009.

wahlwahrscheinlichkeit in Mobilfunkstichprobe) addiert werden. Diese klassische Form der Designgewichtung ist der Ausgangspunkt für die später vorgestellten Erweiterungen. Angemerkt sei noch, dass die gemeinsame Auswahlwahrscheinlichkeit (in beiden Stichproben zu sein) an dieser Stelle vernachlässigt wurde, da diese bei bundesweiten Stichproben extrem gering ist.

Bestimmender Faktor für die Auswahlwahrscheinlichkeit der Festnetzstichprobe ist die Anzahl der Zielpersonen im Haushalt (reduzierte Haushaltsgröße), da bei mehreren Zielpersonen eine Zufallsauswahl (Last-Birthday-Verfahren) erfolgte und entsprechend die Auswahlwahrscheinlichkeit bestimmt werden muss. Zudem definiert auch die Anzahl der Festnetz- und Mobilfunknummern die Auswahlwahrscheinlichkeit eines Haushalts bzw. einer Person. Über je mehr Nummern ein Haushalt bzw. eine Person erreichbar ist, desto größer ist auch die Chance, in die Stichprobe zu gelangen.

Die **Auswahlwahrscheinlichkeit für Personen**, in der ADM-Stichprobe enthalten zu sein, lässt sich bestimmen über:

$P\,(\text{Personen}) = ((k_F * (m_F/M_F) * (1/z)) + (k_c * (m_c/M_c))$, wobei:

k_F = Anzahl Festnetznummern, über die der Haushalt erreichbar ist,

m_F = realisierte Stichprobe im Festnetz,

M_F = Anzahl Festnetznummern im Auswahlrahmen (bereinigt um ungültige Nummern),

z = Anzahl Zielpersonen im Haushalt,

k_c = Anzahl Mobilfunknummern, über die der Befragte persönlich erreichbar ist,

m_c = realisierte Stichprobe im Mobilnetz,

M_c = Anzahl Mobilfunknummern im Auswahlrahmen (bereinigt um ungültige Nummern)-

Im Unterschied zur ursprünglichen Gabler-Häder-Formel wurden dabei Modifikationen vorgenommen, die mittlerweile dem ADM-Standard entsprechen, der von infas maßgeblich mitentwickelt wurde.[32] So wurde der Auswahlrahmen für Festnetz- und Mobilfunknummern um die jeweils nicht existierenden sowie geschäftlichen resp. Faxnummern bereinigt.[33]

[32] ADM (2012), ADM-Forschungsprojekt „Dual-Frame-Ansätze“ 2011/2012, Forschungsbericht.

[33] Der Anteil dieser Nummern liegt entsprechend dem Ergebnis des ADM-Forschungsprojekts im Festnetzauswahlrahmen bei 60 Prozent und im Mobilfunkauswahlrahmen bei 72 Prozent.

Analog ergibt sich für die **Haushalte die Auswahlwahrscheinlichkeit** über:

P (Haushalte) = ((k_F * (m_F/M_F)) + (k_{ch} * (m_c/M_c)), wobei:

k_{ch} = Anzahl Mobilfunknummern, über die der Haushalt erreichbar ist

3.2.2 Neuerer Ansatz: Erweiterung der Gewichtung um einen Erreichbarkeitsterm

Im ADM-Projekt „Dual Frame 2“, an dem infas aktiv beteiligt war, wurden neue Ansätze der Gewichtung von telefonischen Dual-Frame-Stichproben entwickelt, die neben der Weiterentwicklung der bestehenden Gewichtungsroutinen auch das Ziel haben, die robuste Gewichtungsroutinen für zukünftige Entwicklungen zu gestalten.

Die wesentlichen Änderungen bestehen in der Erweiterung der Gewichtung um einen Erreichbarkeitsterm. Grundlegende Idee bei diesem Vorgehen ist, die Erreichbarkeit über Festnetz und Mobilfunk, unabhängig vom tatsächlichen Befragungsmodus, für den Befragungstag und die davor liegenden 3 Tage abzufragen und die Erreichbarkeit in die Formel für die Dual-Frame-Gewichtung einzubeziehen. Diese Gewichtungsformel wird erweitert um einen Term für die Erreichbarkeit via Festnetz (1/Erreichbarkeit in Tagen Festnetz) und einen Term für die Erreichbarkeit via Mobilfunk (1/Erreichbarkeit in Tagen Mobilfunk). Dabei erhalten selten erreichbare Personen je Mode dann ein höheres Gewicht, während häufig erreichbare Personen je Mode ein geringeres Gewicht erhalten.

Die Berechnung der Auswahlwahrscheinlichkeit der Personen verändert sich durch den Erreichbarkeitsterm wie folgt:

P (Personen) = ((k_F * (m_F/M_F) * (1/z) * (1/errf)) + (k_c * (m_c/M_c)* (1/errm))), wobei:

errf = Erreichbarkeit Festnetz Befragungstag und in den letzten 3 Tagen,

errm = Erreichbarkeit Mobilfunk Befragungstag und in den letzten 3 Tagen.

Die Berechnung der Auswahlwahrscheinlichkeit des Haushalts konnte nicht umgesetzt werden, da – aus Akzeptanzgründen – im Interview nur die Erreichbarkeit des Befragten und nicht die Erreichbarkeit aller Haushaltsmitglieder erfragt wurde.

3.2.3 Neuerer Ansatz: iterative Designgewichtung

Die zweite wesentliche Weiterentwicklung der bestehenden Dual-Frame-Gewichtungsroutine besteht in der „zukunftssicheren“ Berechnung der beiden Auswahlrahmen. In der Formel werden dabei M_f und M_c iterativ berechnet und eingesetzt, sowohl bei der Berechnung der Auswahlwahrscheinlichkeit der Personen als auch bei jener der Haushalte. Nach diesem neuen Verfahren konnten die Auswahlrahmen zusätzlich bestimmt und die so berechneten Designgewichte mit den übrigen für den Viktimisierungssurvey 2017 berechneten Designgewichten nach dem bisherigen Verfahren verglichen werden.

3.2.4 Proportionalisierung des Dual-Use-Anteils

Eine andere Erweiterung der Designgewichtung ist die bereits für den Viktimisierungssurvey 2012 verwendete Proportionalisierung der Dual-Use-Gruppen (sowohl über Festnetz als auch über Mobilfunk erreichbar) in der Stichprobe. Diese folgte im Anschluss an die klassische Designgewichtung. Unter der Annahme, dass sowohl die Festnetz- als auch die Mobilfunkstichprobe für sich in ihrer jeweiligen Zielgruppe repräsentativ ist und damit auch die Potenziale der Dual-Use-Gruppe gleich sein müssen, wurden dabei gleiche Potenziale für die Teilgruppe „Festnetz und Mobil" hergestellt.

Der zusätzliche Transformationsfaktor, mit dem die reinen Designgewichte der integrierten Stichprobe zum Ausgleich der Potenziale multipliziert wurden, berechnet sich über:

Festnetzstichprobe: $F_{Festnetz} = \sqrt{\frac{\beta_M}{\beta_F}}$

Mobilfunkstichprobe: $F_{Mobilfunk} = \frac{1}{F_{Festnetz}}$, wobei:

β_M = Teilgruppe „Dual Use" aus Mobilfunkstichprobe (designgewichtet),

β_F = Teilgruppe „Dual Use" aus Festnetzstichprobe (designgewichtet).

