

# Zensus 2011



## Methoden und Verfahren



# Zensus 2011

Methoden und Verfahren

**Herausgeber:**

Statistische Ämter des Bundes und der Länder

**Herstellung und Redaktion:**

Statistisches Bundesamt

65180 Wiesbaden

Telefon: +49 (0) 611 75-2405

Telefax: +49 (0) 611 75-3330

[www.destatis.de/kontakt](http://www.destatis.de/kontakt)

**Fachliche Informationen zu dieser Veröffentlichung:**

Statistisches Bundesamt

Bereich „Zensus“

Telefon: +49 (0) 611 75-4864

Telefax: +49 (0) 611 75-4000

Erscheinungsfolge: einmalig

Erschienen im Juni 2015

**Weiterführende Informationen:**

[www.statistikportal.de](http://www.statistikportal.de)

Fotorechte:

© Statistische Ämter des Bundes und der Länder

© Statistisches Bundesamt, Wiesbaden, 2015

(im Auftrag der Herausbergemeinschaft)

Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.

## Inhaltsverzeichnis

Abkürzungsverzeichnis .....	5
1 Allgemeine Angaben .....	7
2 Das neue Zensusmodell im Überblick .....	9
2.1 Der Weg zum registergestützten Zensus – Gründe für den Methodenwechsel .....	9
2.2 Zensustest 2001 .....	9
2.3 Datenquellen des Zensus 2011 – Register und Erhebungen .....	11
2.4 Zusammenführung der Datenquellen – Haushaltegenerierung .....	12
3 Datengewinnung und Datenaufbereitung .....	15
3.1 Datensicherheit .....	15
3.2 Anschriften- und Gebäuderegister ( AGR) .....	15
3.2.1 Aufgaben des AGR .....	15
3.2.2 Aufbau des AGR .....	16
3.2.3 Maßnahmen zur Sicherstellung der Vollständigkeit des AGR .....	18
3.3 Daten zur Bevölkerung .....	20
3.3.1 Daten der Einwohnermelderegister .....	20
3.3.1.1 Datengewinnung .....	20
3.3.1.2 Mehrfachfallprüfung im bundesweiten Bestand .....	22
3.3.1.3 Auflösung von Dubletten .....	23
3.3.2 Erwerbsstatistische Registerdaten: Daten der Bundesagentur für Arbeit sowie Daten von öffentlichen Arbeitgebern zu Beamten, Richtern und Soldaten .....	24
3.3.2.1 Datenquellen und Merkmale .....	24
3.3.2.2 Erhebung .....	25
3.3.2.3 Aufbereitung .....	25
3.3.2.4 Zusammenführung mit den Melderegisterdaten .....	25
3.3.3 Haushaltebefragung auf Stichprobenbasis (Haushaltsstichprobe) .....	26
3.3.3.1 Stichprobendesign .....	27
3.3.3.2 Datengewinnung .....	31
3.3.3.3 Datenaufbereitung .....	34
3.3.3.4 Hochrechnungsverfahren für die Ermittlung der Einwohnerzahlen der Gemeinden mit 10 000 und mehr Einwohnern .....	36
3.3.4 Daten aus den Erhebungen an Anschriften mit Sonderbereichen .....	40
3.3.4.1 Datengewinnung .....	40
3.3.4.2 Datenaufbereitung .....	41
3.3.5 Daten für Deutsche im Ausland .....	43
3.3.5.1 Datengewinnung .....	43
3.3.5.2 Mehrfachfallprüfung im bundesweiten Melderegisterdatenbestand .....	44
3.3.6 Daten der „Befragung zur Klärung von Unstimmigkeiten“ (BKU) .....	44
3.3.6.1 Zweck der Erhebung .....	44
3.3.6.2 Auswahl der Klärungsfälle und Durchführung der Erhebung .....	44
3.3.6.3 Datenaufbereitung .....	46

3.3.7	Zusammenführung der Daten aus den verschiedenen Datenquellen im Referenzdatenbestand (RDB) .....	46
3.3.7.1	Aufbau des RDB .....	46
3.3.7.2	RDB als Sammelpunkt von Informationen aus den verschiedenen Erhebungsteilen.....	46
3.3.7.3	Konsolidierung des RDB.....	48
3.4	Gebäude- und Wohnungszählung (GWZ).....	50
3.4.1	Datengewinnung: Vollerhebung als schriftliche Befragung der Eigentümer .....	50
3.4.1.1	Ermittlung der zählungsrelevanten Anschriften .....	50
3.4.1.2	Ermittlung der Auskunftspflichtigen für die zählungsrelevanten Anschriften.....	50
3.4.1.3	Postalische Datenerhebung.....	52
3.4.2	Plausibilisierung und Imputation der GWZ-Daten .....	53
3.4.2.1	Maschinelle Fehlererkennung und -beseitigung .....	54
3.4.2.2	Manuelle Prüfungen und Korrekturen .....	56
3.4.2.3	Nachträgliche Korrekturen .....	56
4	Generierung von Haushalten: die verschiedenen Stufen der Haushaltgenerierung sowie das integrierte Korrekturverfahren .....	57
4.1	Schritt 1: Bildung erster Haushaltszusammenhänge aus Registerinformationen – Verzeigerungen .....	57
4.2	Schritt 2: Auswertung der Wohnungsnutzerinformation und erste Verknüpfung von Haushalten mit Wohnungen .....	58
4.3	Schritt 3: Bildung weiterer Haushaltszusammenhänge aus Registerinformationen – Referenzen .....	58
4.4	Schritt 4: Integriertes Korrekturverfahren .....	58
4.5	Schritt 5: Zuweisung von Haushalten zu Wohnungen beziehungsweise zu bereits bestehenden Haushalten nach statistischen Generierungskriterien .....	59
4.6	Schritt 6: Typisierung der generierten Haushalte .....	60
5	Ergebniserstellung .....	61
5.1	Ermittlung der Einwohnerzahlen .....	61
5.2	Auswertungen - Gesamtheiten und Teilgesamtheiten .....	62
5.2.1	Abweichende Ergebnisse bei personenbezogenen Auswertungen.....	65
5.2.2	Unterschiede bei den Ergebnissen für Gebäude und Wohnungen .....	65
5.3	Geheimhaltung .....	67
5.3.1	Daten aus Vollerhebungen oder Registerauszählungen .....	67
5.3.2	Daten der Haushaltsstichprobe .....	69
6	Hintergrundinformationen .....	71
	Mitwirkende Personen .....	72
	Adressen der Statistischen Ämter des Bundes und der Länder.....	73

## Abkürzungsverzeichnis

AA	Auswärtiges Amt
AGR	Anschriften- und Gebäuderegister
Anschriften-ID	Anschriften-Identifikationsnummer
BKU	Befragung zur Klärung von Unstimmigkeiten
BKW	Befragung zur Klärung des Wohnsitzes
BMI	Bundesministerium des Innern
BMVg	Bundesministerium der Verteidigung
BRRG	Beamtenrechtsrahmengesetz
BStatG	Bundesstatistikgesetz
bzw.	beziehungsweise
CANCEIS	CANadian Census Edit and Imputation System
d. h.	das heißt
DSMeld	Datensatz für das Meldewesen
EDB	Eingangsdatenbank
EG	Europäische Gemeinschaft
ER	Erwerbsregister
ER-BA	Erwerbsregister Bundesagentur für Arbeit
ER-PS	Erwerbsregister Personalbestand
EU	Europäische Union
Eurostat	Statistikamt der Europäischen Union
GAB	Georeferenzierte Adressdaten Bund
GAD	Gesetz über den Auswärtigen Dienst
GWER	Gebäude- und Wohnungseigentümerregister
GWZ	Gebäude- und Wohnungszählung
HW-Personen	Hauptwohnsitz-Personen
i. d. R.	in der Regel
IDEV	Internet Datenerhebung im Verbund
ISCO	International Standard Classification of Occupations (deutsch: Internationale Standardklassifikation der Berufe)
IT	Informationstechnik
KIdB	Klassifikation der Berufe
MR	Melderegister
MR.1	Datenlieferung der Meldebehörden zum 9. Mai 2011
MR.2	Datenlieferung der Meldebehörden zum 9. August 2011
MR.Kons	Stichtagsrelevanter Ausgangsdatenbestand (konsolidierter Melderegisterbestand)
MRRG	Melderechtsrahmengesetz
NW-Personen	Nebenwohnsitz-Personen
o. a.	oben angegebenen
o. g.	oben genannte
PON	personenbezogene Ordnungsnummer
PZA	Postzustellungsauftrag
RDB	Referenzdatenbestand
SAFE	"Sichere Anonymisierung für Einzeldaten"
SAR	Sonderanschriftenregister
sog.	sogenannte
StichprobenV	Stichprobenverordnung Zensusgesetz
u. a.	unter anderem
u. U.	unter Umständen
VÖT1	Erstveröffentlichungstermin am 31. Mai 2013
VÖT2	Zweitveröffentlichungstermin am 10. April 2014
WiSta	Wirtschaft und Statistik
z. B.	zum Beispiel
ZensG	Zensusgesetz
ZensVorbG	Zensusvorbereitungsgesetz





# 1 Allgemeine Angaben

Die Statistischen Ämter des Bundes und der Länder führten eine Bevölkerungs-, Gebäude- und Wohnungszählung (Zensus) zum Stand 9. Mai 2011 (Berichtszeitpunkt) als Bundesstatistik durch.

Die Ermittlung der Einwohnerzahlen für Bund, Länder und Gemeinden zum Zensusstichtag als neue Basiszahlen für die Statistik war zentraler Zweck des Zensus 2011. Aus den statistischen Ergebnissen zu den Einwohnerzahlen werden durch den Verwaltungsakt des Feststellungsbescheids die „amtlichen“ Einwohnerzahlen, die von besonderer Bedeutung unter anderem für den Finanzausgleich, die Einteilung der Wahlkreise oder auch für die Sitzverteilung im Bundesrat sind, festgesetzt.

Der Zensus 2011 ist ebenso wie die früheren Volkszählungen ein wichtiger Eckpfeiler für das Gesamtsystem der Bevölkerungsstatistiken, auf dem andere Systemteile aufbauen. Die Strukturdaten aus dem Zensus 2011 sind die neue Ausgangsbasis für die Fortschreibung der Bevölkerung nach Geschlecht, Alter, Familienstand und Staatsangehörigkeit entsprechend dem Bevölkerungsstatistikgesetz. Die Zensusergebnisse bilden zudem für eine Reihe von Statistiken, wie zum Beispiel den Mikrozensus, die Auswahlgrundlage für die Stichprobenziehung genauso wie die Basis für den (fortgeschriebenen) Hochrechnungsrahmen. Die gebäude- und wohnungsstatistischen Ergebnisse des Zensus 2011 liefern beispielsweise die Basis für die Fortschreibung des Gebäude- und Wohnungsbestandes mit Hilfe der durch die Bautätigkeitsstatistik ermittelten jährlichen Zu- und Abgänge.

Der Zensus 2011 erstreckte sich auf alle nach den melderechtlichen Vorschriften zum Berichtszeitpunkt meldepflichtigen Personen sowie die im Ausland tätigen Angehörigen der Bundeswehr, der Polizeibehörden und des Auswärtigen Dienstes sowie ihre dort ansässigen Familien. Nicht erfasst wurden dagegen unter anderem in Deutschland lebende Mitglieder einer ausländischen diplomatischen Mission oder einer ausländischen konsularischen Vertretung sowie Angehörige ausländischer Streitkräfte und die jeweils mit ihnen im gemeinsamen Haushalt lebenden Familienmitglieder.

Erhebungseinheiten der Gebäude- und Wohnungszählung (GWZ) waren Gebäude mit Wohnraum, bewohnte Unterkünfte und Wohnungen. Nicht einbezogen wurden Gebäude, Unterkünfte und Wohnungen, die von ausländischen Staaten oder Angehörigen ausländischer Streitkräfte, diplomatischer oder berufskonsularischer Vertretungen genutzt werden und aufgrund internationaler Vereinbarungen unverletzlich sind.

Die nationalen Rechtsgrundlagen des Zensus 2011 bildeten das Zensusvorbereitungsgesetz (ZensVorbG 2011), das Zensusgesetz 2011 (ZensG 2011), die Stichprobenverordnung Zensusgesetz 2011 (StichprobenV) sowie die jeweiligen landesrechtlichen Regelungen.

Zusätzlich musste der Zensus 2011 der Erfüllung der Berichtspflichten nach der Verordnung (EG) Nr. 763/2008 des Europäischen Parlaments und des Rates vom 9. Juli 2008 über Volks- und Wohnungszählungen (ABl. L 218 vom 13. August 2008, S. 14) nachkommen. Zur Durchführung der Verordnung Nr. 763/2008 waren für den Zensus 2011 drei weitere Verordnungen zu beachten. Das war die Verordnung der Europäischen Union (EU) Nr. 519/2010 der Kommission vom 16. Juni 2010 zur Annahme des Programms der statistischen Daten und der Metadaten für Volks- und Wohnungszählungen, die Verordnung Europäische Gemeinschaft (EG) Nr. 1201/2009 der Kommission vom 30. November 2009 in Bezug auf die technischen Spezifikationen für die Themen sowie deren Untergliederungen sowie die Verordnung Europäische Union (EU) Nr. 1151/2010 der Kommission vom 8. Dezember 2010 in Bezug auf die Modalitäten und die Struktur der Qualitätsberichte sowie das technische Format der Datenübermittlung.



## 2 Das neue Zensusmodell im Überblick

### 2.1 Der Weg zum registergestützten Zensus – Gründe für den Methodenwechsel

Die letzte Volkszählung in der Bundesrepublik Deutschland im Jahr 1987 war eine „traditionelle“ Volkszählung, also eine Vollerhebung, die ebenfalls eine Gebäude- und Wohnungszählung umfasste. Zur Durchführung dieser Volkszählung wurden etwa 600 000 Erhebungsbeauftragte eingesetzt, die die damals 61 Millionen Einwohnerinnen und Einwohner der Bundesrepublik persönlich befragten. Sie war von teilweise erheblichen Akzeptanzproblemen von Seiten der Bevölkerung begleitet. Die regelmäßige Durchführung einer Volkszählung ist jedoch unerlässlich, da die Fortschreibung der Einwohnerzahlen aus den Volkszählungen im Laufe der Zeit immer unpräziser wird. Für weitreichende politische sowie wirtschaftliche Entscheidungen und Planungen, beispielsweise für die Durchführung von Wahlen, für Kommunalplanungen oder den Wohnungsbau werden jedoch zuverlässige und aktuelle Einwohnerzahlen benötigt. Dasselbe gilt für die Finanzausgleichssysteme, mit denen die unterschiedliche Finanzkraft der Körperschaften ausgeglichen wird. Für sie stellen die Einwohnerzahlen eine sehr wichtige Grundlage dar.

Das Bundesverfassungsgericht hat in seinem Urteil vom 15. Dezember 1983 zum Volkszählungsgesetz 1983 die Notwendigkeit von Volkszählungen anerkannt, gleichzeitig aber gefordert, dass sich der Gesetzgeber „vor künftigen Entscheidungen für eine Erhebung erneut mit dem dann erreichten Stand der Methodendiskussion auseinandersetzen muss, um festzustellen, ob und in welchem Umfang die herkömmlichen Methoden der Informationserhebung und -verarbeitung beibehalten werden können. Die Methoden der amtlichen Statistik und der Sozialforschung entwickeln sich stetig weiter. Diese Entwicklung darf der Gesetzgeber nicht unberücksichtigt lassen (...). Ebenso muss er bei der Anordnung einer statistischen Erhebung anhand des erreichbaren Materials prüfen, ob eine Totalerhebung trotz einer inzwischen fortgeschrittenen Entwicklung der statistischen und sozialwissenschaftlichen Methoden noch verhältnismäßig ist. Es reicht insoweit zur Begründung nicht aus, lediglich darauf zu verweisen, dass Volkszählungen schon immer in Form von Totalerhebungen durchgeführt worden seien.“ Dieses Bestreben des Bundesverfassungsgerichtes wurde mit dem Zensus 2011 erstmals umgesetzt. Anders als bei der „traditionellen“ Verfahrensweise wurde ein sogenannter registergestützter Zensus durchgeführt, bei dem bereits bestehende Verwaltungsregister – wie beispielsweise die der Meldebehörden – als Grundlage verwendet wurden und die Bevölkerung nur noch stichprobenweise befragt wurde, um die Über- und Untererfassungen (Karteileichen und Fehlbestände) der Register im Zensusdatenbestand zu bereinigen und um zusätzliche, in Registern nicht vorhandene Informationen, zu erheben.

### 2.2 Zensustest 2001

Zur Erprobung der neuen Methodik wurde von 2001 bis 2003 ein umfangreicher Zensustest<sup>1</sup> durchgeführt, um die Verfahrensweisen weiterentwickeln und in die Praxis umsetzen zu können, für die bis dahin nur ein grobes theoretisches Konzept bestand. Zudem sollte getestet werden, inwiefern die angenommenen Vorteile eines registergestützten Zensus tatsächlich zutreffen, das heißt, ob dieser im Vergleich zu einer als Vollerhebung durchgeführten Volkszählung eine kostengünstigere und bürgerfreundlichere Methode darstellt. Rechtliche Grundlage für den Zensustest war das Gesetz zur Erprobung eines registergestützten Zensus vom 27. Juli 2001 (Zensustestgesetz). Darin war festgelegt, dass „zur Erprobung eines registergestützten Zensusverfahrens (...) Testerhebungen, Untersuchungen von Registern und statistisch-methodische Untersuchungen als Bundesstatistik durchgeführt“ werden sollen. Dazu wurden die Daten der Einwohnermelderegister mit den Ergebnissen

<sup>1</sup> Siehe auch Statistische Ämter des Bundes und der Länder, „Ergebnisse des Zensustests“, in: WiSta 8/2004, S. 813-833.

verschiedener, durch Erhebungsbeauftragte durchgeführte Befragungen und der postalischen Gebäude- und Wohnungszählung in Beziehung gesetzt und verglichen. Die Überprüfung der Melderegister auf Über- und Untererfassungsfehler, der sogenannte Registertest, war neben dem Verfahrenstest, in dem verschiedene statistische Verfahren bewertet wurden, ein wesentliches Ziel des Zensustests. Bundesweit wurden für den Registertest etwa 38 000 Anschriften in 555 Gemeinden unterschiedlicher Größe für eine Haushaltebefragung auf Stichprobenbasis (Haushaltsstichprobe) ausgewählt. Bei einer Unterstichprobe im Umfang von rund 16 000 Anschriften wurden die Ergebnisse dieser Befragung weiterhin mit der postalisch durchgeführten Gebäude- und Wohnungszählung abgeglichen, um die dort getroffenen Angaben auf ihre Validität zu überprüfen. Auswahlgrundlage waren Melderegisteranschriften, daher sind unbemeldete Anschriften nicht in den Zensustest einbezogen worden. Ferner wurde ein Teil der Sonderbereiche (Gemeinschafts- und Anstaltsunterkünfte), darunter insbesondere Anstalten, von vornherein aus der Erhebung ausgeschlossen.<sup>2</sup>

Die Resultate des Zensustests zeigten, dass ein registergestützter Zensus in Deutschland grundsätzlich durchführbar und für die Gewinnung von zuverlässigen Einwohnerzahlen und Strukturdaten über die Bevölkerung geeignet ist. Jedoch schwankte die Qualität der Melderegisterdaten abhängig von der Gemeindegröße deutlich, obwohl in der Bundesrepublik Deutschland die allgemeine Meldepflicht in § 11 Melderechtsrahmengesetz (MRRG) festgehalten ist. Die dezentrale Führung der Melderegister durch die Einwohnermeldeämter und die unterschiedlichen Meldefristen in den Bundesländern können z. B. dazu führen, dass eine Person gleichzeitig mit zwei Hauptwohnsitzen oder auch nur mit Nebenwohnsitz(en) in Deutschland gemeldet ist. In der Realität kann es zum Beispiel vorkommen, dass eine Person in Gemeinde A eine sogenannte „Karteileiche“ ist – also eine Person, die in der Gemeinde A gemeldet ist, dort aber nicht lebt – während sie in Gemeinde B einen sogenannten „Fehlbestand“ darstellt – also in B lebt, dort aber nicht gemeldet ist. Der Zensustest ergab, dass die Anzahl von Über- und Untererfassungen in kleineren Gemeinden mit weniger als 10 000 Einwohnern wesentlich geringer ausfällt als in Gemeinden mit mindestens 10 000 Einwohnern. Im Nachgang zum Zensustest wurden daher für Gemeinden mit weniger als 10 000 Einwohnern und Gemeinden mit mindestens 10 000 Einwohnern unterschiedliche Modelle zur Korrektur der Über- und Untererfassungsfehler in den Melderegistern entwickelt. Für Gemeinden mit mindestens 10 000 Einwohnern wurde eine Haushaltebefragung auf Stichprobenbasis zur Ermittlung und statistischen Korrektur dieser Registerfehler vorgeschlagen und für die Gemeinden unter 10 000 Einwohnern eine gezielte Rückfrage an Anschriften, bei denen die Melderegisterinformationen nicht mit den Angaben der Gebäudeeigentümer zu den in den Wohnungen lebenden Personen aus der Gebäude- und Wohnungszählung übereinstimmen (= Befragung zur Klärung von Unstimmigkeiten). Das vorgeschlagene Verfahren gewährleistet insgesamt die Ermittlung von zuverlässigen Einwohnerzahlen, die gemessen an den Ergebnissen des Zensustests in vergleichbarer Qualität für alle Gemeinden vorliegen.

Weiterhin ergab der Zensustest, dass mit einer postalischen Befragung von Gebäude- bzw. Wohnungseigentümern zuverlässige Daten zur Struktur des Gebäude- und Wohnungsbestandes sowie zur Versorgung der Bevölkerung mit Wohnraum bereitgestellt werden können.

Mit dem aus dem Zensustest weiterentwickelten Modell eines registergestützten Zensus wird zum einen der finanzielle Aufwand des Zensus reduziert, zum anderen sinkt auch die Belastung der Bevölkerung deutlich. So wurden beim Zensus 2011 nur ca. 19 Millionen Eigentümer und Verwalter von Gebäuden und Wohnungen und ca. 10,2 Millionen Personen durch weitere primärstatistische Erhebungen befragt, wobei einige Bürgerinnen und Bürger dabei mehrfach Auskunft geben mussten, da sie beispielsweise als Wohnungseigentümer befragt wurden und unabhängig davon in der Stichprobenerhebung Auskunft geben mussten.

<sup>2</sup> Abgrenzungsprobleme führten dennoch dazu, dass an den ausgewählten Anschriften u. a. noch Anstalten, leerstehende oder gewerblich genutzte Gebäude vorgefunden wurden, was zu einem Ausfall von rund 3 % der ausgewählten Anschriften führte.

## 2.3 Datenquellen des Zensus 2011 – Register und Erhebungen

Für die Durchführung eines registergestützten Zensus war die Erstellung spezieller Datengrundlagen von elementarer Bedeutung. Resultierend aus den Erkenntnissen des Zensus-tests erfolgte für den Zensus 2011 der Aufbau eines Anschriften- und Gebäuderegisters, eines Gebäude-, Wohnungs- und Eigentümerregisters sowie eines Sonderanschriftenregisters. Dabei wurde u. a. auf verschiedene, bereits bestehende Verwaltungsregister zurückgegriffen. Schließlich schreibt das Zensusgesetz 2011 folgende Datenquellen für den Zensus 2011 fest, die im Rahmen dieser Veröffentlichung detailliert erläutert werden:

- **Daten gemäß Übermittlung durch die Meldebehörden und durch oberste Bundesbehörden (§ 3 ZensG 2011)**

Seitens der Meldebehörden wurden die jeweiligen Melderegister zum Stichtag 1. November 2010, zum Zensusstichtag (9. Mai 2011) sowie zu einem drei Monate später liegenden Stichtag (9. August 2011) an die Statistischen Ämter der Länder übermittelt. Dabei diente die dritte Datenlieferung der Erfassung sogenannter „temporärer Mehrfachfälle“. Das sind Personen, die sich erst nach dem Zensusstichtag rückwirkend in einer Gemeinde angemeldet haben und deshalb in der Melderegisterlieferung vom 9. Mai 2011 noch nicht enthalten waren. Aus der zweiten und dritten Datenlieferung wurde ein konsolidierter Melderegisterbestand erzeugt, der schließlich die Grundlage für den Zensus 2011 bildete: Er enthielt demzufolge alle stichtagsrelevanten gemeldeten Personen.

- **Daten gemäß Übermittlung durch die Bundesagentur für Arbeit (§ 4 ZensG 2011)**

Seitens der Bundesagentur für Arbeit wurden Registerangaben zu sozialversicherungspflichtig Beschäftigten, arbeitsuchenden und arbeitslosen Personen geliefert.

- **Daten gemäß Übermittlung durch die nach dem Finanz- und Personalstatistikgesetz auskunftspflichtigen Stellen (§ 5 ZensG 2011)**

Durch alle öffentlichen Arbeitgeber wurden Daten zu Beamten, Richtern, Soldaten sowie zu Dienstordnungsangestellten übermittelt.

- **Daten aus der Gebäude- und Wohnungszählung (§ 6 ZensG 2011)**

Da für die Bundesrepublik Deutschland kein zentrales Register über Gebäude und Wohnungen existiert, war es ein Ziel des Zensus 2011, alle Gebäude mit Wohnraum zum Stichtag 9. Mai 2011 zu erfassen und verschiedene Grunddaten – wie beispielsweise das Baujahr oder die Wohnfläche – zu erheben. Dazu wurden die Eigentümer beziehungsweise Verwalter von Gebäuden und Wohnungen postalisch befragt.

- **Daten aus der Haushaltebefragung auf Stichprobenbasis (§ 7 ZensG 2011)**

Diese Befragung in Gemeinden mit mindestens 10 000 Einwohnern<sup>3</sup> diente – wie bereits erwähnt – der statistischen Bereinigung der Melderegisterinformationen um Karteileichen und Fehlbestände, die auf der Grundlage der Haushaltebefragung auf Stichprobenbasis hochgerechnet wurden. Zudem wurden im Rahmen dieser Befragung zusätzliche Merkmale erhoben, die in den Registern nicht zur Verfügung stehen. Auch in kleineren Gemeinden wurde die Haushaltsstichprobe durchgeführt, um die zusätzlichen Merkmale auf der Kreisebene flächendeckend nachweisen zu können.

<sup>3</sup> Stichtag für diese Einteilung in Gemeinden mit mehr oder weniger als 10 000 Einwohnern war laut Zensusgesetz (§ 2 ZensG 2011 Abs. 6) der 31. Dezember 2009 (abweichend aufgrund von Gebietsstandsänderungen in einigen Bundesländern auch der 1. Januar 2010), auch wenn der Stichtag, für den die Zensusergebnisse festgestellt wurden, der 9. Mai 2011 war.

- **Daten aus der Wiederholungsbefragung (§ 17 Abs. 2 und 3 ZensG 2011)**

Nach § 17 Abs. 2 ZensG 2011 wurden durch die Statistischen Ämter der Länder etwa 5 % der Personen aus der Haushaltebefragung auf Stichprobenbasis durch andere Erhebungsbeauftragte erneut befragt, um die Qualität der beim Zensus 2011 für die Gemeinden ab 10 000 Einwohnern ermittelten Einwohnerzahlen zu überprüfen. Die Ergebnisse aus dieser Erhebung dienen außerdem zur Evaluation des Zensus 2011 und damit auch der Vorbereitung des Zensus 2021. In Gemeinden unter 10 000 Einwohnern wurden zur Überprüfung der Ergebnisse gemäß § 17 Abs. 3 ZensG 2011 bei 0,3 % der Bevölkerung die Ergebnisse der Haushaltsstichprobe mit den Feststellungen aus dem Melderegister verglichen.

- **Daten aus der Erhebung an Anschriften mit Sonderbereichen (§ 8 ZensG 2011)**

An allen Anschriften mit Sonderbereichen fand eine Vollerhebung statt. Sonderbereiche sind unter anderem Gemeinschafts- und Anstaltsunterkünfte, in denen Personen in der Regel längerfristig untergebracht bzw. wohnhaft sind. Dabei wird zwischen sensiblen Sonderbereichen (z. B. Behindertenwohnheimen) und nicht-sensiblen Sonderbereichen (z. B. Studentenwohnheimen) unterschieden. Eine Hochrechnung der festgestellten Karteileichen und Fehlbestände war nicht notwendig, da für alle Personen an den betreffenden Anschriften Daten erhoben wurden. An nicht-sensiblen Sonderbereichen wurden ebenso wie bei der Haushaltebefragung auf Stichprobenbasis Zusatzmerkmale erhoben. Für die Personen in sensiblen Sonderbereichen wurde keine Erhebung der Zusatzmerkmale durchgeführt.

- **Daten aus der Mehrfachfalluntersuchung (§ 15 ZensG 2011)**

Im Rahmen der Mehrfachfalluntersuchung wurde zunächst bundesweit geprüft, ob Personen am Zensusstichtag mit mehr als einer alleinigen Wohnung oder Hauptwohnung (Mehrfachfälle) gemeldet waren. Diese Mehrfachfälle wurden, sofern ausschließlich Gemeinden mit mindestens 10 000 Einwohnern betroffen waren, anschließend anhand des jeweiligen Einzugsdatums der betroffenen Personen maschinell bereinigt. War jedoch mindestens eine Gemeinde mit weniger als 10 000 Einwohnern betroffen, so wurde der tatsächliche Wohnungsstatus der Personen mit Hilfe eines Fragebogens zur Klärung des Wohnsitzes primärstatistisch festgestellt. Außerdem wurden alle Meldedatensätze ermittelt, nach denen eine Person bundesweit nur mit Nebenwohnung(en) gemeldet war, aber keine Hauptwohnung oder alleinige Wohnung hatte. Auch in diesem Fall wurde der tatsächliche Wohnstatus mit Hilfe des Fragebogens abgeklärt. Durch dieses Vorgehen war es möglich, allen Personen zum Stichtag 9. Mai 2011 eine Hauptwohnung bzw. eine alleinige Wohnung zuzuordnen.

- **Daten aus der Befragung zur Klärung von Unstimmigkeiten (§ 16 ZensG 2011)**

Diese Erhebung in Gemeinden mit weniger als 10 000 Einwohnern fand statt, wenn Unstimmigkeiten zwischen den Daten der Melderegister und den Angaben zu den Bewohnern in der Gebäude- und Wohnungszählung vorlagen. Auch hier fand keine Hochrechnung statt. Im Fokus stand die Klärung von Einzelfällen. Ebenso wie bei der Mehrfachfallprüfung handelte es sich um eine qualitätssichernde Maßnahme bei der Ermittlung der Einwohnerzahlen.

## 2.4 Zusammenführung der Datenquellen – Haushaltegenerierung

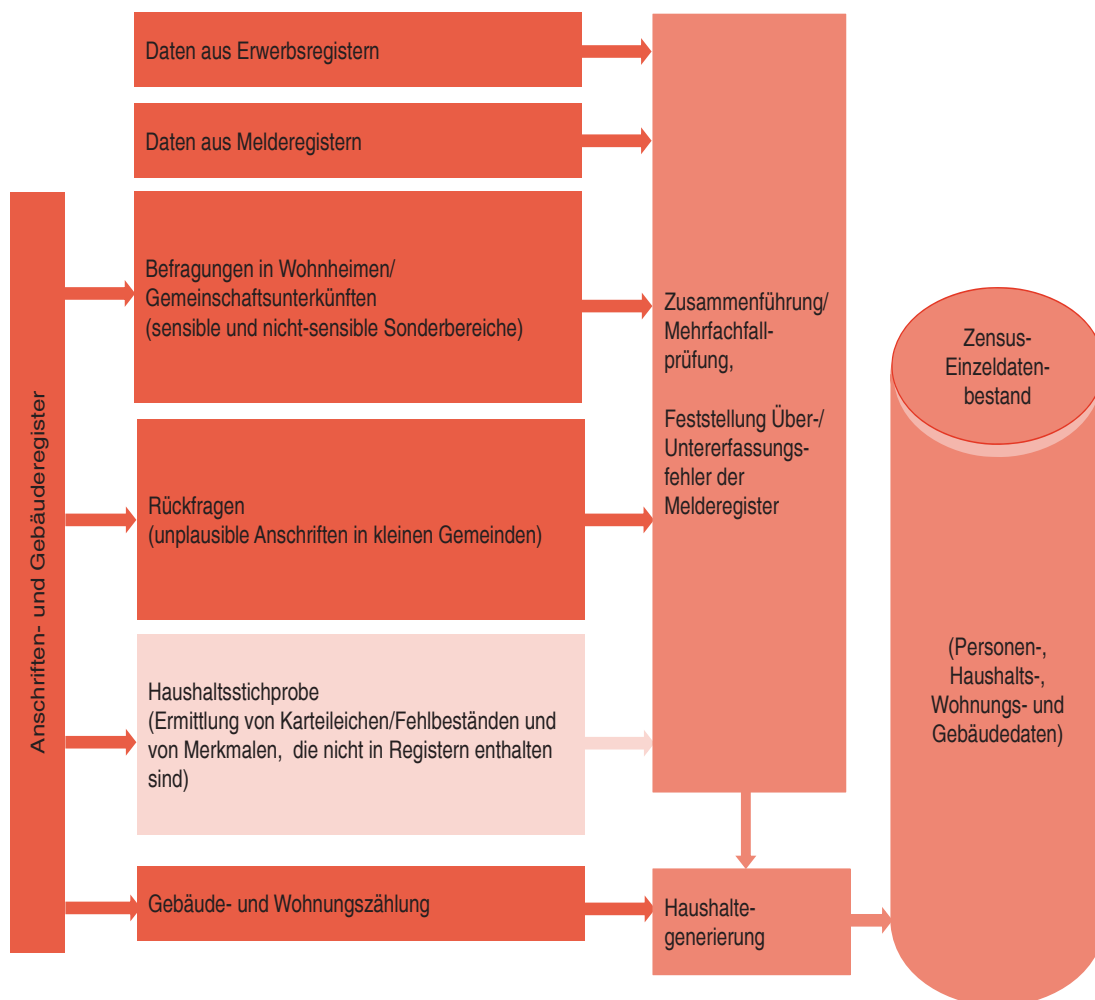
Die Daten aus allen Erhebungsteilen des Zensus 2011 wurden gemäß § 9 Abs. 3 ZensG 2011 zu kombinierten Datensätzen zusammengeführt und aus diesen anschließend im Rahmen der sogenannten Haushaltegenerierung statistische Haushalte gebildet. Dabei handelt es sich um ein automatisiertes Verfahren, das Informationen aus den Melderegistern mit Angaben aus der Gebäude- und Wohnungszählung anschriftenweise miteinander verknüpft und so statistische Wohnhaushalte generiert. Insgesamt sind folgende Schritte der Haushaltegenerierung von zentraler Bedeutung:

- Bildung von statistischen Haushalten unter Verwendung sogenannter Verzeigerungen, also Verbindungen bzw. Beziehungen zwischen Personen, die im Melderegister festgehalten sind (z. B. Familienstand oder Eltern-Kind-Beziehungen).

- Zuordnung der statistischen Haushalte zu Wohnungen auf Basis von Melderegisterinformationen und von Angaben zu Wohnungsnutzern im Rahmen der Gebäude- und Wohnungszählung (Bildung von Wohnhaushalten).
- Korrekturverfahren zur statistischen Bereinigung des Personen-Registerbestands um die Karteileichen und Fehlbestände, die aus der Haushaltebefragung auf Stichprobenbasis hochgerechnet worden waren.

Durch die Haushaltegenerierung können nicht mehr nur Aussagen darüber getroffen werden, welche Eigenschaften Personen, Gebäude und Wohnungen aufweisen, sondern auch, wie die Personen in Haushalten und Familien leben und wie ihre individuelle Wohnsituation aussieht. Zudem bietet die Verknüpfung der Vielzahl von Angaben und Daten schließlich einen einzigen, vollständigen zensus-typischen Datensatz, der als „großes Ganzes“ ausgewertet werden kann. Die Haushaltegenerierung ist also die Synthese der verschiedenen Erhebungsteile des Zensus 2011. Dadurch ermöglicht sie fachlich und regional tiefgegliederte, erhebungsteilübergreifende Kombinationsauswertungen.

