

Dipl.-Ökonom Bernd Michel

Ergebnisse der Mehrfachfallprüfung im Rahmen des Zensustests

Die mit einem Zensus ermittelte amtliche Einwohnerzahl ist Grundlage für zahlreiche Entscheidungen, u. a. für den Finanzausgleich zwischen Bund, Ländern und Gemeinden. Vor einem Methodenwechsel zu einem registergestützten Zensus war es daher notwendig, die Qualität der Melderegister und damit auch die erreichbare Qualität der Zensusergebnisse zu untersuchen. Die Statistischen Ämter des Bundes und der Länder haben diese Untersuchungen im Rahmen des Zensustests durchgeführt und die zusammengefassten Ergebnisse veröffentlicht.¹⁾ In diesem Beitrag stehen die Ergebnisse der „Mehrfachfallprüfung“ im Rahmen des Zensustests im Mittelpunkt.²⁾ Aufgabe der „Mehrfachfallprüfung“ war es, Verfahren zu entwickeln und zu testen, mit denen fälschlich mehrfach in den Melderegisterdateien geführte Personen entdeckt und die Dateien um solche Fälle statistisch bereinigt werden können.

1 Konzeption und Ziele

Die Melderegister werden in Deutschland gemeindeweise geführt. Bei dezentral geführten Melderegistern ist nicht auszuschließen, dass Personen in mehreren Gemeinden gleichzeitig mit alleiniger Wohnung oder mit Hauptwohnung gemeldet sind, oder ausschließlich mit Nebenwohnung registriert sind. Solche Fehler können durch das Meldeverhalten der Bürger und/oder durch die Bearbeitung in den Meldebehörden entstehen. Bei einer Nutzung der Melde-

daten zu Zensuszwecken ohne weitere Prüfung der Angaben besteht daher die Gefahr, dass Personen mehrfach, am falschen Ort oder mit falschem Wohnungsstatus gezählt werden.

Um die Einwohnerzahl von Doppelzählungen frei zu halten, sollten im Rahmen des mit Stichtag 5. Dezember 2001 durchgeführten Zensustestes Verfahren der Mehrfachfallprüfung stichprobenartig getestet werden. In den folgenden Ausführungen wird dargestellt,

- wie die Gesamtzahl der Mehrfachfälle in den Melderegistern ermittelt wurde,
- welche Möglichkeiten einer maschinellen Aufklärung bestehen und
- welche Informationspotenziale die Analyse von zwei zeitlich versetzten Datenlieferungen bietet.

Als Stichprobenverfahren wurde die Geburtstagsauswahl gewählt, da sie sicherstellt, dass jeder in die Auswahl gelangte Einwohner mit seinen in den Melderegistern gespeicherten Angaben einbezogen wird. Das Gesetz zur Vorbereitung eines registergestützten Zensus (Zensusvorbereitungsgesetz) vom 27. Juli 2001³⁾ sah daher in § 2 eine Stichprobenerhebung bei allen Meldebehörden vor, in die alle Einwohner, die am 1. Januar, 15. Mai oder 1. September Geburtstag hatten, einbezogen wurden.⁴⁾ Zusätzlich

1) Siehe Statistische Ämter des Bundes und der Länder: „Ergebnisse des Zensustests“ in WiSta 8/2004; S. 813 ff.

2) In kommenden Ausgaben dieser Zeitschrift werden weitere Berichtsteile ausführlich dargestellt.

3) (BGBl. I S. 1882), Artikel 1: Gesetz zur Erprobung eines registergestützten Zensus (Zensusvorbereitungsgesetz – ZensTeG).

4) Für weitergehende Informationen siehe Bihler, W., in Heft 2 der Schriftenreihe „Methodenberichte“: „Das Stichprobenverfahren der Testerhebungen für einen registergestützten Zensus“.

nahmen die Einwohner teil, die mit unvollständigem Geburtsdatum in den Melderegistern enthalten waren, um die Verfahren der Mehrfachfallprüfung unter erschwerten Bedingungen testen zu können.

Da die Fristen für die An- und Abmeldung erfahrungsgemäß häufig überschritten werden, war es zudem erforderlich, Vorkehrungen zu treffen, um auch die Einwohner am richtigen Wohnort zählen zu können, die sich nach dem Stichtagsdatum 5. Dezember 2001 bei den Meldebehörden rückwirkend an- oder abgemeldet hatten. Gelöst wurde dieses Problem mittels einer zweiten Lieferung von Daten der Melderegister zum 31. März 2002.

Die Geburtstagsauswahl betraf alle Gemeinden. Bereits im Vorfeld der Erhebung wurden daher alle Gemeinden über die anstehenden Arbeiten im Rahmen der Mehrfachfallprüfung unterrichtet. Da die Gemeinden oftmals Dritte mit der technischen Führung der Melderegister beauftragt haben bzw. bei der Führung der Melderegister auf extern entwickelte Software zurückgreifen, wurden auch Gebietsrechenzentren, kommunale Rechenzentren usw. und Anbieter von Software zur Führung von Melderegistern zeitgleich über das Vorhaben einschließlich des Aufbaus der Lieferdatensätze informiert.

Aus den Melderegistern wurden zum Zweck der Mehrfachfallprüfung folgende Merkmale erhoben:

1. als Erhebungsmerkmale:

Geburtsmonat und -jahr; Geschlecht; Staatsangehörigkeiten; bei im Ausland Geborenen: Geburtsstaat; Familienstand; Wohnort; Status der Wohnung (alleinige Wohnung, Haupt- oder Nebenwohnung);

2. als Hilfsmerkmale:

Namen, Vornamen; gegenwärtige Anschriften; Tag der Geburt; Geburtsort; Standesamt und Nummer des Geburtseintrags; Anschrift und Status der künftigen Wohnung oder der Wohnung, in die der Einwohner laut Rückmeldung verzogen ist; Anschrift und Status der Wohnung in der Gemeinde, aus der der Einwohner zugezogen ist; Zuzug aus dem Ausland; Anschrift der zuletzt bewohnten Wohnung in der Gemeinde; Datum des Beziehens der Wohnung; Datum des Auszugs aus der Wohnung; Datum des Fortzugs ins Ausland; Datum der Anmeldung bei der Meldebehörde; Datum der Abmeldung bei der Meldebehörde; Datum des Wohnungsstatuswechsels.

Die Gemeinden wurden gebeten, dem für sie zuständigen Statistischen Landesamt die Daten auf elektronischem Weg (Datenfernübertragung oder Datenträgeraustausch) und in dem vorgegebenen Datensatzformat zu übermitteln. Grundlage des Datenaustausches war der einheitliche Datensatz für das Meldewesen (DSMeld).

In den Statistischen Ämtern der Länder wurden die Datenlieferungen der Gemeinden nach einem vorgegebenen bundeseinheitlichen Verfahren auf Plausibilität und Vollständigkeit geprüft. Die Daten wurden anschließend an das Statistische Bundesamt übermittelt, wo die weiteren Prüf- und Auswertungsarbeiten erfolgten.

Da die amtliche Statistik vor der Durchführung der Testerhebungen keine Erfahrungen mit der Methode und Leistungsfähigkeit von Verfahren zur Aufdeckung von Dubletten in großen Datenbeständen hatte, wurden externe Anbieter von Software zur Dublettensuche damit beauftragt, die Mehrfachfallprüfung durchzuführen. Parallel dazu sollten eigene praktische Erfahrungen gesammelt werden.

Als Ergebnis der Mehrfachfallprüfung entstand eine Auflistung von Einwohnern, die mehrfach mit Hauptwohnung, alleiniger Wohnung oder ausschließlich mit Nebenwohnung gemeldet waren. In diesen unplausiblen Fällen wurden die Einwohner zum Ort ihrer Hauptwohnung am Stichtag des Zensus-tests befragt.

2 Erfahrungen mit der Mehrfachfallprüfung

2.1 Datentechnische und organisatorische Gegebenheiten

Da es sich bei den Datenbeständen um Lieferungen aus Melderegistern handelt, die mit unterschiedlichen Software-Versionen auf unterschiedlichen Plattformen geführt werden, wurden im Vorfeld der Erhebung in Anlehnung an die Datenübermittlungsverordnungen Format und Zeichensatz der Datenlieferungen festgelegt sowie in den Prüfungen der Eingangsplausibilität auf Vereinheitlichung hingearbeitet. Es ließ sich nicht verhindern, dass Daten in nicht vereinbarten Zeichensätzen geliefert wurden. Besonders problematisch war es, wenn eine Datenlieferung ein und derselben Gemeinde mit verschiedenen Zeichensätzen erfolgte, wenn es datentechnische Änderungen zwischen den beiden Datenlieferungen gab, die nicht mitgeteilt wurden, wenn Zeichen bei der Übernahme der Daten auf Großrechner verloren gingen bzw. falsch umgesetzt wurden oder wenn Steuerzeichen zu Abbrüchen bei der Datenfernübertragung führten. Beim Zensus selbst werden die Standardisierungs- und Prüfmaßnahmen entsprechend ausgeweitet.