Diese Proportionalisierung erfolgte sowohl für Personen als auch für Haushalte. Abschließend wurde die Summe der so adjustierten Designgewichte auf die ungewichtete Fallzahl normiert. Das Ergebnis der Berechnungen sind adjustierte Designgewichte für die ADM-Stichprobe für Personen und für Haushalte, in der auch ausschließlich über Mobilfunk erreichbare Personen bzw. Haushalte enthalten sind.

3.2.5 Auswahlwahrscheinlichkeit für die Onomastikstichprobe

Bei der Onomastikstichprobe handelt es sich um eine reine Listenauswahl, das heißt die Auswahl erfolgte aus eingetragenen Telefonnummern mit Angaben zu den Namen. Die gelisteten Nummern sind mehrheitlich Festnetznummer, es finden sich aber auch gelistete Mobilfunknummern. Daher befindet sich in der Onomastikstichprobe auch ein geringer Anteil Mobilfunknummern. Die Stichprobe war nach Bundesländern geschichtet. Die Auswahlwahrscheinlichkeit für die realisierte Onomastikstichprobe (m_{FO}) wurde entsprechend bundeslandweise über den Auswahlrahmen (M_{FO}) je Bundesland für Personen berechnet über:

P_j (Personen, Onomastik)= ((k_F * ($m_{FO,j}/M_{FO,j}$) * (1/z))

für j = 1, 2, ... 16 (Bundesländer).

Für Haushalte erfolgte die Berechnung über:

P_j (Haushalte, Onomastik) = k_F * ($m_{FO,j}$ /$M_{FO,j}$)

3.2.6 Integration der Onomastikstichprobe

Die Auswahl der im Telefonbuch eingetragenen Festnetz- und Mobilfunknummern wurde für die Onomastikstichprobe über die eingetragenen Namen vorgenommen. Für die generierten Telefonnummern der ADM-Stichprobe liegen keine Namen vor. Daher lässt sich die Auswahlwahrscheinlichkeit der ADM-Stichprobeneinheiten für die Onomastikstichprobe nicht bestimmen. Die Berechnung gemeinsamer Auswahlwahrscheinlichkeiten ist hierbei nicht möglich.

Ist die Voraussetzung erfüllt, dass zwei Stichproben vorliegen, die erwartungstreue Schätzungen für den wahren Parameter erlauben, dann ist jede Konvexkombination, d.h. die Berechnung von Konvexgewichten, die sich zur Gesamtheit summieren, der erwartungstreuen Schätzer ebenfalls erwartungstreu.[34] Eine Konvexverbindung führt demnach zu erwartungstreuen Schätzern, kann aber ggf. die Varianz der Schätzer deutlich erhöhen.

Für die Konvexkombination wurde nur der Teil der ADM-Stichprobe verwendet, für den eine mehrfache Auswahlwahrscheinlichkeit unterstellt werden kann. Deshalb wurden also nur Personen mit türkischem Migrationshintergrund mit der Onomastikstichprobe konvex verbunden, d.h. nur deren Designgewichte wurden angepasst. Alle anderen behalten ihre adjustierten Designgewichte für Personen und Haushalte. Da rund 85 Prozent der Onomastikstichprobe tatsächlich auch einen türkischen Migrationshintergrund hatten, ist der damit verbundene Fehler relativ gering.

Bei der Konvexgewichtung werden die ursprünglichen Ausgangsgewichte jeder Stichprobe mit einem beliebigen Faktor multipliziert, wobei die Gewichte der Stichprobe 1 (türkische Migranten der ADM-Stichprobe) mit einem beliebigen Faktor α, die Gewichte der Stichprobe 2 (Onomastikstichprobe) mit $1-\alpha$ multipliziert wird. Im vorliegenden Fall wurde die Stichprobe 1 multipliziert mit: $\alpha=n_1/(n_1+n_2)$ und Stichprobe 2 wurde multipliziert mit: $1-\alpha=n_2/(n_1+n_2)$.

Dies entspricht der einfachen Regel, die ursprünglichen Gewichte mit dem Anteil der jeweiligen Teilstichprobe am Gesamtstichprobenumfang zu multiplizieren.

Nach diesem letzten Integrationsschritt wurden alle Stichproben in eine gemeinsame Stichprobe integriert. Allen Fällen der gemeinsamen Stichprobe ist ein integriertes adjustiertes Designgewicht auf Personenebene und ein integriertes adjustiertes Designgewicht auf Haushaltsebene zugeordnet.

[34] Rendtel, U. (1999). The Application of the Convex Weighting Estimator to Household Panel Surveys. Mimeo, Frankfurt; Spiess, M., Rendtel, U. (2000): Combining an ongoing panel with a new cross-sectional sample, DIW Discussion Paper No. 198, Berlin.

3.2.7 Unterschiedliche Varianten der Designgewichtung

Die genannten Möglichkeiten der Designgewichtung wurden in verschiedenen Kombinationen im Viktimisierungssurvey 2017 angewendet. Es liegen insgesamt 6 unterschiedliche Berechnungsvarianten der Designgewichte vor:

a) Klassische Designgewichtung plus Proportionalisierung

b) Iterative Designgewichtung plus Erreichbarkeitsterm

c) Klassische Designgewichtung

d) Iterative Designgewichtung plus Proportionalisierung

e) Klassische Designgewichtung plus Erreichbarkeitsterm

f) Iterative Designgewichtung

Für alle genannten Designgewichtungen wurden die folgenden Schritte jeweils für Personen und Haushalte durchgeführt. Weiterhin wurde bei allen Gewichtungen die onomastische Stichprobe, wie oben beschrieben, integriert. Zusätzlich wurden die Varianten a) und c) auch ohne die Integration der onomastischen Stichprobe berechnet.

Diese Vielzahl an notwendigen Gewichtungsschritten und die Mehrzahl an Auswahlrahmen sowie die Integration der disproportionalen Stichproben in eine gemeinsame Stichprobe führt im Ergebnis zu einer großen Varianz der Gewichtungsfaktoren und mithin einer geringen Effizienz der Gewichte. Aufgrund der sehr hohen Fallzahl der realisierten Stichprobe bewegen sich allerdings die korrigierten Stichprobenfehler immer noch in engen Grenzen.

3.2.8 Trimmen der Designgewichte

Das Hauptziel der Gewichtung ist der Ausgleich von Verzerrungen, die durch das Stichprobendesign (mit ungleichen Auswahlwahrscheinlichkeiten) und durch selektive Teilnahme bzw. Nichtteilnahme entstehen. Unter Verwendung der Gewichte lassen sich Populationswerte aus der Stichprobe erwartungsgetreu schätzen. Weisen die Gewichte allerdings eine starke Streuung auf, so kann dies zu einer großen Varianz der Schätzfunktionen führen. Hier tritt der in der Statistik immer wieder anzutreffende Trade-off zwischen Bias und Varianz auf. Die Gewichtung reduziert den Bias, eine allzu große Steigerung der Varianz durch die Gewichtung sollte allerdings auch vermieden werden. Aus diesem Grund wird versucht, besonders große Gewichtungsfaktoren (und nachrangig auch besonders kleine Faktoren) möglichst zu umgehen und ggf. geeignete Korrekturen an den Gewichten vorzunehmen.