### Das Zensusmodell im Überblick







## 3 Datengewinnung und Datenaufbereitung

### 3.1 Datensicherheit

Eine der zentralen Herausforderungen des Zensus 2011 war die Gewährleistung des Datenschutzes und der Informationssicherheit, was mit großen Anstrengungen und Aufwand im föderalen Zusammenwirken der Statistischen Ämter des Bundes und der Länder verbunden war.

Frühzeitig waren der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit sowie seine Kolleginnen und Kollegen in den einzelnen Bundesländern eingebunden und begleiteten das Projekt aktiv während der gesamten Laufzeit.

Bei der Datenübermittlung im Wege der Datenfernübertragung wurden dem jeweiligen Stand der Technik entsprechende Maßnahmen zur Gewährleistung von Datenschutz und Datensicherheit getroffen, um insbesondere die Vertraulichkeit, Integrität und Authentizität der Daten zu gewährleisten. Mussten allgemein zugängliche Netze genutzt werden, wurden entsprechende Verschlüsselungsverfahren eingesetzt.

Die Datenhaltung in den Statistischen Ämtern des Bundes und der Länder war besonders geschützt, überwacht und so angelegt, dass nur unbedingt notwendige Zugriffe auf die Daten durch einen eng begrenzten Nutzerkreis möglich waren. Die Erhebungsunterlagen und die Hilfsmerkmale<sup>4</sup> wurden jeweils nach Abschluss der entsprechenden Aufbereitungsschritte vernichtet bzw. gelöscht.

Beim Zensus 2011 wurde – wie bei jeder anderen amtlichen Statistik auch – das Rückspielverbot strikt beachtet und damit die Trennung von amtlicher Statistik und öffentlicher Verwaltung sichergestellt. Die erhobenen Einzeldaten wurden weder an staatliche noch an private Institutionen weitergegeben. Dies galt auch für diejenigen Fälle, in denen durch den Zensus 2011 ermittelt wurde, dass Personen fehlerhaft in den Melderegistern verzeichnet waren.

### 3.2 Anschriften- und Gebäuderegister (AGR)

#### 3.2.1 Aufgaben des AGR

Eine der großen Herausforderungen des Zensus 2011 bestand u. a. darin, Daten aus verschiedenen Erhebungsteilen so zu verbinden, dass am Ende für jede mit dem Zensus zu erfassende Person ein eindeutiger Datensatz mit allen geforderten Zensusmerkmalen gebildet werden konnte. Wegen der Heterogenität der verwendeten Datenquellen – Register und Vollerhebungen, Stichprobenerhebung und maschinelle Datengenerierung – war es notwendig, die Kommunikation zwischen den Erhebungsteilen zu ermöglichen. Dazu wurde ein Basisregister aufgebaut, das letztlich alle Anschriften mit Wohnraum und bewohnten Unterkünften enthielt, die es am 9. Mai 2011 in Deutschland gab. Dieses Anschriften- und Gebäuderegister (AGR) war über den gesamten Zensusablauf das Bindeglied zwischen allen Erhebungsteilen.

Die Aufgaben des AGR waren im Zensusvorbereitungsgesetz festgelegt worden:

Das Anschriften- und Gebäuderegister diente

1. der Steuerung des Ablaufs der Gebäude- und Wohnungszählung sowie der Ablaufkontrolle aller primärstatistischen<sup>5</sup> Erhebungen des Zensus,
2. zur Vorbereitung und als Auswahlgrundlage für die beim Zensus vorgesehenen Stichprobenerhebungen,

<sup>4</sup> Hilfsmerkmale sind Merkmale, die der technischen Durchführung einer Statistik dienen, z. B. der Name des Befragten. Sie müssen nach Abschluss der Überprüfung der Erhebungs- und Hilfsmerkmale auf Schlüssigkeit und Vollständigkeit wieder gelöscht werden.

<sup>5</sup> Eine Primärstatistik ist eine Datenerhebung, die eigens für die betreffende Untersuchung erhoben wird.

3. dazu, die Erhebungen für den Zensus zu koordinieren, im Rahmen der Durchführung des Zensus die aus verschiedenen Quellen stammenden Daten zusammenzuführen und die in den Zensus einzubeziehenden Gebäude, Wohnungen und Personen auf Vollständigkeit zu prüfen,
4. der Entwicklung eines Systems der raumbezogenen Analysen und Darstellungen von statistischen Ergebnissen und der Schaffung einer Grundlage für eine kleinräumige Auswertung des Zensus.

Das AGR hatte insgesamt koordinierende und standardisierende Funktionen. So wurden die verschiedenen Datenquellen durch die Bereitstellung eines umfassenden Anschriftenbestandes koordiniert. Außerdem legte der nach einheitlichen Kriterien standardisierte Anschriftenbestand die Basis dafür, dass in allen Erhebungsteilen auf die standardisierten Schreibweisen des AGR zurückgegriffen wurde. Das Datenmodell für den Zensus 2011 war so festgelegt, dass die Pflege der Anschriften nur einmal zentral im AGR stattfinden konnte. Das AGR stand damit als Leitdatei im Zensus an zentraler Stelle und war sowohl bei der Zensusvorbereitung, als auch bei der Zensusdurchführung relevant. Jeder Anschrift war eine Ordnungsnummer zugeordnet, die diese eindeutig identifizierte. Weitere Erkenntnisse und Informationen zu einer Anschrift konnten über diese Ordnungsnummer, die sogenannte Anschriften-Identifikationsnummer (Anschriften-ID), auch später eindeutig zugeordnet und abgespeichert werden.

### 3.2.2 Aufbau des AGR

Da das AGR bereits für zentrale Vorbereitungsarbeiten zum Zensus 2011 gebraucht wurde, musste mit seinem Aufbau frühzeitig begonnen werden. Zum Aufbau des AGR wurde auf drei verschiedene, bereits vorhandene Verwaltungsregister zurückgegriffen. Von diesen wurden ab 2008 an die Statistischen Ämter des Bundes und der Länder Daten übermittelt, die dann plausibilisiert, standardisiert, gegebenenfalls auf Anschriftenebene zusammengeführt und schließlich zu einem Gesamtbestand aggregiert wurden<sup>6</sup>. Das AGR enthielt alle am Zensusstichtag existierenden Anschriften mit Wohnraum und bewohnten Unterkünften, und somit alle Anschriften, an denen sich zum Zensusstichtag zensusrelevante Gebäude befanden und zensusrelevante Personen lebten. Zu den entsprechenden Anschriften wurden verschiedene Kennzeichen (z. B. Informationen darüber, in welcher der drei herangezogenen Datenquellen die Anschrift zu finden war) und Merkmale (beispielsweise geografische Hauskoordinaten) gespeichert, die weitere Informationen über die Anschrift enthielten.

Für den Aufbau des AGR wurden ab 2008 folgende Verwaltungsregister genutzt: die Georeferenzierten Adressdaten Bund (GAB) der Vermessungsbehörden, die Daten der Meldebehörden und Daten der Personenregister der Bundesagentur für Arbeit. Die liefernden Stellen übermittelten ihre Daten elektronisch zu festgelegten Stichtagen an die Statistischen Ämter des Bundes und der Länder.

Die Datenbestände der drei genannten Quellen unterschieden sich im Aufbau und in den darin enthaltenen Informationen. Die Registerauszüge der Vermessungsbehörden enthielten Anschriften aller Gebäude, die durch die Katasterbehörden der Länder vermessen worden waren, einschließlich der geografischen Hauskoordinaten. Diese Hauskoordinaten definieren dabei die genaue geografische Position eines Hauses. Die GAB-Datei enthielt sowohl Gebäude mit Wohnraum als auch rein gewerblich genutzte Gebäude und lieferte damit den umfangreichsten Bestand an Anschriften. Die Vermessungsdaten wurden durch das Bundesamt für Kartographie und Geodäsie auf Vollständigkeit überprüft, gegebenenfalls wurden fehlende Daten, insbesondere fehlende Hauskoordinaten, ergänzt und in vier jährlichen Lieferungen elektronisch an das Statistische Bundesamt übermittelt. Sie enthielten zu allen vermessenen Gebäuden in Deutschland diejenigen Angaben, die für die Zusammenführungen mit den Daten der Meldebehörden und der Bundesagentur für Arbeit benötigt wurden. Die erste Lieferung erfolgte zum 1. April 2008. Sie enthielt die Merkmale der Anschriften, die am 1. April 2007 gültig waren. Die zweite, dritte und vierte Lieferung erfolgte jeweils bis zum 31. Juli der Jahre 2008, 2009 und 2010.

Die zweite Quelle für den Aufbau des AGR waren Registerabzüge der Meldebehörden. Diese Daten wurden im Zensus 2011 für mehrere Zwecke verwendet. Zum einen wurden die Anschrifteninforma-

<sup>6</sup> Siehe auch Kleber, B. u. a., „Aufbau des Anschriften- und Gebäuderegisters für den Zensus 2011“, in: WiSta 7/2009, S. 629-640.

tionen aus den Registern für den Aufbau des AGR genutzt. Dazu wurden die Daten zu den Wohnanschriften der gemeldeten Personen ausgewertet. Zum anderen wurden im weiteren Zensusablauf demografische und anschriftenbezogene Daten aus den Lieferungen der Melderegister für methodische Untersuchungen zur Festlegung des Stichprobendesigns der Haushaltsstichprobe des Zensus 2011 ausgewertet. Die Melderegister enthalten – im Unterschied zu dem Register der Vermessungsbehörden – nur Anschriften, an denen Personen gemeldet sind. Die im Melderegister enthaltenen Anschriften bildeten dementsprechend in der Regel eine Untermenge der Anschriften der Vermessungsbehörden. Von den Meldebehörden wurden zwei Datenlieferungen zum Aufbau des AGR herangezogen: zu den Stichtagen 1. April 2008 und 1. April 2010. Hierbei lieferten die Gemeinden für jede gemeldete Person Angaben zur Anschrift und zu ausgewählten demografischen Merkmalen an das jeweils zuständige Statistische Amt des Landes. Dort wurden die Daten auf Plausibilität geprüft, gegebenenfalls bereinigt und in die beim Statistischen Bundesamt vorgehaltene Datenbank eingespielt.

Von der Bundesagentur für Arbeit wurde nur eine Datenlieferung zum Aufbau des AGR genutzt, die im April 2008 übermittelt wurde. Diese Lieferung enthielt alle am 30. September 2007 sozialversicherungspflichtig Beschäftigten und alle am 13. März 2008 arbeitslos gemeldeten Personen. Auch diese Daten dienten mehreren Verwendungszwecken: die Anschriftendaten zum Aufbau des AGR und die Angaben zum Erwerbsstatus der Personen (sozialversicherungspflichtig beschäftigt oder arbeitslos gemeldet) für die Stichprobenmodellierung der Haushaltsstichprobe.

Die übermittelten personenbezogenen Daten aus den Registern der Meldebehörden und der Bundesagentur für Arbeit wurden von den für den AGR-Aufbau benötigten Anschriftenmerkmalen getrennt und lediglich als anschriftenbezogene Aggregate (z. B. Zahl der unter einer Anschrift gemeldeten Personen) gespeichert.

Für die Erstellung des AGR mussten die drei Registerbestände auf Anschriftenebene zusammengeführt werden.<sup>7</sup> In einem ersten Schritt wurden die Datenquellen unabhängig voneinander auf Anschriftenebene aggregiert, also alle Datensätze, die zu einer Anschrift gehörten, zusammengefasst. Da in Deutschland für Anschriften kein eindeutiges und in allen Registern geführtes Verknüpfungsmerkmal (z. B. eine einheitliche Anschriften-Identifikationsnummer) existiert, musste für die Zusammenführung auf eine Kombination von Anschriftenmerkmalen zurückgegriffen werden. Da jede Datenquelle eine eigene Definition einer Anschrift verwendet und die entsprechenden Informationen in unterschiedlichen Merkmalen enthalten waren, wurden in einem zweiten Schritt die für die Zusammenführungen zu verwendenden Anschriftenmerkmale festgelegt und damit die Definition einer Anschrift harmonisiert. Als gemeinsame Merkmale, über die eine eindeutige Identifikation einer Anschrift in allen drei Registerbeständen möglich war, wurden folgende Angaben verwendet:

- Amtlicher Gemeindeschlüssel
- Postleitzahl
- Ortsteil
- Straße
- Hausnummer
- Hausnummernzusatz

Um bei der Zusammenführung der Datenquellen bessere Ergebnisse zu erzielen, wurden in einem dritten Schritt die einzelnen Anschriftenkomponenten standardisiert. Dabei wurden die Anschriftenmerkmale sowohl auf fehlerhafte Schreibweisen als auch auf veraltete Eintragungen geprüft. Fehlerhafte Schreibweisen traten insbesondere bei Straßennamen auf. Diese wurden kontrolliert, wenn bei Zusammenführungen bis zur Ebene der Straße einzelne Straßen durch keine zweite Quelle bestätigt werden konnten. Veraltete Eintragungen konnten insbesondere bei der Postleitzahl und dem amt-

<sup>7</sup> Siehe auch Ziprik, K., „Qualitätsaspekte des Anschriften- und Gebäuderegisters im Zensus 2011“, in: WiSta 11/2012, S. 947-956.

lichen Gemeindeschlüssel aufgedeckt werden. In diesen Fällen halfen Abgleiche mit aktuellen Verzeichnissen sowie die Verwendung von sogenannten Umsteigerlisten dabei, die Eintragungen anzupassen.

In einem vierten Schritt wurden schließlich die drei Datenquellen zu einem Gesamtbestand zusammengeführt. Das so entstandene Basisregister enthielt jede Anschrift, die in mindestens einem der drei Register enthalten war, einmal. Außerdem war bei jeder Anschrift vermerkt, in welcher der drei Basisdatenquellen sie enthalten war. Jede Anschrift, die über die oben aufgeführten Merkmale eindeutig definiert war, erhielt eine Anschriften-ID als Identifikator.

Das Ergebnis der Zusammenführungen der drei Verwaltungsregister wurde von den Statistischen Ämtern der Länder geprüft und gegebenenfalls korrigiert bzw. ergänzt. Diese Prüfungen bestanden im Wesentlichen darin,

- diejenigen Anschriften festzustellen, die nur in einem der drei Datenbestände enthalten waren und die Gründe hierfür zu klären (z. B. Umbenennungen von Straßen, die nicht in allen Registern vollzogen wurden). Im Sinne der Vollzähligkeit des Zensus waren dabei diejenigen Fälle von besonderer Bedeutung, in denen aus den Melderegistern keine Angaben zu gemeldeten Personen vorlagen (siehe hierzu Kapitel 3.2.3).
- den Meldebehörden die Anschriftenbereiche (z. B. Straßenzüge) mitzuteilen, für die Anhaltspunkte auf unvollständige oder fehlerhafte Datenlieferungen vorlagen. Die Meldebehörden mussten dann entweder bestätigen, dass die gelieferten Daten vollständig waren, oder den Datenbestand entsprechend ergänzen. Einzelprüfungen vor Ort waren hierbei nicht vorgesehen.

### 3.2.3 Maßnahmen zur Sicherstellung der Vollständigkeit des AGR

Insbesondere für die Nutzung des AGR als Auswahlgrundlage für die Haushaltsstichprobe des Zensus und damit für die Qualität der Zensusergebnisse war es besonders wichtig, dass das AGR alle Anschriften von Personen enthielt, die am Zensusstichtag zur zensusrelevanten Bevölkerung gehörten. Für die Sicherstellung der Vollständigkeit des AGR wurden folgende Maßnahmen ergriffen:

#### a) Weitere Prüfung der Anschriften, die nur in einem der drei Datenbestände enthalten waren

Bereits die Nutzung von drei Verwaltungsregistern für den Aufbau des Anfangsbestandes des AGR stellte eine Qualitätssicherung dar, da Anschriften, die durch alle drei Quellen bestätigt wurden, mit großer Wahrscheinlichkeit korrekt enthalten waren. Aus diesem Grund mussten Anschriften, die nur in einer der drei Quellen zu finden waren, daraufhin untersucht werden, ob es sich tatsächlich um existierende Anschriften mit Wohnraum oder bewohnten Unterkünften handelte. Diese Prüfung wurde gemäß § 14 ZensG 2011 von den Statistischen Ämtern der Länder durchgeführt. Zur Prüfung derjenigen Anschriften, für die entsprechende Prüfmarkierungen im AGR vorlagen, war ein gestuftes Verfahren vorgesehen, wobei in einer ersten Stufe ausschließlich in den Statistischen Ämtern der Länder vorhandene Unterlagen und allgemein zugängliche Quellen verwendet werden durften. Führt dies zu keinem Ergebnis, durften die Statistischen Ämter der Länder Unterlagen der nach Landesrecht für die Bauleitplanung, für das Meldewesen, für die Grundsteuer und für die Führung der Liegenschaftskataster zuständigen Stellen verwenden. Konnten keine Aussagen zum Vorhandensein von Wohnraum getroffen werden, war als letzte Prüfmöglichkeit eine schriftliche Erhebung bei den Eigentümern oder eine Begehung der Anschrift vorgesehen. Die Ergebnisse der Prüfungen wurden von den Statistischen Ämtern der Länder in das AGR integriert.

Die Feststellung von Anschriften, an denen es zwar ein Gebäude mit Wohnraum gab, für die aber keine Personen im Melderegister verzeichnet waren, war erforderlich, um diesen Anschriften eine Auswahlchance für die Haushaltsstichprobe des Zensus zu geben. Nur so war sichergestellt, dass mit der Haushaltsstichprobe auch sogenannte „komplette Fehlbestands-Anschriften“ erfasst werden konnten. Dabei handelte es sich um Anschriften, an denen zwar mindestens eine Person wohnte, für die aber im Melderegister keine Person gemeldet war. Eine Nichtberücksichtigung dieser Anschriften in der Haushaltsstichprobe hätte eine systematische Unterschätzung der Einwohnerzahlen zur Konsequenz gehabt.

### **b) Aktualisierung des AGR durch Update-Datenlieferungen der Vermessungsbehörden sowie der Meldebehörden**

Um die Aktualität des AGR-Datenbestandes sicherzustellen, wurden die weiteren Datenlieferungen der Vermessungsbehörden (vier Datenlieferungen mit jeweils ca. 21 Millionen Datensätzen) und der Meldebehörden (drei Datenlieferungen mit jeweils ca. 85 Millionen Datensätzen), die nach der Primärlieferung aus 2008 übermittelt wurden, verwendet. Jede Lieferung musste dazu wieder plausibilisiert und standardisiert werden. Durch die Updatelieferungen konnte sowohl sichergestellt werden, dass Änderungen in Anschriftenmerkmalen erkannt wurden und zu Korrekturen im Anschriftenbestand führen konnten als auch, dass neu hinzugekommene Anschriften (z. B. neu erschlossene Wohngebiete) im AGR ergänzt wurden.

### **c) Integration von Erkenntnissen aus den primärstatistischen Erhebungen in das AGR**

Die Vollständigkeit und Aktualität wurde schließlich auch durch die Integration von Erkenntnissen aus den primärstatistischen Erhebungen sichergestellt.

Für jede Anschrift im AGR, die Wohnraum enthielt, musste für die Durchführung der Gebäude- und Wohnungszählung (GWZ) mindestens ein Auskunftspflichtiger ermittelt werden. Im Rahmen dieser Ermittlung wurde das AGR vervollständigt, wenn über die Recherchen für die GWZ weitere, bisher im AGR nicht bekannte Anschriften mit Wohnraum entdeckt wurden. Gleichzeitig wurden so schon bei der Zensusvorbereitung die Anschriften der GWZ und des AGR abgeglichen.

Auch bei der Vorbereitung der Erhebungen an Anschriften mit Sonderbereichen fand eine ähnliche, frühzeitige Anbindung der ermittelten Sonderanschriften an die Anschriften des AGR statt, was zur Aufnahme neuer Anschriften führen konnte. Diejenigen Anschriften, an denen sich Sonderbereiche befanden, wurden im AGR entsprechend gekennzeichnet. Durch diese frühzeitige Markierung war es möglich, eine im Zensusablauf gesonderte Behandlung dieser Anschriften in den Erhebungsteilen GWZ und Haushaltsstichprobe sicherzustellen und z. B. für Anschriften von sensiblen Sonderbereichen bereits in der Zensusvorbereitungsphase Schutzvorkehrungen zu treffen und sie von der Auswahlgrundlage für die Haushaltsstichprobe auszuschließen.

Basierend auf den Anschrifteninformationen des AGR zu den zensusrelevanten Anschriften mit Wohnraum oder mit bewohnten Unterkünften ohne die sensiblen Sonderanschriften wurde in der Zensusvorbereitungsphase die Anschriftengrundgesamtheit für die Stichprobenziehung ermittelt. Diese stichprobenrelevanten Anschriften waren als solche im AGR markiert und wurden für die Ziehung in Schichten<sup>8</sup> unterteilt, wobei die Schichten mit Hilfe von Daten aus dem AGR (Gemeindegröße, Anschriftengröße) gebildet wurden. Die Ziehung der Anschriften erfolgte außerhalb des AGR, die gezogenen Anschriften wurden aber im AGR als Stichprobenanschriften markiert. Damit lag diese Information zentral allen weiteren Erhebungsteilen frühzeitig zur Verfügung.

Im Zensusverlauf konnte sich die Rolle einer Anschrift durch weitere Informationen aus den verschiedenen Erhebungsteilen ändern. Anschriften aus dem Anfangsbestand des AGR, für die die nachfolgenden Prüfungen ergeben hatten, dass es sich beispielsweise um die Anschrift eines rein gewerblich genutzten Gebäudes oder um die Anschrift eines unbebauten Grundstücks handelte, wurden lediglich als „nicht zensusrelevant“ markiert, aber nicht aus dem AGR gelöscht. Gab es später, beispielsweise im Rahmen der Vorbereitung der GWZ, gegenteilige Erkenntnisse zur Nutzungsart des Gebäudes, dann konnte die Anschrift durch einfache Änderung der Markierung wieder auf „zensusrelevant“ gesetzt werden.

Neben der Markierung der Anschriften in der Zensusvorbereitungsphase wurde das AGR auch in der Zensusdurchführungsphase fortlaufend durch die Integration von Erhebungsergebnissen aktualisiert; neue Anschriften und Anschriftenkorrekturen auf Basis von Registerlieferungen der Meldebehörden zum Zensusstichtag und drei Monate danach wurden im AGR berücksichtigt. So konnte sich beispielsweise durch die Stichprobenerhebung ergeben, dass an einer Anschrift kein Wohnraum oder aber kein Gebäude vorhanden ist. Diese Informationen wurden im AGR abgespeichert und standen so allen

<sup>8</sup> Siehe hierzu Kapitel 3.3.3.1 c) Auswahlverfahren.

Erhebungsteilen zur Verfügung. Die Nutzung der Anschrift als zentrales Koordinationsinstrument ermöglichte damit ein ergebnisorientiertes Arbeiten mit dem Ziel, einen plausiblen Gesamtdatenbestand zu erzeugen.

### 3.3 Daten zur Bevölkerung

Ein primäres Ziel des registergestützten Zensus 2011 war es, die Einwohnerzahlen für Bund, Länder und Gemeinden zum Zensusstichtag 9. Mai 2011 zu ermitteln. Zusätzlich sollte der Zensus 2011 grundlegende demografische Informationen zu der in Deutschland lebenden Bevölkerung liefern. Da es in Deutschland kein zentrales Register gab, in welchem alle erforderlichen Informationen vollständig enthalten waren, mussten Informationen aus verschiedenen Datenquellen in einem zentralen Datenbestand zusammengeführt werden. Die Basis des zentralen Datenbestands bildeten Informationen aus verschiedenen Registern, die um Informationen aus primärstatistischen Erhebungen ergänzt und zu einem Referenzdatenbestand (RDB) zusammengefasst wurden. Der Referenzdatenbestand hatte auch eine koordinierende Rolle. Er war zugleich der Grundstein für die Verknüpfung zwischen weiteren Erhebungsteilen und die Basis der Kommunikation zwischen den einzelnen Datenquellen. Er bildete unterschiedliche Bezugsgrößen auf Anschriften- und Personenebene ab. Über Identifikatoren ist jeder Personendatensatz eindeutig einer Anschrift aus dem AGR zugeordnet.

#### 3.3.1 Daten der Einwohnermelderegister

##### 3.3.1.1 Datengewinnung

Die Einwohnermelderegister, als grundlegende Datenquelle des Zensus 2011, werden in Deutschland in der Regel gemeindeweise geführt. Wegen dieser dezentralen Registerführung ist nicht sichergestellt, dass Anmeldungen in der Zuzugsgemeinde zeitgleich mit der Verarbeitung der elektronischen Abmeldungen in der Fortzugsgemeinde stattfinden. Ferner stehen den Meldepflichtigen Fristen zur Verfügung, innerhalb derer sie ihren Zuzug oder auch nur eine Veränderung in den persönlichen Meldemerkmale der zuständigen Meldebehörde anzeigen müssen. Hierdurch kommt es zu zeitlichen Verzögerungen (Time-Lags) bei der Verbuchung von zusammengehörigen Meldevorgängen. Um diese Verzögerungen beim Zensus 2011 berücksichtigen zu können, wurden die melderegisterführenden Stellen verpflichtet, den Statistischen Ämtern der Länder nicht nur für den Zensusstichtag (9. Mai 2011), sondern auch zusätzlich zum 9. August 2011 jeweils alle zu diesem Zeitpunkt gemeldeten Personen in elektronischer Form zu übermitteln. Aus diesen beiden Datenbeständen wurde für jede Gemeinde der stichtagsrelevante Bestand an gemeldeten Personen ermittelt.

Bei der Datenlieferung der Meldebehörden zum 9. Mai 2011 (MR.1) handelte es sich um den Ausgangsdatenbestand zum Aufbau des stichtagsrelevanten Zensus-Melderegisterbestandes. Aufgrund der bereits erwähnten Time-Lags bei den Meldevorgängen und zur Kompensation von deren Folgen erfolgte eine zweite Datenlieferung zum 9. August 2011 (MR.2). Diese diente dazu, diejenigen Veränderungen zu erfassen, die erst nach der Datenlieferung für den 9. Mai 2011 in das Melderegister eingetragen wurden, aber am 9. Mai 2011 schon wirksam waren. Das bedeutet, dass paarige Datensätze (das sind Datensätze zu identischen Personen, die in den Datenlieferungen MR.1 und MR.2 unter derselben Anschrift enthalten waren) aktualisiert und Datensätze zu Personen, die erstmals mit der zweiten Datenlieferung stichtagsrelevant enthalten waren, neu aufgenommen wurden.

In den Statistischen Ämtern der Länder wurden beide Datenlieferungen der Meldebehörden in einem ersten Schritt zunächst darauf überprüft, ob die gelieferten Dateien den formalen Vorgaben entsprachen. Dies umfasste:

1. Prüfung, ob für jede Gemeinde eine Datenlieferung erfolgt ist
2. Vergleich der Angaben auf dem Begleitschein mit der Anzahl der Datensätze in der Datenlieferung
3. Prüfung, ob der Datensatzaufbau der vorgegebenen Datensatzbeschreibung entspricht
4. Prüfung, ob der verwendete Zeichensatz der Datenlieferung zulässig ist



Im Anschluss wurden die von den Meldebehörden übermittelten Daten auf Vollständigkeit und Vollzähligkeit sowie Plausibilität geprüft und ggf. bearbeitet. Im Rahmen der Prüfungen auf Vollständigkeit und Vollzähligkeit wurde zuerst überprüft, ob die gelieferten Datensätze den technischen Vorgaben entsprachen, die für die sichere, vertrauliche und rechtsverbindliche Übertragung digitaler Daten der öffentlichen Verwaltung erforderlich sind.<sup>9</sup> Im Rahmen der Prüfung auf Vollständigkeit wurde dabei geprüft, ob die Pflichtmerkmale in jedem Datensatz befüllt waren, und im Rahmen der Vollzähligkeitsprüfung, ob für jede Gemeinde Datensätze im erwarteten Umfang geliefert wurden. Die Referenz zur Ermittlung der Vollzähligkeit bildete dabei ein eigens generiertes Verzeichnis, in dem alle Gemeinden und die Anzahl der zu erwartenden Hauptwohnsitzpersonen, differenziert nach Geschlecht, enthalten waren. Grundlage hierfür bildeten die jeweils aktuell vorliegenden Ergebnisse der Bevölkerungsfortschreibung. Die datenliefernden Stellen wurden anhand einer Quittung insbesondere über die Anzahl der gelieferten Datensätze je Gemeinde informiert. Waren Datenlieferungen unvollständig oder enthielten sie falsche Inhalte, forderten die Statistischen Ämter der Länder Neulieferungen an. Nach erfolgreicher Plausibilisierung der Datenlieferungen wurden die Daten gemeindeweise zum Zwecke der Weiterverarbeitung für das Statistische Bundesamt freigegeben.

Nach der Freigabe der Daten erfolgte die Berechnung der personenbezogenen Ordnungsnummer (PON). Durch das Ordnungsnummernsystem konnte jede Person direkt an einer Anschrift identifiziert werden. Damit wurde durch die PON eine maschinelle Zusammenführung der Datenlieferungen der Melderegister erleichtert. Sie baute auf den im Normalfall unveränderlichen Merkmalen Geburtsname, Vornamen, Geburtsdatum, Geschlecht und Geburtsort der jeweiligen Person auf. Sie wurde nach der Berechnung auch dann nicht mehr verändert, wenn es Änderungen in den – in der Regel unveränderlichen – Merkmalen gab, da ihr in allen Phasen des Zensus eine zentrale Rolle bei der Identifikation einer Person unter einer Anschrift zukam. Bei der Berechnung und Verschlüsselung der PON waren sehr hohe datenschutzrechtliche Anforderungen zu beachten.

Im Statistischen Bundesamt wurden die von den Statistischen Ämtern der Länder zur Weiterverarbeitung freigegebenen Melderegisterdatensätze (einschl. PON) automatisiert in eine zentrale Datenbank integriert. Jedem Personendatensatz wurde dabei eine Anschriften-ID basierend auf dem AGR zugeordnet. Damit war jeder Datensatz über die Kombination aus Anschriften-ID und PON eindeutig bezeichnet. Mit den Melderegisterdatensätzen zum Zensusstichtag (MR.1) wurde so der Anfangsbestand des stichtagsrelevanten Melderegisterbestandes gebildet. Bei der späteren Integration der MR.2-Daten in den MR.1-Bestand wurde zunächst immer geprüft, ob ein Datensatz aus MR.2 bereits in MR.1 enthalten war. Dies war immer dann der Fall, wenn sich für eine Person die Meldedaten zwischen den MR-Ständen nicht geändert hatten, sie also nicht umgezogen war und es auch sonst keine Veränderungen an den Meldedaten gab.

Wurde der Datensatz einer Person aus MR.1 nicht durch MR.2 bestätigt, blieb der MR.1-Satz bestehen, da davon auszugehen war, dass die Gemeinde stichtagsgenau (9. Mai 2011) geliefert hatte, die Person also zum Stichtag unter der in den MR.1-Daten enthaltenen Anschrift zu zählen war. Diese Konstellation konnte verschiedene Ursachen haben. War die Person innerhalb Deutschlands umgezogen, war sie aufgrund MR.1 mit ihrer bisherigen Anschrift im Bestand und sie war zusätzlich mit ihrer neuen Anschrift im MR.2-Bestand (in der gleichen Gemeinde oder in einer anderen Gemeinde) vorhanden. Dieser MR.2-Datensatz wurde dann in den stichtagsrelevanten Ausgangsdatenbestand (MR.Kons) aufgenommen. In späteren Arbeitsschritten wurde geprüft, wo und wie Personen bei mehrfachem Vorhandensein zu zählen waren.

Auch für Datensätze von Personen, bei denen die (in der Regel unveränderlichen) Grundmerkmale Geburtsname, Vornamen, Geschlecht, Geburtsdatum und Geburtsort in MR.2 anders waren, wurde der MR.2-Datensatz in MR.Kons als „neu“ aufgenommen. Innerhalb einer Gemeinde konnten deshalb Dubletten nicht nur durch innergemeindliche Umzüge, sondern auch durch Änderungen in den o. a.

<sup>9</sup> Gefordert wurden die Vorgaben des Datensatzes für das Meldewesen (DSMeld). Der DSMeld regelt Inhalt und Format von in den Melderegistern bundeseinheitlich zu speichernden Informationen. Er soll zudem die Datenspeicherung und -übermittlung transparent machen und die organisatorischen Voraussetzungen dafür schaffen, dass Datenübermittlungen zwischen den Meldebehörden und anderen Behörden technisch übersichtlich und richtig zu realisieren sind.

Merkmale entstehen, z. B. dann, wenn die Reihenfolge von Vornamen geändert wurde oder wenn Geburtsorte in der einen Lieferung erläuternde Anmerkungen enthielten und in der anderen nicht. Handelte es sich dabei um Dubletten innerhalb einer Anschrift, so wurden diese in einem weiteren Arbeitsschritt zusammengefasst.

Für das Verständnis der unterschiedlichen Vorgehensweise in den dezentral geführten Melderegistern und in einem zentralen Meldebestand, der über zeitlich versetzte Datenlieferungen aktualisiert wurde, ist entscheidend, dass die Verarbeitung im Meldewesen der Gemeinde bei Veränderungen an den Grundmerkmalen eine andere ist, als die im konsolidierten Meldebestand. In der Gemeinde wird der Datensatz selbst geändert und kein neuer Datensatz angelegt. Selbst bei innergemeindlichen Umzügen sollte im Melderegister der Gemeinde sichergestellt sein, dass für eine Person keine zwei aktiven Datensätze existieren.

Beim Zensus 2011 war die Vorgehensweise aufgrund der Notwendigkeit, Datenlieferungen zu zwei Stichtagen zu verarbeiten, eine andere. Um keinen Datensatz vorzeitig abzuweisen, wurde zunächst alles, was den Anschein hatte „neu“ zu sein, aufgenommen. Bei Umzug einer Person innerhalb einer Gemeinde waren im stichtagsrelevanten Ausgangsdatenbestand (MR.Kons) für diese Person zwei Datensätze (= Dublette) vorhanden. Während im Meldewesen die Änderung des Datensatzes einer innerhalb der Gemeinde umziehenden Person synchron erfolgt (Anmeldung an der neuen Anschrift und gleichzeitig Abmeldung an der alten Anschrift), verlief dies beim Zensus asynchron, da zwei zeitversetzt gelieferte Datensätze erst nachträglich synchronisiert wurden. In MR.Kons wurde erst im Rahmen der Mehrfachfallprüfung die Dublette aufgelöst, indem die Person an der Fortzugsanschrift gelöscht und damit nur an der Zuzugsanschrift gezählt wurde. Die Mehrfachfallprüfung holte in MR.Kons also das nach, was im Meldewesen durch synchrone Bearbeitung in einem Arbeitsgang passiert. Diese Vorgehensweise wurde auch bei Änderungen an den Grundmerkmalen einer Person angewandt.

Wenn eine Person nach dem Zensusstichtag ins Ausland verzogen oder verstorben war, gab es in der Regel keinen MR.2-Datensatz für diese Person. Gleichwohl war die Person zum Stichtag mit ihren MR.1-Daten im Bestand vorhanden und war zu zählen. Personen, die noch vor dem Stichtag 9. Mai 2011 verstorben oder ins Ausland gezogen waren, ohne dass dies am Stichtag 9. Mai 2011 schon im Melderegister verzeichnet war – für sie lag also ebenfalls ein MR.1-Datensatz und in der Regel kein MR.2-Datensatz vor –, konnten in MR.Kons nicht als Personen erkannt werden, die nicht zu zählen waren. Eine etwaige Korrektur dieser Übererfassungen in den Meldedaten erfolgte über die Befragung an Anschriften mit Sonderbereichen, über die Haushaltsstichprobe in Gemeinden mit mindestens 10 000 Einwohnern und über die Befragung zur Klärung von Unstimmigkeiten in den Gemeinden mit weniger als 10 000 Einwohnern.

### 3.3.1.2 Mehrfachfallprüfung<sup>10</sup> im bundesweiten Bestand

Bei der dezentralen Führung der Melderegister ist nicht auszuschließen, dass Personen in mehreren Gemeinden gleichzeitig mit mehr als einer alleinigen Wohnung oder Hauptwohnung oder ausschließlich mit (einer oder mehreren) Nebenwohnung(en) gemeldet waren. Davon zu unterscheiden sind mehrfache Eintragungen zum Wohnsitz bei einer Haupt- und einer oder mehreren Nebenwohnungen, die laut Melderecht zulässig sind. Dubletten (zwei Personendatensätze, die für dieselbe Person stehen) mussten also nicht grundsätzlich problematisch sein, erschwerten aber die Verarbeitung insofern, als unzulässige von zulässigen Mehrfachanmeldungen zu unterscheiden waren.