In einigen Datenlieferungen wurde von den Codierungen des Datensatzes für das Meldewesen abgewichen, beispielsweise beim Merkmal der Staatsangehörigkeit. Gründe hierfür waren auch Wechsel beim Softwareanbieter bzw. bei der Softwareversion.

In der ersten Datenlieferung wurden 971 037 Datensätze geliefert. Damit waren etwas weniger als 1,2% der Bevölkerung Deutschlands in die Stichprobe zur Mehrfachfallprüfung einbezogen. Das einwohnerstärkste Bundesland Nordrhein-Westfalen stellte in der ersten Datenlieferung mit fast 235 000 Sätzen (etwa 1,3% der Bevölkerung Nordrhein-Westfalens) beinahe ein Viertel der Testdatensätze. Die Stadtstaaten lagen – gemessen an ihrer Einwohnerzahl – mit 1,5% (Bremen), 1,6% (Hamburg) und 1,7% (Berlin) deutlich über dem Bundesdurchschnitt. Dies hängt mit dem überdurchschnittlichen Ausländeranteil in Großstädten und der Tatsache zusammen, dass in anderen Kulturen dem Geburtsdatum nicht die Bedeutung zukommt wie in Deutschland oder anderen EU-Staaten. Infolgedessen ist die Kategorie „Personen mit unvollständigem Geburtsda-

tum“, aber auch die Kategorie „Personen mit Geburtsdatum 1. Januar“ in Bundesländern mit hohem Ausländeranteil stärker besetzt.

Die zweite Datenlieferung der Gemeinden lag mit 969 901 Datensätzen unwesentlich unter der Zahl der Datensätze der ersten Lieferung. Dies gilt auch für die einzelnen Bundesländer.

Personen in der Stichprobe mit mehr als einer alleinigen Wohnung oder Hauptwohnung und Personen, die ausschließlich mit Nebenwohnung gemeldet waren, wurden durch die Statistischen Landesämter zum tatsächlichen Hauptwohnsitz befragt. Es kamen zwei Erhebungsunterstützungsprogramme zur Anwendung:

- Das Statistische Landesamt Sachsen hatte eine PC-gestützte Erhebungsorganisation erstellt (Access-Lösung), mit der die Rücklaufkontrolle, die Erfassung der Erhebungsunterlagen und der Datenaustausch zwischen den Bundesländern sowie die Durchführung des postalischen Mahnwesens gesteuert wurden.
- Das Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen hatte für die in der zweiten Mahnstufe durchgeführten Telefoninterviews eine PC-gestützte Erhebungsorganisation auf Basis der Programmiersprache Blaise entwickelt. Durch das Landesamt für Datenverarbeitung und Statistik Nordrhein-Westfalen wurde ferner eine Verknüpfung der Telefonnummernrecherche über eine Telefonauskunfts-CD-ROM hergestellt, sodass die Suchergebnisse automatisch in die CATI-Datenbank eingespielt werden konnten. Die eigentliche telefonische Befragung wurde mit Hilfe der CATI-Module durchgeführt. Hierzu mussten die aus dem Erhebungsunterstützungsprogramm für die postalische Befragung übernommenen Datensätze verwaltet werden (z.B. Wiedervorlage zu bestimmten Tageszeiten und Wochentagen, Treffen von Verabredungen usw.).

Die Befragung fand im Zeitraum August bis Dezember 2002 statt. Bundesweit wurden 9 159 Personen postalisch befragt. Die Befragung erfolgte grundsätzlich durch das Statistische Landesamt, in dessen Zuständigkeitsbereich die jeweils jüngste Zuzugsadresse lag.

2.2 Verfahren der Mehrfachfallprüfung

Grundannahme bei der Suche von Dubletten in Datenbeständen von Personen ist, dass es Merkmale in den Datensätzen gibt, anhand derer Daten eindeutig der gleichen Person zugewiesen werden können. Gesucht waren also Merkmale, die sich für Personen im Zeitablauf in der Regel nicht ändern und die miteinander kombiniert Schlüsselcharakter erhalten. Es wurde davon ausgegangen, dass Datensätze sich dann auf die gleiche Person beziehen, wenn die Datensätze in den Schlüsselmerkmalen Geschlecht, Geburtsjahr, Geburtsmonat, Geburtstag, Geburtsstaat, Geburtsort, Geburtsname und Vorname übereinstimmen. Wenn diese Übereinstimmung beispielsweise in zwei Datensätzen vollständig gegeben ist, ist die Wahrscheinlichkeit, dass eine Person doppelt gemeldet ist, sehr hoch. Wird berücksichtigt, dass es Fehler bei der Datenaufnahme,

unterschiedliche Schreibweisen – auch bei der Übertragung von fremdsprachlichen Namen ins Deutsche – oder Abkürzungen von Gemeindennamen gibt, können Datensätze auch dann noch der gleichen Person zugeschrieben werden, wenn die Übereinstimmung in den Schlüsselmerkmalen nicht 100% beträgt. Andererseits können marginale Abweichungen in der Schreibweise von Namen (z. B. „oe“ anstatt „ö“) auch bei ansonsten gleichen Merkmalsausprägungen auf Datensätze von zwei Personen hinweisen.

Die amtliche Statistik hatte vor der Durchführung des Zensusustes keine Erfahrung mit einer weitgehend maschinell ausgerichteten Suche nach Dubletten in großen Datenbeständen. Daher hat das Statistische Bundesamt diese Aufgabe an externe Dienstleister vergeben. Um die Leistungsfähigkeit unterschiedlicher Methoden der Dublettensuche beurteilen zu können, wurden verschiedene Unternehmen mit dieser Aufgabe beauftragt. Diese Unternehmen erhielten die auf die Schlüsselmerkmale reduzierten Dateien und hatten die Aufgabe, die Datensätze zu markieren, die zu gleichen Personen gehörten. Die Ergebnisse wurden im Statistischen Bundesamt auf Vollständigkeit und Plausibilität geprüft. Hierzu wurden maschinelle Gruppierungs- und Vergleichsverfahren eingesetzt, die sich als sehr effizient erwiesen haben. Maschinell nicht entscheidbare „Mehrfachfälle“ wurden einer visuellen Kontrolle unterzogen.

Im Ergebnis zeigte sich, dass die in den Test einbezogenen externen Verfahren der Dublettensuche bei einem Zensus zu umfangreichen Prüfungen im Dialog-Betrieb führen würden und dieser Aufwand keine wesentliche Verbesserung in der Dublettenerkennung bringt als das für Kontrollzwecke vom Statistischen Bundesamt entwickelte Konzept des maschinellen Gruppierungs- und Vergleichsverfahrens auf der Grundlage eines Bit-weisen Vergleichs mit stark eingeschränkter Ausnahmeregelung. Von daher wurden Verfahrensschritte zur Dublettenerkennung entwickelt, die externe Lösungen entbehrlich machen. Diese Verfahrensschritte müssen für den Zensus optimiert und an die zu erwartende Datenmenge angepasst werden.

Nicht alle in der Mehrfachfallprüfung gekennzeichneten Dubletten sind Karteileichen, das heißt Übererfassungen. Eine Person, die in der Zeit um den Stichtag umgezogen ist, sich in der neuen Gemeinde zwar vor dem Stichtag angemeldet hat, in ihrer Fortzugsgemeinde aber in der ersten Datenlieferung noch enthalten ist, wird zunächst als Dublette ausgewiesen. Ihr Meldeverhältnis kann aber aufgrund der zweiten Datenlieferung, wenn sie in der Fortzugsgemeinde nicht mehr zum Stichtag geführt wird, richtig gestellt werden (Klärung im Zeitablauf).

In der Übersicht sind die Nutzungsmöglichkeiten von zwei Datenlieferungen aufgrund der Erfahrungen der Mehrfachfallprüfung beim Zensus zusammenfassend dargestellt. Die Übersicht verdeutlicht, dass die Dublettensuche nur *ein* Ergebnis der Mehrfachfallprüfung ist. Die oben dargestellten Vergleiche der beiden Datenlieferungen können für verschiedene Konstellationen weitere Karteileichen aufdecken, in der Regel über maschinelle Verfahren. Die entsprechenden Fallkonstellationen sind in der Übersicht kursiv hervorgehoben.