Ausschließlich für die Kalibrierung wurden deshalb die Eingangsgewichte (die beschriebenen finalen Designgewichte) vor der Kalibrierung getrimmt, d.h. sie wurden durch neue Eingangsgewichte ersetzt. Unter- bzw. Obergrenze der getrimmten Designgewichte wurden durch bestimmte Perzentile der Verteilung, in Abhängigkeit von der Verteilung der Designgewichte, ermittelt. Die Perzentile

wurden dabei so gewählt, dass die Trimmung keine erheblichen Veränderungen der gewichteten Verteilungen bewirkte, gleichzeitig aber die Gewichtungsfaktoren mit den höchsten und niedrigsten Werten gekappt wurden. Die Personen- und Haushaltsgewichte wurden getrimmt und anschließend so reskaliert, dass ihre Summe wieder die Summe der ungetrimmten Designgewichte ergab. Die jeweilige Trimmung sah wie folgt aus:

Tabelle 23 Trimmungsgrenzen der Designgewichte

Designgewichtung	Personen	Haushalte
a1) Klassisch+Prop	5%/95%	2%/96%
a2) Klassisch+Prop ohne Onomastik	5%/95%	2%/96%
b) Iterativ+Erreichbarkeit	Ohne Trimmung	Ohne Trimmung
c1) Klassisch	5%/95%	2%/96%
c2) Klassisch ohne Onomastik	5%/95%	2%/96%
d) Iterativ+Prop	5%/95%	2%/96%
e) Klassisch+Erreichbarkeit	Ohne Trimmung	Ohne Trimmung
d) Iterativ	5%/95%	2%/96%

3.2.9 Kalibrierung und Hochrechnung

Anschließend wurden die getrimmten Designgewichte als Ausgangsgewichte für die Randanpassung verwendet. Die Kalibrierung der Designgewichte erfolgte getrennt auf Personen- und auf Haushaltsebene via GREG (generalized regression estimator).[35]

Kalibriert wurde an die Eckwerte des Mikrozensus 2016 des Statistischen Bundesamts für die Personen am Ort der Hauptwohnung in Privathaushalten im Alter von 16 Jahren oder älter sowie an die Eckwerte des Mikrozensus 2016 für Haushalte am Ort der Hauptwohnung (hier ist auch die äußerst geringe Anzahl der Haushalte enthalten, in denen ausschließlich Personen unter 16 Jahre leben).

Um die Varianz der Gewichte zu begrenzen, wurde neben der Trimmung das Intervall der Gewichte begrenzt, d.h. es wurde eine obere und eine untere Grenze für die Gewichte festgelegt. Festgelegt wurde dabei die Spannbreite der reinen Kalibrierungsgewichte (sog. g-weights), die sich aus dem Verhältnis von ursprünglichen Gewichten (getrimmte Eingangsgewichte der Kalibrierung) zum kalibrierten Endgewicht ergibt. Bei dieser Begrenzung der Gewichte ist allerdings zu beachten, dass engere Grenzen für die Gewichte in geringerer Streuung und damit geringerer Varianz der Schätzungen resultieren, bei zu eng gesetzten Grenzen kann allerdings die Kalibrierung aller Eckwerte[36] schlicht unmöglich werden.

[35] Deville, J.C., Särndal, C.-E. (1992): Calibration Estimators in Survey Sampling. Journal of the American Statistical Association, 82, 376-381; Särndal, C.-E., Swensson, B., Wretman, J. (1992): Model Assisted Survey Sampling, Springer.

[36] Z.B. Sollverteilung aus dem Mikrozensus.

Um die Kalibrierung durchzuführen, muss jede Eckwertvariable für jeden Fall im Datensatz einen gültigen Wert besitzen. Deshalb wurde der geringe Item-Nonresponse vor der Kalibrierung imputiert. Die Imputation erfolgte über den Mittelwert bzw. den Modalwert der jeweiligen Variable. Da die Imputation einzig der Durchführbarkeit der Kalibrierung dient, wurden die imputierten Werte nach der Kalibrierung wieder auf fehlende Werte zurückgesetzt.

Bei Bundesland und BIK-Gemeindegrößenklasse erfolgte keine Imputation, hier wurde die Kategorie „fehlende Werte" als zusätzliche Eckwertkategorie definiert und der Anteil der gewichteten fehlenden Angaben als Zielwert festgelegt.[37] Die Kalibrierung verändert also nicht den (designgewichteten) Anteil der fehlenden Angaben. Die Anteile der Bundesländer und BIK-Gemeindegrößenklassen wurden proportional um den Anteil der fehlenden Angaben reduziert, so dass sich in der Summe wieder 100 Prozent ergaben.

Personen

Bei den Personengewichten wurde das Verhältnis der Gesamthochrechnungsfaktoren nach Kalibrierung zu getrimmten Designgewichten nach unten auf minimal 0,3 und nach oben auf maximal 3,5 begrenzt. Nachfolgende Merkmale wurden für die Kalibrierung berücksichtigt:

- **Geschlecht x Altersgruppen** mit den Altersgruppen 16-24 Jahre, 25-34 Jahre, 35-44 Jahre, 45-54 Jahre, 55-64 Jahre, 65-74 Jahre und 75 Jahre und älter.
- **Höchster allgemeinbildender Schulabschluss x Altersgruppen** mit den Kategorien: niedrig = Förderschul- oder Volksschul- oder Hauptschulabschluss oder Abschluss der POS8 oder kein Schulabschluss, mittel = Mittlere Reife oder Abschluss POS10 oder noch in schulischer Ausbildung oder sonstiger Schulabschluss, hoch = Fachhochschul- oder Hochschulreife.
- **Erwerbstätigkeit** dichotomisiert in erwerbstätig vs. nicht erwerbstätig.
- **Arbeitslosigkeit** dichotomisiert in arbeitslos vs. nicht arbeitslos.
- **Migrationshintergrund** mit den Kategorien: kein Migrationshintergrund, türkischer Migrationshintergrund, Migrationshintergrund ehemalige Sowjetunion und sonstiger Migrationshintergrund. Dabei wurde Migrationshintergrund definiert über eigene Migrationserfahrung und/oder ausschließlich nicht deutsche Staatsangehörigkeit.
- **Haushaltsgröße** mit den Kategorien Personen in Einpersonenhaushalten, Personen in Zweipersonenhaushalten, Personen in Dreipersonenhaushalten und Personen in Vier- und Mehrpersonenhaushalten.
- **Bundesland**
- **BIK-Gemeindegrößenklasse**

[37] Fehlende Werte bei Bundesland und BIK-Gemeindegrößenklasse ergeben sich durch fehlende oder ungültige Angaben der Postleitzahl in den Interviews der Mobilfunkstichprobe. In der Festnetzstichprobe kann dagegen die Vorwahl für die regionale Zuordnung genutzt werden. Im Interview haben 661 Mobilfunkfälle keine Angaben zur Postleitzahl gemacht (Weiß nicht/verweigert); in weiteren 362 Mobilfunkfällen war die genannte Postleitzahl ungültig.