Bei den im Zensus 2011 näher zu untersuchenden Dubletten handelte es sich somit um Datensätze von Personen, die im stichtagsrelevanten Melderegisterbestand zum 9. Mai 2011 mehrfach mit dem Wohnungsstatus Hauptwohnung und/oder alleinige Wohnung enthalten waren. Deren Existenz ist – wie bereits erwähnt – in der dezentralen Melderegisterführung zu suchen, aber auch in den gesetzlichen

<sup>10</sup> Siehe auch: Michel, B., „Ergebnisse der Mehrfachfallprüfung im Rahmen des Zensustests“, in: WiSta 1/ 2004, S. 979-991 und Diehl, E.-M., „Methoden der Mehrfachfallprüfung im Zensus 2011“, in: WiSta 6/2012, S. 473-484.



Fristen zur Anmeldung nach Einzug in eine Wohnung und im Konzept zum Aufbau des konsolidierten Melderegisterbestandes zum 9. Mai 2011. Alle diese Dubletten konnten unterschieden werden in

1. temporäre Dubletten, die im Rahmen der normalen Fristen der An- und Ummeldung entstehen können, und
2. dauerhafte Dubletten, die über die normalen Fristen der An- und Ummeldung hinaus im Bestand enthalten sind.

In beiden Fällen ist eine Person entweder bereits in MR.1 zweimal an verschiedenen Anschriften oder in MR.1 an der Anschrift A und in MR.2 an der Anschrift B gemeldet. Die Frage, ob eine Dublette eine temporäre oder eine dauerhafte Dublette ist, ist für das Zensusergebnis irrelevant, da die Dublettenprüfung im stichtagsrelevanten Ausgangsdatenbestand (MR.Kons) durchgeführt wurde. Aufgabe war, den Stichtagsbestand abzubilden und festzulegen, ob eine Person zum Stichtag unter einer Anschrift entweder zu zählen ist oder nicht.

Damit mussten zunächst alle Dubletten ermittelt und im Anschluss klassifiziert werden in solche, die zulässig waren (Haupt-/Nebenwohnsitzdubletten) und solche, die aufgelöst werden mussten. Hierzu wurden in einem ersten Arbeitsschritt die Namensfelder aller Datensätze standardisiert.<sup>11</sup> Diese standardisierten Felder bildeten in Verbindung mit dem Geburtsdatum und dem Geschlecht die Grundlage für die sich anschließende Dublettensuche.<sup>12</sup> Ziel der Dublettensuche musste es sein, diejenigen Dubletten zu identifizieren, die die Dubletteneigenschaft auch mit überwiegender Wahrscheinlichkeit erfüllten. Hierzu kam ein zweistufiges maschinelles Verfahren zur Anwendung, um aus dem 85 Millionen Datensätze umfassenden stichtagsrelevanten Ausgangsdatenbestand (MR.Kons) die potenziellen Dubletten zu ermitteln. Zur Qualitätssicherung dieser Dublettensuche gab es schließlich noch einen maschinellen Abgleich und eine manuelle Sichtprüfung. Im Rahmen des maschinellen Abgleichs erfolgte die Aufteilung des „Dublettenpools“ in Dubletten mit einer hundertprozentigen Wahrscheinlichkeit der Übereinstimmung und in Dubletten, die noch einer manuellen Qualitätskontrolle bedurften. Bei diesem letzten Prüfschritt standen den Bearbeiterinnen und Bearbeitern für alle nicht zu 100 % übereinstimmenden Dublettenpaare nicht nur die Grundmerkmale der Mehrfachfallprüfung zur Verfügung, sondern zusätzlich alle weiteren relevanten Merkmale, auf deren Grundlage sich eine eindeutige Entscheidung treffen ließ, ob es sich um eine Dublette handelte oder nicht. Hierzu zählten unter anderem der Familienstand, das Eheschließungs-/Scheidungsdatum, Namensbestandteile, Informationen zur derzeitigen und zur Zuzugsanschrift, zum Ehe-/Lebenspartner und zu den Kindern. Lagen hier eindeutige Informationen vor, dass es sich um dieselbe Person handelte, waren die beiden Personendatensätze als einem Menschen zugehörig (paarig) zu bewerten.

Ferner ließen sich durch Abgleich mit dem stichtagsrelevanten Ausgangsdatenbestand (MR.Kons) diejenigen Datensätze herausfiltern, die als Wohnungsstatus Nebenwohnung verzeichnet hatten, zu denen aber im Rahmen der Dublettensuche kein korrespondierender Hauptwohnungssatz ermittelt werden konnte.

### 3.3.1.3 Auflösung von Dubletten

Nach Abschluss der maschinellen und manuellen Dublettenprüfung wurden die ermittelten Dublettenpaare (nur Haupt- und/oder alleinige Wohnungsdubletten) sowie Datensätze von Nebenwohnungen, die mit keiner Haupt- oder alleinigen Wohnung verknüpft werden konnten, zur weiteren Klärung an die Statistischen Ämter der Länder übergeben. Diese stellten für alle Personen, die nur mit Nebenwohnungen gemeldet waren, sowie für Personen mit mehr als einer alleinigen Wohnung oder Hauptwohnung, die in Gemeinden mit weniger als 10 000 Einwohnern gemeldet waren, den Wohnungsstatus fest (§ 15 Abs. 3 und 4 ZensG 2011). Die Befragung der hinter den vermuteten Dublettenelementen stehenden Personen wurde in der Regel als schriftliche Befragung durchgeführt (BKW – Befragung

11 Zu Einzelheiten der Standardisierung siehe Diehl, E.-M., „Methoden der Mehrfachfallprüfung im Zensus 2011“, in: WiSta 6/2012, S. 474 ff.

12 Zu Einzelheiten der Dublettensuche siehe Diehl, E.-M., „Methoden der Mehrfachfallprüfung im Zensus 2011“, in: WiSta 6/2012, S. 476 ff.

zur Klärung des Wohnsitzes). Der beziehungsweise die Auskunftspflichtige musste den Wohnungstatus bestätigen oder diesen ändern und die Hauptwohnung, die tatsächlich zum 9. Mai 2011 bestand, beziehungsweise weitere zum Stichtag bestehende Wohnungen angeben.

Parallel zur Verarbeitung der Rücklaufergebnisse aus dieser primärstatistischen Klärung wurden Personendatensätze aus Hauptwohnungsdubletten, welche sich ausschließlich in Gemeinden mit mindestens 10 000 Einwohnern befanden, durch das Statistische Bundesamt rein maschinell geklärt. Hier wurde anhand der Einzugs- beziehungsweise Anmeldedaten, welche zu den entsprechenden Datensätzen hinterlegt waren, die aktuelle Anschrift ermittelt. Abhängig davon, ob diesen Dubletten-elementen außerdem eine Nebenwohnung zugeordnet werden konnte, wurde auch hier abschließend die korrekte Ausprägung des Wohnungsstatus vergeben.

Sonstige Dubletten wurden unter der jeweiligen Anschrift maschinell zusammengefasst. Als Besonderheit war hierbei zu beachten, dass bei der Zusammenfassung immer die aktuellsten zensusrelevanten Informationen übernommen wurden.

### **3.3.2 Erwerbsstatistische Registerdaten: Daten der Bundesagentur für Arbeit sowie Daten von öffentlichen Arbeitgebern zu Beamten, Richtern und Soldaten**

#### **3.3.2.1 Datenquellen und Merkmale**

Ein umfassendes (zentrales) Register mit Informationen zu allen erwerbsstatistisch relevanten Personengruppen existiert in Deutschland nicht. Im Zensus 2011 wurden deshalb die vorhandenen und für den Zensus 2011 geeigneten erwerbsstatistischen Registerdaten ausgezählt und die nicht in diesen Registern vorhandenen Merkmale sowie die nicht von diesen Registern erfassten Personengruppen über die Haushaltsstichprobe ermittelt.

Für die Gewinnung von erwerbsstatistischen Informationen wurden beim Zensus 2011 folgende Datenquellen genutzt:

- Registerangaben der Bundesagentur für Arbeit zu allen sozialversicherungspflichtig Beschäftigten (ausgenommen ausschließlich geringfügig Beschäftigte) für rund 28,2 Millionen Personen
- Registerangaben der öffentlichen Arbeitgeber zu Beamtinnen und Beamten, Richterinnen und Richtern und Soldatinnen und Soldaten sowie zu den Dienstordnungsangestellten für rund 1,97 Millionen Personen
- Angaben der Haushaltebefragung (erwerbsstatistische Zusatzmerkmale) auf Stichprobenbasis mit etwa 7,9 Millionen Befragten einschließlich Angaben der Befragung an Adressen mit nicht-sensiblen Wohnheimen und Gemeinschaftsunterkünften mit rund 122 000 befragten Personen

Mit den vorhandenen Registerdaten konnten die EU-Pflichtmerkmale „Erwerbstätigkeit“, „Stellung im Beruf“ und „Wirtschaftszweig“ bedient werden. Das Merkmal „(derzeit ausgeübter) Beruf“ war zwar auch in den Registerdaten der Bundesagentur für Arbeit zu den sozialversicherungspflichtig Beschäftigten enthalten, eine Klassifizierung nach der geforderten internationalen Berufsklassifikation ISCO 08 wäre allerdings nur mit einem erheblichen Qualitätsverlust umsetzbar gewesen.<sup>13</sup> Das Merkmal „Beruf“ wurde deswegen ausschließlich über die Haushaltsstichprobe als Klartextangabe erhoben.<sup>14</sup>

Für die nicht in den Registerdaten enthaltenen Erwerbstätigen, dazu gehören die Selbstständigen und die mithelfenden Familienangehörigen sowie Personen, die in der Haushaltsstichprobe eine bezahlte Tätigkeit angegeben hatten, jedoch nicht bei der Bundesagentur für Arbeit gemeldet waren, wurden alle erwerbsstatistischen Grunddaten immer über die Haushaltsstichprobe erhoben.

<sup>13</sup> Es hätte von der älteren nationalen Klassifizierung der Berufe 1988 (KldB 88) der Bundesagentur für Arbeit (3-Steller) in die neue internationale Berufsklassifikation ISCO 08 (1-Steller) umgeschlüsselt werden müssen. Dies hätte einen erheblichen Qualitätsverlust zur Folge gehabt. Über die Haushaltsstichprobe kann der Beruf nach der aktuellen nationalen Klassifikation der Berufe 2010 (5-Steller) nachgewiesen und qualitativ besser in die ISCO 08 (1-Steller) umgeschlüsselt werden.

<sup>14</sup> Für eine ausführliche Methodenbeschreibung siehe auch Loos, C. u. a., „Das Verfahren der Berufskodierung im Zensus 2011“, in WiSta 3/ 2013, S.173-184.

### 3.3.2.2 Erhebung

Bei der Erhebung der erwerbsstatistischen Registerdaten waren die Daten der Bundesagentur für Arbeit aus einer Hand verfügbar, die Daten der personalführenden Stellen der öffentlichen Arbeitgeber jedoch nur über die Erhebung bei einer Vielzahl (über 14 000) auskunftspflichtiger Stellen. Bei den Daten der personalführenden Stellen der öffentlichen Arbeitgeber war die Verantwortlichkeit für die Datenerhebung und -aufbereitung nach Zuständigkeit auf die Statistischen Ämter des Bundes und der Länder verteilt, während diese für die Daten der Bundesagentur für Arbeit allein beim Statistischen Bundesamt lag.

### 3.3.2.3 Aufbereitung

Bedingt durch das Übermittlungsverfahren waren die Registerdaten der Bundesagentur für Arbeit zunächst als Rohdaten ohne inhaltliche Prüfungen anzusehen. Im Gegensatz dazu wurden die Daten der personalführenden Stellen der öffentlichen Arbeitgeber teilweise, je nach verwendetem Übermittlungsverfahren, bereits während der Übertragung einer ersten inhaltlichen Prüfung unterzogen. Dieses Vorgehen gewährleistete einen qualitativen Mindeststandard für eine erfolgreiche Datenübermittlung.

Um den Qualitätsanforderungen im Zensus 2011 gerecht zu werden, war die Entwicklung eines Softwaresystems notwendig, das die vollständige Erfassung aller Erwerbsregisterdaten sowie die Datenprüfung und -aufbereitung beinhaltete und diese für anschließende Folgeprozesse bereitstellte. Das oberste Ziel war dabei, valide Daten über den Bestand und die Strukturen der Erwerbstätigen in Deutschland zu gewinnen, die personenscharf mit den Daten der Melderegister verbunden werden können. Hierfür wurden zwei elektronische Register, das Erwerbsregister – Teil Bundesagentur für Arbeit (ER-BA) und das Erwerbsregister – Teil Personalstand (ER-PS) konzipiert.

Die ersten Datenprüfungen innerhalb der elektronischen Register wurden bereits während des Imports der Daten durchgeführt. Zum einen wurde automatisiert geprüft, ob komplett identische Datensätze (sogenannte Dubletten) vorhanden waren. Zum anderen wurde stichprobenweise eine erste manuelle Sichtprüfung vorgenommen, um einzuschätzen, ob die Merkmale in der richtigen Reihenfolge vorlagen, die Daten das korrekte Format aufwiesen und die Merkmalsausprägungen im zulässigen Wertebereich lagen. Anschließend wurde überprüft, ob für alle Gemeinden Daten geliefert wurden und ob die Lieferung die erwartete Anzahl an Datensätzen je Gemeinde enthielt. Vorhandene Dubletten wurden gelöscht und Fehler in den Datensätzen – sofern möglich – bereinigt. War eine Bereinigung der Fehler nicht möglich, wurde eine korrigierte Neulieferung angefordert. Ebenso wurde bei fehlenden Gemeinden oder Datensätzen eine Nachlieferung veranlasst.

Bei der sich anschließenden maschinellen Prüfung auf Plausibilität wurden alle Datensätze auf Fehler geprüft und die aufgetretenen Fehler soweit möglich automatisiert bereinigt. Geprüft wurde in erster Linie, ob die übermittelten Merkmale einen Wert aufwiesen und die Ausprägungen innerhalb des zulässigen Wertebereichs lagen. Merkmale, die konform zu einer bestimmten Klassifikation sein mussten, beispielsweise der amtliche Gemeindeschlüssel oder der Wirtschaftszweig, wurden mit entsprechenden Referenzdateien abgeglichen und gegebenenfalls korrigiert. Insgesamt hatten die Erwerbsregisterdaten hinsichtlich der im Rahmen der Plausibilisierung überprüften Sachverhalte eine hohe Qualität.

### 3.3.2.4 Zusammenführung mit den Melderegisterdaten

Um die erwerbsstatistischen Merkmale, die im Rahmen des Zensus 2011 erhoben wurden, in den Zensusdatenbestand zu integrieren, mussten die Erwerbsregisterdaten mit den bis dato aktuellsten Melderegisterdaten (konsolidierter Melderegisterbestand MR.Kons) zusammengeführt werden.<sup>15</sup> Für die Zusammenführung mit MR.Kons war ein Zugriff der elektronischen Register ER-BA und ER-PS auf MR.Kons sowie auf das AGR notwendig. Ziel war es, die in den Erwerbsregistern enthaltenen Personendatensätze jeweils eindeutig in MR.Kons zu identifizieren und möglichst vollständig miteinander zu verbinden.

<sup>15</sup> Für eine ausführliche Methodenbeschreibung siehe auch Sedmíhradský, D. u. a., „Erhebung, Aufbereitung und Zusammenführung der Erwerbsregisterdaten im Zensus 2011“, in WiSta 12/2012, S. 1052-1064.

Die Zusammenführung wurde in zwei Phasen durchgeführt. In der ersten Phase erfolgte die eindeutige Identifikation der Erwerbsregisterpersonen über personen- und anschriftenbezogene Merkmale des Erwerbsregisters, also Namen, Geschlecht, Geburtsdatum, amtlicher Gemeindeschlüssel, Postleitzahl, Straße und Hausnummer. Hierzu wurde eine Trennung und Standardisierung der einzelnen Namens- und Anschriftenbestandteile des Erwerbsregisters durchgeführt. Anschließend erfolgte die Suche der Erwerbsregisterpersonen in MR.Kons über eine umfangreich abgestufte und weit verzweigte Reihenfolge von sinnvollen Kombinationen von Personen- und Anschriftenmerkmalen. Ziel war es, eine möglichst hohe automatisierte Zusammenführungsquote unter Berücksichtigung der Genauigkeit und Laufzeit zu erzielen. Bei erfolgreicher Suche, das heißt, wenn eine Person aus dem Erwerbsregister eindeutig einer Person in MR.Kons zugeordnet werden konnte, wurden die beiden Datensätze miteinander verknüpft. Als Bindeglied zwischen den bis zu diesem Zeitpunkt völlig unabhängig voneinander agierenden Systemen Erwerbsregister und Melde- sowie Anschriften- und Gebäuderegister dienten die erzeugte Ordnungsnummer der Anschrift (Anschriften-ID) und die personenbezogene Ordnungsnummer (PON), die zusammengekommen eine eindeutige Zuordnung zu einer Person ermöglichten. Diese beiden Ordnungsnummern wurden nach erfolgreicher Suche den entsprechenden Datensätzen des Erwerbsregisters hinzugefügt.

Anschließend erfolgte in der zweiten Phase die eigentliche Übertragung der Erwerbsregisterdaten entsprechend der personenbezogenen Ordnungsnummer und der Ordnungsnummer der Anschrift in den Bestand der Melde-, Anschriften- und Gebäuderegister, welche damit den sogenannten Referenzdatenbestand bildeten. Die in den Erwerbsregisterdaten enthaltenen Erhebungsmerkmale, wie Erwerbsstatus, Stellung im Beruf, Arbeitsort oder Wirtschaftszweig, waren somit in den Referenzdatenbestand integriert.

Der Anteil der über diese Vorgehensweise automatisiert verknüpften Datensätze war mit 98,9 % (Daten der Bundesagentur für Arbeit) beziehungsweise 98,5 % (Daten des Personalstands öffentlicher Arbeitgeber auf Bundesebene) sehr hoch. Gleiches galt mit rund 99 % für die automatisiert verknüpften Datensätze des Personalstands auf Landesebene. Die verbleibenden Restfälle wurden manuell geprüft und – sofern möglich – verknüpft. In den Fällen, die bei der automatisierten Zusammenführung mehrere Treffer ergeben hatten, konnte direkt auf die Verknüpfungsvorschläge des Abgleichs zurückgegriffen werden. Die manuellen Zusammenführungsarbeiten konzentrierten sich auf die Restfälle der sozialversicherungspflichtig Beschäftigten sowie der Beamtinnen und Beamten, Richterinnen und Richter, Soldatinnen und Soldaten und Dienstordnungsangestellten. Die Anbindung der Erwerbsdaten führte zu keiner Veränderung der Personenzahl im Referenzdatenbestand (RDB), da durch diese Quellen keine neuen Personen im RDB aufgenommen wurden. Die Erwerbsregister hatten damit keinen Einfluss auf die Ermittlung der Einwohnerzahl in Deutschland.

### 3.3.3 Haushaltebefragung auf Stichprobenbasis (Haushaltsstichprobe)

Die Haushaltebefragung auf Stichprobenbasis im Zensus 2011 hatte zwei Ziele: Zum einen diente sie der Feststellung sowie der statistischen Korrektur von Über- und Untererfassungsfehlern in den Melderegistern für alle Gemeinden mit mindestens 10 000 Einwohnern (Ziel 1). Zum zweiten wurden Informationen, die nicht aus Registern entnommen werden konnten, mit Hilfe der Haushaltsstichprobe erhoben (Ziel 2).

#### **Ziel 1: Statistische Bereinigung von Über- und Untererfassungen in den Melderegistern**

Die Angaben in den Melderegistern sind nicht fehlerfrei: Zum einen gibt es Personen, die laut Melderegister an einer Anschrift gemeldet sind, aber nicht (mehr) tatsächlich dort wohnen (sogenannte „Karteileichen“). Zum anderen gibt es auch Personen, die an einer Anschrift wohnen, ohne dort gemeldet zu sein (sogenannte „Fehlbestände“). Diese Fehler konnten mit statistischen Methoden bereinigt werden. Bei dieser statistischen Bereinigung ging es ausschließlich darum, Übererfassungen der Melderegister (Karteileichen) und Untererfassungen (Fehlbestände) zu erkennen und mit Hilfe geeig-

neter statistischer Verfahren bei der Feststellung der amtlichen Einwohnerzahl für alle Gemeinden mit mindestens 10 000 Einwohnern zu berücksichtigen.<sup>16</sup>

Ein Ergebnis des Zensus 2011 war, dass es in großen Gemeinden mehr Ungenauigkeiten in den Melderegistern gab als in kleinen Gemeinden. Der Prozentsatz an Karteileichen und an Fehlbeständen war tendenziell umso größer, je größer eine Gemeinde war. Das ZensusG 2011 hat daher festgelegt, dass die Ergebnisse der Haushaltsstichprobe nur in Gemeinden mit mindestens 10 000 Einwohnern zur statistischen Registerkorrektur genutzt werden.

## **Ziel 2: Erhebung von Zusatzmerkmalen**

Einige Informationen über die Bevölkerung, die nach dem ZensusG 2011 erhoben wurden, konnten gar nicht oder nicht vollständig aus Registern entnommen werden. Dies gilt z. B. für den Bildungsstand oder für die Erwerbstätigkeit von Selbstständigen. Diese zusätzlichen Merkmale wurden deshalb in der Haushaltsstichprobe erhoben und dann so hochgerechnet, dass darüber Angaben für die gesamte Bevölkerung in den Gemeinden mit mindestens 10 000 Einwohnern und in den Kreisen gewonnen werden konnten.

### **3.3.3.1 Stichprobendesign<sup>17</sup>**

#### **a) Festlegung des Stichprobenumfangs**

Für Ziel 1 der Haushaltsstichprobe mussten immer alle zum Stichtag an einer Anschrift lebenden Personen befragt werden, da sonst zwar Karteileichen, aber keine Fehlbestände hätten festgestellt werden können, d. h. die Einwohnerzahl wäre systematisch unterschätzt worden. Die Auswahl dieser Anschriften erfolgte durch eine mathematisch-statistische Zufallsauswahl. Damit hatte jede Anschrift mit Wohnraum, mit Ausnahme von sensiblen Sonderbereichen, in Deutschland die Chance, für die Haushaltsstichprobe ausgewählt zu werden. Dies ist eine zentrale Voraussetzung dafür, dass sich aus der Haushaltsstichprobe – die ja nur eine Stichprobe der Bevölkerung umfasst – verlässliche, repräsentative Rückschlüsse auf die gesamte Bevölkerung ziehen lassen.

Auch bei größter Sorgfalt können Stichprobenbefragungen kein ganz exaktes Bild über die gesamte Bevölkerung wiedergeben. Vielmehr weisen die Ergebnisse der Haushaltsstichprobe – wie bei jeder Stichprobenbefragung – einen sogenannten „Zufallsfehler“ auf. Die Höhe dieses Zufallsfehlers ist allerdings berechenbar, das heißt man konnte im Nachgang der Erhebung feststellen, wie präzise die Ergebnisse sind. Und man konnte bei der Festlegung des Umfangs der Stichprobe und der Wahl des Verfahrens für die Stichprobenziehung bereits eine zu erwartende Präzision der Stichprobenergebnisse festlegen. Es gelten hierfür zwei Grundregeln:

#### **1. Je größer die Stichprobe ist, umso genauer ist das Ergebnis**

Je mehr Menschen befragt werden, desto exakter wird das Bild über die gesamte Bevölkerung. Dabei ist es in der Regel wichtiger, dass eine ausreichend große Anzahl von Personen befragt wird, als dass ein bestimmter Prozentanteil der Bevölkerung einbezogen wird. Je mehr Menschen in einer Gemeinde leben, umso geringer kann der Prozentsatz sein, der in der Stichprobe berücksichtigt werden muss. Das bedeutet, dass in kleineren Gemeinden in der Regel ein größerer Prozentsatz der Bevölkerung befragt werden muss als in größeren Gemeinden, wenn die Ergebnisse für beide Gemeinden eine vergleichbare Präzision haben sollen.

<sup>16</sup> Eine Korrektur in den Melderegistern selbst fand nicht statt, da die Statistischen Ämter des Bundes und der Länder keine Erkenntnisse über einzelne Personen oder Wohnungen an die Verwaltungen zurückgeben dürfen (sogenanntes Rückspielverbot). Dies würde dem Statistikgeheimnis widersprechen. Die Bereinigung erfolgte ausschließlich mittels statistischer Verfahren im Bereich der Statistischen Ämter des Bundes und der Länder.

<sup>17</sup> Siehe auch Dr. Berg, A./Bihler, W., „Das Stichprobendesign der Haushaltsstichprobe des Zensus 2011“ in: WiSta 4/2011, S. 317-328.



## 2. Je mehr sich die befragten Personen bei denjenigen Merkmalen unterscheiden, die in der Stichprobenerhebung erfragt werden, umso größer muss die Stichprobe sein

Je mehr Menschen befragt werden, umso besser ist die Unterschiedlichkeit und Vielfalt der Bevölkerungsgruppen in den Stichprobenergebnissen erkennbar. Je größer die Stichprobe ist, umso höher ist die Wahrscheinlichkeit, dass auch kleinere Bevölkerungsgruppen in der Stichprobe ausreichend vertreten sind. Ist man beispielsweise am Merkmal „Anzahl der Arbeiter in einer Gemeinde“ interessiert und ist die Anzahl der Arbeiter in fast jedem Haushalt annähernd gleich, dann wird die Wahrscheinlichkeit eine ungünstige Stichprobe zu ziehen, d. h. eine Stichprobe, die die Realität sehr schlecht abbildet, stark verringert. In solch einem Fall starker Homogenität der Population reicht auch ein vergleichsweise sehr geringer Stichprobenumfang aus, um die Bevölkerung hinreichend genau abzubilden. Im Extremfall, wenn in jedem Haushalt exakt gleich viele Arbeiter leben, müsste man sogar nur einen einzigen Haushalt in die Stichprobe aufnehmen.

Generell gilt, dass die Variabilität in den Ausprägungen derjenigen Merkmale, die in der Stichprobe erfragt werden, erst nach der Erhebung bekannt ist. Für die Planung des Stichprobenverfahrens behilft man sich daher in der Regel mit Vergleichswerten aus früheren Erhebungen – im Falle des Zensus 2011 aus dem unter anderem zu diesem Zweck durchgeführten Zensustest im Jahre 2001, aus dem Ergebnisse über das Ausmaß von Karteileichen und Fehlbeständen bekannt sind.

In § 7 Abs. 1 ZensG 2011 war als anzustrebendes Qualitätsziel festgelegt, dass die durch den Zensus 2011 festgestellten amtlichen Einwohnerzahlen mit 95 %iger Sicherheit von den tatsächlichen Einwohnerzahlen möglichst nur um maximal 1 %<sup>18</sup> abweichen sollten.<sup>19</sup> Auf der Grundlage dieses Präzisionsziels und auf der Grundlage des Wissens über die Zusammenhänge von Präzision, Stichprobengröße und Vielfalt der Bevölkerung hat der Gesetzgeber den Gesamtstichprobenumfang per Rechtsverordnung auf 9,6 % der Wohnbevölkerung festgelegt<sup>20</sup> und Regeln für die Verteilung der Stichprobe auf die einzelnen Gemeinden aufgestellt.

### b) Auswahlgrundlage und Auswahlseinheiten

Auswahlgrundlage für die Haushaltsstichprobe war das AGR.

Um Anschriften und damit alle darin lebenden Personen für die Haushaltsstichprobe auszuwählen, wurde das AGR herangezogen. An einer Anschrift können unterschiedlich viele Menschen wohnen, da es sich beispielsweise um ein Einfamilienhaus oder um ein Hochhaus mit sehr vielen Wohnungen oder auch um mehrere Gebäude an einer Anschrift handeln kann. Das bedeutet, dass an einer für die Haushaltsstichprobe ausgewählten Anschrift beispielsweise nur eine Person befragt werden musste, wenn diese zum Stichtag allein in einem Einfamilienhaus lebte, während an einer anderen Anschrift mehr als 100 Personen zu befragen waren, wenn es sich um ein Hochhaus mit sehr vielen Wohnungen handelte.

<sup>18</sup> Die Stichprobenverteilung nähert sich mit wachsendem Stichprobenumfang gemäß dem Zentralen Grenzwertsatz einer Normalverteilung an. Unter Annahme einer Normalverteilung liegen 95 % der Stichprobenergebnisse in einem Bereich von  $\pm 1,96$  mal Standardfehler um den wahren, aber unbekannten Wert. Der angestrebte relative Standardfehler beträgt 0,5 %. Um die Schwankungsbreite zu ermitteln, muss demnach der Wert 0,5 mit 1,96 multipliziert werden. Aus Gründen der Vereinfachung wurde er verdoppelt. Den Wert von 1,96 erhält man aus dem 97,5 %-Quantil der Normalverteilung.

<sup>19</sup> Diese Formulierung entspricht der in § 7 Abs. 1 ZensG 2011 aufgestellten Forderung, dass die Ermittlung der amtlichen Einwohnerzahl mit einer angestrebten Genauigkeit eines einfachen relativen Standardfehlers von höchstens 0,5 % zu erfolgen hat. Bei der Festlegung dieser Obergrenze für den Stichprobenzufallsfehler ist der Gesetzgeber davon ausgegangen, dass die mit dem Zensus 2011 ermittelten Einwohnerzahlen die gleiche Genauigkeit aufweisen sollen wie die Ergebnisse früherer Volkszählungen. Auch die in den beiden Volkszählungen 1970 und 1987 ermittelten Einwohnerzahlen haben Untererfassungsfehler (z. B. weil Personen trotz mehrmaliger Versuche nicht erreicht werden konnten) und Übererfassungsfehler (z. B. weil Personen mit Haupt- und Nebenwohnsitz zweimal mit Hauptwohnsitz gezählt wurden) aufgewiesen. Das Ausmaß dieser sogenannten „systematischen Fehler“ wurde bei den Volkszählungen 1970 und 1987 durch nachträgliche Stichprobenbefragungen festgestellt. Bei der Volkszählung 1970 wurden beispielsweise für das Bundesgebiet ein Untererfassungsfehler von 1,4 % und ein Übererfassungsfehler von 0,8 % ermittelt. Die Anwendung des Stichprobenverfahrens beim Zensus 2011 hat hierbei den Vorteil, dass durch qualifizierte Befragungen an den Stichprobenanschriften wohnende Haushalte solche Fehler eher verringert werden können.

<sup>20</sup> Die Wohnbevölkerung umfasst nur die Personen, die mit Hauptwohnsitz gemeldet sind oder die nur einen Wohnsitz haben, während für die wohnberechtigte Bevölkerung zusätzlich die mit Nebenwohnsitz gemeldeten Personen mitzählen. Bezogen auf die wohnberechtigte Bevölkerung liegt der Gesamtstichprobenumfang bei 9,15 %.

Eine besondere Kennzeichnung mussten alle diejenigen Anschriften erhalten, an denen sich sogenannte „Sonderbereiche“ befanden. Diese wurden noch einmal unterschieden in „sensible Sonderbereiche“ und „nicht-sensible Sonderbereiche“. Zu den sensiblen Sonderbereichen zählten z. B. Behindertenwohnheime oder Justizvollzugsanstalten. Die „nicht-sensiblen Sonderbereiche“, das sind z. B. Studentenwohnheime oder Alten- und Pflegeheime, wurden dagegen in die Haushaltsstichprobe einbezogen, da für die dort lebenden Personen auch diejenigen Informationen erhoben werden sollten, die nicht aus Registern entnommen werden konnten (vgl. Kapitel 3.3.3, Ziel 2).

### c) Auswahlverfahren

#### 1. Ganz Deutschland wurde in Erhebungsgebiete aufgeteilt

Um die Genauigkeit der Stichprobenergebnisse zu erhöhen, wurde ein spezielles statistisches Verfahren – die sogenannte „Schichtung“ – angewandt. Dazu wurde vor der Stichprobenziehung ganz Deutschland in sogenannte „Erhebungsgebiete“ aufgeteilt. Dabei musste gewährleistet sein, dass die Einteilung sowohl lückenlos als auch überschneidungsfrei war, dass also jede Anschrift genau einem Erhebungsgebiet angehörte. Dies führte zu folgender Einteilung:

Typ 1: Stadtteile mit durchschnittlich mehr als 200 000 Einwohnern aus denjenigen Städten, die mindestens 400 000 Einwohner hatten

Typ 2: Gemeinden und Städte mit mindestens 10 000 Einwohnern, sofern sie nicht zum Typ 1 gehörten

Typ 3: Zusammenfassung kleiner Gemeinden (unter 10 000 Einwohnern) innerhalb eines Kreises, wenn sie zu einem Gemeindezusammenschluss<sup>21</sup> gehörten und wenn sie in der Summe mindestens 10 000 Einwohner hatten

Typ 4: Zusammenfassung aller Gemeinden eines Kreises, die bis dahin noch keinem Typ zugeordnet wurden

Die Erhebungsgebiete Typ 3 und Typ 4 sollten insbesondere gewährleisten, dass es auch in ländlichen Gebieten belastbare Ergebnisse zu denjenigen Merkmalen gibt, die nur aus der Haushaltsstichprobe stammen (vgl. Kapitel 3.3.3, Ziel 2).

#### 2. Alle Anschriften wurden in verschiedene Größenklassen eingeteilt

Um die Genauigkeit der Stichprobenergebnisse innerhalb der Erhebungsgebiete zu erhöhen, wurde eine zweite Schichtung eingeführt und in allen Erhebungsgebieten alle Anschriften in verschiedene Größenklassen eingeteilt. Diese Größenklassen bezogen sich darauf, wie viele Menschen an einer Anschrift gemeldet waren.

Wie bereits erwähnt, liefert die Hochrechnung umso genauere Ergebnisse, je mehr sich die in der Stichprobe befragten Personen bei denjenigen Merkmalen ähneln, die im Zusammenhang mit dem hochzurechnenden Merkmal stehen. Eine möglichst hohe Genauigkeit der amtlichen Einwohnerzahlen für Bund, Länder und Gemeinden – als wichtigstes Zensusergebnis – wird erreicht, wenn die Schichtung so erfolgt, dass sich die Anschriften, die zu einer Schicht gehören, im Hinblick auf ihre Größe ähneln, sprich die Personenzahl laut Register an den Anschriften in einer Schicht sich möglichst wenig unterscheidet.

Im Zensustest 2001 war deutlich geworden, dass die Häufigkeit von Karteileichen und von Fehlbeständen in der Regel zunimmt, je mehr Personen an der Anschrift gemeldet sind. Trotzdem steht die Anschriftengröße laut Register in sehr starkem Zusammenhang mit der wahren Anschriftengröße. Daher wurden im Stichprobenverfahren die Schichten auf Basis von Anschriftengrößenklassen gebildet. Es hat sich gezeigt, dass für jedes Erhebungsgebiet eine Einteilung in acht hinsichtlich der Personenzahl gleich große Schichten gute Ergebnisse liefert.

<sup>21</sup> Zu den Gemeindezusammenschlüssen gehören Verbandsgemeinden, Verwaltungsgemeinschaften, Ämter, Verwaltungsverbände, Erfüllende Gemeinden, Kirchspiellandgemeinden und Samtgemeinden.

Vor der Einteilung der Anschriften in die verschiedenen Größenklassen wurden noch diejenigen Anschriften separiert, an denen sich die oben beschriebenen „sensiblen Sonderbereiche“ befinden. Die Anschriften mit „nicht-sensiblen Sonderbereichen“ wurden für jedes Erhebungsgebiet in einer eigenen Schicht zusammengefasst, um bei der Stichprobenziehung gesondert berücksichtigt werden zu können.