Nutzungsmöglichkeiten von zwei Datenlieferungen aufgrund der Erfahrungen der Mehrfachfallprüfung beim Zensusstest

Beschreibung	Gründe	Vorgehensweise, falls Person Karteileiche/Fehlbestand ist, in ...	
		Mehrfachfallprüfung	Registertest
A-Fälle: Personen sind in beiden Datenlieferungen mit gleicher Anschrift gemeldet.	Zwischen den beiden Datenlieferungen haben sich keine meldepflichtigen Änderungen beim Wohnort für die Personen ergeben, Änderungen wurden vom Bürger nicht gemeldet oder von der Meldebehörde noch nicht verarbeitet.	Personen mit mehreren Wohnsitzen werden auffällig in der Dublettenprüfung. Wohnsitzfrage wird geklärt.	Keine. Falls unter Meldeadresse nicht wohnhaft, wird die Person als Karteileiche, aber nicht als Mehrfachfall erkannt.
C-Fälle: In der ersten Lieferung enthaltene Personen sind in der zweiten Lieferung für die gleiche Anschrift nicht mehr enthalten.	Zwischen den beiden Datenlieferungen sind meldepflichtige Änderungen beim Wohnort stichtagsrelevant wirksam geworden. Personen sind nach dem Stichtag innerhalb Deutschlands umgezogen, in das Ausland verzogen oder verstorben.	Durch den Vergleich der beiden Datenlieferungen werden „temporäre“ Karteileichen gleichermaßen erkannt und geklärt.	
B-Fälle: In der zweiten Lieferung sind an Adressen Personen gemeldet, die in der ersten Lieferung dort nicht gemeldet waren.	Person hat sich stichtagsrelevant am neuen Wohnort angemeldet und: – an ihrem bisherigen Wohnort bereits abgemeldet und die Abmeldung/Rückmeldung ist bereits verarbeitet. – Abmeldung/Rückmeldung ist noch nicht erfolgt/verarbeitet	– temporäre Karteileiche (B-Fall) – Person wird auffällig durch Vergleich der Teilmengen der B-Fälle mit A-Fällen (ohne Dubletten).	– temporäre Karteileiche (B-Fall) – Keine. Karteileichen (in A-Fällen) sind überhöht.
D-Fälle: In der zweiten Lieferung sind an Adressen Personen gemeldet, die in der ersten Lieferung dort nicht gemeldet waren.	Person hat sich <i>nach dem Stichtag</i> am neuen Wohnort angemeldet.	Die D-Fälle lassen sich beim Zensus nutzen, um Fälle maschinell zu klären, die in anderen Verfahren als Karteileichen erkannt werden.	

3 Ergebnisse der Mehrfachfallprüfung

3.1 Unterschiede in den beiden Datenlieferungen

Bezogen auf die Bevölkerung am Ort der Hauptwohnung waren die Datensätze hochgerechnet für 79,3 Mill. Bürger in der ersten (MR1) und zweiten Datenlieferung (MR2) gleich.

Knapp 2,9 Mill. Personen sind zwischen erster und zweiter Datenlieferung um- bzw. fortgezogen oder verstorben. Beim Projektteil Registertest⁵⁾ waren es 2,7 Mill. Personen (nach Bereinigung). Die Analyse der Personengruppe, die im Zeitraum zwischen den beiden Datenlieferungen Änderungsmeldungen zu verzeichnen hatte, gibt Hinweise auf Möglichkeiten der weiteren Reduktion von Registerfehlern. Bei der Durchführung eines Zensus wird zu klären sein, bei welcher Gemeinde diese Personen zu streichen sind, aber auch, welcher Gemeinde sie zuzuschreiben sind.

Tabelle 1: Übereinstimmung der stichtagsbezogenen Datenlieferungen¹⁾ auf Satzebene nach Bundesländern
Hochgerechnetes Ergebnis der Mehrfachfallprüfung vom 5. Dezember 2001

Bundesland	Haupt-/alleinige Wohnung		Nebenwohnung		Insgesamt	
	MR1 + MR2	nur MR1	MR1 + MR2	nur MR1	MR1 + MR2	nur MR1
Baden-Württemberg	10 155 094	327 523	545 170	26 136	10 700 264	353 659
Bayern	11 915 073	404 964	822 519	32 947	12 737 592	437 911
Berlin	3 219 156	129 566	122 536	5 584	3 341 691	135 151
Brandenburg	2 429 562	137 630	143 694	11 358	2 573 256	148 988
Bremen	628 642	26 620	16 150	373	644 792	26 993
Hamburg	1 669 774	60 039	30 895	2 798	1 700 669	62 837
Hessen	5 892 131	180 326	298 333	9 357	6 190 463	189 683
Mecklenburg-Vorpommern	1 655 071	84 014	93 980	9 168	1 749 050	93 183
Niedersachsen	7 665 186	276 621	413 248	11 925	8 078 434	288 546
Nordrhein-Westfalen	17 463 147	594 485	679 384	32 022	18 142 531	626 507
Rheinland-Pfalz	3 924 541	123 409	203 121	8 616	4 127 662	132 025
Saarland	1 017 296	50 591	60 026	1 078	1 077 322	51 670
Sachsen	4 213 190	141 202	211 030	11 250	4 424 220	152 452
Sachsen-Anhalt	2 465 443	106 498	113 033	6 444	2 578 475	112 942
Schleswig-Holstein	2 659 528	158 612	200 992	8 041	2 860 519	166 653
Thüringen	2 312 385	76 713	107 086	6 901	2 419 471	83 615
Deutschland ...	79 285 218	2 878 814	4 061 196	183 998	83 346 415	3 062 811

1) Datenlieferung 1 (MR1): Stichtag 5. Dezember 2001; Datenlieferung 2 (MR2): Stichtag 31. März 2002.

5) Beim „Registertest“ wurden in einer von der Mehrfachfallprüfung unabhängigen Stichprobe für ausgewählte Adressen die Angaben in den Melderegistern mit den Ergebnissen einer primärstatistischen Erhebung (Haushaltbefragung) verglichen.

3.2 Untersuchung von Haupt-/Nebenwohnungsbeziehungen

Im Rahmen der Dublettenprüfung wurden auch Personen als Dubletten gekennzeichnet, die mit Hauptwohnung und Nebenwohnung gemeldet waren.

Im Rahmen der Plausibilitätskontrolle wurde dieser Personenkreis daraufhin untersucht, ob Wohnungsstatus und Anschriften der Wohnungen einander entsprachen. Bei korrespondierenden Haupt- und Nebenwohnungsanschriften fand keine Befragung statt. Anderenfalls war zu prüfen, in welche Prüfkategorie der Fall fällt. Beispielsweise konnte der Fall „Nebenwohnung ohne Hauptwohnung“ entstehen, dieser zog eine Befragung mit dem Ziel der Feststellung des Hauptwohnsitzes in 2 316 Fällen nach sich.

Die insgesamt 33 825 Dubletten, die zunächst mit dem Status „Hauptwohnung und Nebenwohnung“ herausgefiltert wurden, wurden daraufhin untersucht, inwieweit sie echte Dubletten sind. Als Instrumentarium wurde hier das für die Dublettensuche entwickelte Instrumentarium der gestaffelten Stringvergleiche⁶⁾ angewandt. Um die Qualität des maschinellen Verfahrens beurteilen zu können, wurde das Ergebnis manuell überprüft. So wurde festgestellt, dass von den 33 825 zu untersuchenden Dubletten 31 856 (94,2%) echte Dubletten waren und 1 969 Dubletten (5,8%) fälschlich zusammengeführte Dubletten.

Die herausgefilterten echten Dubletten wurden maschinell auf Korrespondenz der verzeigten Adressen untersucht. Um bei den bereits standardisierten Datensätzen noch eventuell abweichende Schreibweisen und Abkürzungen zu kompensieren, wurden weiterhin schrittweise die Felder Ort und Straße verkürzt. Letztendlich hat sich gezeigt, dass man mit einer Verkürzung des Feldes Ort auf die 5 ersten Zeichen und der Straßenbezeichnung auf die 15 ersten Zeichen ein hinreichend genaues Ergebnis beim Vergleich der Adressen erreichen kann. So ist es gelungen, von den 31 856 zu untersuchenden Dubletten bei 30 740 (96,5%) eine Korrespondenz nachzuweisen, bei 1 116 Dubletten (3,5%) konnte der Nachweis nicht erbracht werden.