Das Ergebnis der Kalibrierung sind zum einen fallzahlnormierte kalibrierte Endgewichte für die Personengewichtung und zum anderen Hochrechnungsfaktoren auf die Grundgesamtheit der 16-jährigen und älteren Bevölkerung.

In der folgenden Tabelle ist das Ergebnis der Gewichtung zusammengefasst. Aufgrund der verschiedenen Arten der Designgewichtung (siehe Abschnitt 3.2.6) werden die Ergebnisse separat aufgelistet. Die kalibrierten Ergebnisse wiederum unterscheiden sich zwischen den verschiedenen Gewichtungsmethoden lediglich bei Bundesland und BIK-Größenklasse, da diese Variablen Missings aufweisen. Daher werden die Ergebnisse der Kalibrierung auf Basis der klassischen Designgewichtung mit Onomastik aufgelistet. Dargestellt sind in der folgenden Tabelle die Prozentverteilungen der folgenden aktiven Gewichtungsmerkmale:

- Ungewichtet,
- a1: klassisch designgewichtet mit Proportionalisierung,
- a2: klassisch designgewichtet mit Proportionalisierung, ohne Onomastik,
- b: iterativ designgewichtet mit Erreichbarkeit,
- c1: klassisch designgewichtet,
- c2: klassisch designgewichtet, ohne Onomastik,
- d: iterativ designgewichtet mit Proportionalisierung,
- e: klassisch designgewichtet mit Erreichbarkeit,
- f: iterativ designgewichtet,
- kalibriert (auf Basis von c1).

Tabelle 24 Prozentualer Verteilungsvergleich (Personen)

Merkmale	Unge-wichtet	a1	a2	b	c1	c2	d	e	f	Kalibriert nach c1	MZ 2016 (Soll)
Spalten%	%	%	%	%	%	%	%	%	%	%	%
Geschlecht x Altersgruppe											
Mann 16-24 Jahre	4,0	6,7	6,5	6,0	5,4	5,2	6,8	5,8	5,6	5,9	5,9
Mann 25-34 Jahre	4,4	7,5	7,7	6,4	5,3	5,3	7,7	6,1	5,6	7,7	7,7
Mann 35-44 Jahre	6,0	7,2	7,3	6,2	6,0	6,0	7,3	6,0	6,1	7,1	7,1
Mann 45-54 Jahre	9,6	10,0	9,9	9,5	9,4	9,3	10,0	9,4	9,5	9,5	9,5
Mann 55-64 Jahre	10,6	10,1	10,2	9,7	10,0	10,1	10,1	9,7	10,0	8,0	8,0
Mann 65-74 Jahre	8,7	7,1	7,2	7,0	8,0	8,2	7,0	7,1	7,9	5,5	5,5
Mann 75 Jahre u. älter	6,0	4,4	4,4	4,6	5,5	5,6	4,2	4,9	5,4	5,3	5,3
Frau 16-24 Jahre	3,1	4,6	4,3	4,6	4,0	3,8	4,7	4,5	4,1	5,4	5,4
Frau 25-34 Jahre	3,7	5,3	5,4	5,3	4,3	4,3	5,5	5,0	4,5	7,3	7,2
Frau 35-44 Jahre	6,3	6,5	6,3	6,8	6,4	6,2	6,5	6,8	6,4	6,9	6,9
Frau 45-54 Jahre	10,9	10,6	10,3	11,2	11,2	11,1	10,5	11,3	11,2	9,3	9,3
Frau 55-64 Jahre	11,6	9,7	9,8	10,8	11,0	11,2	9,6	11,0	10,8	8,2	8,2
Frau 65-74 Jahre	8,6	6,2	6,3	6,8	7,7	7,8	6,1	7,1	7,5	6,1	6,1
Frau 75 Jahre und älter	6,7	4,3	4,4	5,1	5,8	6,0	4,1	5,4	5,6	7,9	7,9
Bildung x Altersgruppe											
Niedrig 16-24 Jahre	0,8	1,3	1,3	1,2	1,0	1,0	1,3	1,1	1,0	1,8	1,8
Niedrig 25-34 Jahre	0,8	1,4	1,4	1,3	1,0	1,0	1,5	1,2	1,1	3,2	3,2
Niedrig 35-44 Jahre	1,5	2,0	1,9	1,9	1,6	1,6	2,0	1,8	1,7	3,7	3,7
Niedrig 45-54 Jahre	2,9	3,5	3,3	3,1	3,2	3,0	3,5	3,1	3,2	5,8	5,8
Niedrig 55-64 Jahre	4,4	4,4	4,4	4,4	4,4	4,4	4,4	4,4	4,4	6,7	6,7
Niedrig 65-74 Jahre	4,4	3,7	3,7	3,7	4,2	4,3	3,6	3,8	4,1	6,6	6,5
Niedrig 75 Jahre und älter	5,0	3,6	3,7	3,9	4,7	4,8	3,5	4,1	4,5	9,7	9,7
Mittel 16-24 Jahre	3,1	4,6	4,3	4,4	4,1	3,8	4,7	4,3	4,2	5,6	5,5
Mittel 25-34 Jahre	2,2	3,4	3,4	3,0	2,5	2,5	3,4	2,8	2,6	4,5	4,5
Mittel 35-44 Jahre	4,0	4,7	4,7	4,4	4,2	4,2	4,8	4,4	4,3	4,7	4,7
Mittel 45-54 Jahre	7,7	8,0	8,0	7,9	8,0	8,0	8,0	7,9	8,1	7,0	7,0
Mittel 55-64 Jahre	8,4	7,6	7,7	7,9	8,2	8,3	7,5	7,9	8,1	5,3	5,3
Mittel 65-74 Jahre	5,9	4,7	4,8	4,8	5,5	5,6	4,6	5,0	5,4	2,7	2,7
Mittel 75 Jahre und älter	3,6	2,4	2,4	2,8	3,2	3,3	2,3	2,9	3,1	1,7	1,7
Hoch 16-24 Jahre	3,2	5,3	5,2	5,0	4,3	4,2	5,4	4,8	4,5	3,9	3,9
Hoch 25-34 Jahre	5,1	8,0	8,2	7,5	6,1	6,1	8,2	7,1	6,3	7,3	7,3
Hoch 35-44 Jahre	6,7	7,0	7,0	6,7	6,5	6,4	7,0	6,6	6,5	5,6	5,6
Hoch 45-54 Jahre	9,8	9,1	9,0	9,7	9,5	9,4	9,1	9,7	9,4	6,0	6,0
Hoch 55-64 Jahre	9,4	7,8	7,9	8,2	8,4	8,5	7,7	8,4	8,3	4,3	4,3
Hoch 65-74 Jahre	6,9	5,0	5,0	5,3	6,0	6,1	4,9	5,5	5,8	2,4	2,4
Hoch 75 Jahre und älter	4,0	2,7	2,7	3,1	3,5	3,6	2,6	3,3	3,4	1,8	1,8