### 3. Die Aufteilung der Gesamtstichprobe auf die Erhebungsgebiete und die Größenklassen verfolgte das Ziel einer möglichst großen Präzision – man spricht hierbei von einer Optimierung der Aufteilung der Gesamtstichprobe

Nach der Schichteinteilung folgte die Aufteilung des Gesamtstichprobenumfangs auf diese Schichten. Bei der Aufteilung des Gesamtstichprobenumfangs mussten verschiedene Bedingungen beachtet werden. Zunächst wurden in allen Sonderbereichsschichten 10 % der Anschriften, mindestens aber zwei Anschriften, in die Stichprobe gezogen. Damit flächendeckend für ganz Deutschland belastbare Ergebnisse zu den Zusatzmerkmalen gewonnen werden konnten, wurde dann auf alle Anschriftengrößenklassen der Erhebungsgebiete vom Typ 3 und vom Typ 4 ein Stichprobenumfang von 5 % realisiert.<sup>22</sup>

Die weitere Aufteilung des Gesamtstichprobenumfangs auf die Anschriftenschichten der Erhebungsgebiete der Typen 1 und 2 erfolgte nach einem komplexen mathematischen Optimierungsverfahren, welches simultan alle Schichten in allen Erhebungsgebieten miteinbezieht. Dieses Verfahren ermittelte sukzessive diejenige Aufteilung des Stichprobenumfangs auf die Schichten, bei der die höchstmögliche Präzision für die Ermittlung der amtlichen Einwohnerzahl für alle Kreise und alle Gemeinden mit mindestens 10 000 Einwohnern erreicht war.<sup>23</sup>

Das gewählte Optimierungsverfahren hätte theoretisch in manchen Schichten zu Auswahlätzen von über 100 % führen können, während aus anderen Schichten nur eine sehr kleine Stichprobe gezogen worden wäre. Aus diesem Grund und da sich eine extrem starke Streuung negativ auf die Hochrechnung weiterer Merkmale auswirken würde, wurden Grenzen für die Auswahlätze in den verschiedenen Erhebungsgebieten eingeführt. Folgende Grenzen wurden festgelegt:

Gemeinden mit	Auswahlsatz für die einzelnen Anschriftengrößenklassen
10 000 bis unter 30 000 Einwohner	zwischen 5 % und 50 %
30 000 bis unter 100 000 Einwohner	zwischen 4 % und 40 %
ab 100 000 Einwohner	zwischen 2 % und 40 %

Diese Ober- und Untergrenzen bedeuten, dass beispielsweise in einer Gemeinde mit 20 000 Einwohnern in jeder der acht Anschriftengrößenklassen mindestens 5 % und maximal 50 % der Anschriften für die Stichprobe ausgewählt wurden. Dass diese Ober- und Untergrenzen je nach Gemeindegrößenklasse variierten, liegt daran, dass die Präzision von Stichprobenergebnissen mehr von der Zahl der einbezogenen Personen und weniger vom Prozentsatz der Bevölkerung, der in die Stichprobe einbezogen wird, abhängt. In den Gemeinden mit knapp über 10 000 Einwohnern benötigt man in der Regel einen höheren durchschnittlichen Auswahlatz als in den größeren Gemeinden, wenn man für alle Gemeinden ähnlich genaue Ergebnisse erreichen will.

<sup>22</sup> Weil es in Rheinland-Pfalz sehr viele kleine Gemeinden gibt, die in Verbandsgemeinden zusammengeschlossen sind, erhielten dort nur die Erhebungsgebiete des Typs 4 den Stichprobenumfang von 5 %, während der Prozess der optimierten Verteilung in den Erhebungsgebieten der Typen 1 bis 3 erfolgte.

<sup>23</sup> Um diese Präzision zu messen, wurden für alle amtlichen Einwohnerzahlen aller Erhebungsgebiete vom Typ 1 und vom Typ 2 die oben bereits beschriebenen Stichprobenzufallsfehler zunächst quadriert und dann aufsummiert. Die maximal mögliche Präzision ist dann erreicht, wenn diese Summe durch eine andere Aufteilung des Stichprobenumfangs nicht mehr kleiner wird.



Im Ergebnis führte das Optimierungsverfahren dazu, dass für jede Anschriftengrößenklasse in jedem Erhebungsgebiet – und damit für jede Gemeinde – ein maßgeschneiderter, individueller Auswahlatz berechnet wurde. Für vergleichbar große Gemeinden konnte dies zu unterschiedlichen Auswahlätzen – auch innerhalb der einzelnen Schichten – führen, wenn sie sehr unterschiedliche Anschriftenstrukturen hatten. Dies war beispielsweise dann der Fall, wenn eine Gemeinde sehr viele große Anschriften hatte, während eine andere überwiegend aus Ein- oder Zweifamilienhäusern bestand.

#### 4. Berücksichtigung von Neuzugängen

Die Stichprobenziehung erfolgte nach dem oben beschriebenen Verfahren zum Datenstand 1. September 2010. Dabei gelangten bundesweit 1 948 371 „Normalanschriften“ und 2 049 nicht-sensible Sonderanschriften in die Stichprobe. Später neu zur Grundgesamtheit hinzugekommene Anschriften wurden in zwei Nachziehungen berücksichtigt, wobei die erste (Neuzugangsziehung) mit AGR-Stand zum 31. März 2011 regional nach Regierungsbezirken und fachlich nach vier Anschriftengrößenklassen geschichtet war, die zweite (Ergänzungsziehung) mit Stand vom 19. September 2011 – aufgrund des geringen Umfangs – nur noch regional nach Bundesländern. Über die erste Nachziehung gelangten – ebenfalls bundesweit – 9 675 Anschriften, über die zweite weitere 7 073 Anschriften in die Stichprobe.<sup>24</sup>

##### 3.3.3.2 Datengewinnung

###### a) Einrichtung von Erhebungsstellen und Einsatz von Erhebungsbeauftragten

Nach § 10 ZensG 2011 waren die Länder für die Einrichtung von Erhebungsstellen zuständig. Die zentrale Aufgabe der Erhebungsstellen war die Organisation und Durchführung der Haushaltsstichprobe vor Ort. Diese Zuständigkeit beinhaltete die Gewinnung, die Schulung und die Koordination des Einsatzes von Erhebungsbeauftragten. Im Rahmen der Erhebungstätigkeit waren die Erhebungsstellen für die Eingangskontrolle der Erhebungsunterlagen, das Erinnerungs- und Mahngeschäft und für die Existenzfeststellung von Personen an Stichprobenanschriften verantwortlich.<sup>25</sup> Wichtige Voraussetzung für den Betrieb der Erhebungsstellen war es, diese räumlich, organisatorisch und personell von anderen Verwaltungsstellen zu trennen. Die für die Erhebungsstellen tätigen Personen wurden verpflichtet, das Statistikgeheimnis (§ 16 Bundesstatistikgesetz) zu wahren und auch solche Erkenntnisse über Auskunftspflichtige geheim zu halten, die bei ihrer Tätigkeit gewonnen werden konnten. Bundesweit wurden 545 Erhebungsstellen eingerichtet. Landkreise und größere Kommunen wie z. B. kreisfreie Städte verfügten in der Regel über eine eigene Erhebungsstelle.

Für die Durchführung der Haushaltsstichprobe wurden Erhebungsbeauftragte nach § 11 ZensG 2011 eingesetzt. Die Aufgaben der Erhebungsbeauftragten umfassten die Begehung der Erhebungsbezirke<sup>26</sup> vor der eigentlichen Befragung einschließlich Terminvergabe, die Kontaktaufnahme mit den Auskunftspflichtigen, die eigentliche Befragungstätigkeit einschließlich der Existenzfeststellung sowie die abschließenden Arbeiten einschließlich der Rückgabe der vollständig ausgefüllten Erhebungsunterlagen an die Erhebungsstelle. Im Zuge der Organisation der Erhebungstätigkeit waren die Erhebungsbeauftragten an gesetzliche Vorschriften (beispielsweise § 7 Abs. 6 und § 11 ZensG 2011) sowie an die von der Erhebungsstelle vorgegebenen Bearbeitungsfristen gebunden. Als Richtgröße wurde für alle Erhebungsbeauftragten ein Umfang von rund 100 zu befragenden Personen empfohlen. Basis für die Zuweisung der Erhebungsbeauftragten zu ihren jeweiligen Erhebungsbezirken war eine Registerdatenlieferung der Meldebehörden zum Bezugsdatum 1. November 2010. Anschriften, die aufgrund ihrer Größe nicht von einem Erhebungsbeauftragten alleine bearbeitet werden konnten (sog. Großanschriften), wurden im Vorfeld mittels einer Begehung durch die Erhebungsstellen identifiziert und ggf. in zwei oder mehrere Erhebungsbezirke geteilt.

24 Siehe dazu Dr. Berg, A./Bihler, W., „Der Auswahlplan für die Ziehung der Neuzugänge der Haushaltsstichprobe des Zensus 2011“ in: WiSta 3/2014, S. 151-154.

25 Dies war z. B. der Fall, wenn es der bzw. dem Erhebungsbeauftragten trotz mehrerer Versuche nicht gelungen war, Kontakt mit dem Auskunftspflichtigen aufzunehmen.

26 Dies sind die Stichprobenanschriften, die den Erhebungsbeauftragten durch die Erhebungsstelle jeweils zugewiesen wurden.

### b) Feststellung der zu befragenden Personen (Existenzfeststellung)

Bevor die eigentliche Befragung durchgeführt bzw. der Fragebogen personalisiert und ausgegeben wurde, mussten die Erhebungsbeauftragten klären, welche Personen an den Stichprobenanschriften ihres Erhebungsbezirks existent und auskunftspflichtig waren, d. h. wer von den dort wohnhaften Personen zur Bevölkerung zum Stichtag 9. Mai 2011 zählte. Hierfür wurden den Erhebungsbeauftragten Informationen aus den Melderegistern zur Verfügung gestellt. Zum einen erhielten sie Namenslisten, die alle an der zu erhebenden Anschrift in der Melderegisterlieferung vom 1. November 2010 gemeldeten Personen enthielten. Diese konnten damit als Näherungswerte für die Menge der anzutreffenden Auskunftspflichtigen verwendet werden. Zum anderen wurde die Information aus den Melderegistern zur Zahl der an einer Anschrift gemeldeten Personen verwendet, um Erhebungsunterlagen im benötigten Umfang bereitzustellen. Im Rahmen der Haushaltebefragung auf Stichprobenbasis gemäß § 7 ZensG 2011 erfolgte diese sogenannte Existenzfeststellung ausschließlich an Anschriften, an denen sich kein Sonderbereich (Gemeinschafts-, Anstalts- und Notunterkunft, Wohnheim etc.) befand und damit ausschließlich für Personen, die in Privathaushalten wohnten (zur Existenzfeststellung von Personen in Gemeinschaftsunterkünften gemäß § 8 ZensG 2011 siehe Kapitel 3.3.4). Ergänzend zur Existenzfeststellung war noch festzuhalten, welche Auskunftspflichtigen in welcher Wohnung zusammenwohnten. Im Zensus 2011 zählt als privater Haushalt jede zusammenwohnende Personengemeinschaft sowie Personen, die alleine wohnen. Die Informationen zu den Auskunftspflichtigen wurden in einer separaten Erhebungsliste festgehalten. Sollten Wohnungen an den zu erhebenden Anschriften zum Zensusstichtag leer gestanden haben bzw. gewerblich genutzt worden sein, war dies ebenfalls in die Erhebungsliste einzutragen.

Der Regelfall der Existenzfeststellung war die Erhebung der personenbezogenen Daten der Auskunftspflichtigen durch den Erhebungsbeauftragten selbst vor Ort. Nach schriftlicher Ankündigung (mittels einer Terminankündigungskarte) suchte der Erhebungsbeauftragte jede Wohnung an den zu erhebenden Anschriften auf. Für die Feststellung der Existenz einer Person war es nicht erforderlich, dass diese an dem angekündigten Termin anwesend war. Vielmehr war es ausreichend, wenn die erforderlichen Angaben von einer volljährigen, auskunftspflichtigen Person der gleichen Wohnung gemacht wurden. Die zur Existenzfeststellung erforderlichen Angaben „Name“, „Vorname/-n“, „Geburtsdatum“ und „Geschlecht“ wurden von den Erhebungsbeauftragten handschriftlich auf den Erhebungslisten festgehalten, auf der auch die Existenz bzw. Nichtexistenz der Bewohner vermerkt wurde. In den Erhebungsstellen wurden diese Erhebungslisten der Erhebungsbeauftragten in die sogenannte elektronische Erhebungsliste übertragen. In dieser waren im Resultat alle Personen enthalten, deren Existenz zum Zensusstichtag festgestellt worden war.

Nicht in allen Fällen war eine Existenzfeststellung durch Erhebungsbeauftragte vor Ort möglich. Wurde die Auskunft verweigert oder wurden Haushalte (also die Personen in zum Zensusstichtag von auskunftspflichtigen Personen bewohnten Wohnungen) nicht angetroffen, übertrug sich die Zuständigkeit für die Existenzfeststellung von den Erhebungsbeauftragten auf die Erhebungsstellen. An die nicht angetroffenen bzw. die Auskunft verweigernden Haushalte versendete die Erhebungsstelle in der Regel fünf Fragebogen. Nach dem Rücklauf von ausgefüllten Fragebogen musste von der Erhebungsstelle geprüft werden, ob diese für real existierende Haushaltsmitglieder und nicht für fiktive Personen ausgefüllt wurden. Dieser Prüfprozess erstreckte sich im Bedarfsfall über mehrere Prüfstufen:

Zuerst prüften die Erhebungsstellen, ob die angefragten Personen im Melderegisterbestand zum Zensusstichtag 9. Mai 2011 an der Stichprobenanschrift verzeichnet waren. Diese Information erhielten sie auf Anfrage vom jeweiligen Statistischen Amt des Landes aus der dort vorliegenden Melderegisterlieferung zum Zensusstichtag 9. Mai 2011. Die Paarigkeit von Zweifelsfällen mit dem aktuellen Melderegistereintrag zum Stichtag an der Stichprobenanschrift galt als hinreichendes Indiz für die Existenz der Person an der Anschrift. Bei Unpaarigkeit, wenn also Fragebogen bei der Erhebungsstelle für Personen ohne bisherige Existenzfeststellung eingegangen waren und diese Personen im aktuellen Melderegister unter der Anschrift nicht vorzufinden waren, erfolgte die Existenzfeststellung durch die Erhebungsstelle mittels Postzustellungsauftrag (PZA). Hierbei wurde durch die Erhebungsstellen ein weiteres Informationsschreiben an die nicht im Melderegister verzeichneten Personen, von denen ein Fragebogen vorlag, mittels PZA verschickt. Ein erfolgreicher PZA war dabei ein hinreichendes Indiz für

die Existenz der Person an der Anschrift. Alternativ zur Existenzfeststellung mittels PZA war auch eine Existenzfeststellung im Zuge einer persönlichen Kontaktaufnahme mit der Person oder eine postalische Kontaktaufnahme per Einschreiben mit den Optionen „Eigenhändig“ bzw. „Rückschein“ möglich.

Bei positiver Existenzfeststellung von Zweifelsfällen wurden die benötigten Angaben in der elektronischen Erhebungsliste nachgetragen. Eine nicht erfolgreiche persönliche bzw. schriftliche Kontaktaufnahme durch PZA oder Einschreiben führte zur Feststellung der Nichtexistenz. Die Ergebnisse der Existenzfeststellung wurden an das Folgeverfahren des sogenannten Melderegisterabgleichs weitergeleitet (vgl. Kapitel 3.3.7.2). In diesem Bearbeitungsschritt wurden die Über- und Untererfassungsfehler in den Melderegistern von Gemeinden mit mindestens 10 000 Einwohnern festgestellt.

### c) Durchführung der Befragung

Neben der Korrektur der Über- und Untererfassungsfehler in den Melderegistern zur Ermittlung der Einwohnerzahl einer Kommune diente die Haushaltebefragung auf Stichprobenbasis der Erhebung von Zensusmerkmalen, die nicht aus Verwaltungsregistern gewonnen werden können. Hierzu wurde ein umfangreicher Fragebogen entwickelt. Die Erhebung selbst erfolgte mittels Befragung durch die Erhebungsbeauftragten oder – bei Selbstausfüllwunsch – durch elektronische bzw. durch schriftliche Beantwortung des ausgehändigten Fragebogens.

Bei der mündlichen Auskunftserteilung durch die auskunftspflichtigen Personen trugen die Erhebungsbeauftragten die Angaben der Auskunftspflichtigen in die zuvor durch die Erhebungsbeauftragten personalisierten Papierfragebogen ein. Die Befragungen selbst erfolgten im Anschluss an die Feststellung der Auskunftspflicht der Befragten sowie der Existenzfeststellung für alle Personen der jeweiligen Haushalte. Für auskunftspflichtige, nichtanwesende Personen war ein neuer Befragungstermin zu vereinbaren, da sogenannte Proxy-Interviews<sup>27</sup> laut § 18 Abs. 4 ZensG 2011 nur für die Existenzfeststellung, nicht aber für die Beantwortung des Fragebogens zugelassen waren. Das jeweilige Befragungsdatum und die zur Personalisierung der Fragebogen verwendeten Fragebogennummern wurden in der Erhebungsliste festgehalten; die Fragebogen mussten von den Erhebungsbeauftragten zeitnah bei den Erhebungsstellen abgegeben werden. Zur Erhebungsunterstützung wurden den Erhebungsbeauftragten Übersetzungshilfen in 13 verschiedenen Sprachen (Englisch, Arabisch, Kroatisch/Serbisch, Französisch, Griechisch, Italienisch, Spanisch, Polnisch, Portugiesisch, Rumänisch, Türkisch, Vietnamesisch, Russisch) zur Verfügung gestellt. Diese fremdsprachigen Erhebungsunterlagen dienten ausschließlich der Erhebungsunterstützung bei den Befragungen. Die auszufüllenden Fragebogen waren immer in deutscher Sprache.

Neben der mündlichen Auskunftserteilung bestand auch die Möglichkeit, die Fragen schriftlich zu beantworten. Dies konnte durch Ausfüllen des Papierfragebogens oder elektronisch mittels der Internetanwendung IDEV (Internet Datenerhebung im Verbund) erfolgen. Voraussetzung, die Fragen selbst beantworten zu können, war aber grundsätzlich immer die vorherige Feststellung der Existenz an der Anschrift und die daraus folgende Ausgabe eines personalisierten Papierfragebogens durch Erhebungsbeauftragte. Für Selbstausfüllende wurden die Fragebogen personalisiert, der Selbstausfüllwunsch inklusive Übergabedatum sowie die Fragebogennummer in der Erhebungsliste vermerkt.

Die Erhebungsphase sollte in der Regel die gesetzlich vorgeschriebene Dauer von zwölf Wochen ab Zensusstichtag nicht überschreiten (vgl. § 7 Abs. 6 ZensG 2011). In Ausnahmefällen, z. B. im Rahmen des Erinnerungs- und Mahngeschäfts, konnte davon abgewichen werden.

<sup>27</sup> Ein Proxy-Interview ist ein Interview, bei dem Dritte über die eigentlichen Zielpersonen befragt werden.

### d) Kontrollarbeiten der Erhebungsstellen

Mit der Durchführung der Haushaltsstichprobe durch die Erhebungsstellen wurden umfangreiche Arbeiten zur Sicherstellung der Qualität der zu erhebenden Daten durchgeführt. Bei der Eingangskontrolle nahmen die Erhebungsstellen die schriftlichen Erhebungsunterlagen von den Erhebungsbeauftragten und die Papierfragebogen der Selbstausfüllenden entgegen (die Online-Fragebogen gingen direkt bei den Statistischen Ämtern der Länder ein und wurden von diesen im Zuge der Eingangskontrolle registriert). Die von den Erhebungsbeauftragten in Papierform erstellten Erhebungslisten wurden in den Erhebungsstellen in eine elektronische Fassung überführt. Auf Basis dieser elektronischen Erhebungsliste war die IT-basierte Vollzähligkeit der Fragebogen sicherzustellen. Für alle existenten Personen musste entweder ein Papierfragebogen in der Erhebungsstelle oder ein von den Statistischen Ämtern der Länder vergebener Eingangsvermerk bei elektronischer Übermittlung vorliegen. Mittels Sichtkontrolle waren alle eingehenden Fragebogen auf ihre Verwertbarkeit hin zu überprüfen. Die verwertbaren Fragebogen wurden gemeinsam mit den übrigen Erhebungsunterlagen erhebungsbezirksweise nach der Registrierung in den Erhebungsstellen vorgehalten, bis diese durch die Statistischen Ämter der Länder abgeholt wurden.

### 3.3.3.3 Datenaufbereitung

#### a) Aufbereitungsschritte und Plausibilitätsprüfungen

Im ersten Schritt der Datenaufbereitung hatten die Erhebungsstellen die Vollzähligkeit aller eingehenden Erhebungsunterlagen sicherzustellen. Nach Übermittlung an die Statistischen Ämter der Länder wurden dort die DV-technisch generierten Lieferscheine mit den tatsächlich übernommenen Erhebungspapieren verglichen. Im Anschluss erfolgte die Digitalisierung der Papierfragebogen. Die danach vorliegenden Bilder (Bit-Images) wurden gelesen, interpretiert und falls notwendig in einem manuellen Bearbeitungsschritt nachbearbeitet. Klartextangaben wurden im Anschluss – mit Ausnahme des Merkmals „Beruf“ – maschinell verarbeitet.<sup>28</sup> Danach wurden die Personenangaben aus der elektronischen Erhebungsliste (Liste aller Personen, für die die Existenz an einer Stichprobenanschrift festgestellt wurde) und die gelesenen Belege zusammengeführt. Nach Ausschluss derjenigen Fragebogen, für die im Rahmen der Existenzfeststellung keine Person in der elektronischen Erhebungsliste an der betroffenen Anschrift registriert war, verblieb die Gesamtmenge aller Datensätze, die zu plausibilisieren waren. Hierbei wurden die Datensätze durch maschinelle Prüfroutinen auf fehlende bzw. unplausible Einträge überprüft. Aufgedeckte Fehler wurden mit Hilfe von Imputationsverfahren plausibel ersetzt. Bei allen Aufbereitungsschritten wurden Qualitätskennzeichen angelegt. Diese ermöglichten u. a. Analysen zu den im Zuge der Aufbereitung erfolgten Datenkorrekturen.

Da bei der Online-Erhebung mit Ausnahme der Identifikation der Auskunftspflichtigen, der korrekten Abfolge der zu beantwortenden Fragen und der Prüfung, dass es keine unzulässigen Mehrfachantworten bei einzelnen Fragen gab, selbst keine Plausibilitätsprüfungen stattfanden, erfolgte die Plausibilitätsprüfung der online erhobenen Datensätze gemeinsam mit den mittels Papierfragebogen erhobenen Datensätzen.

Folgende Kriterien wurden hinsichtlich der Plausibilität geprüft:

- Vollständigkeit: Überprüfung der Datensätze auf fehlende Angaben.
- Mehrfachnennungen: Überprüfung, ob unzulässige und damit unplausible Mehrfachangaben vorlagen. Dies war nur bei beleggelesenen Datensätzen notwendig, wenn Auskunftspflichtige auf dem Papierfragebogen mehrere Kreuze zu einer Frage gesetzt hatten.
- Strukturplausibilität: Überprüfung, ob unzulässige Werte oder Codes eingetragen worden waren.

<sup>28</sup> Für den Zensus 2011 wurde ein eigenständiges Verfahren zur Kodierung der Berufsangaben entwickelt. Für eine detaillierte Beschreibung siehe Loos, C., Eisenmenger, M., Bretsch, D., „Das Verfahren der Berufskodierung im Zensus 2011“ in: WiSta 3/2013, Seite 173 ff.

- Interplausibilität: Überprüfung des Zusammenhanges zwischen einzelnen Merkmalen. Dabei wurden Fehler aufgrund logischer Widersprüche zwischen den Merkmalen wie z. B. Fehler, die auf Nichtbeachtung der korrekten Fragenabfolge zurückzuführen waren, identifiziert.

## b) Imputationsverfahren

Zur Korrektur und Ersetzung (Imputation) von unplausiblen Ausprägungen (darunter fallen auch fehlende Angaben) kamen bei der Haushaltsstichprobe drei Verfahren zum Einsatz:

- Cold-Deck-Imputation
- Deterministische Imputation
- Hot-Deck-Imputation

### 1. Cold-Deck-Imputation

Mit dem Cold-Deck-Verfahren wurden unplausible oder fehlende Angaben ersetzt, indem Angaben aus anderen Datenquellen genutzt werden. Es wurde bei denjenigen Merkmalen angewandt, für die Daten aus der elektronischen Erhebungsliste bzw. Registermerkmale aus der zur Erhebungsunterstützung angeforderten Meldedatenlieferung vom November 2010 herangezogen werden konnten. Zunächst wurde geprüft, ob Daten aus der elektronischen Erhebungsliste verwendet werden konnten. Sofern dort keine bzw. keine plausiblen Angaben vorlagen, konnte für Personen, die im Melderegister verzeichnet waren (paarige Personen), auf diese Melderegisterangaben im RDB zurückgegriffen werden. Gründe für unplausible Ausprägungen waren beispielsweise fehlende Angaben, unzulässige Mehrfachangaben, Unplausibilität zu anderen Merkmalen, Strukturunplausibilität (= Wertebereichsfehler oder Fehler bei der Kodierung von Merkmalen) oder Nichtbeachtung der korrekten Fragenreihenfolge.

Folgende Merkmale wurden in die Cold-Deck-Imputation einbezogen:

- Geschlecht (elektronische Erhebungsliste, Melderegister)
- Geburtsdatum (elektronische Erhebungsliste, Melderegister)
- Zugehörigkeit zu einer öffentlich-rechtlichen Religionsgesellschaft (Melderegister)
- Straße (elektronische Erhebungsliste)
- Hausnummer (elektronische Erhebungsliste)
- PLZ (elektronische Erhebungsliste)
- Ort (elektronische Erhebungsliste)
- Staatsangehörigkeit (Melderegister)
- Familienstand (Melderegister)
- Wohnungsstatus (Melderegister)

### 2. Deterministische Imputation

Sofern auch das Melderegister keine plausiblen Merkmalsausprägungen enthielt bzw. die Person nicht in der Meldedatenlieferung enthalten war, wurde, wenn möglich, deterministisch imputiert.

Bei der deterministischen Imputation wurde eine unplausible Ausprägung eines Merkmals über eine eindeutige Beziehung zu einer Ausprägung eines anderen Merkmals bzw. zu mehreren Ausprägungen anderer Merkmale korrigiert. Dieses Verfahren konnte nur bei wenigen Merkmalen eingesetzt werden. Wurden beispielsweise die Fragen nach der besuchten Schulart und der Klassenstufe plausibel beantwortet, zur Eingangsfrage nach dem Schulbesuch aber keine Angaben gemacht, führte die deterministische Imputation dazu, dass diese Frage bejaht wurde. Das Merkmal „Schulbesuch“ wurde somit mit der einzig plausiblen Ausprägung aus diesem Fragebogen befüllt. Sofern zwei oder mehr plausible Ausprägungen zur Imputation möglich waren, schied das Verfahren der deterministischen Imputation aus. Stattdessen wurden die betroffenen Datensätze für das im Anschluss stattfindende Hot-Deck-Imputationsverfahren gekennzeichnet.



### 3. Hot-Deck-Imputation

Die noch fehlenden oder unplausiblen Merkmale wurden im letzten Schritt über eine Hot-Deck-Imputation ergänzt bzw. korrigiert. Die Hot-Deck-Imputation nutzt den Datenbestand aller Personen, indem „ähnliche“ Personen gesucht und deren Angaben übernommen werden. Die verwendete einfache Nearest-Neighbour-Spenderimputation suchte auf Grundlage bestehender Antworten des jeweils vorliegenden, unplausiblen Empfängerdatensatzes einen Spenderdatensatz, der die geringsten Unterschiede (Distanz) aufwies. Die Ähnlichkeit von Empfänger- und Spenderdatensatz wurde im Sinne gemeinsamer Merkmalsausprägungen betrachtet. Je mehr Merkmalsausprägungen zwischen Spender- und Empfängerdatensatz übereinstimmten, desto geringer waren die Unterschiede.<sup>29</sup> Die Ähnlichkeit beschränkte sich dabei auf vier Fragen- bzw. Merkmalsblöcke. Die unplausiblen Merkmale im Empfängerdatensatz wurden anschließend blockweise in einem Schritt durch die Ausprägungen des Spenders ersetzt. Im Anschluss an die Hot-Deck-Imputation wurde geprüft, ob der jeweilige Block nach diesen Ersetzungen plausibel vorlag. Sollten vorher plausible Merkmale durch die Ersetzungen unplausibel geworden sein, wurden diese durch die entsprechenden Angaben des Spenderdatensatzes ersetzt.

Die Angaben zu Name, Vorname/-n, Anschrift und Telefonnummer wurden nicht per Hot-Deck imputiert. Fehlende Angaben wurden unbefüllt belassen, da sie für die Ergebnisbereitstellung keine Rolle spielten und nach Abschluss der Datenaufbereitung ohnehin gelöscht werden mussten. Die Angaben zu Geschlecht und Geburtsdatum waren Pflichtfelder. Leere oder unplausible Angaben wurden durch Cold-Deck-Imputation oder durch deterministische Imputation befüllt. Im Rahmen der anschließenden Hot-Deck-Imputation konnten diese Angaben nicht mehr verändert werden. Die freiwillige Angabe zum Glaubensbekenntnis wurde nicht imputiert, sondern im Falle einer fehlenden Angabe unbefüllt gelassen.

Das Imputationskonzept sah vor, dass ein Block eines Datensatzes mehrfach als Spender verwendet werden konnte. Zur Begrenzung der Spenderfunktion wurde eine maximale Verwendungszahl bestimmt. Zusätzlich wurde eine Maximaldistanz zwischen Empfänger- und Spenderdatensatz je Block festgelegt. Es wurde jeweils der Spender gewählt, der die geringste Distanz und bei gleicher Distanz die geringste Verwendungszahl aufwies. Zunächst wurde auf Kreisebene, danach im Bedarfsfall auch auf höheren regionalen Ebenen (Regierungsbezirk, Land) nach einem geeigneten Spender gesucht.

#### 3.3.3.4 Hochrechnungsverfahren für die Ermittlung der Einwohnerzahlen der Gemeinden mit 10 000 und mehr Einwohnern

##### a) Auszählungs- und Hochrechnungsteil

In den Gemeinden, die am 31. Dezember 2009 bzw. am 1. Januar 2010<sup>30</sup> mindestens 10 000 Einwohner hatten („große Gemeinden“), wurde eine statistische Bereinigung des Melderegisterbestandes um Übererfassungen (Karteileichen) und Untererfassungen (Fehlbestände) auf Basis der mit der Haushaltsstichprobe des Zensus 2011 erhobenen Daten durchgeführt. Dennoch ist die Einwohnerzahl dieser Gemeinden nicht ein reines Ergebnis der Hochrechnung dieser Registerfehler aus der Stichprobe, vielmehr setzen sich die Anschriften der Gemeinden aus einem sogenannten Auszählungsteil und einem Hochrechnungs- oder Korrekturteil zusammen:

Zum Auszählungsteil zählten

- alle Anschriften, die am 31. Dezember 2009 bzw. am 1. Januar 2010 noch zu einer Gemeinde mit weniger als 10 000 Einwohnern („kleine“ Gemeinden) gehörten (vgl. § 2 Abs. 6 ZensG 2011) sowie
- alle Sonderanschriften (egal ob in „kleinen“ oder „großen“ Gemeinden).

<sup>29</sup> Mathematisch bedeutet die Suche nach den „ähnlichsten“ Datensätzen, dass über einen Algorithmus die Datensätze mit den geringsten Distanzen ermittelt werden. Dies erfolgt über eine oder mehrere Distanzfunktionen, bei denen die Abweichungen zwischen den einzelnen Merkmalen verglichen bzw. aufaddiert werden.

<sup>30</sup> Aufgrund von Gebietsstandsänderungen zum 1. Januar 2010 wurde in einigen Bundesländern dieses Datum als Stichtag für die Gemeindegrößenzuordnung (große und kleine Gemeinden) verwendet.

Im Auszählungsteil erfolgte keine Bereinigung der Über- und Untererfassungen der Melderegister über die Haushaltsstichprobe.

Alle übrigen Anschriften gehörten zum Hochrechnungsteil. Im Hochrechnungsteil erfolgte eine Bereinigung des Registerbestands um die aus der Haushaltsstichprobe hochgerechneten Übererfassungen (Karteileichen) und Untererfassungen (Fehlbestände).<sup>31</sup>

Gemäß den Empfehlungen aus dem Stichprobenforschungsprojekt<sup>32</sup> wurden die Karteileichen und Fehlbestände nicht direkt hochgerechnet.<sup>33</sup> Vielmehr wurde zunächst jeweils die Zahl der existenten (durch die Zahl der in der Stichprobe angetroffenen) und die Zahl der paarigen (durch die Zahl der in der Stichprobe angetroffenen und im Melderegister gemeldeten) Personen geschätzt. Karteileichen und Fehlbestände ergaben sich dann durch Differenzbildung gemäß den Zusammenhängen

Karteileichen = Melderegisterbestand – paarige Personen,

Fehlbestände = existente Personen der Haushaltsstichprobe – paarige Personen

und der Einwohnerzahlbeitrag des Hochrechnungsteils ist dann

Melderegisterbestand + Fehlbestände – Karteileichen.

Dieser Umweg ist aus stichprobentheoretischer Sicht günstiger, da die Fallzahlen in der Stichprobe, die anschließend hochgerechnet werden, größer sind (es gibt mehr existente und paarige Personen als Karteileichen und Fehlbestände), was die Fehlerrechnung für die Karteileichen und Fehlbestände stabiler macht. Aus den o. g. Beziehungen folgt aber, dass der Einwohnerzahlbeitrag des Hochrechnungsteils identisch ist zu den hochgerechneten existenten Personen. Somit kann man als Fazit festhalten: Im Hochrechnungsteil werden die Einwohner aus den in der Stichprobe als existent festgestellten Personen direkt geschätzt und zwar unter Berücksichtigung des beobachteten statistischen Zusammenhangs zwischen den an den Anschriften als existent festgestellten und den im Melderegister verzeichneten Personen.

Die beiden resultierenden Teilsummen aus Auszählungs- und Hochrechnungsteil ergeben die Einwohnerzahl der Gemeinde.

## b) Hochrechnung im Allgemeinen und Regressionsschätzung im Speziellen

Wird eine Stichprobe von Personen gezogen, steht jede dieser Personen stellvertretend für andere Personen, die nicht in die Stichprobe gelangt sind. Für wie viele andere Personen sie ein Stellvertreter ist, hängt im Grundsatz davon ab, welchen Prozentsatz von allen Personen man für die Stichprobe ausgewählt hat; man spricht vom Auswahlatz oder – da die Stichprobe zufällig ausgewählt wird – von der Auswahlwahrscheinlichkeit. Wird jede zehnte Person ausgewählt, repräsentiert jede Stichprobenperson 10 Personen insgesamt. Diesen Faktor 10 – der Kehrwert der Auswahlwahrscheinlichkeit – nennt man Hochrechnungsfaktor. Ein hochgerechnetes Ergebnis aus einer Stichprobe erhält man, indem man die Befunde für die einzelnen Stichprobenpersonen mit deren jeweiligem Hochrechnungsfaktor multipliziert. Drei Ausländerinnen und Ausländer in unserem Stichprobenbeispiel ergeben also hochgerechnet – d. h. multipliziert mit dem Hochrechnungsfaktor – 30 Ausländerinnen und Ausländer insgesamt.

Dieses Vorgehen ist das einfachste denkbare Hochrechnungsverfahren und heißt „freie Hochrechnung“. Liegen Hilfsinformationen für die Gesamtheit, für die man Ergebnisse aus der Stichprobe hochrechnen will, vor, kann man diese nutzen um die Schätzung zu verbessern. Ein Spezialfall dieser dann „gebunden“ statt „frei“ bezeichneten Hochrechnung ist die sogenannte Regressionsschätzung.

31 Der Hochrechnungsrahmen umfasst nicht nur die Nicht-Sonderanschriften aus großen Gemeinden der Auswahlgesamtheit, sondern auch Zugänge, die nach den Stichprobenziehungen (auch nach den Nachziehungen) als Meldeanschriften oder Anschriften mit Wohnraum ins AGF aufgenommen wurden. Für diese Zugänge, die keine Auswahlchance hatten, werden dadurch implizit die Strukturen der Stichprobenanschriften unterstellt.

32 Siehe auch Münnich, R. u. a., „Stichprobenoptimierung und Schätzung im Zensus 2011“, Statistik und Wissenschaft, Band 21, Wiesbaden, Juli 2012.