Um die derzeit 3,5% nicht sicher maschinell klärbaren Dubletten aufklären zu können, sind in der Zeit bis zum Zensus weitere Prüfverfahren zu entwickeln. Sollten diese Untersuchungen zu keiner weiteren Verbesserung des Verfahrens führen, würden derartige Dubletten vorerst in diesem Arbeitsschritt unerkannt bleiben. Letztendlich würden diese Dubletten bekannt werden nach der Befragung aller Nebenwohnungen ohne korrespondierende Hauptwohnung oder alleinige Wohnung. Eine Nebenwohnung könnte so zu einer Hauptwohnung oder alleinigen Wohnung verzeigert oder auch umgewandelt werden, entsprechend dem durch den Befragten bekannt gegebenen Wohnungsstatus. Sollte

eine bisher unbekannt Hauptwohnung benannt werden, ist diese zu erfassen und als Fehlbestand zu kennzeichnen. Alle anderen Nebenwohnungen wären Karteileichen. Bei 24 994 Fällen korrespondierten die Merkmale Amtlicher Gemeindeschlüssel, Postleitzahl, Ort, Straße und Hausnummer zu 100%, was 81,3% von allen korrespondierenden Dubletten bzw. 78,5% aller Dubletten entspricht.

3.3 Ergebnisse der Befragung

Aus den Sätzen der ersten Datenlieferung der Geburtstagsauswahl wurden in einem abgestuften Verfahren die Personen extrahiert, die mit mehr als einer Hauptwohnung bzw. ausschließlich mit Nebenwohnung(en) gemeldet waren. Nach der Befragung wurden im Statistischen Bundesamt abschließend die befragten Dubletten anhand von Geburtsnamen, Nachnamen, Vornamen, Geschlecht, Geburtsdatum und Geburtsort noch einer inhaltlichen Prüfung unterzogen. Damit wurden auch in den Befragungsketten noch enthaltene Mängel aufgrund falscher Zuordnungen externer Anbieter beseitigt. Die auf diese Weise aufgelösten „Dubletten“ sind in den nachfolgend dargestellten Ergebnissen nicht mehr enthalten. Aufgenommen wurden dagegen die Datensätze der Personen, die im Melderegister von ein und derselben Gemeinde mehrmals unter der gleichen Anschrift geführt wurden (672 Fälle), aber aus der Befragung aussteuert worden waren. Die Datei, die den folgenden Tabellen zugrunde liegt, umfasst vor Hochrechnung insgesamt 11 370 Fälle (Karteileichen).

Das Hochrechnungsverfahren sieht vor, die Personen mit unvollständigem Geburtsdatum und die am 1. Januar Geborenen als „Totalschicht“ zu betrachten, also nicht hochzurechnen. Die beiden anderen Geburtstagstage (15. Mai und 1. September) stehen repräsentativ für die verbleibenden Tage des Jahres.⁷⁾

Die hier dargestellten Ergebnisse sind aufgrund konservativer Annahmen bei der Entscheidung Dublette/Nicht-Dublette Mindestzahlen. Aufgrund anderer Unzulänglichkeiten in der Registerführung, die nach Durchführung des Tests zum Teil behoben sind, werden die Dubletten im Rahmen der Testerhebung generell leicht unterschätzt.

Hochgerechnet beträgt die Zahl der in der Mehrfachfallprüfung aufgedeckten Karteileichen für die Bevölkerung am Ort der Hauptwohnung 485 499 Personen. Ihre Verteilung auf die Bundesländer zeigt Tabelle 2.

Es ist festzustellen, dass die Zahl der dauerhaften unzulässigen Dubletten in den Melderegistern – im Unterschied zu den Meinungen im Vorfeld des Zensus – mit rund 485 000 Personen unter den Erwartungen bleibt. Hierbei ist allerdings zu berücksichtigen, dass das bei der Mehrfachfallprüfung angewandte Verfahren der Bereinigung des

⁶⁾ Unter einem String (Zeichenkette) wird eine Aneinanderreihung von einzelnen Zeichen (Buchstaben, Ziffern, Sonderzeichen) verstanden.

⁷⁾ Grundsätzlich ließe sich die Genauigkeit der Ergebnisse bei kleinen Fallzahlen verbessern, wenn die am 1. Januar in Deutschland geborenen Deutschen wie die am 1. September bzw. 15. Mai geborenen Personen hochgerechnet würden und der Rest als „Totalschicht“ behandelt würde. Der durchschnittliche Hochrechnungsfaktor läge dann bei 365,25/3. Diese wünschbare differenzierte Hochrechnung, die der eingangs angeführten Abgrenzung Rechnung trägt, konnte nur nicht durchgeführt werden, da für die am 1. Januar in Deutschland geborenen Deutschen die erforderlichen Eckwerte nicht vorliegen. Das gewählte Hochrechnungsverfahren führte bei einigen Analysen zu nicht interpretationsfähigen Ergebnissen. Bei einer Änderung des Hochrechnungsverfahrens würde die Zahl der in der Mehrfachfallprüfung durch die Befragung am Ort der Hauptwohnung aufgedeckten Karteileichen bundesweit um etwa 45 000 Personen höher liegen. Von größerer Bedeutung sind dabei Änderungen in den Ergebnissen einzelner Bundesländer.

Tabelle 2: Aufgedeckte Karteileichen am Ort der Hauptwohnung nach Verfahren und Bundesländern
Hochgerechnetes Ergebnis der Mehrfachfallprüfung vom 5. Dezember 2001

Bundesland	Personen im Melderegister	Aufgedeckte Karteileichen am Ort der Hauptwohnung durch ...						
		Befragung			Vergleich ¹⁾		insgesamt	
		1 000	Anzahl	% ²⁾	Anzahl	% ²⁾	Anzahl ³⁾	% ²⁾
		1	2	3	4	5	6	7
Baden-Württemberg	10 479,1	39 241	0,37	6 032	0,06	45 273	0,43	
Bayern	12 323,4	34 526	0,28	13 863	0,11	48 389	0,39	
Berlin	3 349,2	14 939	0,45	2 303	0,06	17 242	0,51	
Brandenburg	2 568,1	13 437	0,52	6 009	0,23	19 446	0,75	
Bremen	655,3	3 493	0,53	357	0,06	3 850	0,59	
Hamburg	1 730,1	11 069	0,64	384	0,02	11 453	0,66	
Hessen	6 073,5	38 719	0,64	12 323	0,20	51 042	0,84	
Mecklenburg-Vorpommern	1 739,3	20 855	1,20	3 584	0,20	24 439	1,40	
Niedersachsen	7 946,1	48 267	0,61	15 203	0,19	63 470	0,80	
Nordrhein-Westfalen	18 060,6	76 246	0,42	17 878	0,10	94 124	0,52	
Rheinland-Pfalz	4 048,8	15 578	0,38	11 558	0,29	27 136	0,67	
Saarland	1 068,6	5 558	0,52	715	0,07	6 273	0,59	
Sachsen	4 355,1	20 022	0,46	3 481	0,08	23 503	0,54	
Sachsen-Anhalt	2 575,4	9 459	0,37	5 568	0,21	15 027	0,58	
Schleswig-Holstein	2 818,7	17 630	0,63	5 020	0,17	22 650	0,80	
Thüringen	2 389,1	9 195	0,38	2 987	0,13	12 182	0,51	
Deutschland ...	82 180,3	378 234	0,46	107 265	0,13	485 499	0,59	

1) Vergleich der Datenlieferung 2 (MR2) vom 31. März 2002 mit der Datenlieferung 1 (MR1) vom 5. Dezember 2001. – 2) Anteil der Spalte 1. – 3) Spalte 2 + Spalte 4.

Bestandes um etwa 370 000 „temporäre Dubletten“ die Befragungsfälle bereits reduziert hat.

In dem von den Gemeinden zum Stichtag 5. Dezember 2001 berichteten Bestand der Melderegister sind damit etwa 856 000 Personen⁸⁾ zuviel – weil dauerhaft oder temporär doppelt – enthalten. In Verbindung mit dem Ergebnis der Auszählung der Melderegister aller Gemeinden zum Stichtag des Zensusstests gemäß § 2 Zensusstestgesetz, das bereits um etwa 410 000 Personen niedriger lag als die von

der Bevölkerungsfortschreibung ermittelte Einwohnerzahl, ist damit festzustellen, dass die Bevölkerungsfortschreibung, bezogen auf die Bevölkerung am Ort der Hauptwohnung, um mindestens 1,3 Mill. Personen über einer Registerauswertung mit Bereinigung der Mehrfachfälle liegt.