Merkmale	Unge-wichtet	a1	a2	b	c1	c2	d	e	f	Kalibriert nach c1	MZ 2016 (Soll)
Spalten%	%	%	%	%	%	%	%	%	%	%	%
Erwerbstätigkeit											
Erwerbstätig	50,4	53,1	53,2	52,3	50,3	50,4	53,2	51,9	50,6	59,1	59,1
Nicht erwerbstätig	49,6	46,9	46,8	47,7	49,7	49,6	46,8	48,1	49,4	40,9	40,9
Arbeitslosigkeit											
Arbeitslos	2,2	2,9	2,9	2,6	2,4	2,4	3,0	2,5	2,5	3,6	3,6
Nicht arbeitslos	97,8	97,1	97,1	97,4	97,6	97,6	97,0	97,5	97,5	96,4	96,4
Migrationshintergrund											
Kein Migrationshintergrund	88,2	85,5	87,2	86,4	87,1	88,9	85,3	86,7	86,9	79,5	79,5
Migrationshintergrund türkisch	2,9	3,4	1,4	3,0	3,1	1,1	3,5	3,0	3,2	3,0	3,0
Migrationshintergrund ehem. SU	2,9	3,4	3,5	3,5	3,2	3,3	3,4	3,4	3,3	3,7	3,7
Migrationshintergrund sonstiger	6,0	7,7	7,9	7,2	6,6	6,7	7,8	6,9	6,7	13,9	13,9
Haushaltsgröße											
1 Person	31,1	22,2	22,8	23,7	23,0	23,5	22,1	23,5	22,7	23,5	23,5
2 Personen	35,9	35,2	35,8	35,8	36,8	37,5	34,9	36,1	36,5	38,6	38,6
3 Personen	13,6	16,2	16,1	15,4	15,5	15,4	16,3	15,4	15,6	17,3	17,3
4 und mehr Personen	19,3	26,4	25,3	25,1	24,7	23,7	26,7	25,1	25,2	20,6	20,6
Bundesland											
Schleswig-Holstein	3,3	3,0	3,0	2,9	3,1	3,2	3,0	3,0	3,1	3,3	3,5
Hamburg	2,7	2,3	2,3	2,5	2,5	2,6	2,3	2,5	2,5	2,1	2,2
Niedersachsen	8,3	7,9	7,9	8,2	8,2	8,3	7,9	8,3	8,2	9,2	9,6
Bremen	0,8	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,8	0,8
Nordrhein-Westfalen	20,6	20,0	19,6	19,4	20,4	20,0	19,9	19,5	20,3	20,8	21,7
Hessen	6,6	6,5	6,4	6,3	6,6	6,5	6,5	6,3	6,6	7,2	7,5
Rheinland-Pfalz	4,3	4,1	4,0	4,1	4,2	4,2	4,1	4,1	4,2	4,7	4,9
Baden-Württemberg	11,0	10,8	10,4	11,4	11,3	10,9	10,8	11,5	11,3	12,6	13,2
Bayern	16,0	15,4	15,5	16,3	16,0	16,1	15,4	16,3	16,0	15,0	15,6
Saarland	1,1	1,1	1,1	1,1	1,2	1,2	1,1	1,1	1,2	1,2	1,2
Berlin	6,6	5,5	5,6	6,0	6,0	6,1	5,5	6,0	5,9	4,1	4,3
Brandenburg	3,1	3,0	3,0	2,9	3,0	3,1	3,0	2,9	3,0	2,9	3,1
Mecklenburg-Vorpommern	1,7	1,8	1,9	1,7	1,8	1,8	1,9	1,7	1,8	1,9	2,0
Sachsen	5,4	5,4	5,5	5,5	5,5	5,6	5,4	5,5	5,5	4,8	5,0
Sachsen-Anhalt	2,7	2,7	2,8	2,7	2,8	2,8	2,7	2,6	2,8	2,7	2,8
Thüringen	2,6	2,6	2,7	2,5	2,7	2,7	2,6	2,5	2,7	2,6	2,7
Fehlend	3,3	7,2	7,6	6,1	4,1	4,3	7,3	5,7	4,3	4,1	-

Merkmale	Unge-wichtet	a1	a2	b	c1	c2	d	e	f	Kalibriert nach c1	MZ 2016 (Soll)
Spalten%	*%*	*%*	*%*	*%*	*%*	*%*	*%*	*%*	*%*	*%*	*%*
BIK-Gemeindegrößenklasse											
Unter 2.000 EW	1,0	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,1	1,7	1,8
2.000 bis unter 5.000 EW	2,1	2,2	2,2	2,1	2,2	2,2	2,2	2,1	2,2	2,5	2,6
5.000 bis unter 20.000 EW	7,3	7,3	7,2	7,1	7,5	7,3	7,3	7,2	7,5	8,2	8,6
20.000 bis unter 50.000 EW	9,8	9,8	9,6	10,1	10,1	10,0	9,8	10,1	10,1	10,5	11,0
50.000 bis unter 100.000 EW Rand	6,8	6,5	6,4	6,7	6,9	6,9	6,4	6,8	6,8	7,6	7,9
50.000 bis unter 100.000 EW Kern	2,2	2,0	2,0	2,0	2,1	2,1	1,9	2,0	2,1	2,1	2,2
100.000 bis unter 500.000 EW Rand	12,9	12,8	12,7	12,8	13,2	13,2	12,8	12,8	13,2	14,1	14,7
100.000 bis unter 500.000 EW Kern	14,4	14,3	14,4	14,3	14,3	14,3	14,3	14,3	14,3	14,4	15,0
500.000 und mehr EW Rand	9,6	8,9	9,0	9,0	9,4	9,6	8,9	9,1	9,4	9,2	9,6
500.000 und mehr EW Kern	30,4	27,8	27,9	28,7	29,0	29,0	27,8	28,7	28,9	25,5	26,7
Fehlend	3,4	7,3	7,6	6,2	4,3	4,3	7,5	5,8	4,5	4,3	-

Quelle: Mikrozensus, eigene Berechnungen

Erläuterung:
a1/a2: klassisch designgewichtet mit Proportionalisierung / klassisch designgewichtet mit Proportionalisierung, ohne Onomastik
b: iterativ designgewichtet mit Erreichbarkeit
c1/c2: klassisch designgewichtet / klassisch designgewichtet, ohne Onomastik
d: iterativ designgewichtet mit Proportionalisierung
e: klassisch designgewichtet mit Erreichbarkeit
f: iterativ designgewichtet

Bei der Betrachtung der Ergebnisse der verschiedenen Varianten der Designgewichtung zeigt sich, dass die Unterschiede zwischen den Berechnungsvarianten sehr gering sind. Lediglich beim Anteil an fehlenden Werten bei Bundesland und BIK-Größenklasse zeigen sich deutlichere Unterschiede. Ansonsten unterscheiden sich die Werte lediglich um bis zu 2 Prozentpunkte.