33 Synonymer Begriff für Hochrechnung: Schätzung.

Mit Regression beschreibt man in der Statistik ein Modell, das zwei beobachtete Größen zueinander in Beziehung setzt, wobei die Kenntnis über diesen Zusammenhang nur aus einer Reihe einzelner Beobachtungen besteht. Ein triviales Beispiel: bei Kleinkindern kann man in der Regel davon ausgehen, dass sie umso schwerer sind, je größer sie sind. Würde man nun für eine Stichprobe von Kleinkindern – z. B. 100 Kinder – in einem Diagramm auf der x-Achse die Größe und auf der y-Achse das Gewicht abtragen, ergibt sich eine Punktwolke von 100 Punkten, die grob von links unten (kleine Größe und kleines Gewicht) nach rechts oben (große Größe und großes Gewicht) verläuft. Es ist nun möglich, durch diese Punktwolke nach definierten Kriterien eine „optimale“ Gerade zu legen, die dann eine mathematische Beziehung zwischen Größe und Gewicht darstellt. Diese Gerade heißt Regressionsgerade.

Wäre man an der Durchschnittsgröße aller Kleinkinder einer Gemeinde interessiert (= Untersuchungsvariable oder Zielvariable), könnte man unter den Kleinkindern eine Stichprobe ziehen und für die Stichprobenpersonen die Größe bestimmen. Wenn für alle Kleinkinder – nicht nur die in der Stichprobe – das Gewicht (= Bezugsvariable) bereits bekannt wäre und aus der Stichprobe die Regressionsgerade zwischen Gewicht und Größe bestimmt werden kann, kann damit die freie Hochrechnung verbessert werden. Stark vereinfacht ausgedrückt geschieht dies, indem man für alle Kleinkinder aus ihrem Gewicht mit Hilfe der Regressionsgeraden ihre Größe schätzt und daraus den Mittelwert bildet.

Beim Zensus 2011 funktioniert dies im Prinzip genau so, außer dass man sich für alle Personen und nicht nur für Kleinkinder interessiert, und dass „Größe“ durch „existente“ bzw. „paarige“ Personen zu ersetzen ist und es nicht nur eine, sondern eine ganze Reihe von Bezugsvariablen gibt.

### c) Das Schätzverfahren für den Hochrechnungsteil

Gemäß den Empfehlungen des Stichprobenforschungsprojekts sollte die Schätzung der existenten und paarigen Personen eine spezielle Form der Regressionsschätzung sein, und zwar der sogenannte GREG-Schätzer (verallgemeinerter Regressionsschätzer). Dies ist ein Verfahren, mit dem sich – bei dem gegebenen hinreichend großen Stichprobenumfang – die Einwohnerzahl praktisch ohne Verzerrung („Bias“) beziehungsweise – um einen Begriff aus der Stichprobentheorie zu gebrauchen – asymptotisch „erwartungstreu“ hochrechnen lässt. Diesen Vorteil der freien Hochrechnung erreicht damit annähernd auch die Regressionsschätzung. Im Vergleich zur freien Hochrechnung nutzt die Regressionsschätzung aber Zusatzinformationen und ist damit genauer als die freie Hochrechnung.

Regressionsschätzung meint beim Zensus 2011, dass man für jede betroffene Gemeinde getrennt für eine vorab definierte Liste von Merkmalsausprägungen anschriftenweise Register- und Stichprobenauszahlung miteinander in Bezug setzt (bildlich: y-Achse: Stichprobeninformation = Untersuchungsvariable, x-Achse: Registerinformation = Bezugsvariable<sup>34</sup>). Das Melderegister stellt also die für die Regressionsschätzung benötigten Hilfsinformationen in Form unterschiedlicher Bezugsvariablen bereit. Für die Einwohnerzahl ist die Zielvariable die Zahl der laut Stichprobe mit Hauptwohnsitz existenten Personen.<sup>35</sup> Die Liste der Bezugsvariablen wurde – für die Ermittlung der Einwohnerzahl – wie folgt festgelegt:

<sup>34</sup> Synonyme Bezeichnung für Bezugsvariable: Hilfsvariable.

<sup>35</sup> Die Zielvariable „existente HW-Personen insgesamt“ ergibt direkt den Einwohnerzahlbeitrag des Hochrechnungsteils. Die Hochrechnung von Karteileichen und Fehlbeständen durch Differenzbildung und die anschließende Korrektur des Registerbestands ergibt kein anderes Ergebnis.



= 1 (Konstante)	
Gemeldete Personen insgesamt	
Geschlecht/Staatsangehörigkeit	Deutsch, männlich
	Deutsch, weiblich
	Nicht-deutsch, männlich
Alter	Unter 6
	6 bis unter 18
	18 bis unter 25
	25 bis unter 30
	30 bis unter 40
	40 bis unter 50
	50 bis unter 60
	60 bis unter 65

Eine Bezugsvariable ist somit nicht definiert über ein Merkmal, sondern über eine spezielle Merkmalsausprägung (z. B. „unter 6 Jahre“) oder eine Kombination von Merkmalsausprägungen (z. B. „deutsch, männlich“). Das Regressionsmodell zur Ermittlung der amtlichen Einwohnerzahlen arbeitet mit zwölf Bezugsvariablen; hinzu kommt als 13. Variable formal eine Konstante, da man vorab nicht wissen kann, ob die Regressionsgerade durch den Nullpunkt verläuft oder nicht. Anschriftenweise „in Bezug setzen“ meint, dass für die o. g. Bezugsvariablen (Registerinformation) nicht nur der Eckwert pro Gemeinde, sondern auch die Personensumme je Stichprobenanschrift ausgezählt wird und der Personensumme der Untersuchungsvariablen je Stichprobenanschrift (Stichprobeninformation) gegenüber gestellt wird.

Technisch können sich die Regressionsschätzungen u. a. jetzt noch darin unterscheiden, in welcher fachlichen und regionalen Differenzierung die Regressionsgerade – der mathematische Zusammenhang zwischen Untersuchungs- und Bezugsvariablen – bestimmt wird.

Im Zensus 2011 wurden die Regressionskoeffizienten – das sind die Parameter, die wie in obigem ganz einfachen Beispiel von Größe und Gewicht der Kleinkinder die Regressionsgerade durch die Punktwolke im Diagramm beschreiben – getrennt für jede Gemeinde bzw. jeden Stadtteil geschätzt (der Fachausdruck dafür lautet „nicht gruppiert“) und nicht auf übergeordneter regionaler Ebene, etwa für das ganze Bundesland („gruppiert“). Die Simulationsrechnungen des Stichprobenforschungsprojekts haben gezeigt, dass diese Entscheidung nicht mit einem Verlust an Präzision verbunden ist. Die Regressionsschätzung wurde – wiederum gemäß wissenschaftlicher Empfehlung – aber nicht je fachlicher Schicht (Anschriftengrößenklasse) durchgeführt, sondern für die gesamte regionale Nachweisungseinheit Gemeinde. In einem solchen Fall nennt man die Regressionsschätzung „kombiniert“ (alle Schichten einer Gemeinde werden gemeinsam betrachtet). Bei Simulationsrechnungen mit Zensus-Echtdaten wurde bestätigt, dass die kombinierte, nach Gemeinden bzw. Stadtteil durchgeführte Regressionsschätzung die robustesten Hochrechnungsergebnisse liefert.

Die (auf Anschriften bezogenen) Hochrechnungsfaktoren beinhalten folgende Komponenten:

- Die Auswahlwahrscheinlichkeiten der Stichprobenanschriften (der Teil, den man auch bei einer freien Hochrechnung hätte)
- Einen Korrekturfaktor für Anschriftenzusammenfassungen und Antwortausfälle (echte Ausfälle verkleinern den Netto-Stichprobenumfang)
- Die „Korrektur“ der freien Hochrechnung mittels der Hilfsinformationen (Bezugsvariablen) im Rahmen des Regressionsmodells

Weitere in die Hochrechnung integrierte Prozesse umfassen sowohl die erforderliche ganzzahlige Rundung auf „volle“ Einwohner als auch die Fehlerrechnung (Varianzschätzung).

Die Varianzschätzung erlaubt die Überprüfung, inwieweit die angestrebten Präzisionsziele für die Einwohnerzahl erreicht werden konnten. Dabei ist es ausreichend, den Zufallsfehler nur für die existenten Personen am Hauptwohnsitz zu berechnen.

### 3.3.4 Daten aus den Erhebungen an Anschriften mit Sonderbereichen

#### 3.3.4.1 Datengewinnung

Bei der Analyse der Ergebnisse der Volkszählung 1987 war festgestellt worden, dass die damaligen Melderegister insbesondere für Gemeinschafts-, Anstalts- und Notunterkünfte sowie für Wohnheime erhebliche Über- und Untererfassungsfehler aufgrund relativ hoher Fluktuation und/oder unzureichendem Meldeverhalten (z. B. in Studentenwohnheimen) enthielten. Zudem bestehen in den Bundesländern zum Teil unterschiedliche Melderechtsvorschriften für Personen, die in derartigen Unterkünften wohnen. Damit war eine hinreichend zuverlässige Ermittlung der an Sonderbereichsanschriften wohnenden Personen allein mit den Meldedaten nicht möglich. Das ZensG 2011 schreibt deshalb vor, dass an allen Anschriften mit Wohnheimen und Gemeinschaftsunterkünften (sogenannte Sonderbereiche) die dort wohnenden Personen durch eine primärstatistische Befragung festgestellt werden mussten. Diese Vollerhebung wurde bundesweit in allen Gemeinden unabhängig von ihrer Größe durchgeführt.

Für die bei einer Vollerhebung im Vorfeld notwendige Identifikation der Anschriften mit Sonderbereichen konnte nicht auf ein vorhandenes Register zurückgegriffen werden. Die Feststellung der Anschriften erfolgte in zwei Schritten: In einem ersten Schritt wurden – neben einer breiten Recherche in öffentlich zugänglichen Quellen – die Kommunen und andere wichtige Träger von Sonderbereichseinrichtungen auf freiwilliger Basis befragt. Daran schloss sich eine schriftliche Vorbefragung der bisher ermittelten Träger und Einrichtungen von Sonderbereichen an.

Primäres Ziel der Vorbefragung war es, Sonderbereiche im Sinne des ZensG 2011 zu identifizieren, nach Sonderbereichsarten zu klassifizieren sowie weitere Informationen zu erhalten, die für die nachfolgende Vollerhebung im Zensus 2011 von Bedeutung waren (z. B. Zahl der Plätze der Einrichtung). Mit Hilfe dieser Angaben wurde sukzessive erstmalig ein Sonderanschriftenregister (SAR) aufgebaut, das mit dem AGR verknüpft wurde.

Auch nach Abschluss der Vorbefragung wurden noch weitere Sonderanschriften identifiziert, z. B. im Rahmen der Erhebungen der Haushaltsstichprobe und der Gebäude- und Wohnungszählung sowie durch die Vor-Ort-Kenntnisse der Erhebungsstellen. An diesen nachträglich gefundenen Anschriften mit Sonderbereichen fand dann ebenfalls eine Erhebung aller Personen statt, die zum 9. Mai 2011 dort wohnhaft waren.

Bei der Durchführung der Erhebung an Anschriften mit Sonderbereichen wurde zwischen sensiblen Sonderbereichen<sup>36</sup> und nicht-sensiblen Sonderbereichen<sup>37</sup> unterschieden. Nach § 2 Abs. 5 ZensG 2011 in Verbindung mit § 18 Abs. 5 ZensG 2011 wurden solche Sonderbereiche als sensibel eingestuft, bei denen die Information über die Zugehörigkeit für Betroffene die Gefahr einer sozialen Benachteiligung hervorrufen konnte. Für Personen in sensiblen Sonderbereichen war die Leitung dieser Einrichtungen auskunftspflichtig. In nicht-sensiblen Sonderbereichen waren die Bewohnerinnen und Bewohner selbst auskunftspflichtig, sofern sie auskunftsfähig und volljährig waren.

Die Unterscheidung spiegelt sich auch in den beiden Fragebogenvarianten wider: Die Einrichtungsleitungen von sensiblen Sonderbereichen erhielten den sogenannten Erfassungsbogen, auf dem die geforderten Angaben in Listenform für mehrere Personen einzutragen waren. Die Bewohnerinnen und Bewohner von nicht-sensiblen Sonderbereichen erhielten einen Individualfragebogen, auf dem die Antworten nur für sich persönlich eingetragen werden konnten.

<sup>36</sup> Zu sensiblen Sonderbereichsarten zählen Behindertenwohnheime, Krankenhäuser, Flüchtlingsunterkünfte, Justizvollzugsanstalten, Notunterkünfte/Unterkünfte für Wohnungslose sowie fallweise Mutter-Kind-Heime und Kinder- und Jugendheime.

<sup>37</sup> Zu den nicht-sensiblen Sonderbereichsarten gehören Studentenwohnheime, Arbeiterheime, Alten- und Pflegeheime, Internate, Schulen des Gesundheitswesens, Klöster, Kasernen der Bundeswehr/Bundes- und Landespolizei, sonstige Wohnheime/Heime, Normaler Bereich innerhalb einer Sonderanschrift (Hausmeister-, Personal- oder Privatwohnungen) sowie fallweise Mutter-Kind-Heime und Kinder- und Jugendheime.

Sowohl der Erfassungsbogen für die Anschriften mit sensiblen Sonderbereichen als auch der individuelle Fragebogen für nicht-sensible Sonderbereichsanschriften konnte direkt ausgefüllt oder online mit Hilfe eines besonders geschützten Übertragungsverfahrens für amtliche Statistiken (IDEV) übermittelt werden.

Da die Erfassungseinheit beim Zensus 2011 generell die Anschrift war, wurden an Anschriften mit Sonderbereichen durch die Erhebung auch diejenigen Personen erfasst, die nicht in der Gemeinschaftsunterkunft wohnten, sondern in einer privaten Wohnung, z. B. in einer Hausmeister- oder in einer Personalwohnung. An Mischanschriften, also Anschriften, an denen sowohl sensible als auch nicht-sensible und/oder Normalbereiche zu finden sind, wurde je Bereich der hier entsprechend vorgesehene Fragebogen eingesetzt.

Anschriften mit nicht-sensiblen Sonderbereichen konnten zudem für die Haushaltebefragung auf Stichprobenbasis ausgewählt werden. In diesem Fall kam ein umfassenderer Fragenkatalog zum Einsatz, der zusätzlich zu den Erhebungsmerkmalen nach § 8 ZensG 2011 (Wohnverhältnis, demografische Merkmale) noch Fragen zur Zuwanderung und zur Erwerbs- und Bildungsbeteiligung nach § 7 ZensG 2011 enthielt. In diesen Fällen wurde ein „Kombinierter Fragebogen zur Erhebung über die Bevölkerung an Anschriften mit Wohnheimen/Gemeinschaftsunterkünften und zur Haushaltebefragung auf Stichprobenbasis“ verwendet.

Eine Sonderform stellte die Sonderbereichsart „Kasernen“ dar. Diese wurden wie sensible Sonderbereiche behandelt.<sup>38</sup> Eine persönliche Befragung der Leitungen an den Bundeswehrkasernen durch Erhebungsbeauftragte wurde vom Bundesministerium der Verteidigung mit Hinweis auf Sicherheitsbelange abgelehnt. Daher wurde eine zentrale Lieferung der erforderlichen Daten durch das Bundesministerium der Verteidigung an das Statistische Bundesamt vereinbart. In gleicher Weise wurden die Daten für Personen in den Kasernen der Bundespolizei und bei ausgewählten Landespolizeidirektionen zentral an das Statistische Bundesamt übersandt. Eine Haushaltebefragung auf Stichprobenbasis nach § 7 ZensG 2011 fand an diesen Anschriften nicht statt.

Sonderfälle innerhalb der Sonderbereichserhebung stellten die Bereichsarten „Seeleute/Binnenschiffer“ und „Ausländische Streitkräfte, Diplomaten“ dar. Hier wurde auf eine Vor-Ort-Befragung verzichtet und die mehrfachfallgeprüften Melderegisterdaten für die Einwohnerzahlermittlung übernommen.

Mit der Durchführung der Erhebung vor Ort wurden in der Regel die von den kommunalen Erhebungsstellen ausgewählten Erhebungsbeauftragten betraut.<sup>39</sup>

### 3.3.4.2 Datenaufbereitung

Die Aufbereitung der erhobenen Daten fand in mehreren Schritten in den Statistischen Ämtern der Länder, in Teilen aber auch im Statistischen Bundesamt statt. In der Aufbereitung wurden alle Prozesse durchgeführt, die notwendig waren, um insbesondere die erhobenen Personendaten mit den Melderegisterbeständen abzugleichen, die Daten zu plausibilisieren und – falls notwendig – zu vervollständigen. Zur Datenaufbereitung gehörten die Schritte Zusammenführung von Fragebogen mit der elektronischen Erhebungsliste<sup>40</sup>, Melderegisterabgleich bzw. -anbindung, Mehrfachfallprüfung, Wohnstatusfeststellung, Plausibilisierung und Imputation sowie die abschließende Datenübermittlung an den Referenzdatenbestand. Je nach Erhebungsart (sensibler bzw. nicht-sensibler Sonderbereich, Kaserne) waren die Prozessabläufe etwas unterschiedlich.

<sup>38</sup> Siehe Gesetzesbegründung zum ZensG 2011 (BR-Drucks. Nr. 3/09 vom 2. Januar 2009, S. 87-88).

<sup>39</sup> In den Stadtstaaten war die Erhebungsstelle für die Sonderbereichsbefragungen Teil des jeweiligen Statistischen Amtes des Landes.

<sup>40</sup> In der elektronischen Erhebungsliste wurden im Wesentlichen die demografischen Kernangaben und das Befragungsergebnis je Person durch die Erhebungsstellen gespeichert. Die Angaben stammten aus der papiernen Erhebungsliste der Erhebungsbeauftragten, die damit die Existenzfeststellung dokumentierten.

Im ersten Schritt fand die Zusammenführung der Erhebungslisten mit den eingegangenen Fragebögen statt. Ziel der Zusammenführung war es, anschriftenweise die Fragebogen und die Personen der elektronischen Erhebungsliste zu verknüpfen, also sicherzustellen, dass jeder Person auf der elektronischen Erhebungsliste auch der richtige Fragebogen zugewiesen wurde. Berücksichtigt wurden dabei alle Sonderbereichsanschriften, für die eine elektronische Erhebungsliste vorhanden war, also alle nicht-sensiblen Sonderbereiche mit ausschließlich individueller Befragung mit oder ohne Zusatzmerkmale nach § 7 ZensG 2011. Die Zusammenführung wurde auch an Mischanschriften durchgeführt, allerdings dort nur für die vorhandenen nicht-sensiblen Sonderbereiche. Für sensible Sonderbereiche wurde die Zusammenführung nicht durchgeführt, da für diese keine elektronische Erhebungsliste vorlag. Hier konnten die durch die Statistischen Ämter der Länder elektronisch erfassten Daten aus dem Erfassungsbogen direkt an den konsolidierten Melderegisterbestand (MR.Kons) angebunden werden.

Im zweiten Schritt fand mit den Datensätzen aus den sensiblen und den nicht-sensiblen Sonderbereichen sowie den Datensätzen der Kasernen der Melderegisterabgleich statt, wobei die im konsolidierten Melderegisterbestand (MR.Kons) und der elektronischen Erhebungsliste enthaltenen Personen anschriftenweise (für Anschriften mit Sonderbereichen) abgeglichen und verknüpft wurden.

Der nächste Schritt, die Mehrfachfallprüfung, war ein bundesweiter Abgleich der an Sonderbereichsanschriften gemeldeten Personen mit dem Meldedatenbestand MR.Kons, um Doppelerfassungen bei der Ermittlung der Einwohnerzahlen zu vermeiden. Die Mehrfachfallprüfung diente der Vorbereitung der Wohnstatusfeststellung. Für die Feststellung des Wohnungsstatus war nicht nur der Wohnstatus einer Person an der Sonderanschrift relevant, sondern daneben mussten auch weitere vorhandene Wohnsitze berücksichtigt werden. Bei mehreren Wohnsitzen musste sichergestellt werden, dass genau ein Wohnsitz die Hauptwohnung war und alle anderen die Nebenwohnsitze. Die Ergebnisse der Mehrfachfallprüfung zu Haupt- und Nebenwohnsitzen wurden daher in das Verfahren der Wohnstatusfeststellung integriert.

Die Wohnstatusfeststellung war für die Ermittlung der Einwohnerzahl innerhalb der Datenaufbereitung entscheidend. Diese lief nach eindeutigen Regeln – orientiert an den Melderechtsvorschriften <sup>41</sup> – automatisiert ab. Für die Wohnstatusfeststellung wurden dabei Informationen aus den Sonderbereichserhebungen, dem Melderegister und der Mehrfachfallprüfung hinzugezogen. Bei der Wohnstatusfeststellung wurde für jede Person, die im konsolidierten Melderegisterbestand an der Sonderanschrift aufgeführt war und/oder erhoben bzw. nicht erhoben wurde, festgestellt, ob sie an der Sonderanschrift oder an einer anderweitigen Anschrift mit alleiniger Wohnung, Hauptwohnung, Nebenwohnung zu zählen war. Wurde für eine an einer Sonderanschrift erhobene Person keine weitere Anschrift festgestellt, so wurde diese mit alleinigem Wohnsitz an der Sonderanschrift gezählt. Wurden für eine an einer Sonderanschrift erhobene Person mehrere Wohnsitze ermittelt, musste geprüft werden, welcher der Wohnsitze als Hauptwohnsitz gezählt wird. Die anderen Wohnsitze wurden dann zu Nebenwohnsitzen.

Die Daten zu Bewohnerinnen und Bewohnern von nicht-sensiblen Sonderbereichen wurden durch einen persönlichen Fragebogen ermittelt, der auch Fragen zu ihren Wohnverhältnissen (u. a. zu weiteren Wohnungen und deren Status) und zum Familienstand enthielt. Lagen aus den Fragebogen die hierzu erforderlichen Angaben vor, so waren diese Daten für die Feststellung des Wohnstatus an der Sonderanschrift ausschlaggebend und die weiteren über die Mehrfachfallprüfung festgestellten Wohnsitze mussten entsprechend angepasst werden. Ansonsten wurden in dem automatisierten Verfahren zur Wohnstatusfeststellung die vorhandenen Melderegisterangaben nach fest vorgegebenen Regeln hinzugezogen.

---

<sup>41</sup> Zum Beispiel ist für Verheiratete der Familienwohnsitz der Hauptwohnsitz und für alle anderen Personen ist es der übliche Aufenthaltsort.

Für Bewohnerinnen und Bewohner von sensiblen Sonderbereichen waren nach § 18 ZensG 2011 die Einrichtungsleitungen auskunftspflichtig. Es konnte jedoch nicht davon ausgegangen werden, dass die Einrichtungsleitungen Informationen zu allen eventuell vorhandenen Wohnsitzen ihrer Bewohnerinnen und Bewohner hatten (wurde daher auch im Fragebogen nicht gefordert). Zudem hatte der Gesetzgeber festgelegt, dass sich ihre Auskunftspflicht nur über die ihnen bekannten Daten erstreckte. Zentrale Ausgangspunkte der Wohnstatusfeststellung waren hier die Angabe im Erfassungsbogen zum Datum des Bezuges der Wohnung bzw. Einrichtung und die Ergebnisse der Mehrfachfallprüfung zu Haupt- und Nebenwohnsitzen. Ziel war hier ebenfalls – wie bei den nicht-sensiblen Sonderbereichen – die Feststellung des alleinigen Wohnsitzes oder des Hauptwohnsitzes am oder außerhalb der Sonderanschrift.

Die Wohnstatusfeststellung für Personen an Kasernenanschriften orientierte sich, da hier die Daten zentral geliefert wurden und außer dem Datum des Bezugs der Kaserne keine weiteren Angaben zu den Wohnverhältnissen vorlagen, eng an der Vorgehensweise für die sensiblen Sonderbereiche. Auch hier war der Melderegistereintrag zur Festlegung eines eindeutigen Hauptwohnsitzes maßgeblich.

Der letzte Schritt der Datenaufbereitung war die Plausibilisierung der Fragebogenangaben der Personen an nicht-sensiblen Sonderanschriften. Die Plausibilisierung einschließlich der Korrektur und Ersetzung (Imputation) von unplausiblen Ausprägungen erfolgte gemeinsam mit den Datenbeständen aus der Haushaltsstichprobe (vgl. Kapitel 3.3.3.3).

### 3.3.5 Daten für Deutsche im Ausland

#### 3.3.5.1 Datengewinnung

Die der EU-Zensusverordnung zugrunde liegende Definition des Begriffs „Bevölkerung“ sah vor, dass die ins Ausland entsandten Angehörigen der Streitkräfte und die für die Streitkräfte im Ausland tätigen Personen, die Angehörigen der deutschen Polizeibehörden und des Auswärtigen Dienstes sowie ihre dort ansässigen Familien für die Feststellung der Einwohnerzahl im Heimatland zu zählen waren. Zusätzlich waren zeitlich befristet in den Auswärtigen Dienst übernommene und in das Ausland entsandte Angehörige anderer Bundesbehörden (§ 31 Abs. 1 Gesetz über den Auswärtigen Dienst [GAD]) zu erfassen. Zu erfassen waren auch Angehörige des Auswärtigen Dienstes, die das Auswärtige Amt nach den Entsenderichtlinien des Bundes zu internationalen Organisationen beurlaubte und entsandte sowie Angehörige des Auswärtigen Dienstes, die nach § 123a des Beamtenrechtsrahmengesetzes (BRRG) einer öffentlichen Einrichtung außerhalb des Anwendungsbereiches des BRRG einer vorübergehenden Tätigkeit zugewiesen wurden. So wurde sichergestellt, dass auch die Personen zur amtlichen Bevölkerung von Deutschland gezählt wurden, die aufgrund ihrer zeitweisen Verwendung außerhalb des deutschen Staatsgebiets in Deutschland nicht meldepflichtig waren.

Um diese Personengruppe zu erfassen, war in § 3 Abs. 3 und 4 ZensG 2011 eine einmalige Datenlieferung durch das Auswärtige Amt (AA), das Bundesministerium der Verteidigung (BMVg) und das Bundesministerium des Innern (BMI) angeordnet. Die Daten der ins Ausland entsandten Personen entstammten Verwaltungsunterlagen, die durch die in § 3 Abs. 4 ZensG 2011 benannten Auskunftspflichtigen geführt werden. Diese Unterlagen wurden angelegt, um den internen Datenbedarf der Auskunftspflichtigen zu bedienen. Dementsprechend konnte nicht von einer Detaillierung und Standardisierung der Merkmale wie im Datensatz für das Meldewesen (DSMeld) ausgegangen werden. Zusätzlich unterlagen Teile der Daten einem besonderen Schutzbedürfnis, welches sich aus den Aufgaben der Personen ergab. Aus diesem Grund wurde bei der Datenerhebung nur auf einen reduzierten Merkmalskranz zurückgegriffen, der zur Ermittlung der amtlichen Einwohnerzahl unabdingbar war.

Die durch die drei Stellen übersandten Daten wurden vor der Weiterverarbeitung einer Eingangskontrolle unterzogen, die folgende Prüfungen umfasste:

1. Vergleich der Angaben auf dem Begleitschein mit Satzanzahl in der Datenlieferung
2. Überprüfung des Datensatzaufbaus gemäß Datensatzbeschreibung
3. Überprüfung des Zeichensatzes der Datenlieferung

Wenn diese Daten die Eingangsprüfung erfolgreich durchlaufen hatten, wurden sie plausibilisiert. Hier war zu prüfen und ggf. zu korrigieren, ob

- die gelieferten Daten stichtagstreu geliefert wurden.
- die Daten vollständig geliefert wurden.
- der Aufbau der Datensätze eine weitere maschinelle Verarbeitung und Auswertung zuließ.
- der aktuelle Staatenschlüssel für das Verwendungsland genutzt wurde, ggf. Korrektur bzw. Vercodung der Klartexteintragungen.
- Personen mehrfach geliefert wurden.

### 3.3.5.2 Mehrfachfallprüfung im bundesweiten Melderegisterdatenbestand

Durch die datenliefernden Stellen konnte nicht sichergestellt werden, dass nur Angaben zu Personen ohne Hauptwohnsitz auf dem Territorium der Bundesrepublik Deutschland geliefert wurden. Um Doppelzählungen zu vermeiden, war es daher notwendig zu prüfen, ob die ins Ausland entsandten Personen oder die sie begleitenden Familienangehörigen noch mit einer Haupt- oder Nebenwohnung in Deutschland erfasst waren. Hierzu wurden die Namensfelder analog der im Kapitel 3.2.1 beschriebenen Vorgehensweise standardisiert und im Anschluss auf Treffer im MR.Kons geprüft. Für den Fall, dass eine Person auch noch im Inland mit einer Hauptwohnung gemeldet war, war diese im Inland zu zählen und aus der Datenlieferung des Auskunftspflichtigen zu entfernen.

### 3.3.6 Daten der „Befragung zur Klärung von Unstimmigkeiten“ (BKU)

#### 3.3.6.1 Zweck der Erhebung

Für Gemeinden mit weniger als 10 000 Einwohnern wurde neben der Mehrfachfallprüfung sowie der Erhebung an Anschriften mit Sonderbereichen die sogenannte Befragung zur Klärung von Unstimmigkeiten (BKU) zur Qualitätssicherung der Einwohnerzahlen durchgeführt. Anders als in Gemeinden mit mindestens 10 000 Einwohnern, bei der die Haushaltebefragung auf Stichprobenbasis zur Aufdeckung und Korrektur der Über- und Untererfassungsfehler in den Melderegistern eingesetzt wurde, war in den Gemeinden mit weniger als 10 000 Einwohnern die BKU die geeignete Methode, um diese Fehler aufzudecken und zu korrigieren. Simulationsrechnungen mit den Daten aus dem Zensus 2001 hatten gezeigt, dass die BKU bei Anschriften mit einer bewohnten Wohnung (in der Regel Einfamilienhäuser), die in Gemeinden unter 10 000 Einwohnern den überwiegenden Teil der Anschriften ausmachten, effizient eingesetzt werden kann. Eine Haushaltsstichprobe hätte dagegen in Gemeinden mit bis zu 10 000 Einwohnern nur mit einem sehr hohen Auswahlsatz bis hin zu einer Totalerhebung effizient durchgeführt werden können. Bei der im Zensus 2001 festgestellten tendenziell besseren Registerqualität der kleinen Gemeinden wurde daher in Gemeinden mit weniger als 10 000 Einwohnern die Methode der BKU angewandt.

#### 3.3.6.2 Auswahl der Klärungsfälle und Durchführung der Erhebung

Zur Überprüfung und ggf. Korrektur der Melderegisterfehler in Gemeinden mit weniger als 10 000 Einwohnern waren diejenigen Anschriften zu identifizieren, zu denen widersprüchliche Angaben zur Zahl der an einer Anschrift wohnenden Personen aus dem Melderegister sowie aus der GWZ vorlagen. Diese Identifikation erfolgte anhand einer erhebungsteilübergreifenden Plausibilitätsprüfung zwischen den Melderegisterdaten und den Angaben aus der GWZ. Auf Basis der Angaben der Melderegister wurde eine vorläufige Haushaltegenerierung durchgeführt (siehe Kapitel 4). Im Zensus 2011 zählt als privater Haushalt jede zusammenwohnende Personengemeinschaft sowie Personen, die alleine wohnen. Da somit ein Haushalt immer alle Bewohner einer Wohnung umfasste, konnten den Ergebnissen der GWZ die Ergebnisse der vorläufigen Haushaltegenerierung gegenübergestellt werden.

Als unplausibel und damit zu klärende Fälle galten Anschriften, bei denen Anzeichen für eine Über- oder Untererfassung von Personen laut Melderegister vorlag. Die Kriterien für eine Auswahl der Klärungsfälle waren:



1. An der Anschrift existierte laut GWZ eine Wohnung, und es war nach vorläufiger Haushaltegenerierung kein Haushalt vorhanden, der mit der Wohnung verknüpft werden könnte. Dieser Fall war ein Hinweis auf eine Untererfassung im Melderegister.
2. An der Anschrift existierte eine Wohnung, die nicht mit einem Haushalt verknüpft werden konnte. Es gab an der Anschrift zudem mindestens einen unverknüpften Haushalt und die Angabe der an der Anschrift wohnhaften Personenzahl laut GWZ wich von der Angabe laut Melderegister ab. Dieser Fall war ein Hinweis auf eine Übererfassung im Melderegister.
3. An der Anschrift existierte eine Wohnung, die noch nicht mit einem Haushalt verknüpft werden konnte. Es gab jedoch mindestens zwei unverknüpfte Haushalte und zur Personenzahl lag aus der GWZ keine Angabe vor. Dieser Fall war ein Hinweis auf eine Übererfassung im Melderegister.
4. Die laut GWZ an der Anschrift befindliche Wohnung wurde mit einem Haushalt verknüpft. Es existierten jedoch noch weitere, nicht mit der Wohnung verknüpfte Haushalte. Zudem wich die Personenzahl der an der Anschrift laut GWZ wohnenden Personen von der im Melderegister ab. Dieser Fall war in aller Regel ein Hinweis auf das Vorliegen einer Übererfassung im Melderegister, wenn die GWZ-Personenzahl kleiner war als die Personenzahl laut Melderegister.
5. Die laut GWZ an der Anschrift befindliche Wohnung wurde mit einem Haushalt verknüpft. Es existierten aber weitere Haushalte, die nicht mit der Wohnung verknüpft werden konnten. Aus der GWZ lag keine Angabe über die an der Anschrift wohnhaften Personen vor. Dieser Fall war ein Hinweis auf eine Übererfassung im Melderegister.
6. Die laut GWZ an der Anschrift befindliche Wohnung wurde mit einem Haushalt verknüpft und es existierte keine weitere unverknüpfte Person. Die nach GWZ-Angaben an der Anschrift befindliche Personenzahl überstieg die Anzahl der laut Melderegister gemeldeten Personen. Dieser Fall war ein Hinweis auf eine Untererfassung im Melderegister.

Nach §16 ZensG 2011 waren die Statistischen Ämter der Länder für die Bereinigung von Unstimmigkeiten in Bezug auf die identifizierten Anschriften zuständig. Die Erhebung wurde – wie auch die Haushaltebefragung auf Stichprobenbasis und die Erhebung an Anschriften mit Sonderbereichen – von Erhebungsbeauftragten durchgeführt und in den Statistischen Ämtern der Länder weiterverarbeitet.

Es war nicht ausgeschlossen, dass für die Befragung zur Klärung von Unstimmigkeiten Anschriften identifiziert werden konnten, die bereits im Zuge der Haushaltebefragung auf Stichprobenbasis erhoben wurden. Um Doppelbefragungen zu vermeiden, erfolgte zunächst ein Abgleich mit den Anschriften der Haushaltsstichprobe. Bei Übereinstimmung entfiel die BKU. Dies war möglich, da analog zur Haushaltsstichprobe alle zum Stichtag an der Anschrift wohnhaften Personen zu erheben waren und die in der Befragung zur Klärung von Unstimmigkeiten zu erhebenden Merkmale auch Teil des Erhebungsprogramms der Haushaltsstichprobe waren. Damit folgte die im Zuge der BKU durchzuführende Existenzfeststellung denselben Regeln, die auch für die Haushaltsstichprobe galten (vgl. Kapitel 3.3.3.2 b) ). Um die tatsächlich an der Anschrift wohnenden Personen zu ermitteln, wurden an den betroffenen Anschriften für jede dort wohnende Person die Erhebungsmerkmale „Monat und Jahr der Geburt“, „Geschlecht“, „Familienstand“, „Wohnungsstatus“, „Staatsangehörigkeiten“ und „Zahl der in der Wohnung wohnhaften Personen“ sowie die Hilfsmerkmale „Familiennamen“, „frühere Namen und Vorname/-n“, „Tag der Geburt“ und „Anschrift“ erhoben.

Die zu überprüfenden Anschriften wurden von den Statistischen Ämtern der Länder sukzessive an die Erhebungsstellen übermittelt und für die Befragung durch die Erhebungsbeauftragten freigegeben. Dieses Vorgehen wurde deshalb gewählt, da der Identifikation der Anschriften Plausibilitätsprüfungen vorausgingen, für die das Vorliegen von Ergebnissen der GWZ notwendig war. Damit die Befragung möglichst zeitnah zum Zensusstichtag stattfinden konnte, wurde nicht die Vollzähligkeit der GWZ-Ergebnisse abgewartet, sondern mit den Plausibilitätsprüfungen für die in Frage kommenden Anschriften begonnen, sobald die GWZ-Ergebnisse hierfür jeweils vorlagen.