Gleichwohl könnte vermutet werden, dass die Verfahren der Dublettensuche zu wenige Mehrfachfälle aufgedeckt hätten. Hier zeigen die aufwändig durchgeführten Prüfarbeiten, dass alle Verfahren in der im Zensusstest zur Anwendung

Tabelle 3: Relativer Standardfehler bei den aufgedeckten Karteileichen am Ort der Hauptwohnung nach Bundesländern
Hochgerechnetes Ergebnis der Mehrfachfallprüfung vom 5. Dezember 2001

Bundesland Gemeindegrößenklasse von ... bis unter ... Einwohnern	Personen im Melderegister ¹⁾	Aufgedeckte Karteileichen durch Befragung insgesamt			
		Anzahl	Einfacher relativer Standardfehler	Ergebnis liegt mit einer Wahrscheinlichkeit von 68% zwischen ... und ...	
Baden-Württemberg	10 479 051	45 273	6,37	42 389	48 157
Bayern	12 323 356	48 389	6,94	45 030	51 748
Berlin	3 349 202	17 242	10,03	15 513	18 971
Brandenburg	2 568 105	19 446	11,53	17 205	21 687
Bremen	655 263	3 850	21,46	3 024	4 676
Hamburg	1 730 132	11 453	11,91	10 089	12 817
Hessen	6 073 471	51 042	6,44	47 752	54 332
Mecklenburg-Vorpommern	1 739 272	24 439	8,93	22 256	26 622
Niedersachsen	7 946 141	63 470	5,92	59 713	67 227
Nordrhein-Westfalen	18 060 575	94 124	4,59	89 807	98 441
Rheinland-Pfalz	4 048 844	27 136	10,30	24 342	29 930
Saarland	1 068 605	6 273	17,29	5 188	7 358
Sachsen	4 355 074	23 503	9,36	21 303	25 703
Sachsen-Anhalt	2 575 403	15 027	13,58	12 987	17 067
Schleswig-Holstein	2 818 671	22 650	9,85	20 420	24 880
Thüringen	2 389 100	12 182	14,13	10 461	13 903
Deutschland ...	82 180 267	485 499	2,09	475 360	495 638
unter 10 000	23 071 006	149 915	3,93	144 028	155 802
10 000 – 50 000	26 928 133	153 321	3,81	147 486	159 156
50 000 – 800 000	24 839 085	139 268	3,76	134 035	144 501
800 000 und mehr	7 342 043	42 995	6,34	40 268	45 722

1) Geprüfte Registersätze im Rahmen der Mehrfachfallprüfung (einschl. B- und D-Fällen).

8) Zusätzlich zu dem Ergebnis der primärstatistischen Klärung von Dubletten ist das Ergebnis der maschinellen Auflösung von temporären Dubletten (etwa 370 000 Personen) mit in Abzug zu bringen.

Tabelle 4: Karteileichen nach Mehrfachfallprüfung nach Bundesländern und Gemeindegrößenklassen
Hochgerechnetes Ergebnis des Zensusstests vom 5. Dezember 2001
Bevölkerung am Ort der Hauptwohnung

Bundesland Gemeindegrößenklasse von ... bis unter ... Einwohnern	Personen im Melderegister ¹⁾	Karteileichen ¹⁾ , ohne „temporäre“ Karteileichen		Durch Mehrfachfallprüfung geklärte Karteileichen ²⁾		Verbleibende Karteileichen ohne „temporäre“ und nach Mehrfachfallprüfung	
		1 000	%	1 000	%	1 000	%
Baden-Württemberg	10 307,1	219,3	2,1	45,3	0,4	174,1	1,7
Bayern	11 957,5	307,9	2,6	48,4	0,4	259,5	2,2
Berlin	3 272,3	205,1	6,3	17,2	0,5	187,8	5,8
Brandenburg	2 542,4	63,7	2,5	19,4	0,8	44,3	1,9
Bremen	648,2	27,2	4,2	3,9	0,6	23,3	3,6
Hamburg	1 629,4	97,9	6,0	11,5	0,7	86,4	5,4
Hessen	5 801,2	209,0	3,6	51,0	0,8	157,9	2,9
Mecklenburg-Vorpommern ...	1 742,1	45,9	2,6	24,4	1,4	21,4	1,4
Niedersachsen	7 772,0	201,3	2,6	63,5	0,8	137,8	1,8
Nordrhein-Westfalen	17 408,8	517,7	3,0	94,1	0,5	423,5	2,4
Rheinland-Pfalz	3 972,3	85,1	2,1	27,1	0,7	58,0	1,6
Saarland	1 050,9	72,5	6,9	6,3	0,6	66,2	6,3
Sachsen	4 299,6	63,7	1,5	23,5	0,5	40,2	0,9
Sachsen-Anhalt	2 510,4	66,7	2,7	15,0	0,6	51,7	2,2
Schleswig-Holstein	2 724,1	99,6	3,7	22,7	0,8	76,9	2,9
Thüringen	2 346,4	38,3	1,6	12,2	0,5	26,2	1,1
Deutschland ...	79 984,9	2 320,8	2,9	485,5	0,6	1 835,3	2,3
unter 10 000	22 947,5	459,5	2,0	149,9	0,7	309,6	1,4
10 000 – 50 000	26 112,7	643,4	2,5	153,3	0,6	490,1	1,9
50 000 – 800 000	23 944,5	801,6	3,4	139,3	0,6	662,3	2,8
800 000 und mehr	6 980,2	416,3	6,0	43,0	0,6	373,3	5,4

1) Auszählung Registertest – ohne Bevölkerung in Anstalten. – 2) Hochrechnung und Berechnung der Anteilswerte auf Basis der Bevölkerung einschließlich der Anstaltsbevölkerung.

kommenden Spezifikation grundsätzlich zu viele „Dubletten“ nachgewiesen haben. Entsprechend sind auch Dubletten in die Befragung eingegangen, die keine waren und die zum Teil durch Reaktionen der Befragten („An der anderen Adresse habe ich nie gewohnt.“), aber vor allem durch die im Nachgang der Befragung erfolgten Prüfungen im Statistischen Bundesamt aufgelöst wurden. Derartige Fälle sind in den Ergebnistabellen nicht mehr enthalten.

Tabelle 3 zeigt den einfachen relativen Standardfehler für die durch die Mehrfachfallprüfung aufgedeckten Karteileichen. Für Deutschland beträgt der einfache relative Standardfehler bei den aufgedeckten Karteileichen 2,09%, das heißt für die im Rahmen der Mehrfachfallprüfung aufgedeckte Zahl der Karteileichen insgesamt in Höhe von 485 499 Fällen liegt mit einer Wahrscheinlichkeit von 68% das „exakte“ Ergebnis zwischen 475 360 und 495 638 Fällen. In den Bundesländern Bremen und Saarland ist aufgrund der Werte des einfachen relativen Standardfehlers davon auszugehen, dass die Aufdeckung von Mehrfachfällen wegen der zu geringen Fallzahl nicht belastbar ist.

In der Tabelle 4 werden die im Registertest festgestellten Karteileichen um die Fälle bereinigt, die im Rahmen der Befragung sowie im Anschluss daran durch die im Statistischen Bundesamt durchgeführten Kontrollen des Befragungsergebnisses aufgelöst wurden.

Die Zahl der ungeklärten Karteileichen sinkt damit bundesweit deutlich unter die 2-Millionen-Grenze auf 2,3%. Die Befragung im Rahmen der Mehrfachfallprüfung mit anschließenden Kontrollmaßnahmen der insgesamt im

Melderegister enthaltenen Personen filtert damit knapp 0,6 Prozentpunkte aus den Karteileichen heraus.⁹⁾ In den Bundesländern variiert dieser Wert zwischen 0,4 Prozentpunkten (Bayern) und 1,4% (Mecklenburg-Vorpommern).

Als Ergebnis der bisherigen Analysen der Mehrfachfallprüfung ist festzustellen:

Von den Karteileichen in den Melderegistern sind deutlich weniger als 1% des gesamten Registerbestandes Mehrfachfälle.

Offensichtlich funktionieren die im deutschen Meldewesen zur Anwendung kommenden Verfahren, insbesondere das für die Mehrfachfallprüfung bedeutende Verfahren der Rückmeldung. Probleme liegen im Meldeverhalten der Bürgerinnen und Bürger.

3.4 Weiterentwicklungspotenzial der Mehrfachfallprüfung

3.4.1 Untersuchung der nichtstichtagsrelevanten Fälle

In die bisherigen Betrachtungen sind ausschließlich die stichtagsrelevanten Anmeldungen einbezogen worden (siehe die Übersicht in Kapitel 2).

Mit der zweiten Datenlieferung wurden aber auch Daten zu Personen geliefert, die sich nach dem Stichtag 5. Dezember 2001 am neuen Wohnort angemeldet haben (so genannte D-Fälle).

⁹⁾ Das Ergebnis der maschinellen Auflösung von temporären Dubletten (etwa 370 000 Personen) kann hier nicht mehr in Abzug gebracht werden, da die Tabelle bereits um die Fälle bereinigt ist, die über die zweite Datenlieferung beim Registertest geklärt werden konnten.

Diese Fälle können entstehen durch:

- Geburt nach dem Stichtag
- Zuzug aus dem Ausland nach dem Stichtag
- Anmeldung am neuen Wohnort (Umzug innerhalb Deutschlands) nach dem Stichtag.