Insgesamt betrachtet, sind die meisten untersuchten Merkmale in der realisierten (ungewichteten) Stichprobe im Vergleich zur Grundgesamtheit (Sollverteilungen nach Mikrozensus, vgl. Tabelle 24) sehr gut repräsentiert. Lediglich bei der formalen Schulbildung zeigt sich, dass Personen mit niedriger formaler Bildung erkennbar unter- und Personen mit hoher formaler Bildung überrepräsentiert sind. Dieses Phänomen ist in sozialwissenschaftlichen Befragungen oft zu beobachten. Zudem haben sich allgemein Personen mit Migrationshintergrund seltener an der Befragung beteiligt. Auch dieses Ergebnis entspricht den Erfahrungen aus diversen anderen Befragungen. Des Weiteren sind Einpersonenhaushalte über- und Dreipersonenhaushalte unterrepräsentiert. Die Berücksichtigung einer Mobilfunkstichprobe dürfte dies mit verursacht haben, da die Wahrscheinlichkeit, eine Person aus einem Einpersonenhaushalt zu erreichen, in der Mobilfunkstichprobe größer ist als in der Festnetzstichprobe. Die Differenzen zur Sollverteilung laut Mikrozensus sind durch die beschriebenen Gewichtungsschritte ausgeglichen worden.

Zur Beurteilung der Gewichte werden im Folgenden neben dem Mittelwert und der Standardabweichung auch das Effektivitätsmaß der Kalibrierung angegeben (E). Das Effektivitätsmaß E basiert auf der Varianz des Gewichtungsfaktors. Das Effektivitätsmaß gibt in Prozent der realisierten Fallzahl an, wie groß bei Verwendung des Gewichts die effektive Fallzahl (n') bei einem passiven Merkmal ist, das mit den aktiven Merkmalen nicht korreliert.[38] Die effektive Fallzahl entspricht der Anzahl Befragter, die bei einer uneingeschränkten Zufallsauswahl gegeben die Varianz des Merkmals in der Stichprobe den gleichen Stichprobenfehler produziert hätte. Das Effektivitätsmaß drückt das Verhältnis von n zu n' als Prozentwert aus und wird berechnet über:

$E = (n'/n) * 100$, wobei: $n' = ((\sum_i g_i)^2/\sum_i g_i^2)$

Die Effektivität sowie die effektive Fallzahl sind in der folgenden Tabelle für alle Berechnungsvarianten aufgelistet, da hier die Unterschiede zwischen den Gewichtungsvarianten deutlich werden.

Tabelle 25 Effektivitätsmaße und effektive Fallzahl (Personen)

	Gewichtungsvarianten							
	a1	**a2**	**b**	**c1**	**c2**	**d**	**e**	**f**
Effektivitätsmaß	39,3	38,0	29,9	51,4	49,8	37,8	32,6	49,5
Effektive Fallzahl	12.262	11.468	9.321	16.017	15.029	11.793	10.175	15.433

Quelle: eigene Berechnungen

Erläuterung:
a1/a2: klassisch designgewichtet mit Proportionalisierung/klassisch designgewichtet mit Proportionalisierung, ohne Onomastik
b: iterativ designgewichtet mit Erreichbarkeit
c1/c2: klassisch designgewichtet / klassisch designgewichtet, ohne Onomastik
d: iterativ designgewichtet mit Proportionalisierung
e: klassisch designgewichtet mit Erreichbarkeit
f: iterativ designgewichtet

Die Effektivitäten der Kalibrierungen auf Basis der verschiedenen Arten der Designgewichtung unterscheiden sich sehr stark. Die beste Variante ist die klassische Designgewichtung ohne Proportionalisierung oder Erreichbarkeit (c1 bzw. c2). Die Effektivität der iterative Designgewichtung ohne Proportionalisierung oder Erreichbarkeit liegt jedoch nur leicht darunter (f). Die Erreichbarkeitsfragen (b und e) in die Gewichtung aufzunehmen, bei gleichzeitigem Verzicht auf eine Trimmung der Daten, führt zu den schlechtesten Effektivitäten. Allerdings sorgt auch die Proportionalisierung für einen niedrigeren Effektivitätswert (a und d).

Die Effektivität ist insgesamt relativ gering bzw. der Designfaktor ist auch bei der Gewichtung auf Basis der klassischen oder iterativen Designgewichtung vergleichsweise hoch. Dies bedeutet eine deutliche Reduktion der effektiven

[38] *Aktive* Merkmale sind Merkmale, die bei der Kalibrierung eingesetzt wurden. Alle übrigen erhobenen Informationen sind *passive* Merkmale.

Fallzahl und mithin eine Vergrößerung des Standardfehlers bei gewichteten Populationsschätzern. Insgesamt ist das ein Ergebnis des Stichprobendesigns, da in die Gesamtstichprobe unterschiedliche Frames und disproportionale Stichproben integriert wurden. Der Ausgleich der daraus resultierenden unterschiedlichen Auswahlwahrscheinlichkeiten und Disproportionalitäten erhöht zwangsläufig die Varianz der Gewichte.

Es bleibt allerdings festzuhalten, dass aufgrund der sehr hohen Fallzahl die effektive Fallzahl immer noch sehr groß, der Stichprobenfehler, trotz Designeffekt, entsprechend gering ist.

Aufgrund der Ergebnisse hinsichtlich der Effizienz der Gewichte empfiehlt infas, für die Gewichtung und Hochrechnung der Daten des DVS 2017 die auf Basis der klassischen Designgewichtung mit Onomastik berechneten Gewichtungs- und Hochrechnungsfaktoren zu verwenden. Aus diesem Grund beschränkt sich die Darstellung in der folgenden Tabelle auf die Eckdaten zum kalibrierten Gewicht und Hochrechnungsfaktor auf diese Gewichtungsvariante.

Tabelle 26 Kennwerte der kalibrierten Gewichte und Hochrechnungsfaktoren Personen (Gewichtungsvariante c1 – klassische Designgewichtung mit Onomastik)

Kennwert (Personen)	Kalibriertes Gewicht	Hochrechnungsfaktor
Mittelwert	1	2266,55
Standardabweichung	0,97	2206,12
Minimum	0,08	172,24
Maximum	9,21	20870,19
Fallzahl	31.192/30.180 (mit/ohne Onomastik)	

Quelle: eigene Berechnungen

Haushalte

Bei den Haushaltsgewichten wurde das Verhältnis der Gesamthochrechnungsfaktoren nach Kalibrierung zu getrimmten Designgewichten nach unten auf minimal 0,3 und nach oben auf maximal 4,0 begrenzt. Da keine Daten über den Haupteinkommensbezieher eines Haushalts vorliegen, konnten die Haushaltsdesigngewichte analog zu 2012 an nur vergleichsweise wenigen Verteilungen aus dem Mikrozensus kalibriert werden. Für folgende Merkmale wurde kalibriert:

- **Haushaltsgröße** mit den Kategorien Einpersonenhaushalte, Zweipersonenhaushalte, Dreipersonenhaushalte und Vier- und Mehrpersonenhaushalte.
- **Bundesland**
- **BIK-Gemeindegrößenklasse**

Das Ergebnis der Kalibrierung sind zum einen fallzahlnormierte kalibrierte Endgewichte für die Haushaltsgewichtung und zum anderen Hochrechnungsfaktoren auf die Grundgesamtheit aller Haushalte. Sinnvoll verwendbar sind diese ausschließlich für Merkmale, die für den gesamten Haushalt erhoben wurden,

also haushaltsbezogene Fragestellungen. In der folgenden Tabelle ist das Ergebnis der Gewichtung zusammengefasst.