### 3.3.6.3 Datenaufbereitung

Die Verfahren der Aufbereitung der Daten der BKU folgten denen der Haushaltebefragung auf Stichprobenbasis. Aufgrund des reduzierten Merkmalskranzes der BKU entfiel die Verarbeitung von Klartextangaben. Somit unterteilte sich die Datenaufbereitung in die Prozesse Vollständigkeitskontrolle, Digitalisierung der Papierfragebogen, das Lesen und Interpretieren der digitalisierten Fragebogen, Zusammenführung der Daten der elektronischen Erhebungsliste (Liste aller Personen, für die die Existenz an einer BKU-Anschrift festgestellt wurde) mit denen der digitalisierten Fragebogen und die Plausibilitätsprüfung einschließlich Imputation (vgl. Kapitel 3.3.3.3).

### 3.3.7 Zusammenführung der Daten aus den verschiedenen Datenquellen im Referenzdatenbestand (RDB) <sup>42</sup>

#### 3.3.7.1 Aufbau des RDB

Aufgrund der Heterogenität der beim Zensus 2011 verwendeten Datenquellen – Register, Vollerhebungen und Stichprobenerhebungen – war es notwendig, die Erhebungsbefunde der verschiedenen Datenquellen zusammenzuführen und abzustimmen. Hierzu wurde der Referenzdatenbestand (RDB) erstellt.

In den §§ 9 und 12 ZensG 2011 wird dem RDB als Zusammenschluss des AGR, der Melderegisterdaten und der erwerbsstatistischen Register eine koordinierende Rolle im Zensus 2011 zugewiesen. Der RDB bildete unterschiedliche Bezugsgrößen ab, wie Informationen auf Anschriften- und Personenebene. Informationen zur Anschrift enthielt das AGR. Informationen auf Personenebene stammten aus den Angaben der Melderegister (MR) und der Erwerbsregister (ER) (vgl. Kapitel 3.3.2). Über ein Ordnungsnummernsystem war jeder Personendatensatz eindeutig einer Anschrift aus dem AGR zugeordnet.

Die zentrale Quelle für die Personendatensätze waren die Melderegisterdaten, wobei drei Datenlieferungen für den Aufbau des RDB genutzt wurden. Sie stellten den Grundstock des RDB dar und wurden sowohl zur Vorbereitung und Organisation der primärstatistischen Erhebungen als auch zur Ermittlung der Einwohnerzahl verwendet. In diesen Ausgangsdatenbestand des RDB wurden in einem nächsten Schritt die Daten aus den Erwerbsregistern integriert (vgl. Kapitel 3.3.2).

#### 3.3.7.2 RDB als Sammelpunkt von Informationen aus den verschiedenen Erhebungsteilen

Der RDB als zentrale Datenbank ermöglichte die Zusammenführung der Ergebnisse der einzelnen Erhebungsteile und steuerte damit das Zusammenspiel der Datenquellen. Insbesondere die primärstatistischen Erhebungen

- Haushaltebefragung auf Stichprobenbasis,
- Erhebungen an Anschriften mit Sonderbereichen,
- Gebäude- und Wohnungszählung,
- Befragung zur Klärung von Unstimmigkeiten und
- Befragung zur Klärung des Wohnsitzes

bedienten sich bei der Durchführung und der Zusammenführung der Personendatensätze des RDB und ergänzten den Registerdatenbestand mit ihren Ergebnissen.

Da die Abweichungen zwischen den Melderegistern und dem Ergebnis der Haushaltebefragung auf Stichprobenbasis (Karteileichen und Fehlbestände in den Melderegistern) maßgeblichen Einfluss auf die mit dem Zensus ermittelten Einwohnerzahlen hatten, wurden die Befragungsergebnisse der Haushaltsstichprobe vor deren Übernahme in den RDB noch einmal intensiv auf Plausibilität geprüft. Damit sollten insbesondere potenzielle Erhebungsfehler oder Fehler bei der Abgrenzung der Stichprobenanschriften – wenn z. B. die Erhebung an der falschen Anschrift (z. B. an der Anschrift „Birkenweg 2“ statt

<sup>42</sup> Siehe auch Hirner, S./Stiglmayr, S., „Der Referenzdatenbestand im Zensus 2011“, in: WiSta 1/ 2013, S. 30-41.

an der Anschrift „Birkenweg 2a“) oder nur an Teilen an einer Anschrift (z. B. nur an einem Gebäudeteil bei einer aus drei Gebäudeteilen bestehenden Anschrift) durchgeführt wurde – aufgedeckt und korrigiert werden. Eine Korrektur konnte auf Anschriftenebene oder auf Personenebene erfolgen, indem die Existenzkennzeichnungen einzelner Personen geändert wurden. Wurde der Erhebungsbefund für die ganze Anschrift geändert, wurden diese Veränderungen auf die Personenebene übertragen.

Neben den Ergebnissen der Haushaltebefragung auf Stichprobenbasis konnten auch die Ergebnisse aus der Erhebung der Anschriften mit Sonderbereichen und der Befragung zur Klärung von Unstimmigkeiten von den Daten der Melderegister abweichen. Die Meldedaten für die Sonderanschriften und für die Anschriften der Befragung zur Klärung von Unstimmigkeiten wurden für die Plausibilitätsprüfung der Erhebungsdaten genutzt. Damit wurden auch hier die Abweichungen zwischen der Anzahl der gemeldeten Personen im Melderegister und der Anzahl der existenten Personen nach Erhebung aufgedeckt. Auffällige Anschriften konnten erneut geprüft und ggf. korrigiert werden. In diesen Fällen mussten manuelle Korrekturen im RDB durchgeführt werden. Wurden Personendatensätze entdeckt, die an einer Anschrift sowohl als Karteileiche als auch als Fehlbestand enthalten waren, sich jedoch in den demografischen Merkmalen nicht unterschieden, wurden sie zu einem Personendatensatz zusammengefasst. Auf diese Weise konnte ein überhöhter Registerfehler für einzelne Anschriften korrigiert werden.

Zu diesen „erhebungsteilübergreifenden“ Plausibilitätsprüfungen wurden auch Informationen aus der GWZ herangezogen. So konnten z. B. Anschriften, für die große Abweichungen zwischen der in der GWZ angegebenen Zahl der bewohnten Wohnungen und den an der Anschrift gemeldeten Personen vorlagen, noch einmal auf Erhebungs- bzw. Antwortfehler überprüft werden. Dies konnte dazu führen, dass auch auf Seiten der GWZ nachrecherchiert werden musste. Die Aufbereitungsprogramme der GWZ griffen bei der Verarbeitung der Fragebogen für Zwecke der Plausibilisierung sowie für die Imputation unplausibler oder fehlender GWZ-Merkmale auf Informationen aus AGR und RDB zu. Stellte sich bei der Erhebung der GWZ heraus, dass sich an einer Anschrift kein Wohnraum befand, wurden diese Ergebnisse aus der GWZ wiederum in das AGR integriert, um so für andere Erhebungsteile relevante Informationen zentral an einer Stelle bereitstellen zu können.

Die Zusammenführung von Befragungsergebnissen aus den verschiedenen Erhebungsteilen im RDB diente vor allem dazu, Karteileichen und Fehlbestände der Melderegister zu identifizieren. Die drei Erhebungsteile (Haushaltsstichprobe, Erhebungen an Anschriften mit Sonderbereichen, Befragung zur Klärung von Unstimmigkeiten) wurden alle durch Erhebungsbeauftragte durchgeführt. Die Befragungsergebnisse konnten sowohl die gesamte Anschrift (z. B. Gebäude steht leer) als auch einzelne Personen (z. B. Person wohnt unter der Anschrift) betreffen. Diese Erkenntnisse wurden ebenfalls in den RDB integriert. Auf Personenebene wurden daher die Existenzfeststellungen durch die Erhebungsbeauftragten mit dem aktuellen Datenbestand im RDB abgeglichen und Datensätze der Melderegister wurden entweder bestätigt (paariger Datensatz), nicht bestätigt (Karteileiche im Melderegister) oder es wurden neue Datensätze aus den genannten Erhebungsteilen in den RDB aufgenommen (Fehlbestand der Melderegister). Als Ergebnis des Abgleichs erhielten die Datensätze des RDB eine Kennzeichnung, ob sie von einem Erhebungsbeauftragten in einer Befragung als existent oder nicht-existent festgestellt wurden. Diese Markierung erfolgte für alle Datensätze an der erhobenen Anschrift.

Auch war es möglich, Informationen in Bezug auf die gesamte Anschrift, wie beispielsweise den Grund für einen Befragungsausfall, im AGR abzubilden. Dies wurde notwendig, wenn Erhebungsbeauftragte z. B. an einer zu erhebenden Anschrift aus den genannten Erhebungsteilen feststellten, dass sich an dieser kein Wohnraum befand (Baulücke, Anschrift ausschließlich gewerblich genutzt). In diesen Fällen konnten keine erhobenen Personen mit den Datensätzen des RDB abgeglichen werden, die Anschrift wurde im AGR entsprechend markiert. Dies führte im weiteren Prozess dazu, dass alle gemeldeten Personen an dieser Anschrift als nicht-existent gekennzeichnet und damit als Karteileichen gewertet wurden. Die Stichprobenanschrift war mit der Eintragung dieses Erhebungsbefundes abschließend bearbeitet. Ähnlich war das Vorgehen bei Erhebungsergebnissen aus der Befragung zur Klärung von Unstimmigkeiten. Auch hier wurden Ergebnisse, die die Eigenschaft der gesamten Anschrift betrafen, direkt im AGR vermerkt. Wurden bei der Erhebung an Anschriften mit Sonder-

bereichen Erkenntnisse gewonnen, die die gesamte Anschrift betrafen, so wurden diese direkt ins AGR eingetragen. Dabei konnte zum Beispiel festgestellt werden, dass es sich bei der betroffenen Anschrift nicht um einen Sonderbereich oder dass es sich nicht um einen nicht-sensiblen, sondern um einen sensiblen Sonderbereich handelte.

### 3.3.7.3 Konsolidierung des RDB

Während der RDB mit weiteren Informationen aus den einzelnen Erhebungsteilen befüllt wurde, war zu gewährleisten, dass die integrierten Informationen widerspruchsfreie Ergebnisse für die Anschrift und die Personen enthielten. Die zentrale Sammlung dieser relevanten Informationen im RDB ermöglichte dabei frühzeitige Abgleiche der Daten. Um widersprüchliche Befragungsergebnisse feststellen zu können, wurden die Ergebnisse der Erhebungsteile in zwei Stufen geprüft. Bereits direkt bei der Integration in den RDB wurden in einer ersten Stufe die Informationen aus den Erhebungen mit den vorhandenen Daten der Melderegister abgeglichen und plausibilisiert. In einer zweiten Stufe konnten durch Gegenüberstellung der Ergebnisse der verschiedenen Erhebungen die Resultate untereinander abgestimmt und Unplausibilitäten bereinigt werden. Ziel der Prüfungen war es, widerspruchsfreie Datensätze zu erhalten. Dazu mussten für einzelne Merkmale Vorfahrtsregeln definiert werden, um bei widersprüchlichen Ergebnissen zu einem Endergebnis zu gelangen. Dies war für die Ermittlung zuverlässiger Einwohnerzahlen zwingend erforderlich.

Die Feststellung der Karteileichen und Fehlbestände konnte endgültig erst mit Abschluss der Mehrfachfallprüfung ausgewiesen werden. Erst der Vergleich mit den bereinigten Melderegistern lieferte die endgültige Zahl an Karteileichen und Fehlbeständen. Damit war es z. B. auch möglich, dass die durch Erhebungsbeauftragte festgestellten Fehlbestände gegenüber der Melderegisterlieferung vom 1. November 2010 durch die zensusrelevanten Melderegisterlieferungen zum 9. Mai 2011 und zum 9. August 2011 zu paarigen Datensätzen wurden.

Vom Grundsatz her sollten beim Zensus 2011 Mehrfachbefragungen an ein und derselben Anschrift ausgeschlossen werden. Dazu wurden Prioritäten für die Auswahl der Anschriften für die verschiedenen Erhebungen festgelegt. So wurden bei der Identifikation der Anschriften für die Befragung zur Klärung von Unstimmigkeiten Sonderanschriften herausgefiltert, um Doppelbefragungen zu vermeiden. Gesteuert wurde diese Einschränkung bereits bei der Markierung der zu befragenden Anschriften zur Klärung von Unstimmigkeiten, da im AGR für diese Erhebung nur Anschriften gekennzeichnet werden konnten, bei denen es sich nicht um Sonderanschriften handelte. Damit wurde sichergestellt, dass diese Anschriften nur von den Erhebungsbeauftragten der Sonderbereichserhebung befragt wurden, da diese zeitlich gesehen früher angesiedelt war.

Auch bei Stichprobenanschriften in Gemeinden mit weniger als 10 000 Einwohnern galt, dass diese zusätzlich für die Befragung zur Klärung von Unstimmigkeiten identifiziert werden konnten. Auch hier wurde mit Hilfe der Angaben des AGR gesteuert, dass die entsprechenden Anschriften nur einmal befragt wurden. Da die Haushaltsstichprobe zeitlich vor der Befragung zur Klärung von Unstimmigkeiten stattfand, wurden die dort erzielten Ergebnisse in Bezug auf die Existenzfeststellung an den Anschriften sowohl in die Ergebnisfelder der Haushaltsstichprobe als auch der Befragung zur Klärung von Unstimmigkeiten aufgenommen.

Anschriften und deren Zuordnung zu einem Erhebungsbereich konnten sich im Zensusablauf jedoch ändern, wenn zu einzelnen Anschriften Rückläufe vermeldet wurden, die nicht den erwarteten entsprachen oder wenn trotz entsprechender Markierungen keine Rückläufe im RDB vermerkt waren. Zu solch unplausiblen Rückläufen konnte es kommen, wenn sich die Bereichszuordnung der Anschriften im Zeitablauf änderte. Die Kennzeichnung der Anschriften mit Sonderbereichen erfolgte in der Regel schon einige Monate vor dem Zensusstichtag. Während dieser Kennzeichnungsphase wurden die Anschriften für die Haushaltebefragung auf Stichprobenbasis gezogen. Kurz vor dem Start der Erhebungsphase wurden die entsprechenden Anschriften dann an die jeweiligen Erhebungsunterstüt-

zungssysteme für Sonderanschriften bzw. für die Haushaltsstichprobe übermittelt. Die Anschriften, an denen die Befragung zur Klärung von Unstimmigkeiten durchzuführen war, konnten dagegen erst nach der Verarbeitung der Rückläufe aus der GWZ identifiziert und im AGR markiert werden. Stellte sich bei der Erhebung heraus, dass es sich bei einer Sonderanschrift doch um eine „Normalanschrift“ handelt, wurde die Sonderanschriftenmarkierung im AGR entfernt. Auch konnte sich bei der Erhebung herausstellen, dass es sich nicht wie vermutet um eine nicht-sensible, sondern um eine sensible Sonderanschrift handelte, die so nicht zur Grundgesamtheit der stichprobenrelevanten Anschriften zählte. Ebenso war es auch möglich, dass Sonderanschriftenmarkierungen erst im späteren Verlauf ergänzt wurden. Ursache konnte hier beispielsweise sein, dass erst bei der Haushaltsstichprobe erkannt wurde, dass es sich um eine Sonderanschrift handelte. Auch weitere zusätzliche Recherchen konnten dazu führen, dass Anschriften auch nach dem Zensusstichtag noch als Sonderanschriften markiert wurden. Sogar Anschriften für die Befragung zur Klärung von Unstimmigkeiten wurden nachträglich noch als Sonderanschriften markiert.

Insbesondere durch die relativ große zeitliche Spanne zwischen Beginn und Ende der Markierung von Sonderanschriften war es möglich, dass am Ende Kombinationen von unzulässigen oder unvollständigen Markierungen im AGR und im RDB enthalten waren. Ursache war hier auch, dass zu befragende Anschriften je nach Markierung im AGR nur in einem bestimmten Erhebungsunterstützungssystem, z. B. im System für die Erhebung an Anschriften mit Sonderbereichen, integriert wurden und diese separat geführten Erhebungsunterstützungssysteme nicht regelmäßig mit dem AGR synchronisiert werden konnten. Das führte dazu, dass neue Informationen zu Markierungen der Anschriften nicht einen Wechsel der Art der Erhebung nach sich zogen.

Auch lagen die Ergebnisse der primärstatistischen Erhebungen zu unterschiedlichen Zeiten vor und wurden damit nicht in einer festgelegten Reihenfolge in den RDB integriert. Damit war zwar der direkte Abgleich mit den Inhalten des RDB aus den Registerlieferungen möglich, eine kontinuierliche Plausibilisierung der Ergebnisse untereinander war jedoch nicht realisierbar.

Um trotzdem sicherstellen zu können, dass nach Abschluss der primärstatistischen Erhebungen plausible Ergebnisse weiterverarbeitet wurden, wurden Regelungen festgelegt, wie bei einem unplausiblen Rücklauf vorzugehen ist. Folgende Grundprinzipien wurden dabei eingehalten.

- War eine Anschrift als Sonderanschrift markiert, hatten immer die Ergebnisse der Sonderbereichserhebung Vorrang.
- Für alle Anschriften der Haushaltebefragung auf Stichprobenbasis lagen am Ende Ergebnisse vor. Folgende Formen waren zulässig:
  - Kennzeichnung der existenten Personen im RDB
  - Aufnahme eines Ausfallgrundes für die Anschrift
  - Rücklauf über die Erhebung an nicht-sensiblen Sonderanschriften
- Konnte eine Stichprobenanschrift, die an einer nicht-sensiblen Sonderanschrift gezogen wurde, nicht über die Sonderbereichserhebung erhoben werden, weil sie entweder inzwischen sensible Sonderanschrift oder Normalanschrift geworden war, so erhielt die Anschrift einen speziellen Ausfallgrund zur Kennzeichnung.
- Wurde bei der Erhebung an einer Stichprobenanschrift festgestellt, dass es sich um eine Sonderanschrift handelte, wurde diese über die Erhebungsunterstützungssysteme der Sonderbereiche erhoben. Gleichzeitig erhielt die Anschrift einen Ausfallgrund als Ergebnis der Haushaltebefragung auf Stichprobenbasis.
- Für alle Anschriften zur Klärung von Unstimmigkeiten lagen am Ende Ergebnisse vor. Folgende Formen waren zulässig:
  - Kennzeichnung der existenten Personen im RDB
  - Aufnahme eines Ausfallgrundes für die Anschrift

- Handelte es sich bei einer Rückfrageanschrift zur Klärung von Unstimmigkeiten auch um eine Stichprobenanschrift, wurde diese nur einmal befragt und die Ergebnisse wurden in beiden Ergebnisfeldern im RDB vermerkt.
- Wurde eine Rückfrageanschrift zur Klärung von Unstimmigkeiten nachträglich als Sonderanschrift markiert, erhielt sie einen Ausfallgrund. Für diese Anschrift wurden die Daten aus der Erhebung an Anschriften mit Sonderbereichen übernommen.

Durch die Festlegung auf die genannten Grundprinzipien war es möglich, trotz unzulässiger oder unvollständiger Rückläufe sicherzustellen, dass zur weiteren Verarbeitung konsistente Ergebnisse verwendet wurden.

### 3.4 Gebäude- und Wohnungszählung (GWZ)

#### 3.4.1 Datengewinnung: Vollerhebung als schriftliche Befragung der Eigentümer

Da in Deutschland keine flächendeckenden Register über Gebäude und Wohnungen existieren, waren die Daten zu Gebäuden und Wohnungen mittels Befragung zu erheben. Dabei gaben die nationalen und EU-Anforderungen nach regional tief gegliederten Ergebnissen bis zur Gemeindeebene zu Gebäuden, Wohnungen und Haushalten den Ausschlag dafür, die Gebäude- und Wohnungszählung (GWZ) nicht als Stichprobe, sondern als Vollerhebung durchzuführen. Der Vorgabe der Politik folgend, ein kostengünstigeres Verfahren zu entwickeln und Auskunftspflichtige zu entlasten, sollten – analog zur 1995 durchgeführten GWZ in den neuen Bundesländern – die Eigentümer bzw. Verwalter der Gebäude und Wohnungen schriftlich befragt werden.

##### 3.4.1.1 Ermittlung der zählungsrelevanten Anschriften

Zur Vorbereitung der postalischen GWZ mussten zunächst die Anschriften der zählungsrelevanten Gebäude, d.h. derjenigen Gebäude, die Wohnraum enthalten, festgestellt werden. Dies geschah im Rahmen des Aufbaus des Anschriften- und Gebäuderegisters (AGR). Die Arbeiten für dieses Register begannen bereits im Jahre 2008 mit dem Einzug der Gebäudeanschriften aus verschiedenen Datenquellen. Die erforderlichen gesetzlichen Grundlagen sind im Zensusvorbereitungsgesetz (ZensVorbG 2011) geregelt.

Ziel der GWZ war die Erhebung von Gebäude- und Wohnungsmerkmalen für alle zählungsrelevanten Erhebungseinheiten: Gebäude mit Wohnraum und bewohnte Unterkünfte <sup>43</sup> sowie Wohnungen.

Um im AGR diejenigen Anschriften, an denen sich Wohnraum befindet, von solchen abzugrenzen, an denen sich beispielsweise nur Gewerberäume befinden, wurde die Bedingung formuliert, dass eine Anschrift immer dann GWZ-relevant ist (also potenziell Wohnraum beinhaltet), wenn dort mindestens eine Person laut Melderegister oder laut Register der Bundesagentur für Arbeit wohnhaft ist oder war, oder wenn eine anderweitige Prüfung der Anschrift durch das zuständige Statistische Amt der Landes mit dem Ergebnis stattgefunden hat, dass dort Wohnraum vorhanden war (dies kann, muss aber keine Vor-Ort-Begehung gewesen sein). Für die so ermittelten 19,5 Millionen GWZ-relevanten Anschriften mussten anschließend die Auskunftspflichtigen für die Gebäude und Wohnungen ermittelt werden.

##### 3.4.1.2 Ermittlung der Auskunftspflichtigen für die zählungsrelevanten Anschriften

Auskunftspflichtig für die GWZ waren laut § 18 ZensG 2011 die „Eigentümer und Eigentümerinnen, die Verwalter und Verwalterinnen, die sonstigen Verfügungs- und Nutzungsberechtigten der Gebäude oder Wohnungen“. Die Ermittlung dieser Auskunftspflichtigen war nur aus solchen Datenquellen zulässig, die der Gesetzgeber ausdrücklich für diese Aufgabe benannt hatte. In § 10 ZensVorbG 2011 waren folgende Quellen zur Ermittlung der Auskunftspflichtigen vorgesehen:

<sup>43</sup> Bewohnte Unterkünfte sind behelfsmäßige Bauten. Hierzu zählen z. B. Wohnbaracken, Bauwagen, Wohnwagen (z. B. auf Campingplätzen), Gartenlauben, Schrebergartenhütten, Jagdhütten, Weinberghütten, Almhütten, fest verankerte Wohnschiffe und Wohncontainer, sofern diese dauerhaft bewohnt werden.



- Für die Grundsteuer zuständige Stellen
- Für die Führung der Grundbücher zuständige Stellen
- Für die Führung der Liegenschaftskataster nach Landesrecht zuständige Stellen
- Finanzbehörden
- Versorgungs- und Entsorgungsbetriebe

Im Vorfeld der Zensusvorbereitung war die Eignung dieser Datenquellen sorgfältig geprüft worden. Im Ergebnis hatten sich die kommunalen Grundsteuerstellen sowie die Versorgungs- und Entsorgungsbetriebe als am meisten geeignete Quellen herausgestellt, da hier sowohl die elektronische Verfügbarkeit im vorgegebenen Datenformat als auch eine akzeptable Aktualität der Auskunftspflichtigenanschriften gegeben waren. Einige Quellen wie das Grundbuch hatten sich hingegen als ungeeignet erwiesen, da sie gerade diese Kriterien nicht flächendeckend erfüllen konnten.

Mit den ausgewählten Datenlieferanten wurde ein standardisiertes Datenformat vereinbart, in dem die Daten übermittelt werden sollten. Diese Daten waren die Anschrift des zahlungsrelevanten Gebäudes (= Objektanschrift), der Name und die postalische Anschrift des oder der Auskunftspflichtigen. Ziel war ein einfaches elektronisches Einlesen der Daten in die IT-Systeme der amtlichen Statistik unabhängig von der Ursprungssoftware der jeweiligen Gemeinde. Dieses Verfahren ließ sich allerdings nicht wie gewünscht realisieren, da trotz dezidierteter Datensatzbeschreibung eine Vielzahl von Lieferanten diese nur teilweise eingehalten hatten bzw. die Ursprungsdaten nicht in den gewünschten Datenfeldern vorlagen. Beispielsweise waren bei den Angaben zu den Auskunftspflichtigen Vor- und Nachnamensfelder mit beliebigen namensähnlichen Inhalten gefüllt oder die Objektanschriften enthielten statt Straße und Hausnummer Flurstückangaben, sodass eine Verknüpfung über die eigentliche Objektanschrift nicht möglich war. Die Bearbeitung dieser Daten und weitere Recherchen zur Komplettierung der Auskunftspflichtigenangaben haben einen immens hohen, in dieser Form nicht vorhersehbaren Aufwand zur Vorbereitung der GWZ verursacht (April 2009 bis Mitte 2011).

Zentrales IT-Instrument zur Verarbeitung der Auskunftspflichtigenangaben war das als Satellitensystem am AGR angebundene „Gebäude- und Wohnungseigentümerregister“ (GWER), das sämtliche Angaben der Auskunftspflichtigen mit den Objektanschriften der zahlungsrelevanten Gebäudeanschriften im AGR verknüpfte. Das GWER diente als zentrales Steuerungsinstrument für alle Prozesse der Datenerhebung innerhalb der GWZ (Erstellung der Versanddateien, Rücklaufkontrolle, Mahnwesen), sodass die Objektangaben nicht redundant vorgehalten werden mussten. Die Verknüpfung der Auskunftspflichtigen mit den Objekten im AGR war allerdings auch nur dann erfolgreich, wenn die Objektangaben aus den Datenquellen in gleicher Weise vorlagen wie im AGR, wenn also die Schreibweisen der Straßen, Hausnummern, Hausnummernzusätze etc. übereinstimmten. Dafür wurden entsprechende Standardisierungsroutinen eingesetzt.

Ob die Verknüpfung mit dem vermeintlichen Auskunftspflichtigen auch fachlich-inhaltlich korrekt war – sodass später tatsächlich die Fragen zum Objekt beantwortet werden konnten – war zu diesem Zeitpunkt noch unklar. Hintergrund hierfür war, dass die verwendeten Daten der Grundsteuerstellen oder der Ver- und Entsorgungsbetriebe zwar den Einzug von Steuern und Gebühren ermöglichten, aber nichts darüber aussagten, ob die Steuern und Gebühren zahlenden Personen oder Unternehmen weitergehende Angaben zu den Gebäuden und Wohnungen machen konnten. Es kam durchaus vor, dass die formalrechtlich Auskunftspflichtigen das Gebäude faktisch nicht kannten und die Fragen deshalb gar nicht beantworten konnten. Dies war insbesondere bei Gebäuden mit Eigentumswohnungen der Fall, für die von den Datenlieferanten eine Vielzahl von Auskunftspflichtigenangaben übermittelt wurden ohne weitere Angaben darüber, ob es sich dabei um Eigentümer, Verwalter von Eigentümergemeinschaften, Mietverwalter oder gegebenenfalls auch Mieter handelte. Dies führte oftmals dazu, dass für ein und dasselbe Gebäude sowohl der Verwalter als auch die einzelnen Wohnungseigentümer und weitere Personen angeschrieben wurden, was zum einen die Vollzähligkeitskontrolle erschwerte, zum anderen bei den Eigentümergemeinschaften Unmut auslöste, da einige Verwalter für die „Dienstleistung“ des Fragebogenausfüllens nicht unerhebliche Kosten berechnet hatten.

Um die Aktualität der Auskunftspflichtigenanschriften zu überprüfen, wurde ca. sechs Monate vor dem Zensusstichtag eine GWZ-Vorbefragung durchgeführt. Diese diente vor allem dazu, vermeintlichen Auskunftspflichtigen, die das Objekt zwischenzeitlich verkauft hatten, die Möglichkeit zu geben, den neuen Eigentümer anzugeben sowie zu prüfen, ob die Auskunftspflichtigen an der ermittelten Anschrift noch postalisch zu erreichen waren (Zustellbarkeitsprüfung). Darüber hinaus hatte die Vorbefragung den Zweck, die Anzahl der Wohnungen der Auskunftspflichtigen an den Anschriften zu ermitteln, um damit die Anzahl der jeweils benötigten Fragebogen feststellen zu können. Außerdem wurde abgefragt, ob die Nutzung des Online-Fragebogens gewünscht war. Einige Bundesländer (z. B. Sachsen) haben im letzteren Fall gar keinen Papier-Fragebogen mehr, sondern ausschließlich eine Zugangskennung verschickt. Die Vorbefragung wurde in einigen Bundesländern flächendeckend und in einigen Bundesländern für ausgewählte Fälle (z. B. Gebäude mit Eigentumswohnungen) durchgeführt. Insgesamt wurden für die Vorbefragung etwa neun Millionen Fragebogen versandt.

### 3.4.1.3 Postalische Datenerhebung

#### a) Fragebogenentwicklung

Die im ZensG 2011 festgelegten Erhebungsmerkmale wurden in einem Fragebogen umgesetzt, der die folgenden, sich zum Teil widersprechenden Anforderungen erfüllen musste:

- Geringer Papierverbrauch (zur Minimierung der Portokosten), trotzdem die Möglichkeit, möglichst viele Wohnungen platzsparend auf nur einem Bogen zu melden
- Technische Anforderungen an Farbe, Kästchengröße, Randbreiten etc. für die Beleglesung
- Gute inhaltliche Verständlichkeit für die Auskunftspflichtigen, da kein Interviewer persönlich für Rückfragen anwesend sein würde, insbesondere gute Erläuterungen zu den Fragen

Der Kompromiss dieser Anforderungen ergab einen Fragebogen, auf dem Angaben für bis zu sechs Wohnungen gemacht werden konnten zuzüglich eines Deckblattes mit einigen notwendigen sogenannten Screeningfragen, z. B. um Auskunftspflichtigen die Möglichkeit zu geben, bereits hier anzugeben, dass sie die falschen Adressaten waren oder dass das Objekt gar keine Wohnungen enthielt.

Analog zum Papierfragebogen wurde ein elektronischer Fragebogen für die Online-Erhebung vorbereitet. Dazu ist das in der amtlichen Statistik bereits länger eingesetzte Online-Erhebungsinstrument „IDEV“ zur Anwendung gekommen. Dieses bot im Gegensatz zum Papierfragebogen einige Komfort-Funktionen. Z. B. konnten die Auskunftspflichtigen ihre zählungsrelevanten Wohnungen im Online-Fragebogen selbst flexibel anlegen, abhängig davon, wie viele Wohnungen sie melden mussten. Eine fixe Festlegung wie im Papierbogen auf sechs Wohnungen und damit verbunden die Notwendigkeit, ggf. weitere Papierbogen anzufordern, war daher beim Online-Fragebogen nicht notwendig.

#### b) Druck und Versand der Fragebogen

Der Prozess der postalischen Datenerhebung war auf den fristgemäßen Einzug qualitätsgerechter Daten und eine zügige Bearbeitung mit effizienten maschinellen Verfahren ausgerichtet. Ziel war die vollzählige Speicherung der Gebäude- und Wohnungsdaten als Rohdaten in der zentralen GWZ-Datenbank. Um einen fristgerechten Versand der Erhebungsunterlagen vor dem Stichtag zu gewährleisten, haben die Druckdienstleister von den Statistischen Ämtern der Länder frühzeitig Auszüge aus dem GWER mit den aktuell ermittelten und qualitätsgesicherten Angaben zu den Auskunftspflichtigen und den dazugehörigen Objekten erhalten. Die Weitergabe dieser sensiblen Daten an externe Dienstleister erforderte, dass die dort betroffenen Mitarbeiterinnen und Mitarbeiter auf das Statistikgeheimnis verpflichtet werden mussten. Die Gebäude- und Auskunftspflichtigenanschriften sowie die entsprechenden maschinenlesbaren Kennzeichen wurden in die vorab gedruckten Fragebogen eingedruckt. Nach entsprechendem Druck und Konfektionierung der Fragebogen (Festlegung der benötigten Anzahl) wurden diese direkt von der Druckerei kuvertiert und versendet.

Bei der GWZ konnte ein Gebäude mehrere Auskunftspflichtige aufweisen (z. B. Eigentümer von Eigentumswohnungen) und Auskunftspflichtige konnten für mehrere Gebäude zuständig sein. Da alle Auskunftspflichtigen für jedes Objekt einen Fragebogen erhalten hatten, wurden insgesamt 22,5 Millionen Fragebogen versandt.



### c) Rücklauf der Fragebogen

Der Rücklauf der Fragebogen von den Auskunftspflichtigen erfolgte direkt an die beleglesenden Einrichtungen. Dies konnten die Beleglesezentren der Statistischen Ämter der Länder oder externe Dienstleister sein. Letztere mussten auf das Statistikgeheimnis verpflichtet sein. In der Beleglesung wurden zunächst die Belegkennzeichen gescannt, um den generellen Rücklauf („Ist ein Fragebogen eingegangen oder nicht?“) zu dokumentieren. Diese reine Rücklaufdokumentation diente der Steuerung des Mahnwesens. Dann wurden die eigentlichen Inhalte der Fragebogen gescannt und maschinell gelesen. Die gelesenen und in Codes umgesetzten Inhalte wurden über die zentrale Eingangsdatenbank (EDB) in das GWZ-Datenaufbereitungssystem transferiert, wo anschließend weitere Schritte der Aufbereitung (siehe Kapitel 3.4.2) stattfanden.

### d) Dateneingang der elektronischen Meldungen

Allen Auskunftspflichtigen war die Möglichkeit gegeben, die Erhebungsdaten mit dem entsprechenden Online-Fragebogen elektronisch zu übermitteln. Die Zugangskennungen zu den Online-Fragebogen wurden den Auskunftspflichtigen zusammen mit den Papierfragebogen zugesendet. Beim Einloggen in den Online-Fragebogen waren die entsprechenden Objektangaben dort schon vorbelegt. Der Online-Fragebogen enthielt bereits einige Plausibilitätsprüfungen. Der ausgefüllte Online-Fragebogen ging ebenfalls über die zentrale EDB in das GWZ-Datenaufbereitungssystem ein. Analog fand auch hierfür eine Rücklaufkontrolle für das Mahnwesen statt.

Mit Wohnungsunternehmen konnten spezielle Vereinbarungen getroffen werden, in denen die Lieferung der Erhebungsdaten in Dateiform festgelegt wurde. Der Lieferweg hierfür war der in der amtlichen Statistik bereits eingesetzte CORE.reporter, der den Wohnungsunternehmen kostenlos zum Download zur Verfügung gestellt wurde. In den CORE.reporter wurde eine vorher von den Wohnungsunternehmen nach Vorgaben der amtlichen Statistik erstellte csv-Datei hochgeladen, überprüft und anschließend über die EDB an das GWZ-Datenaufbereitungssystem übermittelt. Ca. 6 000 Wohnungsunternehmen haben für 1,8 Millionen Gebäude mittels CORE.reporter gemeldet.

Insgesamt wurden etwa 15 Millionen Meldungen auf Papier bearbeitet, 7,5 Millionen erfolgten online. Für etwa 90 000 Gebäude konnten keine Auskunftspflichtigen ermittelt werden, für 270 000 Gebäude erfolgte kein Rücklauf und für weitere 580 000 Gebäude wurden leere oder fast leere Fragebogen zurückgesandt. Die fehlenden Angaben zu diesen ca. eine Million Gebäuden mussten später imputiert werden.