Im Zeitraum vom 6. Dezember 2001 bis zum 31. März 2002 haben sich annähernd 2,6 Mill. Personen in den Melderegistern angemeldet, davon etwa die Hälfte bis zum 1. Februar 2002.

Im Projektteil Registertest wurde dagegen die Zahl der nichtstichtagsrelevanten Anmeldungen von Personen mit 2,2 Mill. deutlich niedriger ermittelt (siehe Tabelle 5). Der Unterschied ist umso größer, als sich die Zahlen aus der Mehrfachfallprüfung ausschließlich auf die Zeit nach dem Stichtag beziehen, während die Zahlen aus dem Registertest zusätzlich die Abmeldungen vor dem Stichtag mit enthalten. Mit Ausnahme des Saarlands liegen die nichtstichtagsrelevanten Fälle in der Mehrfachfallprüfung über denen des Registertests. In Brandenburg war die hohe Abweichung von 58,08% zurückzuführen auf Verwaltungsreformen und damit einhergehende Gebietsreformen.

Angesichts eines in den beiden Datenlieferungen nachweisbaren zeitlich auseinander fallenden An- und Abmeldeverhaltens sind die D-Fälle auch geeignet, einen qualitätssichernden Beitrag im Rahmen der Verfahren zu liefern, die sich an die Mehrfachfallprüfung anschließen und den verbleibenden Rest der Karteileichen aufdecken sollen. Die Qualitätssicherung kann darin bestehen, dass im Nachgang der Mehrfachfallprüfung aufgedeckte Karteileichen erst dann primärstatistisch abgeklärt werden, wenn ihr stichtagsbezogener Verbleib nicht hinreichend über die zweite Datenlieferung geklärt werden kann.

Tabelle 5: Nichtstichtagsrelevante Fälle in Mehrfachfallprüfung und Registertest
Hochgerechnetes Ergebnis des Zensustests vom 5. Dezember 2001

Bundesland	Nichtstichtagsrelevante Fälle		
	Mehrfachfallprüfung	Registertest	Differenz Mehrfachfallprüfung gegenüber Registertest
	Anzahl		%
Baden-Württemberg	346 341	320 462	+7,47
Bayern	328 554	314 842	+4,17
Berlin	135 574	107 137	+20,98
Brandenburg	127 224	53 326	+58,08
Bremen	27 060	16 150	+40,32
Hamburg	49 814	46 399	+6,86
Hessen	176 782	156 087	+11,71
Mecklenburg-Vorpommern .	61 278	45 838	+25,20
Niedersachsen	250 569	233 442	+6,84
Nordrhein-Westfalen	582 306	514 353	+11,67
Rheinland-Pfalz	112 155	74 875	+33,24
Saarland	26 984	31 844	-18,01
Sachsen	121 298	96 206	+20,69
Sachsen-Anhalt	77 234	60 342	+21,87
Schleswig-Holstein	90 457	80 257	+11,28
Thüringen	68 219	53 050	+22,24
Deutschland ...	2 581 849	2 204 610	+14,61

Aufschluss über die Größenordnung der Zahl der Karteileichen, die im Registertest noch gezählt werden, aber durch den Vergleich der beiden Datenlieferungen (MR1 und MR2) aufgeklärt werden können, geben Untersuchungen über den Verbleib von Personen, die nur in der ersten Datenlieferung enthalten waren (so genannte C-Fälle). Etwa 1,8 Mill. dieser Personen haben sich nach dem Stichtag angemeldet, wie der Vergleich mit der zweiten Datenlieferung MR2 ergab. Da sich insgesamt aber 2,6 Mill. Personen nach dem Stichtag in den Melderegistern angemeldet haben, könnte durch eine flächendeckende Untersuchung der verbleibenden Menge die Herkunft für weitere 800 000 Personen geklärt werden. Erklärungsgrund für die noch offenen 800 000 Personen ist – neben der Geburt nach dem Stichtag (die am 1. Januar 2002 Geborenen wurden bereits ausgesteuert) und dem Zuzug aus dem Ausland nach dem Stichtag (maximal 300 000 Personen für einen Zeitraum von vier Monaten) – der Zuzug von rund 500 000 Personen aus einer anderen Gemeinde.

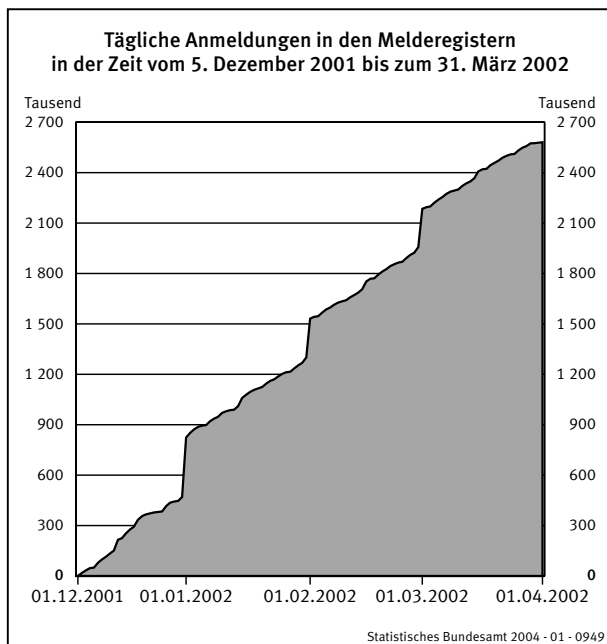
Die durch den Vergleich der beiden Datenlieferungen noch zusätzlich klärbaren Dubletten oder auch aufzudeckenden Fehlbestände liegen somit bei etwa 500 000 Personen (Bevölkerung am Ort der Hauptwohnung). Als wichtige weitere Zusatzinformation steht hier aus den etwa 500 000 Melderegisterdatensätzen die Information „Zuzug von“ zur Verfügung. Hier ist die Gemeinde, in der die betreffende Person zuvor gewohnt hat, nebst dortiger Anschrift sowie Wohnungsstatus verzeichnet. Diese Information kann zum Beispiel ein Bindeglied sein zur Auflösung einer bereits erkannten Dublette, zu der aber weitere Verlaufsinformationen fehlen. Die wichtigste Information aus dem Vergleich der beiden Datenlieferungen wird allerdings die Benennung von Fehlbeständen sein. Ein Fehlbestand lässt sich regelmäßig dann vermuten, wenn weder in MR1 noch in MR2 ein stichtagsrelevanter korrespondierender Eintrag gefunden werden kann. Hier informiert das Hilfmerkmal „Zuzug von“ darüber, in welcher Gemeinde und mit welchem Wohnungsstatus die Person zuvor gewohnt hat. Soweit die Person nicht aus dem Ausland zugezogen ist, ist sie dann als Fehlbestand der inländischen Gemeinde zum Stichtag zuzurechnen. Aufgrund der Anlage des Zensustests lassen sich über den zu erwartenden Umfang von Fehlbeständen keine Angaben machen.

3.4.2 Buchungszeitpunkte der Anmeldungen

Betrachtet man die über die einzelnen Tage kumulierten Anmeldungen (siehe das Schaubild), so ist festzustellen, dass die Anmeldungen zu den Monatsenden sprunghaft ansteigen. Auch im Hinblick auf einzelne Wochentage sind Häufungen von Anmeldungen zu verzeichnen. Diese Sprünge haben einerseits technisch-organisatorische Ursachen im Arbeitsablauf der Meldebehörden (z. B. Öffnungszeiten und die Verarbeitung von angesammelten Abmeldungen zentral zu einem bestimmten Zeitpunkt), sind aber auch in den Meldegewohnheiten der Bürger selbst begründet.

Diese Feststellungen sind in dreierlei Hinsicht von Bedeutung:

1. In Bezug auf die Festlegung des Stichtags für die Durchführung eines Zensus.



2. In Bezug auf Untersuchungen zur Diskrepanz von Datum der Anmeldung und Datum des Bezugs der Wohnung. Im Rahmen des Registertests sind die Fälle, die sich erst nach dem Stichtag angemeldet haben, vom Interviewer aber bereits an der neuen Adresse angetroffen wurden, als „Fehlbestände“ klassifiziert worden. Im Zuge der Vorbereitungen zur Durchführung eines Zensus wird zu prüfen sein, ob und inwieweit ein derartiger Datenbestand für Zwecke der Aufdeckung von „Fehlbeständen“ genutzt werden kann.
3. In Bezug auf die Interpretation der Ergebnisse des Registertests, auch im Hinblick auf angedachte zusätzliche primärstatistische Klärungen. Es stellt sich die Frage, ob ein Teil des Personenkreises, bei dem die Eigenschaft „Karteileiche bzw. Fehlbestand“ vermutet wird, bei einem flächendeckenden Verfahren, wie es der Test der Mehrfachfallprüfung im Unterschied zum Registertest ist, nicht längst gefunden ist.