Tabelle 27 Prozentualer Verteilungsvergleich (Haushalte)

Merkmale	Unge-wichtet	a1	a2	b	c1	c2	d	e	f	Kalibriert nach c1	MZ 2016 (Soll)
	%	%	%	%	%	%	%	%	%	%	%
Haushaltsgröße											
1 Person	31,1	34,7	35,4	41,1	37,9	35,4	34,7	40,9	37,6	41,0	41,0
2 Personen	35,9	34,0	34,6	32,4	34,2	34,6	33,9	32,7	34,2	34,0	34,0
3 Personen	13,6	13,2	13,0	11,3	11,9	13,0	13,2	11,3	12,0	12,3	12,3
4 und mehr Personen	19,3	18,1	17,1	15,3	16,0	17,1	18,2	15,2	16,2	12,7	12,7
Bundesland											
Schleswig-Holstein	3,3	3,1	3,1	3,2	3,2	3,1	3,1	3,3	3,2	3,4	3,5
Hamburg	2,7	2,5	2,5	2,7	2,7	2,5	2,4	2,7	2,7	2,3	2,4
Niedersachsen	8,3	7,6	7,5	7,8	8,0	7,5	7,5	7,9	7,9	9,2	9,6
Bremen	0,8	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,7	0,8	0,9
Nordrhein-Westfalen	20,6	19,9	19,4	19,8	20,1	19,4	19,8	19,9	20,1	20,3	21,3
Hessen	6,6	6,4	6,2	6,4	6,5	6,2	6,4	6,4	6,5	7,1	7,4
Rheinland-Pfalz	4,3	3,9	3,9	3,9	4,0	3,9	3,9	4,0	4,0	4,6	4,8
Baden-Württemberg	11,0	10,3	9,9	10,3	10,6	9,9	10,2	10,4	10,6	12,1	12,7
Bayern	16,0	14,8	14,8	15,1	15,4	14,8	14,7	15,2	15,3	14,8	15,5
Saarland	1,1	1,0	1,1	1,1	1,1	1,1	1,0	1,1	1,1	1,2	1,2
Berlin	6,6	6,0	6,1	6,5	6,6	6,1	6,0	6,6	6,6	4,6	4,8
Brandenburg	3,1	3,0	3,1	3,1	3,1	3,1	3,0	3,1	3,1	2,9	3,0
Mecklenburg-Vorpommern	1,7	2,0	2,1	2,2	1,9	2,1	2,0	2,1	2,0	2,0	2,0
Sachsen	5,4	5,7	5,9	6,0	5,8	5,9	5,7	5,9	5,8	5,1	5,3
Sachsen-Anhalt	2,7	2,9	3,0	3,1	3,0	3,0	3,0	3,1	3,0	2,7	2,9
Thüringen	2,6	2,6	2,7	2,7	2,7	2,7	2,6	2,7	2,7	2,6	2,7
Fehlend	3,3	7,6	8,0	5,4	4,4	8,0	7,9	5,0	4,7	4,4	-
BIK-Gemeindegrößenklasse											
Unter 2.000 EW	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,6	1,6
2.000 bis unter 5.000 EW	2,1	2,1	2,1	2,1	2,1	2,1	2,1	2,1	2,1	2,2	2,3
5.000 bis unter 20.000 EW	7,3	7,3	6,8	7,1	7,1	6,8	6,9	7,0	7,1	7,4	7,7
20.000 bis unter 50.000 EW	9,8	9,8	9,4	9,8	9,8	9,4	9,4	9,8	9,8	10,1	10,5
50.000 bis unter 100.000 EW Rand	6,8	6,8	6,3	6,6	6,7	6,3	6,3	6,5	6,6	7,2	7,5
50.000 bis unter 100.000 EW Kern	2,2	2,2	1,9	2,1	2,1	1,9	1,9	2,1	2,1	2,1	2,2
100.000 bis unter 500.000 EW Rand	12,9	12,9	11,9	12,3	12,4	11,9	11,9	12,2	12,3	13,3	13,9
100.000 bis unter 500.000 EW Kern	14,4	14,4	14,9	15,0	14,9	14,9	14,9	15,2	15,0	15,3	16,0
500.000 und mehr EW Rand	9,6	9,6	8,5	8,9	9,0	8,5	8,3	8,8	8,9	8,7	9,1
500.000 und mehr EW Kern	30,4	30,4	29,2	30,3	30,4	29,2	29,2	30,2	30,3	27,7	29,0
Fehlend	3,4	3,4	8,0	4,8	4,5	8,0	8,1	5,1	4,8	4,5	-

Quelle: Mikrozensus, eigene Berechnungen
Erläuterung:
a1/a2: klassisch designgewichtet mit Proportionalisierung / klassisch designgewichtet mit Proportionalisierung, ohne Onomastik
b: iterativ designgewichtet mit Erreichbarkeit
c1/c2: klassisch designgewichtet / klassisch designgewichtet, ohne Onomastik
d: iterativ designgewichtet mit Proportionalisierung
e: klassisch designgewichtet mit Erreichbarkeit
f: iterativ designgewichtet

Insgesamt ist die Effektivität bei den Haushaltsgewichten, die nur für wenige haushaltsbezogene Fragen relevant ist, etwas höher als bei den Personengewichten (vgl. Tabelle 28). Dies ist eine Folge der geringeren Zahl der kalibrierbaren Merkmale auf Haushaltsebene. Hier sind demnach zwar nicht so viele Merkmale angepasst, die Varianz der Gewichte aber entsprechend geringer.

Erneut zeigt sich, dass die klassische Designgewichtung die höchste Effektivität aufweist. Direkt gefolgt von der neuen Designgewichtung mit iterativer Auswahlgesamtheit. Die Proportionalisierung und die Aufnahme der Erreichbarkeit in die Berechnung bei gleichzeitigem Weglassen der Trimmung führen zu deutlich niedrigeren Effektivitätswerten.

Tabelle 28 Effektivitätsmaß und effektive Fallzahl im Vergleich (Haushalte)

	Gewichtungsvarianten							
	a1	**a2**	**b**	**c1**	**c2**	**d**	**e**	**f**
Effektivitätsmaß	48,4	47,9	47,4	72,0	72,3	43,3	55,1	66,0
Effektive Fallzahl	15.084	14.461	14.781	22.450	21.833	13.503	17.188	20.572

Quelle: eigene Berechnungen

Erläuterung:
a1/a2: klassisch designgewichtet mit Proportionalisierung / klassisch designgewichtet mit Proportionalisierung, ohne Onomastik
b: iterativ designgewichtet mit Erreichbarkeit
c1/c2: klassisch designgewichtet / klassisch designgewichtet, ohne Onomastik
d: iterativ designgewichtet mit Proportionalisierung
e: klassisch designgewichtet mit Erreichbarkeit
f: iterativ designgewichtet

In Tabelle 29 sind die Eckwerte zum kalibrierten Haushaltsgewicht und zum entsprechenden Hochrechnungsfaktor auf Basis der klassischen Designgewichtung (inkl. Onomastik) angegeben.