## 3.4.2 Plausibilisierung und Imputation der GWZ-Daten

Der Plausibilisierung in der GWZ lagen folgende Prinzipien zugrunde:

- **Einsatz überwiegend maschineller Verfahren**

Aufgrund der großen Datenmenge war die Plausibilisierung als maschinelles Verfahren konzipiert. Manuelle Eingriffe waren nur in wenigen Fällen vorgesehen und sollten sich auf einen bestimmten Fehlertyp (Unplausibilitäten zwischen der Zahl der Wohnungen laut GWZ und der Zahl der laut Melderegister wohnhaften Personen) bzw. den manuellen Abgleich zwischen den Namen der Bewohnerinnen und Bewohner laut GWZ und den Namen der gemeldeten Personen beschränken.

- **Einheitliches Verfahren**

Für alle Bundesländer wurde ein einheitliches Verfahren angewendet, von dem die Länder nur in Ausnahmefällen und aufgrund landesspezifischer Besonderheiten abgewichen sind.

- **Kontinuierliche und frühestmögliche Plausibilisierung der Daten**

Anders als in vielen anderen Statistiken konnte mit dem Start der Plausibilisierung nicht gewartet werden bis alle Daten zu allen Gebäuden eingegangen waren. Stattdessen startete die Plausibilisierung für ein Gebäude automatisiert sobald alle Angaben zu diesem Gebäude vorhanden waren. Dies hatte zeitliche Gründe. Die geplanten Rückfragen bei den Auskunftspflichtigen sowie die Arbeiten, in welche die kommunalen Erhebungsstellen eingebunden waren, sollten möglichst stichtagsnah erfolgen. Voraussetzung für diese Schritte waren geprüfte bzw. plausible Daten.

### 3.4.2.1 Maschinelle Fehlererkennung und -beseitigung

#### a) Bearbeitung von Gebäude- und Wohnungsdubletten

Zu Beginn der Plausibilisierung musste sichergestellt werden, dass zu jedem Gebäude genau ein Datensatz mit den zugehörigen Gebäude- und Wohnungsangaben vorhanden war. Da teilweise auch mehrere Auskunftspflichtige die Angaben für ein Gebäude übermittelten – etwa für Gebäude mit Eigentumswohnungen die einzelnen Wohnungseigentümerinnen und -eigentümer für ihre jeweiligen Wohnungen – mussten diese einzelnen Angaben zunächst „zusammengefasst“ werden. Falls mehrere Datensätze zu einem Gebäude vorlagen, wurde die jeweils häufigste Nennung für die Gebäudemerkmale übernommen. Anschließend wurde in einem relativ aufwändigen Verfahren geprüft, ob unter den jeweils übermittelten Wohnungen Wohnungsdubletten vorhanden waren oder ob Wohnungen fehlten.

#### b) Fehlererkennung

Im Anschluss an diesen Schritt konnte die eigentliche Plausibilisierung beginnen. Dabei erfolgte eine Prüfung der Daten hinsichtlich der Kriterien Vollzähligkeit, Vollständigkeit, Strukturplausibilität, Interplausibilität sowie auf Plausibilität der Angaben im Vergleich mit dem Melderegister (die Anzahl der Wohnungen im Gebäude wurde auf starke Abweichungen zur Anzahl der gemeldeten Personen geprüft).

Insgesamt wurden in der GWZ über den Fragebogen 13 Erhebungsmerkmale sowie eine Reihe von Hilfsmerkmalen erfasst. Für die Prüfung wurden bestimmte Plausibilisierungsregeln aufgestellt (insgesamt 109), die alle Datensätze erfüllen mussten. Bei diesen Regeln wurde zwischen Fehlern und Prüfhinweisen unterschieden. Fehler sind Unplausibilitäten von Merkmalen, die in jedem Fall korrigiert werden müssen. Von Prüfhinweisen spricht man, wenn der vorhandene Wert für ein Merkmal falsch sein könnte, dies aber erst geprüft werden muss. Besonders viele Fehler traten bei den Merkmalen „Eigentumsverhältnisse des Gebäudes“ und „Eigentumsverhältnisse der Wohnung“ sowie bei der „Anzahl der Wohnungen im Gebäude“ auf. Ein Beispiel für einen Prüfhinweis waren Unplausibilitäten zwischen der Anzahl der Wohnungen und der Anzahl der gemeldeten Personen an einer Anschrift. Solche Gebäude wurden von den Statistischen Ämtern der Länder geprüft.

#### c) Deterministische Imputation

Das Verfahren der deterministischen Imputation wurde sowohl bei unplausiblen als auch bei fehlenden Merkmalen (Item Non-Response) angewendet. Es konnte immer dann eingesetzt werden, wenn eine eindeutige Beziehung zwischen dem unplausiblen bzw. fehlenden Merkmal und einem oder mehreren plausiblen Merkmalen vorlag. So wurde z. B. in einem Datensatz, bei dem bei der Frage G3 zum Gebäudetyp keine Angabe vorlag und bei Frage G1 nach der Art des Gebäudes „Bewohnte Unterkunft“ angekreuzt war, das Merkmal Gebäudetyp auf „Anderer Gebäudetyp“ gesetzt.

Zu den Verfahren der deterministischen Imputation gehören auch Fixeinsetzungen. So wurde z. B. bei einer fehlenden Angabe zum Merkmal „Wohnungstyp“ (also ob es sich um eine Ferien- oder Freizeitwohnung bzw. Diplomaten- oder Streitkräftewohnung handelte) immer „keines von beiden“ eingesetzt.

Allerdings war der Anteil der Fehler, die auf diese Weise korrigiert werden konnten, relativ gering, da solche eindeutigen Beziehungen zwischen den erhobenen Merkmalen nur in wenigen Fällen vorlagen und insbesondere bei logischen Widersprüchen zwischen zwei Merkmalen nicht immer eindeutig war, welches davon das fehlerhafte Merkmal war. Aus diesem Grund musste für die Korrektur von Merkmalen noch ein weiteres Verfahren eingesetzt werden.

#### d) Imputation mit einem Hot-Deck-Verfahren nach dem Nearest-Neighbour-Prinzip

Als zweites Korrekturverfahren wurde ein Hot-Deck-Imputationsverfahren verwendet, das nach dem Nearest-Neighbour-Prinzip arbeitet (vgl. Kapitel 3.3.3.3. b)). Bei diesem Verfahren wird jedes unplausible oder fehlende Merkmal durch das entsprechende Merkmal eines real beobachteten, plausiblen Datensatzes ersetzt, der dem Datensatz mit fehlenden und/oder unplausiblen Werten am „ähnlichsten“ ist, der also die geringsten Abweichungen unter allen vorhandenen plausiblen Datensätzen aufweist.

In der GWZ wurde hierfür kein eigenes Verfahren programmiert, sondern die bereits bestehende Imputationssoftware CANCEIS <sup>44</sup> verwendet. CANCEIS wurde von Statistics Canada entwickelt und bereits in einer Reihe von Zensen weltweit zur Imputation von Daten eingesetzt. CANCEIS ist ein sog. „Edit-and-Imputation“-Programm. Es prüft die Datensätze über die vorher aufgestellten Regeln zunächst auf ihre Plausibilität und teilt sie auf dieser Basis automatisiert in Spender- und Empfängerdatensätze ein. Im Anschluss sucht das Programm für jeden unplausiblen Datensatz nach den „ähnlichsten“ Spenderdatensätzen, den sogenannten „Nächsten Nachbarn“. <sup>45</sup>

Mit Hilfe von CANCEIS wurden nicht nur einzelne fehlerhafte Merkmale, sondern auch fehlende Wohnungen imputiert. Insgesamt wurden in der GWZ einzelne Merkmale bei rund 51 % der Gebäude deterministisch und bei rund 30 % der Gebäude mit Hilfe von CANCEIS imputiert. <sup>46</sup> Darunter sind allerdings auch Gebäude, die mit Hilfe beider Imputationsverfahren korrigiert wurden.

### e) Imputation vollständig fehlender Gebäude

Wenn zu einem Gebäude keine Auskunftspflichtigen recherchiert werden konnten oder keine Angaben zum Gebäude übermittelt wurden (Unit Non-Response), musste das gesamte Gebäude imputiert werden. Dieser Schritt konnte erst ganz am Schluss nach dem Ende der Erhebungsphase erfolgen.

Hierfür wurde ein eigenes Imputationsverfahren entwickelt. Dabei wurde von der Annahme ausgegangen, dass alle Gebäude, die sich in einem baulich homogen strukturierten Gebiet befinden, grundsätzlich ähnlich sind. Dies bedeutet, dass in solchen Bereichen (z. B. typischerweise Einfamilienhausgebieten) das ausgefallene Gebäude ähnliche bzw. identische Gebäude- und Wohnungsmerkmale wie die Gebäude in der unmittelbaren Umgebung aufweist. In solchen Fällen wurden daher für das ausgefallene Gebäude die Angaben eines Nachbargebäudes imputiert.

Dafür musste in einem ersten Schritt das Gebiet bestimmt werden, das auf seine bauliche Strukturierung untersucht werden sollte. Dies geschah, indem mit Hilfe der Gebäudekoordinaten und der Anschrift des ausgefallenen Gebäudes räumliche Bereiche definiert und so lange schrittweise vergrößert wurden, bis sich eine bestimmte Anzahl von Gebäuden (mind. zehn) innerhalb dieses Bereiches befanden. Diese Gebäude wurden dann anhand der beiden Gebäudemerkmale „Anzahl der Wohnungen“ und „Baujahr“ daraufhin untersucht, ob sie baulich ähnlich waren.

Lag das Gebäude in einem baulich homogen strukturierten Gebiet, wurde eines der plausiblen Gebäude in dem untersuchten Gebiet als Spendergebäude verwendet. Wenn das Gebiet in der Umgebung des ausgefallenen Gebäudes nicht homogen strukturiert war, musste eine Begehung durch Erhebungsbeauftragte der Gemeinde erfolgen, bei der die wichtigsten Gebäudemerkmale durch Inaugenscheinnahme ermittelt und erfasst wurden (sog. Ersatzvornahmen).

Etwa 2,3 % der in der GWZ erfassten Gebäude wurden komplett imputiert und für weitere rund 3 % aller Gebäude wurden die Daten mit Hilfe von Erhebungsbeauftragten der kommunalen Erhebungsstellen nacherhoben.

44 Eine genaue Beschreibung der Funktionsweise von CANCEIS (CANadian Census Edit and Imputation System) liefert auch BANKIER, M./CROWE, S. (2009): Enhancements to the 2011 Canadian Census E & I System. Working Paper No.15, UNECE Work Session on Statistical Data Editing, Neuchâtel.

45 Wie CANCEIS in die Aufbereitung der GWZ-Daten integriert wurde, ist beschrieben in Grunwald, S./Krause, A., „Umgang mit fehlenden Angaben in der Gebäude- und Wohnungszählung 2011“, in: WiSta 8/2014, S. 437-449.

46 Auskunftspflichtige hatten insbesondere Probleme mit dem Merkmal „Eigentumsverhältnisse der Wohnung“ und verstanden häufig nicht richtig, in welchen Fällen dieses Merkmal ausgefüllt werden musste. Schließt man daher die Gebäude, bei denen nur dieses Merkmal unplausibel war von der Auswertung aus, so sinkt der Anteil der Gebäude mit deterministischen Imputationen auf rund 25 % und der Anteil der Gebäude mit durch CANCEIS imputierten Merkmalen auf rund 26 %.

### 3.4.2.2 Manuelle Prüfungen und Korrekturen

Die manuelle Prüfung von Datensätzen war aufgrund der großen Datenmenge auf einige wenige Konstellationen reduziert. In Fällen, in denen aus den Angaben der Auskunftspflichtigen hervorging, dass es sich um ein Gebäude ohne Wohnraum handelte, dort aber Personen gemeldet waren, prüften die Statistischen Ämter der Länder diese Angaben mit den ihnen nach Zensusgesetz zur Verfügung stehenden Informationen. Einige Statistische Ämter der Länder nahmen dazu auch noch einmal Kontakt mit den Auskunftspflichtigen auf oder beauftragten die Erhebungsstellen in den Gemeinden mit einer Prüfung.

Darüber hinaus kam es zu Rückfragen respektive Prüfungen, wenn die Anzahl der Wohnungen in einem Gebäude deutlich von der Anzahl der gemeldeten Personen abwich.

Eine weitere Aufgabe, die von den Statistischen Ämtern der Länder manuell bearbeitet wurde, war der sogenannte manuelle Namensabgleich mit den Daten aus dem Melderegister. Dieser fand im Vorfeld der Haushaltegenerierung statt. In der GWZ wurden grundsätzlich die Namen von zwei Bewohnern bzw. Bewohnerinnen der Wohnungen erhoben. Für jeden dieser Namen wurde geprüft, ob er zu einer an der Anschrift gemeldeten Person passte. Immer wenn dies der Fall war, wurde der aus der GWZ stammende Name mit der gemeldeten Person verknüpft. Dies war notwendig, um später u. a. mit Hilfe der Informationen aus dem Melderegister für diese Wohnung einen Haushalt generieren und die im Zensus erhobenen Informationen für diesen Haushalt auswerten zu können (siehe Kapitel 4). Die Verknüpfung der Namen erfolgte in zwei Schritten. Im ersten Schritt wurde versucht, so viele Namen wie möglich mit Hilfe eines maschinellen Verfahrens zu verknüpfen. Da für einen Großteil der Wohnungen in der GWZ Informationen über einen Papierfragebogen erhoben wurden und hier nicht immer alle Buchstaben korrekt von der Beleglesung erfasst werden konnten bzw. weil teilweise auch ungenaue oder unvollständige Informationen von den Auskunftspflichtigen übermittelt wurden, wurde versucht, die verbleibenden Namen manuell einer an der Anschrift gemeldeten Person zuzuordnen.

### 3.4.2.3 Nachträgliche Korrekturen

Die Plausibilisierung der Daten wurde durch eine Reihe von Auszählungen des GWZ-Datenbestandes begleitet. Dabei wurden die GWZ-Ergebnisse mit bestimmten externen Eckwerten z. B. aus der Wohnungsbestandsfortschreibung, verglichen. Im Zuge dieser Prüfungen konnten – jedoch erst nach Abschluss der Datenaufbereitung – einige weitere Fehler in den Daten entdeckt werden, die dann noch nachträglich beseitigt werden mussten. Dazu gehörten insbesondere Beleglesefehler durch fehlerhafte Zeichenerkennung oder durch unsachgemäße Handhabung des Fragebogens (wenn Auskunftspflichtige leere Fragebogenseiten durchgestrichen haben, was die Beleglesung u. U. als gültige Markierungen erkannt hat.).

Insgesamt waren in der GWZ nur 31 % der Gebäude vollständig fehlerfrei berichtet worden. Weitere 26 % der Gebäude wiesen nur bei dem Merkmal „Eigentumsverhältnisse der Wohnung“ Unplausibilitäten auf. Wie eingangs beschrieben, wurde häufig nicht richtig verstanden, wann Angaben zu diesem Merkmal notwendig waren und wann nicht. Bezieht man alle diejenigen Gebäude ein, bei denen maximal dieses eine Merkmal unplausibel war, waren dies rund 57 % der Gebäude.

## 4 Generierung von Haushalten: die verschiedenen Stufen der Haushaltegenerierung sowie das integrierte Korrekturverfahren

Im Rahmen der Haushaltegenerierung wurden sämtliche Datenquellen des Zensus 2011 zusammengeführt und zu einem zensustypischen Datensatz verknüpft, der fachlich und regional tief gegliederte Auswertungen zulässt. Dieser wurde im Rahmen des in der Haushaltegenerierung eingebetteten Korrekturverfahrens noch mit den berechneten amtlichen Einwohnerzahlen synchronisiert. Hierzu flossen hochgerechnete Ergebnisse aus der Haushaltsstichprobe in das Verfahren ein.

Beim Zensus 2011 wurden Informationen über Haushalte und Familien nicht erhoben, vielmehr wurden durch Zusammenführung von Informationen aller verschiedenen Datenquellen flächendeckend Haushalte generiert. Diese „Haushaltegenerierung“ hat den Vorteil, ohne zusätzliche Erhebung von Merkmalen Haushalts- und Familieninformationen – auch kleinräumig – zur Verfügung zu stellen. Sie ist daher für die Bevölkerung ein belastungsarmes Instrument zur Gewinnung von haushaltsstatistischen Informationen.

Haushaltegenerierung bedeutet zweierlei: Zum einen die anschriftenweise Zusammenführung von Personen zu Familien und Haushalten, zum anderen das Verknüpfen von Haushalten mit konkreten Wohnungen unter der jeweiligen Anschrift. Damit wurden im deutschen Zensus Wohnhaushalte abgebildet. Dies steht in Einklang zu den Vorgaben der EU, bedeutet auf nationaler Ebene jedoch eine eingeschränkte Vergleichbarkeit zum Beispiel zu den Ergebnissen des Mikrozensus, da dort Wirtschaftshaushalte nachgewiesen werden. Für die EU werden entsprechend der Vorgaben bei der Haushaltegenerierung nur Personen mit alleiniger oder Hauptwohnung berücksichtigt, während nach nationalem Konzept auch Nebenwohnsitzpersonen Berücksichtigung finden, eine Person also nach deutschem Haushaltskonzept mehreren Haushalten gleichzeitig angehören kann.

Für die Haushaltegenerierung im Rahmen des Zensus 2011 war wesentlich, dass auch das statistische Korrekturverfahren, also die Bereinigung von Über- und Untererfassungen, deren Umfang und Struktur mit Hilfe der Haushaltsstichprobe des Zensus festgestellt und hochgerechnet wurde, im Zuge der Haushaltegenerierung erfolgte. Damit war gewährleistet, dass die Löschungen und Imputationen im Personen-Registerbestand nicht zu Verzerrungen in den Haushaltsstrukturen führten.

Die Haushaltegenerierung erfolgte in mehreren Stufen, die nachstehend vereinfacht beschrieben werden.<sup>47, 48</sup>

### 4.1 Schritt 1: Bildung erster Haushaltszusammenhänge aus Registerinformationen – Verzeigerungen

Das Melderegister enthält Verweise zwischen Personen, die Verzeigerungen genannt werden. Diese Verzeigerungen geben eindeutige Hinweise auf Beziehungen zwischen an der gleichen Anschrift gemeldeten Personen. Zum einen werden mit Hilfe von Verzeigerungen eingetragene Partnerschaften zwischen zwei Personen gekennzeichnet. Dies sind im Einzelnen die Ehe und die eingetragene gleichgeschlechtliche Lebenspartnerschaft. Zum anderen werden mit Hilfe von Verzeigerungen gesetzliche Vertretungen gekennzeichnet. Dabei kann es sich um Eltern-Kind-Beziehungen, aber auch um sonstige gesetzliche Vertretungen handeln, wobei Kinder üblicherweise lediglich bis zur Vollendung des achtzehnten Lebensjahres verzeigert sind.

Im ersten Schritt der Haushaltegenerierung wurden die Verzeigerungen aus dem Melderegister genutzt, um Partner sowie Eltern und deren Kinder zusammenführen. Nach diesem Schritt war jede Registerperson einem vorläufigen Haushalt zugeordnet, welcher sich im weiteren Verlauf der Haushaltegenerierung immer wieder verändern konnte. Diese Haushalte bestanden entweder aus Einzelpersonen oder aus einfachen Familienstrukturen resultierend aus den Verzeigerungen.

<sup>47</sup> Die Reihenfolge und Beschreibung der Schritte dient dem methodischen Verständnis und spiegelt nicht im Detail die technische Realisierung wider.

<sup>48</sup> Siehe auch Kreuzmair, I./Reisch M., „ZENSUS 2011: Ablauf der Haushaltegenerierung“, in: Bayern in Zahlen 9/2012, S.615-624.

### 4.2 Schritt 2: Auswertung der Wohnungsnutzerinformation und erste Verknüpfung von Haushalten mit Wohnungen

Um die Verknüpfungen zwischen den aus der Gebäude- und Wohnungszählung gelieferten Wohnungen und den Personen aus den Melderegisterdaten herzustellen, fand ein maschinelles Abgleichverfahren von Namensfeldern der beiden Erhebungsteile statt, der sogenannte maschinelle Namensabgleich. Hierfür wurden im Rahmen der GWZ pro Wohneinheit die Namen von bis zu zwei Bewohnern bzw. Bewohnerinnen erfragt.

### 4.3 Schritt 3: Bildung weiterer Haushaltszusammenhänge aus Registerinformationen – Referenzen

Da die im ersten Generierungsschritt ausgewerteten Verzeigerungen lediglich klassische Haushaltsstrukturen abbildeten – und diese auch nur, wenn sie explizit im Melderegister angegeben waren – war es notwendig, personenbezogene Registerinformationen zu nutzen, um Indizien für weitere Haushaltszusammengehörigkeiten zu erhalten. Dabei wurden neben den klassischen Haushaltsstrukturen auch nichteingetragene Partnerschaften oder Großeltern-Enkel-Beziehungen erstellt. Hierzu kam ein mehrstufiges regelbasiertes Verfahren zur Anwendung. Je nachdem, welcher Beziehungstyp gesucht wurde, wurden Kombinationen unterschiedlicher Registerinformationen verwendet. Zum Beispiel wurden nichteingetragene Partnerschaften anhand der Informationen zu Geschlecht, Alter, Familienstand, Zuzugsdatum und Zuzugsanschrift abgeleitet.

Nach diesem Schritt waren bereits 92 % aller Haushalte fertig generiert. 76 % waren bereits mit einer Wohnung verknüpft, die restlichen 16 %, für die keine Übereinstimmung von Wohnungsnutzernamen gefunden werden konnte (v. a. auch, da keine angegeben waren), wurden in Schritt 5 noch mit einer Wohnung verknüpft.

### 4.4 Schritt 4: Integriertes Korrekturverfahren<sup>49</sup>

Um einen qualitativ hochwertigen, fachlich und regional flexibel auswertbaren Zensus-einzeldatensatz zu erhalten, musste der Registerdatenbestand auf der Basis der Einzeldaten um Über- und Untererfassungen bereinigt werden. Für diese statistische Korrektur der Registerfehler wurde eine Differenzierung der mit der Haushaltsstichprobe festgestellten Karteileichen und Fehlbestände nach demografischen Merkmalen sowie nach Haushaltszusammenhängen (entweder wurde für alle Haushaltsmitglieder die Eigenschaft „Karteileiche“ bzw. „Fehlbestand“ festgestellt oder nur für einzelne Haushaltsmitglieder) benötigt, um eine systematische Verzerrung der demografischen bzw. haushaltsstatistischen Zensusergebnisse zu vermeiden.

Zu diesem Zwecke wurde ein Verfahren entwickelt, das die gemeindeweise aggregierten Vorgaben aus der Haushaltsstichprobe möglichst genau umsetzte. Die gemeindeweise Bereinigung der Karteileichen und Fehlbestände fand dabei in einem zwei- bzw. dreistufigen Verfahren statt:

1. Zunächst wurde anhand der aus der Haushaltsstichprobe geschätzten Verteilungen der Karteileichen und Fehlbestände nach einzelnen demografischen Merkmalen (Alter, Geschlecht, Staatsangehörigkeit, Familienstand) sowie nach Erwerbstätigkeit näherungsweise bestimmt, wie viele Personen einer bestimmten Merkmalskombination als Karteileichen zu löschen bzw. als Fehlbestände zu imputieren sind. Die Gesamtzahl der zu löschenden und zu imputierenden Datensätze pro Gemeinde stellte dabei eine strikt einzuhaltende Vorgabe dar.
2. Im nächsten Schritt wurden dann die konkreten Datensätze bestimmt, die zu löschen bzw. zu doppeln waren. Dabei wurden sowohl die in Schritt 1 vorgenommene Klasseneinteilung in Merkmalskombinationen als auch haushaltsstatistische Randbedingungen berücksichtigt. Das bedeutet zum

<sup>49</sup> Siehe auch Hofmeister, K./Dr. Fürnrohr, M., „Das Korrekturverfahren beim Zensus 2011“, in: Bayern in Zahlen 6/2014, S. 310-318.



Beispiel, dass eine Person mit einer „passenden“ Merkmalskombination sinnvollerweise an einer Anschrift gelöscht wird, an der die bis zu diesem Prozessschritt generierten Haushalte bereits „voll“ sind.

3. Für die Fehlbestände musste in einem dritten Schritt entschieden werden, in welchen „Empfängerhaushalt“ oder in welche – gegebenenfalls auch noch nicht belegte – „Empfängerwohnung“ die duplizierte Person hineinimputiert wird. Hierzu wurden Schätzungen aus der Haushaltsstichprobe zur Anzahl verschiedener Haushaltsgrößen und -typen herangezogen, d. h. das Hineinimputieren einer Person in einen Haushalt sollte möglichst bis zu diesem Prozessschritt noch überrepräsentierte Haushaltstypen verringern und die Anzahl unterrepräsentierter Haushaltstypen erhöhen.

Eine solche Korrektur der Einzeldaten konnte nur statistisch erfolgen. Es war nicht die buchhalterisch betrachtete „Richtigkeit“ des Einzelfalls relevant, sondern die Produktion möglichst unverzerrter Strukturdaten über die Bevölkerung in Deutschland.

In 5 % aller Haushalte erfolgte durch die Löschung von Karteileichen oder die Imputation von Fehlbeständen noch eine Änderung.

#### **4.5 Schritt 5: Zuweisung von Haushalten zu Wohnungen beziehungsweise zu bereits bestehenden Haushalten nach statistischen Generierungskriterien**

In den bisher dargestellten Schritten der Haushaltegenerierung wurden an einer Anschrift – abgesehen von den im Korrekturverfahren (Kapitel 4.4) mit reinen Fehlbestandshaushalten belegten Wohnungen – nur diejenigen Wohnungen mit den generierten Haushalten zusammengeführt, für die mindestens ein Wohnungsnutzername aus der GWZ vorlag, der mit den an dieser Anschrift im Melderegister vorhandenen Personen über den Namensabgleich zusammenzuführen war. Die ggf. noch vorhandene Restmenge aller anderen an dieser Anschrift noch vorhandenen, unverknüpften bewohnten Wohnungen, für die entweder keine Wohnungsnutzernamen angegeben wurden oder keine zu der Angabe der Wohnungsnutzer passende Person an der Anschrift gefunden werden konnte, stand noch für eine Zusammenführung mit den ggf. noch an dieser Anschrift vorhandenen und noch nicht mit einer Wohnung verknüpften Haushalten zur Verfügung. Es konnte also noch bewohnte Wohnungen, die noch mit keinem Haushalt verknüpft waren, und/oder Haushalte, die noch mit keiner Wohnung verknüpft waren, geben. In Schritt 5 der Haushaltegenerierung erfolgte zum einen eine Zusammenführung der noch zur Verfügung stehenden Restmengen an unverknüpften Wohnungen und Haushalten. Waren danach immer noch unverknüpfte Haushalte vorhanden, so erfolgte über statistische Kriterien eine Zusammenführung der noch zur Verfügung stehenden Restmengen an unverknüpften mit bereits verknüpften Haushalten; dabei wurden über Annahmen zu sinnvollen Belegungsdichten an einer Anschrift noch zu verteilende Personen so auf die an der Anschrift bereits generierten Haushalte verteilt, dass keine unplausiblen Personenzahlen, gemessen an der verfügbaren Wohnungsgröße, resultierten. Ferner wurde versucht, für die Gemeinde bzw. den Kreis insgesamt möglichst gut aus der Haushaltsstichprobe geschätzte Eckzahlen zur Anzahl bestimmter Haushaltstypen zu treffen.

Neben den bei Schritt 3 bereits erwähnten 16 % aller Haushalte, die bereits fertig generiert waren, für die lediglich noch eine Wohnung benötigt wurde, erfolgte in diesem Schritt die Verarbeitung der letzten 3 % aller Haushalte.



### 4.6 Schritt 6: Typisierung der generierten Haushalte

Sämtliche von der Haushaltgenerierung generierten Haushalte und Familien sowie die sich darin befindlichen Personen erhielten drei verschiedene Typisierungen für folgende Merkmale:

- Typ des Haushalts
- Größe des Haushalts
- Stellung einer Person im Haushalt
- Typ der Familie
- Größe der Familie
- Stellung einer Person in der Familie

Es handelte sich dabei um eine EU-Typisierung<sup>50</sup>, in der nur Hauptwohnsitzpersonen berücksichtigt werden, eine sogenannte nationale EU-Typisierung, die der EU-Typisierung entspricht, zusätzlich jedoch auch Nebenwohnsitzpersonen berücksichtigt sowie um eine nationale Typisierung, die für Haupt- und Nebenwohnsitzpersonen erfolgt und am Bezugspersonenkonzept des Mikrozensus angelehnt ist.

---

<sup>50</sup> EU-Verordnung 1201/2009.

## 5 Ergebniserstellung

Die Ergebnisse des Zensus 2011 wurden in zwei Etappen veröffentlicht. Zum Erstveröffentlichungstermin am 31. Mai 2013 wurden neben den Hauptergebnissen, den Einwohnerzahlen, vorläufige soziodemografische Personenergebnisse und die Ergebnisse der Gebäude- und Wohnungszählung veröffentlicht.

Zum Zweitveröffentlichungstermin im Frühjahr 2014 gab es neben Anpassungen der bereits veröffentlichten Ergebnisse weitere soziodemografische Personenergebnisse sowie Ergebnisse zu Haushalten und Familien. Am 10. April 2014 stellten die Statistischen Ämter des Bundes und der Länder die endgültigen demografischen Ergebnisse bereit, am 28. Mai 2014 erfolgte dann die Veröffentlichung der endgültigen Daten zu Bildung, Migration, Erwerbstätigkeit sowie zu Haushalten, Familien, Gebäuden und Wohnungen. Diese Ergebnisse des Zensus 2011 können kostenfrei unter [www.ergebnisse.zensus2011.de](http://www.ergebnisse.zensus2011.de) abgerufen werden.

Für die korrekte Ergebnisinterpretation sind zusätzliche Kenntnisse über den Prozess der Ergebniserstellung sinnvoll. Dies umfasst Informationen zu der der Auswertung zugrunde liegenden Datenbasis wie auch Informationen zum Verfahren der Ergebniserstellung.

Im Folgenden soll zunächst die Ermittlung der Einwohnerzahlen zusammengefasst dargestellt werden, um anschließend auf die Besonderheiten bei der Ergebniserstellung der endgültigen Zensuszahlen sowie auf die Unterschiede zwischen den Auswertungen des Erst- und Zweitveröffentlichungstermins einzugehen. Abschließend folgt ein kurzer Exkurs, welche Maßnahmen zum Thema Geheimhaltung im Zensus 2011 Anwendung fanden.

### 5.1 Ermittlung der Einwohnerzahlen

Ausgangsbasis der Ermittlung der Einwohnerzahlen waren die Personendatensätze, die von den Meldebehörden nach § 3 ZensG 2011 aus den Melderegistern übermittelt wurden. Die Datensätze sind eindeutig den Kommunen zugeordnet und enthalten die Information, ob ein Datensatz relevant für die Einwohnerzahl der Kommune ist (siehe Kapitel 3.3.1).

Die Ermittlung der Einwohnerzahlen erfolgte jedoch nicht durch eine einfache Auszählung aus den Melderegistern, stattdessen waren die bereits ausführlich beschriebenen Prozessschritte notwendig, um eine korrekte Einwohnerzahl zu ermitteln. Da die Ermittlung der Einwohnerzahl abhängig von der Gemeindegröße erfolgte, ergaben sich dabei zwei unterschiedliche Verfahren:

Die Feststellung der Einwohnerzahl in den „großen“ Gemeinden mit mindestens 10 000 Einwohnern erfolgte durch

- Auszählung von Personen an Nicht-Sonderanschriften aus den Melderegistern,
- eine Vollerhebung aller an Anschriften mit Sonderbereichen wohnenden Personen,
- maschinelle Bereinigung der Melderegisterbestände um unzulässige Mehrfachfälle (für den Bereich der Nicht-Sonderanschriften) sowie
- statistische Korrektur dieses plausibilisierten Melderegisterbestandes um aus der Haushaltsstichprobe hochgerechnete Karteteileichen und Fehlbestände.

In den „kleinen“ Gemeinden mit weniger als 10 000 Einwohnern erfolgte die Feststellung der Einwohnerzahl durch

- Auszählung von Personen an Nicht-Sonderanschriften aus den Melderegistern,
- eine Vollerhebung aller an Anschriften mit Sonderbereichen wohnenden Personen,
- primärstatistische Bereinigung der Melderegisterbestände um unzulässige Mehrfachfälle (an Nicht-Sonderanschriften) sowie
- die Befragung zur Klärung von Unstimmigkeiten an Nicht-Sonderanschriften und Bereinigung um die an diesen Anschriften festgestellten Über- und Untererfassungen der Melderegister.

## 5.2 Auswertungen – Gesamtheiten und Teilgesamtheiten

Bedingt durch die Kombination der verschiedenen Datenquellen des Zensusmodells wird nicht immer bei allen Auswertungen die volle Gesamtheit aller Personen, Haushalte, Wohnungen usw. abgebildet. Je nach Erhebung, aber auch in Abhängigkeit von der Kreuzkombination von Merkmalen in der konkreten Auswertungstabelle, gibt es Einschränkungen, die bei der Interpretation der Zahlen beachtet werden müssen und die im Folgenden dargestellt werden.

In die im Frühjahr 2014 publizierten demografischen Zensusergebnisse und in die Ergebnisse zur Zugehörigkeit zu einer öffentlich-rechtlichen Religionsgesellschaft („Registermerkmale“) wurden alle mit Haupt- oder alleinigem Wohnsitz gemeldeten Personen einbezogen. Für die korrekte Interpretation dieser Ergebnisse muss zusätzlich beachtet werden, dass beim Ergebnis für Deutschland auch die Deutschen im Ausland (Personenkreis gemäß § 3 Abs. 3 ZensG 2011) enthalten sind. Bei den beiden Merkmalen Religionszugehörigkeit und Familienstand sind diese allerdings in der Kategorie „ohne Angabe“ enthalten.

Die Ergebnisse zu den Themenbereichen „Schul- und Berufsbildung“, „Beruf“ und „Migration“ (= Zusatzmerkmale, die nicht aus Registern gewonnen werden konnten) basieren ausschließlich auf den durch die Haushaltebefragung auf Stichprobenbasis gewonnenen Informationen, die auch in nicht-sensiblen Wohnheimen und Gemeinschaftsunterkünften, wie beispielsweise in Studentenwohnheimen, nicht jedoch an sensiblen Gemeinschaftsunterkünften (z. B. Behindertenwohnheime oder Notunterkünfte für Wohnungslose) erhoben wurden. Auch die „Deutschen im Ausland“ sind hier in den Ergebnissen nicht enthalten. Wegen dieser Einschränkung sowie auf Grund der Tatsache, dass bei der Hochrechnung aus der Haushaltsstichprobe bei Kreuzkombination mit demografischen Merkmalen die Antwortausprägung der Erhebung und nicht die Registerausprägung der demografischen Merkmale zugrunde gelegt wird, können – je ausgewerteter regionaler Einheit – Differenzen zwischen der Randsumme der nachgewiesenen Einzelpositionen und der ermittelten Gesamteinwohnerzahl sowie Abweichungen in demografischen Zwischensummen für diese regionale Einheit bestehen.

Bestimmte erwerbsstatistische Ergebnisse sind aus einem Auszählungsteil aus Registern und einem Hochrechnungsteil aus der Haushaltsstichprobe zusammengesetzt. Für die drei Merkmale „Erwerbsstatus“, „Stellung im Beruf“ und „Wirtschaftszweig“ werden die sozialversicherungspflichtig Beschäftigten (ausgenommen geringfügig Beschäftigte) der Bundesagentur für Arbeit sowie Beamtinnen und Beamte, Richterinnen und Richter und Soldatinnen und Soldaten der öffentlichen Arbeitgeber gezählt und die nicht enthaltene Restmenge der Erwerbstätigen wie die Selbstständigen und die mithelfenden Familienangehörigen aus der Haushaltsstichprobe hochgerechnet. Bei einer Kombination der Merkmale „Erwerbsstatus“, „Stellung im Beruf“ und „Wirtschaftszweig“ mit anderen Zusatzmerkmalen der Haushaltsstichprobe basieren aus methodischen Gründen alle Ergebnisse auf der Haushaltsstichprobe, während die Daten der erwerbsstatistischen Register für diese Ergebnisauswertungen unberücksichtigt bleiben mussten. Daraus resultierende Unterschiede zwischen register- und rein stichprobenbasierten Ergebnissen sind methodenbedingt. Während die Haushaltsstichprobe des Zensus eine Selbsteinordnung durch die Befragten darstellt, handelt es sich bei den erwerbsstatistischen Registern um Meldungen durch die Arbeitgeberinnen und Arbeitgeber.