In Hinblick auf weitere primärstatistische Klärungen, wie sie in anderen Projektteilen des Zensus-tests erörtert werden, stellt sich nicht die Frage, wie groß der Personenkreis ist, der durch die Mehrfachfallprüfung bereits geklärt ist, aber wegen der fehlenden Verzahnung des Registertests mit der Mehrfachfallprüfung als „primärstatistisch zu klären“ vorgeschlagen wird. Vielmehr ist es ein zentrales Problem, die Abgrenzung verschiedener primärstatistischer Klärungsmöglichkeiten im Zuge des Gesetzgebungsverfahrens zur Durchführung eines registergestützten Zensus zu begründen. Aus heutiger Sicht werden sich gewisse Überschneidungen bei den Erhebungen zunächst nicht vermeiden lassen. Daraus folgt auch, dass der erste registergestützte Zensus – zusätzlich zu seiner primären statistisch-inhaltlichen Aufgabe – auch im Zensus-test vom 5. Dezember 2001 offen gebliebene methodische Fragen klären muss.

4 Bewertung der Ergebnisse

Gemessen an den Ergebnissen der Mehrfachfallprüfung weisen die Melderegister für die Bevölkerung am Ort der Hauptwohnung im Bundesdurchschnitt eine Karteileichenrate von knapp 0,6% auf (siehe Tabelle 6). Der durchschnittliche Anteil der durch die Mehrfachfallprüfung geklärten Zahl der Karteileichen an der Bevölkerung gemäß Melderegister streut zwischen den Ländern nur unbedeutend. In den Flächenländern schwankt er zwischen 0,4 und 0,8%, ausgenommen Mecklenburg-Vorpommern mit 1,4%. In den Stadtstaaten liegt die Karteileichenrate zwischen 0,5% in Berlin und 0,7% in Hamburg.

Tabelle 6: Personen und Karteileichen nach Bundesländern und Gemeindegrößenklassen
Hochgerechnetes Ergebnis der Mehrfachfallprüfung vom 5. Dezember 2001
Bevölkerung am Ort der Hauptwohnung

Bundesland Gemeindegrößenklasse von ... bis unter ... Einwohnern	Personen im Melderegister	Aufgedeckte Karteileichen		Personen nach Mehrfachfall- prüfung
	1 000	%		1 000
Baden-Württemberg ..	10 479,1	45,3	0,4	10 433,8
Bayern	12 323,4	48,4	0,4	12 275,0
Berlin	3 349,2	17,2	0,5	3 332,0
Brandenburg	2 568,1	19,4	0,8	2 548,7
Bremen	655,3	3,9	0,6	651,4
Hamburg	1 730,1	11,5	0,7	1 718,6
Hessen	6 073,5	51,0	0,8	6 022,5
Mecklenburg- Vorpommern	1 739,3	24,4	1,4	1 714,9
Niedersachsen	7 946,1	63,5	0,8	7 882,6
Nordrhein-Westfalen .	18 060,6	94,1	0,5	17 966,5
Rheinland-Pfalz	4 048,8	27,1	0,7	4 021,7
Saarland	1 068,6	6,3	0,6	1 062,3
Sachsen	4 355,1	23,5	0,5	4 331,6
Sachsen-Anhalt	2 575,4	15,0	0,6	2 560,4
Schleswig-Holstein ...	2 818,7	22,7	0,8	2 796,0
Thüringen	2 389,1	12,2	0,5	2 376,9
Deutschland ...	82 180,3	485,5	0,6	81 694,8
unter 10 000	23 071,0	149,9	0,7	22 921,1
10 000 – 50 000 ...	26 928,1	153,3	0,6	26 774,8
50 000 – 800 000 ...	24 839,1	139,3	0,6	24 699,8
800 000 und mehr	7 342,0	43,0	0,6	7 299,0

Die Betrachtung der Karteileichenrate nach Gemeindegrößenklassen, das heißt der Anteile der Karteileichen an der gesamten Bevölkerung der Gemeinden in der jeweiligen Größenklasse, zeigt nicht die erwartete deutliche Tendenz der höheren Raten in größeren Gemeinden. Die Rate für die Gemeinden mit weniger als 10 000 Einwohnern liegt mit 0,7% – wenn auch nur unwesentlich – über der in Städten mit mehr als 800 000 Einwohnern, hier beträgt die Karteileichenrate 0,6%.

Das im Zensus-test umgesetzte Konzept der Mehrfachfallprüfung war nicht explizit auf die Ermittlung von Fehlbeständen ausgelegt, trotzdem konnten im Rahmen der Mehrfachfallprüfung als Nebeneffekt der Befragung der Dublettenfälle bundesweit hochgerechnet 20 167 Fehlbestände aufgedeckt werden. Eine höhere Aufdeckung von Fehlbeständen wird durch die bereits beschriebene Auswertung der zweiten Datenlieferung erwartet.

Bei Betrachtung der einzelnen Geburtstagsklassen erscheint hingegen das erwartete Bild hinsichtlich der Verteilung.

Tabelle 7: Personen und Karteileichen nach Bundesländern, Gemeindegrößenklassen und Geburtstagen
 Hochgerechnetes Ergebnis der Mehrfachfallprüfung vom 5. Dezember 2001
 Bevölkerung am Ort der Hauptwohnung

Bundesland Gemeindegrößenklasse von ... bis unter ... Einwohnern	Personen im Melderegister								
	mit vollständigem Geburtsdatum ohne die am 1. Januar Geborenen			mit Geburtstag am 1. Januar			mit unvollständigem Geburtsdatum		
	geprüfte Registersätze	Karteileichen		geprüfte Registersätze	Karteileichen		geprüfte Registersätze	Karteileichen	
	1 000	Anzahl	%	1 000	Anzahl	%	1 000	Anzahl	%
Baden-Württemberg	10 419,3	38 514	0,4	53,5	606	1,1	6,3	121	1,9
Bayern	12 266,6	34 027	0,3	50,0	391	0,8	6,8	108	1,6
Berlin	3 311,7	14 240	0,4	18,8	293	1,6	18,7	406	2,2
Brandenburg	2 560,2	13 372	0,5	7,6	59	0,8	0,4	/	/
Bremen	649,4	3 396	0,5	3,9	56	1,5	2,0	41	2,1
Hamburg	1 713,9	10 769	0,6	9,8	174	1,8	6,5	126	1,9
Hessen	6 025,9	37 966	0,6	36,1	510	1,4	11,4	243	2,1
Mecklenburg-Vorpommern	1 734,0	20 789	1,2	5,2	66	1,3	0,1	-	-
Niedersachsen	7 906,0	47 719	0,6	36,3	455	1,3	3,8	93	2,4
Nordrhein-Westfalen	17 937,1	74 272	0,4	94,5	1 221	1,3	29,0	753	2,6
Rheinland-Pfalz	4 029,7	15 351	0,4	16,9	177	1,0	2,2	50	2,2
Saarland	1 063,3	5 489	0,5	4,8	51	1,1	0,5	/	/
Sachsen	4 342,0	19 913	0,5	12,9	97	0,8	0,3	/	/
Sachsen-Anhalt	2 567,5	9 410	0,4	7,8	49	0,6	0,1	-	-
Schleswig-Holstein	2 806,7	17 464	0,6	10,7	146	1,4	1,3	/	/
Thüringen	2 382,4	9 160	0,4	6,6	33	0,5	0,1	/	/
Deutschland ...	81 715,6	371 852	0,5	375,1	4 384	1,2	89,6	1 998	2,2
unter 10 000	22 991,1	110 710	0,5	77,3	628	0,8	2,6	62	2,4
10 000 – 50 000	26 794,3	114 215	0,4	121,4	1 294	1,1	12,5	313	2,5
50 000 – 800 000	24 660,6	110 491	0,5	134,2	1 775	1,3	44,2	1 008	2,3
800 000 und mehr	7 269,5	36 436	0,5	42,3	687	1,6	30,3	615	2,0

lung der Karteileichen (siehe Tabelle 7). In der Klasse der Personen mit „vollständigem Geburtsdatum ohne die am 1. Januar Geborenen“ schwankt die Karteileichenrate um den Bundesdurchschnitt. Da es sich hier auch um die am stärksten besetzte Klasse handelt, wird der Durchschnitt durch diese geprägt. Deutlich höher ist die Karteileichenrate in der Klasse der Personen „mit Geburtstag am 1. Januar“. Im Vergleich zum Bundesdurchschnitt (mit vollständigem Geburtsdatum ohne die am 1. Januar Geborenen) mit 0,5% beträgt in dieser Klasse die Rate durchschnittlich 1,2%, wobei eine beträchtliche Streuung zwischen den Ländern zu beobachten ist. In der Klasse der Personen „mit unvollständigem Geburtsdatum“ beträgt die Karteileichenrate im Durchschnitt 2,2%; auch hier streuen die Anteile in den einzelnen Ländern erheblich, wobei zu einzelnen Ländern aufgrund zu niedriger Fallzahlen keine Aussage gemacht werden kann.