Tabelle 29 Kennwerte der kalibrierten Gewichte und Hochrechnungsfaktoren Haushalte (Gewichtungsvariante c1 – klassische Designgewichtung mit Onomastik)

Kennwert (Personen)	**Kalibriertes Gewicht**	**Hochrechnungsfaktor**
Mittelwert	1	1.313,16
Standardabweichung	0,62	819,43
Minimum	0,12	156,90
Maximum	5,76	7558,26
Fallzahl	31.192/30.180 (mit/ohne Onomastik)	

Quelle: eigene Berechnungen

4 Literaturverzeichnis

The American Association for Public Opinion Research (AAPOR) (2016): Standard Definitions. Final Dispositions of Case Codes and Outcome Rate for Surveys. Revised Version 2016, S. 62.

Arbeitskreis Deutscher Marktforschungsinstitute (ADM) (2012), ADM-Forschungsprojekt „Dual-Frame-Ansätze“ 2011/2012, Forschungsbericht.

Auspurg, K., Hinz, T. und S. Liebig (2009): „Komplexität von Vignetten, Lerneffekte und Plausibilität im Faktoriellen Survey“, in Methoden – Daten – Analysen, Jg. 3, Heft 1, S. 59-96.

Deville, J.C., Särndal, C.E.(1992): Calibration Estimators in Survey Sampling. Journal of the American Statistical Association, 82, 376-381; Särndal, C.-E., Swensson, B., Wretman, J. (1992): Model Assisted Survey Sampling, Springer.

Hoffmann, H. (2007): Kombinierte Stichproben für Telefonumfragen - Ansätze in Europa. In: S. Gabler und S. Häder (Hg.): Mobilfunktelefonie - Eine Herausforderung für die Umfrageforschung. ZUMA-Nachrichten Spezial Band 13. Mannheim, S. 47-57.

Häder, S., Häder, M. (Hrsg.): Telefonbefragungen über das Mobilfunknetz: Konzept, Design und Umsetzung einer Strategie zur Datenerhebung. Wiesbaden: VS Verlag für Sozialwissenschaften 2009.

Häder, S., Gabler, S. (1998): Ein neues Stichprobendesign für telefonische Umfragen in Deutschland. In Gabler S. , S. Häder und J.H.P. Hoffmeyer-Zlotnik (Hrsg.) Telefonstichproben in Deutschland. Opladen, Westdeutscher Verlag.

Humpert, A., Schneiderheinze, K. (2002), Stichprobenziehung für telefonische Zuwandererumfragen – Praktische Erfahrungen und Erweiterung der Auswahlgrundlage, in: Gabler, S. und S. Häder (Hg.), Telefonstichproben - Methodische Innovationen und Anwendungen in Deutschland, Münster: Waxmann, S. 187-208.

Humpert, A., Schneiderheinze, K (2000), Stichprobenziehung für telefonische Zuwandererumfragen – Einsatzmöglichkeiten der Namenforschung (Onomastik), in: ZUMA-Nachrichten 47, S36-63.

Knerr, P., Schiel, S. und Quandt, S. (2017): Sachstandsbericht. Deutscher Viktimisierungssurvey 2017. Bonn.

Rendtel, U. (1999). The Application of the Convex Weighting Estimator to House-hold Panel Surveys. Mimeo, Frankfurt; Spiess, M. & U. Rendtel (2000). Combining an ongoing panel with a new cross-sectional sample, DIW Discussion Paper No. 198, Berlin.

Schiel, S., Dickmann, C., Gilberg, R. und Malina, A. (2013): Repräsentative Bevölkerungsbefragung im Rahmen des BaSiD-Teilvorhabens „Sicherheitsgefährdungen durch Kriminalität“ – Methodenbericht. Bonn.

Schiel, S., Knerr, P. (2017): „Pretestbericht, Deutscher Viktimisierungssurvey 2017 – Repräsentative Bevölkerungsbefragung“. Bonn.

Anhang

Tabelle 30 Änderungen am Fragebogen im Vergleich zur Erhebung 2012

Fragenummern	Inhalte
Gestrichene Inhalte	
Frage 101	Gesundheitszustand
Block 400 komplett (damals Modul 2)	Mediennutzung
Frage 500/501 (damals Modul 5)	Kontrollüberzeugungen
Block 700 komplett (damals Modul 1)	Kriminalitätsfurcht letzte 7 Tage
Fragen 1713-1719, 1728 bis 1734	*Bei Raub:* Fragen zur Berichterstattung über Viktimisierungen in den Medien
Fragen 1913-1919, 1929 bis 1935	*Bei Körperverletzung:* Fragen zu Berichterstattung über Viktimisierungen in den Medien
Fragen 2011 und 2012	Geburtsland der Großeltern
Neue Inhalte	
Frage 103, Item I	Erweiterung der Itembatterie um Vertrauen in die Staatsanwaltschaft
Fragen 309 und 310	Indirekte Erfahrung mit der Polizei
Fragen 311-313, 317, 321-324 (teilweise in Modul 2, teilweise unabhängig davon)	Vertrauen gegenüber Gerichten
Block 3000	Schädigung durch Malware
Block 3100	Phishing: Preisgabe vertraulicher Daten nach Erhalt einer gefälschten E-Mail
Block 3200	Pharming: Preisgabe von Zugangsdaten nach Umleitung auf eine gefälschte Webseite
Fragen 318-320 (in Modul 2)	Soziale und ökonomische Deprivation
Frage 308 (Vignette S in Modul 2)	1 zusätzliche Vignette zur Raub
Frage 308 (Vignette W-T in Modul 2)	3 zusätzliche Vignetten zu Wohnungseinbruch
Fragen 619, Item E/ Frage 708, Item E (in Modul 5)	Einfügung eines Items zur Furcht vor Terroranschlägen in den Batterien zur affektiven und kognitiven deliktspezifischen Kriminalitätsfurcht
Frage 625 (in Modul 5)	Auswirkung der Angst vor Terroranschlägen auf die Lebensqualität
Frage 709	Itembatterie zur konativen Kriminalitätsfurcht
Fragen 1756, 1758	*Bei Raub:* Fragen zur Feststellung, ob Viktimisierungen der Vorurteilskriminalität (hate crime) zuzurechnen sind, sowie auf welches Merkmal sich ggf. die Motivation des Täters bezog
Fragen 1951, 1953	*Bei Körperverletztung:* Fragen zur Feststellung, ob Viktimisierungen der Vorurteilskriminalität (hate crime) zuzurechnen sind, sowie auf welches Merkmal sich ggf. die Motivation des Täters bezog
Frage 2024	Erfassung der Religionszugehörigkeit, falls die Person angegeben hat, aufgrund ihrer Religion Opfer von Vorurteilskriminalität geworden zu sein