Die nachstehende Übersicht zeigt, bei welchen personenbezogenen Auswertungen nur Teilpopulationen berücksichtigt wurden, wobei Personen an sensiblen Sonderbereichen sowie „Deutsche im Ausland“ grundsätzlich nicht enthalten sind:

Teilbevölkerungsgruppe	Auswertungen nach Teilbevölkerungsgruppen		
	Personen mit Migrationshintergrund	Erwerbstätige wie auch Erwerbslose (Mindestalter 15 Jahre)	Erwerbstätige (Mindestalter 15 Jahre)
Themenbereich	Migrationshintergrund nach Zuzugsjahr (aggregiert)	Stellung im Beruf (aus Stichprobe)	Arbeitsort nach Auspendlern
	Migrationshintergrund nach Zuzugsstaat	Wirtschaftszweig	Arbeitsort nach Einpendlern
	Migrationshintergrund nach Aufenthalt in Jahren		

In der Auswertungsdatenbank haben alle Tabellen einen „Untertitel“ erhalten, damit die Nutzerinnen und Nutzer direkt dem Tabellenkopf entnehmen können, auf welcher Datengrundlage und Berechnung die Daten erstellt wurden:

- "Auszählung aus dem bereinigten Registerbestand", wenn die Ergebniserstellung ausgezählt wird
- "Hochrechnung aus der Haushaltsstichprobe", wenn aus der Haushaltsstichprobe hochgerechnet wird
- "Kombinierte Registerauszählung und Hochrechnung aus der Haushaltsstichprobe", wenn nach dem sogenannten Baukastenprinzip ausgewertet wird

Auch bei den haushaltsstatistischen Zensusergebnissen finden nicht immer alle vorhandenen Datensätze Eingang in die Ergebnisdarstellung. Dies ist davon abhängig, was gezeigt werden soll. So wurden beispielsweise für die Typisierung der Haushalte alle Personen – Personen mit Haupt- oder alleinigem Wohnsitz sowie Personen mit Nebenwohnsitz – berücksichtigt. In der Ergebnisdarstellung zu „Anzahl von Personen in einem Haushaltstyp“ werden aber der Auswertungslogik folgend, nur Personen, die mit Haupt- oder alleinigem Wohnsitz gemeldet sind, dargestellt. Die folgende Übersicht zeigt, welche Teilgesamtheiten bei Auswertungen zu Haushalten und Familien, zur Wohnsituation von Haushalten und/oder Familien, zur Haushaltszugehörigkeit von Personen sowie zur Wohnsituation von Personen Berücksichtigung finden. Basis sind zunächst immer alle aktiven Personen-<sup>51</sup>, Haushalts-, Familien-, Wohnungs- oder Gebäudedatensätze, bei denen je nach Auswertungswunsch Teilgesamtheiten gegebenenfalls ausgeschlossen werden:

<sup>51</sup> Es wurden nur die Personen gezählt, die für die Ermittlung der Einwohnerzahlen relevant waren, also Personen, die mit Haupt- oder alleinigem Wohnsitz gemeldet waren.

	In Auswertungen	
	eingeschlossen	ausgeschlossen
Reine Haushalts- oder Familientabellen	<p>Haushalte/Familien in vermieteten oder vom Eigentümer bewohnten Wohnungen</p> <p>Haushalte/Familien in Diplomaten-/Streitkräftewohnungen</p> <p>Haushalte/Familien in Ferienwohnungen</p> <p>Haushalte/Familien in nicht-sensiblen Sonderbereichen mit eigener Haushaltsführung und mit Wohnung verknüpft</p>	<p>Haushalte/Familien ohne Wohnung</p> <p>Personen in sensiblen Sonderbereichen</p> <p>Personen ohne eigene Haushaltsführung in nicht-sensiblen Sonderbereichen</p>
Tabellen zur Wohnsituation von Haushalten und/oder Familien	<p>Haushalte/Familien in vermieteten oder vom Eigentümer bewohnten Wohnungen</p> <p>Haushalte/Familien in Ferien-/Freizeitwohnungen</p> <p>Haushalte/Familien in nicht-sensiblen Sonderbereichen (sofern Privathaushalt und mit Wohnung verknüpft)</p>	<p>Sensible Sonderbereiche und nicht-sensible Sonderbereiche (sofern keine eigene Haushaltsführung erfolgt)</p> <p>Haushalte/Familien in Diplomaten-/Streitkräftewohnungen</p> <p>Haushalte/Familien ohne Wohnung</p>
Personentabellen mit Haushalts-/Familienzugehörigkeit	<p>Personen in vermieteten oder vom Eigentümer bewohnten Wohnungen</p> <p>Personen in Diplomaten-/Streitkräftewohnungen</p> <p>Personen in nicht-sensiblen Sonderbereichen mit eigener Haushaltsführung</p>	<p>Personen in Ferien-/Freizeitwohnungen</p> <p>Personen ohne Haushaltszugehörigkeit</p> <p>Personen in sensiblen Sonderbereichen</p> <p>Personen an nicht-sensiblen Sonderbereichen (sofern keine eigene Haushaltsführung erfolgt)</p> <p>Personen in Haushalten/Familien ohne Wohnung</p>
Tabellen zur Wohnsituation von Personen	<p>Personen in vermieteten oder vom Eigentümer bewohnten Wohnungen</p> <p>Personen in nicht-sensiblen Sonderbereichen mit eigener Haushaltsführung und mit Wohnung verknüpft</p>	<p>Personen in Diplomaten-/Streitkräftewohnungen</p> <p>Personen in Ferien-/Freizeitwohnungen</p> <p>Personen ohne Haushaltszugehörigkeit</p> <p>Personen ohne Wohnung</p> <p>Personen in sensiblen Sonderbereichen und nicht-sensiblen Sonderbereichen (sofern keine eigene Haushaltsführung erfolgt)</p>

Auswertungen zu Gebäuden- und Wohnungen enthalten i. d. R. vermietete Wohnungen, vom Eigentümer bewohnte Wohnungen, Ferien- und Freizeitwohnungen und leerstehende Wohnungen. Diplomaten-/Streitkräftewohnungen sind im Datenmaterial enthalten und können bei Bedarf mit ausgewertet werden, dies erfolgt beispielsweise in den Gemeindeblättern in der „Insgesamt“-Kategorie. In den im Mai 2013 veröffentlichten Ergebnissen waren diese Wohnungen noch aus der Auswertung ausgeschlossen. Gewerblich genutzte Wohnungen sind zwar im Datenmaterial enthalten, werden aber grundsätzlich nicht mit ausgewertet, da es sich hierbei nicht um Erhebungseinheiten lt. ZensG 2011 handelt.

Sobald man einzelne Wohnungen – z. B. Diplomaten-/Streitkräftewohnungen – aus den Auswertungen ausschließt, kann es zu Abweichungen zwischen der Zahl der Gebäude mit genau einer Wohnung und den darin befindlichen Wohnungen kommen.

Neben den zu beachtenden Besonderheiten bei Auswertungen aus dem finalen Datenbestand des Zensus 2011, fallen auch beim Vergleich der Daten des Erst- und Zweitveröffentlichungstermins die nachfolgenden Unterschiede auf.

### 5.2.1 Abweichende Ergebnisse bei personenbezogenen Auswertungen

Die Hauptursache für Abweichungen bei den demografischen Ergebnissen ergibt sich aus dem unterschiedlichen Berechnungsverfahren: Hochrechnung aus der Haushaltsstichprobe zum Erstveröffentlichungstermin am 31. Mai 2013 (VÖT1) einerseits und Auszählung aus statistisch korrigierten Registern zum zweiten Veröffentlichungstermin am 10. April 2014 (VÖT2) andererseits.

Als weitere Ursachen für Abweichungen zwischen vergleichbaren Zensusergebnissen sind zu berücksichtigen:

1. Zum VÖT1 konnte die Verteilung der Karteileichen und Fehlbestände nach demografischen Merkmalen nur in grober Form in das integrierte Korrekturverfahren eingehen. Aus der Haushaltsstichprobe konnten mit hinreichender Genauigkeit nur ein- oder zweidimensionale Verteilungen – zum Beispiel nach fünf zusammengefassten Altersklassen, nach Geschlecht und Staatsangehörigkeit (deutsch/ nicht-deutsch) – als Anpassungsbasis für das integrierte Korrekturverfahren geschätzt werden. Für die Schätzung der Karteileichen und Fehlbestände in tieferer demografischer Differenzierung wurden statistische Verfahren (loglineare Modelle) verwendet, die eine den Datenstrukturen entsprechende Verteilung der Über- und Untererfassungen näherungsweise ermöglichten, aufgrund des unterschiedlichen Berechnungsverfahrens aber zu anderen Ergebnissen kommen als die Auszählung der korrigierten Melderegister zum VÖT2.
2. Für die Hochrechnung der demografischen Merkmale aus der Haushaltsstichprobe wurden die Angaben der Auskunftspflichtigen im Fragebogen genutzt, für die Auszählung der statistisch bereinigten Melderegister dagegen die Angaben im Melderegister. Da Befragungsergebnisse aus der Haushaltsstichprobe vereinzelt von den Melderegistereinträgen abweichen können, kommt es so zu Ergebnisdifferenzen.
3. Die Fehlerrechnung zur Hochrechnung der Haushaltsstichprobe sperrt bei Bedarf andere Tabellenfelder als die Prüfung der Differenzen der durch das Geheimhaltungsverfahren veränderten Werte zu den Originalwerten.

### 5.2.2 Unterschiede bei den Ergebnissen für Gebäude und Wohnungen

Nach der Veröffentlichung der ersten Ergebnisse des Zensus 2011 am 31. Mai 2013 fanden diejenigen Aufbereitungsschritte statt, die die Bildung von Haushalten zum Ziel hatten (vgl. Kapitel 4). Beim Vergleich von GWZ- und Melderegisterangaben konnte es zu widersprüchlichen Konstellationen kommen. Je nach Fallkonstellation konnte es an einer Anschrift zu einer Verringerung oder auch zu einer Erhöhung der Wohnungszahl kommen. Insgesamt hat es im Saldo eine Verringerung der Wohnungszahl von den im Mai 2013 zu dem im Mai 2014 veröffentlichten Ergebnissen gegeben. Anhand der folgenden Beispiele wird verdeutlicht, warum die Ergebnisse des Zensus 2011 nach der Haushaltgenerierung Abweichungen hinsichtlich der Anzahl der Wohnungen, der Gebäude und der bewohnten Wohnungen von den im Mai 2013 veröffentlichten Ergebnissen aufweisen.

Fallkonstellationen, die zu einer Verringerung der Wohnungszahl führten:

- Beispiel 1: An einer Anschrift konnte kein Eigentümer beziehungsweise keine Eigentümerin für das dort vorhandene Gebäude ermittelt werden. Daher konnten zu diesem Gebäude auch keine Daten primärstatistisch erhoben werden. Die fehlenden Angaben für diese Anschrift wurden mittels statistischer Verfahren durch die Duplizierung eines Gebäudes im Bestand (mit vollständigen und plausiblen Angaben im Fragebogen) gewonnen, beispielsweise eines Gebäudes mit sechs Wohnungen (Imputation). An der Anschrift konnten aber aus den dort durch den Zensus ermittelten Personen nur fünf Haushalte gebildet werden. Daher konnte einer der bewohnten Wohnungen kein Haushalt zugeordnet werden. Da die Angabe zur Zahl der Wohnungen hier unsicher war (imputiertes Gebäude), wurde diese Wohnung im Datenbestand gelöscht.
- Beispiel 2: An einer Anschrift waren in der GWZ fünf bewohnte Wohnungen erhoben worden. Diese Anschrift war auch in der Haushaltsstichprobe enthalten. Aus den in dieser Befragung ermittelten Angaben wurden vier Haushalte gebildet. Einer der fünf bewohnten Wohnungen konnte daher kein Haushalt zugeordnet werden. In diesem Fall wurde der Information aus der Haushaltsstichprobe Vorrang gegenüber der Information aus der GWZ eingeräumt. Die Wohnung ohne Haushalt wurde daher im Datenbestand gelöscht.
- Beispiel 3: Es gab auch imputierte Gebäude, in denen gar keine Haushalte gebildet werden konnten, weil unter dieser Anschrift keine Personen gemeldet waren bzw. durch den Zensus ermittelt werden konnten. Auch in diesen Gebäuden wurden wie in Beispiel 1 die imputierten bewohnten Wohnungen, denen kein Haushalt zugeordnet werden konnte, gelöscht. Führt dies dazu, dass in einem Gebäude keine Wohnung mehr vorhanden war, wurde das gesamte Gebäude gelöscht.
- Beispiel 4: Vom gleichen Haushalt bewohnte Wohnungen wurden zusammengelegt. Eine Zusammenlegung von Wohnungen erfolgte, wenn in der Haushaltegenerierung entdeckt wurde, dass im GWZ-Fragebogen ein und derselbe Bewohnername mehreren unterschiedlichen Wohnungen an der gleichen Anschrift zugeordnet war. Gleiches galt für laut Melderegister miteinander verzeigte Personen (als Ehegatten oder minderjährige Kinder). Bei Stichprobenanschriften erfolgte eine Haushaltszuordnung über die Feststellung aus der Haushaltsstichprobe.

Fallkonstellationen, die zu einer Erhöhung der Wohnungszahl führten:

- Beispiel 5: An einer Anschrift konnte kein Eigentümer bzw. keine Eigentümerin für das dort vorhandene Gebäude ermittelt werden. Daher konnten zu diesem Gebäude auch keine Angaben über den Fragebogen erhoben werden. Mit Hilfe statistischer Verfahren wurde deshalb wie in Beispiel 1 an dieser Anschrift ein Gebäude dupliziert (Imputation). Das eingesetzte Gebäude enthielt sechs bewohnte Wohnungen. An der Anschrift konnten aus den dort durch den Zensus ermittelten Personen sieben Haushalte gebildet werden. Einer der Haushalte konnte daher zunächst keiner bewohnten Wohnung zugeordnet werden. Da es sich um ein imputiertes Gebäude handelte, wurde der Information der gemeldeten Personen Vorrang eingeräumt. Deshalb wurde mit Hilfe statistischer Verfahren eine zusätzliche bewohnte Wohnung hinzugefügt.
- Beispiel 6: An einer Anschrift waren in der GWZ fünf bewohnte Wohnungen erhoben worden. Diese Anschrift war auch in der Haushaltsstichprobe enthalten. Aus den in dieser Befragung ermittelten Angaben wurden sechs Haushalte gebildet. Daher stand für einen der Haushalte keine bewohnte Wohnung mehr zur Verfügung. In diesem Fall wurde der Information aus der Haushaltsstichprobe Vorrang gegenüber der Information aus der GWZ eingeräumt. Aus diesem Grund musste mit statistischen Verfahren eine zusätzliche bewohnte Wohnung eingesetzt werden.

Darüber hinaus gibt es auch in den im Mai 2014 veröffentlichten Ergebnissen als bewohnt geltende Wohnungen, die trotzdem mit keinem Haushalt verknüpft werden konnten. Bei Auswertungen zum Wohnungsbestand werden diese Wohnungen mit ausgewiesen, bei Auswertungen zu den Wohnverhältnissen von Haushalten allerdings nicht, weil es zu diesen Wohnungen keinen Haushalt gibt. Diese Fallkonstellation kommt nur an Anschriften vor, die nicht in der Haushaltsstichprobe enthalten waren und ist folgendermaßen begründet:



- Beispiel 7: An einer Anschrift waren in der GWZ sechs bewohnte Wohnungen erhoben worden. Eine der sechs Mietparteien im zugehörigen Gebäude hatte zum Stichtag bereits eine neue Wohnung bezogen und für einen kurzen Übergangszeitraum Mietverträge für zwei Wohnungen. Die Eigentümerin des Gebäudes übermittelte diese Wohnung deshalb korrekt als „vermietet“. Aufgrund des Umzugs konnten aber an der Anschrift aus den dort noch gemeldeten Personen nur fünf Haushalte generiert werden. Für die verbleibende bewohnte Wohnung stand deshalb kein Haushalt mehr zur Verfügung. In diesem Fall wurden die Informationen der gemeldeten Personen und die Information aus der GWZ gleichrangig behandelt. Die verbleibende bewohnte Wohnung wurde auf „bewohnt, aber ohne aktuell geführten Haushalt“ gesetzt.

In der Summe ergibt sich durch die in den Beispielen 1 bis 6 beschriebenen Veränderungen eine bundesweite Differenz zum im Mai 2013 veröffentlichten Wohnungsbestand von – 750 000 Wohnungen, die sich wie folgt zusammensetzt:

- 760 000 Wohnungen durch Löschungen von Wohnungen, ggf. ganzen Gebäuden
- 200 000 Wohnungen durch Zusammenlegungen von Wohnungen
- + 230 000 Wohnungen durch Imputationen von Wohnungen, ggf. ganzen Gebäuden und
- 20 000 Wohnungen durch bestandsrelevante Umwidmungen von Wohnungen.

Darüber hinaus gibt es rund eine Million laut GWZ bewohnte Wohnungen, denen kein Haushalt zugeordnet werden konnte (vgl. Beispiel 7).

### 5.3 Geheimhaltung

Die Einzeldaten des Zensus 2011 unterliegen dem Statistikgeheimnis nach § 16 Bundesstatistikgesetz. Es muss gewährleistet sein, dass aus den Veröffentlichungstabellen keine Rückschlüsse auf einzelne Personen möglich sind.

Die statistische Geheimhaltung im Zensus 2011 verhindert zum einen, dass aus den veröffentlichten Daten erkennbar wird, dass manche statistische Einheiten (Personen, Gebäude, Wohnungen, Familien, Haushalte) eine „einzigartige“ Kombination der Zensus-Merkmale aufweisen und zum anderen, dass aus den veröffentlichten Daten für bestimmte statistische Einheiten Rückschlüsse auf eine bei ihnen vorliegende Merkmalsausprägung möglich sind. Dies wird auf unterschiedliche Weise sichergestellt, und zwar abhängig von der Art der Datenerhebung. Dabei wird nach folgenden Kategorien unterschieden:

1. Daten aus Erhebungsteilen mit Vollerhebungen (alle Auswertungen mit ausschließlich demografischen Merkmalen, Gebäude- und Wohnungszahlen sowie Haushalts- und Familieninformationen) und
2. Daten aus der Haushaltsstichprobe (Auswertungen wie beispielsweise Bildung, Erwerbstätigkeit u. a.), die hochgerechnet werden.

#### 5.3.1 Daten aus Vollerhebungen oder Registerauszählungen

Tabellen, die durch reine Auszählung des Datenmaterials erstellt werden, werden durch das Verfahren SAFE<sup>52</sup> („Sichere Anonymisierung für Einzeldaten“) geheim gehalten. Es handelt sich dabei um ein datenveränderndes Verfahren. Bei SAFE wird ein Rückschluss auf individuelle statistische Einheiten (Personen, Gebäude, Wohnungen, Familien, Haushalte) verhindert, indem die Einzeldaten leicht angepasst werden. Dabei verändert SAFE die Daten so, dass jede in den Originaldaten existierende Merkmalskombination (z. B. aus Alter, Geschlecht, Familienstand, Angaben zur Erwerbstätigkeit) im

52 Höhne, J. (2011) 'SAFE – A method for anonymising the German Census', paper presented at the Joint UNECE/Eurostat work session on statistical data confidentiality, Tarragona, Spain, 26-28 October 2011, available at [http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2011/16\\_Germany.pdf](http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2011/16_Germany.pdf).

Eine Kurzbeschreibung zum SAFE-Verfahren sowie Kennzahlen zu den Abweichungen können auf der Internetseite [www.zensus2011.de](http://www.zensus2011.de) abgerufen werden.

geheim gehaltenen Datenbestand mindestens dreimal oder gar nicht mehr auftritt. Beim Zensus 2011 erfolgt dies auf der räumlichen Ebene der Gemeinde bzw. im Falle der Stadtstaaten Berlin und Hamburg auf Ebene der Stadtteile.

Diese Änderungen werden kontrolliert vorgenommen, sodass sie sich weitgehend untereinander ausgleichen. Dadurch wird erreicht, dass Abweichungen in zentralen Auswertungstabellen (dazu zählen Tabellen, die über die Zensusdatenbank online verfügbar sind, sowie das Tabellenprogramm, das an das Statistikamt der Europäischen Union (Eurostat) zu liefern war) minimiert und alle wichtigen statistischen Ergebnisse verlässlich abgebildet werden. Diese Kontrolle der Abweichungen zwischen den aus den Original-Merkmalsausprägungen und den aus den mit SAFE veränderten Merkmalsausprägungen berechneten Ergebnissen erfolgt beim Zensus 2011 ebenfalls auf der räumlichen Ebene der Gemeinde. Untergemeindliche Zensusergebnisse genügen zwar auch den Anforderungen an die statistische Geheimhaltung; sie können jedoch höhere (nicht-kontrollierte) Abweichungen zu den Originalergebnissen vor Geheimhaltung haben.

Durch die Datenveränderung erzeugt man bei sehr kleinen Fallzahlen zwar eine gewisse Verfälschung, gleichzeitig bleibt aber die Qualität und statistische Aussagefähigkeit der Daten weitestgehend erhalten. Dies ist darin begründet, dass die Daten immer nur so verändert werden, dass wichtige statistische Informationen und Verteilungen nur unwesentlich verändert werden.

Folgende, im Rahmen des Zensus erhobene, Daten werden dem Geheimhaltungsverfahren SAFE unterzogen, sofern sie für Auswertungen vorgesehen sind:

- Daten aus der Vollerhebung der Gebäude- und Wohnungszählung
- Daten zu Personen aus den korrigierten Registern (inkl. der demografischen Daten zu Personen aus der Vollerhebung an Wohnheimen und Gemeinschaftsunterkünften)
- Daten zu Haushalten und Familien aus der Haushaltgenerierung

Aus technischen Gründen wurden die Auswertungsmerkmale zu Personen (also aus den Registern und aus der Vollerhebung an Wohnheimen und Gemeinschaftsunterkünften) und die Auswertungsmerkmale zu Haushalten, Familien, Wohnungen und Gebäuden zunächst in zwei unterschiedliche Datenbestände – sogenannte „Merkmalsscheiben“ – geteilt und anschließend getrennt geheim gehalten. Die vorgenommenen Änderungen sowie deren Kontrolle fanden dementsprechend auch nur jeweils innerhalb der jeweiligen Merkmalsscheibe statt. Eine kreuzkombinierte Auswertung von Merkmalen aus unterschiedlichen Merkmalsscheiben bleibt dennoch möglich: Im nicht-öffentlichen Teil des Auswertungssystems wird durch eine Ähnlichkeitssuche jedem Datensatz nach SAFE eindeutig ein Originaldatensatz vor SAFE mit seinen Identifikatoren (z. B. Anschriften-ID, Haushalts-Nr.) zugeordnet. Über die zugeordneten Identifikatoren werden die Merkmalsscheiben miteinander verknüpft. Da es hierbei jedoch zu höheren Abweichungen als bei „scheibeninternen“ Auswertungen kommen kann, werden die Ergebnisse vor Veröffentlichung auf ihre Abweichung untersucht und bei zu hohen Abweichungen ggf. Angaben nur geklammert dargestellt bzw. von der Darstellung ausgenommen und stattdessen ein Punkt („.“) dargestellt.

Im Unterschied zu dem oben beschriebenen Vorgehen werden die Gesamteinwohnerzahlen der Kommunen immer aus dem Originaldatensatz ermittelt, SAFE findet auf diese Ergebnisse keine Anwendung.

Auch die Berechnung der von den statistischen Ämtern veröffentlichten Verhältniszahlen erfolgt auf den Originalwerten (Original-Zähler und Original-Nenner), da es anderenfalls – bei Verwendung des mit SAFE geheim gehaltenen Zählers und Nenners – durch die SAFE-Abweichungen ggf. zu unverhältnismäßig hohen Verzerrungen der Quotienten kommen kann (insbesondere bei kleinen Fallzahlen im Zähler und/oder im Nenner). Durch geeignete Vergrößerung des Divisionsergebnisses oder – in Einzelfällen – auch durch Abweichen von der kaufmännischen Rundung wird die statistische Geheimhaltung auch bei den aus Originalwerten berechneten Verhältniszahlen gewährleistet. Verhältniszahlen, die Nutzer in den Tabellen des öffentlichen Auswertungssystems selbst erstellen können, werden dagegen durch die Division mit durch SAFE geheim gehaltenen Werten berechnet.

### 5.3.2 Daten der Haushaltsstichprobe

Bei Stichprobenergebnissen ist die statistische Geheimhaltung aufgrund der Tatsache, dass keine vollständig erhobenen Werte ausgezählt werden, sondern eine Hochrechnung aus einer Zufallsstichprobe erfolgt, sichergestellt. Um zu verdeutlichen, dass es sich bei den hochgerechneten Ergebnissen lediglich um Schätzwerte und nicht um exakt ermittelte Häufigkeiten handelt, werden die Werte auf ein Vielfaches von 10 gerundet dargestellt.

Hierbei werden einzelne Ergebnisse gesperrt, wenn sie als statistisch unzuverlässig bewertet werden. Diese Bewertung stützt sich zum einen auf die Zahl der befragten Personen, die zum hochgerechneten Ergebnis beigetragen haben, und zum anderen auf die (näherungsweise bestimmte) Wahrscheinlichkeit, mit der sie für die Stichprobe ausgewählt wurden.

In Anlehnung an die Vorgehensweise beim Mikrozensus werden auch beim Zensus die Ergebnisse der Stichprobenerhebung gesperrt, wenn sie einen zu großen Stichprobenzufallsfehler erwarten lassen. Beim Zensus werden Ergebnisse mit einem relativen Standardfehler von mehr als 15 % gesperrt und in den Tabellen mit folgendem Symbol „/“ gekennzeichnet. Sie gelten dann als „unzuverlässig“ bzw. „unsicher“. <sup>53</sup> Der Standardfehler ist dabei wesentlich vom Stichprobenumfang, daneben aber auch vom Auswahlatz der Stichprobe abhängig. Je größer die Anzahl der Personen, die zum Tabellenfeld beitragen und je größer der Auswahlatz, desto geringer ist der relative Standardfehler. Grundsätzlich lässt sich der Standardfehler mit folgender Formel errechnen:

$$\text{relativer Standardfehler} = \sqrt{\frac{1 - \text{Auswahlatz}}{\text{Anzahl der Stichprobenpersonen, die zum Tabellenfeldwert beitragen}}}$$

Aus den bekannten Größen „relativer Standardfehler“ und „Auswahlatz“ lässt sich je Tabellenfeld die erforderliche Besetzungszahl errechnen. Im Zuge des Abgleichs der tatsächlichen und der erforderlichen Besetzungszahl je Tabellenfeld erfolgt dann ggf. die Sperrung einzelner Felder.

Für die Ergebniserstellung zu den Themenbereichen „Schul- und Berufsbildung“, „Beruf“ und „Migration“ ergibt sich somit folgendes Fallzahlkriterium:

$$\text{Fallzahl} < \frac{1 - \text{Auswahlatz}}{(0.15)^2}$$

<sup>53</sup> Eine Klammerung von Werten mit einem Standardfehler zwischen 10 % und 15 % wie beim Mikrozensus gibt es beim Zensus nicht.

Bei Ergebnissen, die sich aus einem ausgezählten und einem hochgerechneten Teilergebnis zusammensetzen, sind die Zahl der ausgezählten und die Zahl der hochgerechneten Personen entscheidend. Mithilfe dieser Informationen wird der relative Standardfehler des hochgerechneten Ergebnisses (als Maß seiner statistischen Zuverlässigkeit) approximiert. Übersteigt der geschätzte Fehler 15 %, wird das Ergebnis als „unzuverlässig“ gekennzeichnet. Bei Auswertungen zum Wirtschaftszweig, Erwerbsstatus oder der Stellung im Beruf für Gemeinden mit mindestens 10 000 Einwohnern oder bei höheren regionalen Ebenen können in das Ergebnis sowohl ein Auszählungs- als auch ein Hochrechnungsteil einfließen. Da der eine Teil aus dem Erwerbsregister ausgezählt wird und somit keinem relativen Standardfehler unterliegt, schafft dieser Ergebnisbestandteil mehr „Sicherheit“ im Gesamtergebnis. Je größer der Auszählungsanteil am Ergebnis eines Tabellenfelds, desto weniger Fälle der Haushaltebefragung auf Stichprobenbasis müssen vorgefunden werden, damit das Ergebnis als sicher gilt. Es erfolgt daher die nachstehende Erweiterung des bereits genannten Fallzahlkriteriums:

$$Fallzahl < \frac{(1 - \text{Auswahlsatz}) * SP_{\text{Anteil}}^2}{(0.15)^2}$$

Die Rundung auf ein Vielfaches von 10 hat die Vermeidung einer Scheingenauigkeit zum Ziel und wird daher für alle Zensusergebnisse mit einem Hochrechnungsanteil angewendet. Bezüglich der ausgewiesenen Randsummen gilt im Umkehrschluss, dass die Rundung der Randsummen auf der Summe der ungerundeten Einzelwerte basiert. Dadurch können gerundete Teilergebnisse aus der Ergebnistabelle von der gerundeten Gesamtsumme abweichen.

Bei hochgerechneten Ergebnissen ist durch die Kombination aus Zufallsfehler, Rundung und der Kennzeichnung von statistisch nicht belastbaren Werten die statistische Geheimhaltung sichergestellt.

## 6 Hintergrundinformationen

Auf der Internetseite [www.zensus2011.de](http://www.zensus2011.de) sind umfassendes Begleitmaterial und Veröffentlichungen zum und aus dem Zensus 2011 zu finden.

Fragebogen:

[https://www.zensus2011.de/DE/Infothek/Fragebogen/Fragebogen\\_node.html](https://www.zensus2011.de/DE/Infothek/Fragebogen/Fragebogen_node.html)

Merkmalskatalog zu den Auswertungsmerkmalen sowie Definitionen:

[https://www.zensus2011.de/DE/Infothek/Begleitmaterial\\_Ergebnisse/Begleitmaterial\\_node.html](https://www.zensus2011.de/DE/Infothek/Begleitmaterial_Ergebnisse/Begleitmaterial_node.html)

Zensusergebnisse und Glossar:

<https://ergebnisse.zensus2011.de/>



### Mitwirkende Personen

Autorinnen und Autoren sind die Fachverantwortlichen aus dem Bereich Zensus 2011 der Statistischen Ämter des Bundes und der Länder



**Statistisches Bundesamt**  
Gustav-Stresemann-Ring 11  
65189 Wiesbaden  
<http://www.destatis.de>  
Infoservice  
Telefon: 0611 75-2405  
Telefax: 0611 75-3330  
[www.destatis.de/kontakt](http://www.destatis.de/kontakt)

**Statistisches Bundesamt  
Zweigstelle Bonn**  
Graurheindorfer Straße 198  
53117 Bonn  
Telefon: 0611 75-1  
Telefax: 0611 75-8990/-8991  
[poststelle@destatis.de](mailto:poststelle@destatis.de)

**Statistisches Bundesamt  
i-Punkt Berlin**  
Friedrichstraße 50  
(Checkpoint Charlie)  
10117 Berlin  
Telefon: 0611 75-9434  
Telefax: 0611 75-9430  
[i-punkt@destatis.de](mailto:i-punkt@destatis.de)

## Statistische Ämter der Länder

**Statistisches Landesamt  
Baden-Württemberg**  
Böblinger Straße 68  
70199 Stuttgart  
Telefon: 0711 641-2866  
Telefax: 0711 641-2973  
[www.statistik-bw.de](http://www.statistik-bw.de)  
[vertrieb@stala.bwl.de](mailto:vertrieb@stala.bwl.de)

**Hessisches Statistisches Landesamt**  
Rheinstraße 35/37  
65185 Wiesbaden  
Telefon: 0611 3802-802  
Telefax: 0611 3802-890  
[www.statistik-hessen.de](http://www.statistik-hessen.de)  
[info@statistik.hessen.de](mailto:info@statistik.hessen.de)

**Statistisches Amt Saarland**  
Virchowstraße 7  
66119 Saarbrücken  
Telefon: 0681 501-5925  
Telefax: 0681 501-5915  
[www.statistik.saarland.de](http://www.statistik.saarland.de)  
[presse.statistik@lzd.saarland.de](mailto:presse.statistik@lzd.saarland.de)

**Bayerisches Landesamt  
für Statistik**  
Sankt-Martin-Straße 47  
81541 München  
Telefon: 089 2119-3205  
Telefax: 089 2119-3457  
[www.statistik.bayern.de](http://www.statistik.bayern.de)  
[vertrieb@statistik.bayern.de](mailto:vertrieb@statistik.bayern.de)

**Statistisches Amt  
Mecklenburg-Vorpommern**  
Lübecker Straße 287  
19059 Schwerin  
Telefon: 0385 588-56411  
Telefax: 0385 588-56708  
[www.statistik-mv.de](http://www.statistik-mv.de)  
[statistik.auskunft@statistik-mv.de](mailto:statistik.auskunft@statistik-mv.de)

**Statistisches Landesamt  
des Freistaates Sachsen**  
Macherstraße 63  
01917 Kamenz  
Telefon: 03578 33-1423  
Telefax: 03578 33-1598  
[www.statistik.sachsen.de](http://www.statistik.sachsen.de)  
[vertrieb@statistik.sachsen.de](mailto:vertrieb@statistik.sachsen.de)

**Amt für Statistik Berlin-Brandenburg**  
Behlertstraße 3a  
14467 Potsdam  
Telefon: 0331 8173-1777  
Telefax: 030 9028-4091  
[www.statistik-berlin-brandenburg.de](http://www.statistik-berlin-brandenburg.de)  
[info@statistik-bbb.de](mailto:info@statistik-bbb.de)

**Landesamt für Statistik  
Niedersachsen (LSN)**  
Göttinger Chaussee 76  
30453 Hannover  
Telefon: 0511 9898-1134  
Telefax: 0511 9898-4132  
[www.statistik.niedersachsen.de](http://www.statistik.niedersachsen.de)  
[auskunft@statistik.niedersachsen.de](mailto:auskunft@statistik.niedersachsen.de)

**Statistisches Landesamt  
Sachsen-Anhalt**  
Merseburger Straße 2  
06110 Halle /Saale  
Telefon: 0345 2318-0  
Telefax: 0345 2318-913  
[www.statistik.sachsen-anhalt.de](http://www.statistik.sachsen-anhalt.de)  
[info@stala.mi.sachsen-anhalt.de](mailto:info@stala.mi.sachsen-anhalt.de)

**Statistisches Landesamt Bremen**  
An der Weide 14-16  
28195 Bremen  
Telefon: 0421 361-6070  
Telefax: 0421 361-6168  
[www.statistik.bremen.de](http://www.statistik.bremen.de)  
[bibliothek@statistik.bremen.de](mailto:bibliothek@statistik.bremen.de)

**Information und Technik  
Nordrhein-Westfalen**  
Geschäftsbereich Statistik  
Mauerstraße 51  
40476 Düsseldorf  
Telefon: 0211 9449-2495  
Telefax: 0211 9449-2104  
[www.it.nrw.de](http://www.it.nrw.de)  
[statistik-info@it.nrw.de](mailto:statistik-info@it.nrw.de)

**Thüringer Landesamt  
für Statistik**  
Europaplatz 3  
99091 Erfurt  
Telefon: 0361 37-900  
Telefax: 0361 37-84699  
[www.statistik.thueringen.de](http://www.statistik.thueringen.de)  
[auskunft@statistik.thueringen.de](mailto:auskunft@statistik.thueringen.de)

**Statistisches Amt für Hamburg  
und Schleswig-Holstein  
Standort Hamburg**  
Steckelhörn 12  
20457 Hamburg  
Telefon: 040 42831-1766  
Telefax: 040 42831-1700  
**Standort Kiel**  
Fröbelstraße 15-17  
24113 Kiel  
Telefon: 0431 6895-9393  
Telefax: 0431 6895-9498  
[www.statistik-nord.de](http://www.statistik-nord.de)  
[info@statistik-nord.de](mailto:info@statistik-nord.de)

**Statistisches Landesamt  
Rheinland-Pfalz**  
Mainzer Straße 14-16  
56130 Bad Ems  
Telefon: 02603 71-4444  
Telefax: 02603 71-194444  
[www.statistik.rlp.de](http://www.statistik.rlp.de)  
[info@statistik.rlp.de](mailto:info@statistik.rlp.de)