Tabelle 8 zeigt zusammengefasst den Wirkungsgrad der in der Mehrfachfallprüfung entwickelten und getesteten Verfahren. Diese Tabelle zeigt die besondere Bedeutung der zweiten Datenlieferung (MR2) und auch den Anteil der maschinell bereinigbaren Fälle an den primärstatistisch bereinigten Fällen. Die Tabelle lässt zudem einen Ausblick auf den zukünftigen primärstatistischen Aufwand bei einem registergestützten Zensus im Rahmen der Mehrfachfallprüfung zu.

Durch den Vergleich mit der zweiten Datenlieferung konnten bundesweit nach Hochrechnung 2 079 534 temporäre Mehrfachfälle geklärt werden, was einem Anteil von 2,4% am Gesamtbestand an Registerdatensätzen entspricht. Diese

„temporären Dubletten“ haben ihre Ursachen vor allem im Meldeverhalten der Bürger (innerhalb welcher Zeit melden sich die Bürger an bzw. um) und den gemeindespezifischen Verarbeitungszeiten. Der Anteil der „temporären Dubletten“ schwankt zwischen den einzelnen Flächenländern erheblich um den Bundesdurchschnitt von 2,4%, so beträgt der Anteil in Nordrhein-Westfalen nur 1,9%, aber in Schleswig-Holstein und Mecklenburg-Vorpommern über 4%. Das Ergebnis bei den Stadtstaaten liegt hingegen unter dem Bundesdurchschnitt, was auf die dortigen zentralen Melderegister zurückzuführen sein dürfte.

Im Rahmen der primärstatistischen Bereinigung wurden bundesweit hochgerechnet 523 456 Fälle überprüft, was 0,6% des Gesamtbestands an Registerdatensätzen entspricht. Parallel hierzu wurde versucht, durch variable Stringvergleiche die Mehrfachfälle maschinell zu klären. Durch einen Vergleich der Ergebnisse konnte festgestellt werden, dass es mit den entwickelten Instrumentarien möglich gewesen wäre, 95,0% der primärstatistisch bereinigten Fälle maschinell, das heißt ohne Belastung der Bürger und mit deutlich reduziertem Kostenaufwand, zu bereinigen. So blieben bei Anwendung der Instrumentarien zur maschinellen Bereinigung in einem künftigen Zensus nur rund 26 000 Fälle übrig, welche primärstatistisch bereinigt werden müssten, wobei auch hier noch eine Reduzierung aufgrund weiterer Forschungen zu erwarten ist.

Erwartungsgemäß liegt der Anteil der durch Vergleich der beiden Melderegisterabzüge bereinigten Fälle bei den Personen mit unvollständigem Geburtsdatum bzw. mit Geburtsdatum 1. Januar deutlich über dem Bundes-

Tabelle 8: Ergebnis der Mehrfachfallprüfung nach Bundesländern, Gemeindegrößenklassen, Art der Bereinigung und Stichprobenteilzugehörigkeit
Hochgerechnetes Ergebnis der Mehrfachfallprüfung vom 5. Dezember 2001

Land — Gemeindegrößenklasse von ... bis unter ... Einwohnern	Registersätze insgesamt (unabhängig vom Wohnungs- status)	Darunter						
		durch Vergleich mit der zweiten Datenlieferung MR2 maschinell bereinigte Fälle			primärstatistisch bereinigte Fälle		dar.: maschinell bereinigbare Fälle	
		Anzahl		%	Anzahl	%	Anzahl	% ¹⁾
		1	2	3	4	5	6	7
Insgesamt								
Baden-Württemberg	11 053 923	246 647	2,2	66 338	0,6	61 930	93,4	
Bayern	13 175 503	332 250	2,5	61 275	0,5	56 986	93,0	
Berlin	3 476 842	54 148	1,6	23 615	0,7	22 289	94,4	
Brandenburg	2 722 244	92 005	3,4	21 821	0,8	21 448	98,3	
Bremen	671 785	11 558	1,7	4 119	0,6	3 754	91,1	
Hamburg	1 763 506	37 410	2,1	15 461	0,9	14 809	95,8	
Hessen	6 380 146	137 023	2,2	43 490	0,7	40 527	93,2	
Mecklenburg-Vorpommern	1 842 233	77 345	4,2	12 370	0,7	12 266	99,2	
Niedersachsen	8 366 980	188 483	2,2	63 039	0,8	60 889	96,6	
Nordrhein-Westfalen	18 769 038	350 695	1,9	94 647	0,5	86 971	91,9	
Rheinland-Pfalz	4 259 687	89 955	2,1	22 519	0,5	21 454	95,3	
Saarland	1 128 992	45 252	4,0	8 185	0,7	8 025	98,0	
Sachsen	4 576 672	123 636	2,7	28 291	0,6	28 282	100,0	
Sachsen-Anhalt	2 691 417	90 713	3,4	21 174	0,8	21 170	100,0	
Schleswig-Holstein	3 027 172	136 274	4,5	21 691	0,7	21 171	97,6	
Thüringen	2 503 086	66 140	2,6	15 422	0,6	15 420	100,0	
Deutschland ...	86 409 226	2 079 534	2,4	523 456	0,6	497 393	95,0	
unter 10 000	24 369 055	590 398	2,4	141 185	0,6	138 116	97,8	
10 000 – 50 000	28 188 138	712 396	2,5	155 973	0,5	148 718	95,4	
50 000 – 800 000	26 166 774	603 887	2,3	168 173	0,6	156 261	92,9	
800 000 und mehr	7 685 259	172 853	2,2	58 125	0,8	54 298	93,4	
Personen mit vollständigem Geburtsdatum ohne die am 1. Januar Geborenen								
Baden-Württemberg	10 992 502	244 596	2,2	65 404	0,6	61 149	93,5	
Bayern	13 116 321	329 732	2,5	60 631	0,5	56 407	93,0	
Berlin	3 438 806	53 223	1,6	22 787	0,7	21 564	94,6	
Brandenburg	2 713 965	91 553	3,4	21 725	0,8	21 357	98,3	
Bremen	665 877	11 343	1,7	3 977	0,6	3 633	91,3	
Hamburg	1 747 128	37 082	2,1	15 154	0,9	14 538	95,9	
Hessen	6 331 641	135 433	2,1	42 589	0,7	39 770	93,4	
Mecklenburg-Vorpommern	1 836 729	77 024	4,2	12 329	0,7	12 229	99,2	
Niedersachsen	8 325 765	187 115	2,2	62 298	0,8	60 208	96,6	
Nordrhein-Westfalen	18 643 341	347 102	1,9	92 371	0,5	85 071	92,1	
Rheinland-Pfalz	4 239 894	89 374	2,1	22 239	0,5	21 216	95,4	
Saarland	1 123 464	44 807	4,0	8 090	0,7	7 944	98,2	
Sachsen	4 563 003	123 050	2,7	28 161	0,6	28 161	100,0	
Sachsen-Anhalt	2 683 232	90 309	3,4	21 099	0,8	21 099	100,0	
Schleswig-Holstein	3 014 560	135 548	4,5	21 479	0,7	20 984	97,7	
Thüringen	2 496 024	65 849	2,6	15 381	0,6	15 381	100,0	
Deutschland ...	85 932 252	2 063 140	2,4	515 713	0,6	490 713	95,2	
unter 10 000	24 285 813	587 386	2,4	140 344	0,6	137 353	97,9	
10 000 – 50 000	28 050 798	707 106	2,5	153 947	0,6	146 967	95,5	
50 000 – 800 000	25 984 191	597 920	2,3	164 765	0,6	153 371	93,1	
800 000 und mehr	7 611 450	170 728	2,2	56 657	0,7	53 022	93,6	

1) Anteil Spalte 6 an Spalte 4.

durchschnitt, ebenso der Anteil der primärstatistisch zu bereinigenden Fälle an der jeweiligen Schicht, während der Anteil der maschinell klärbaren Fälle unter dem Bundesdurchschnitt liegt. Ursache hierfür ist, dass der Anteil an ausländischen Bürgern in diesen beiden Schichten besonders hoch ist. Häufig ist diesem Personenkreis das genaue Geburtsdatum nicht bekannt, sodass in den Melderegistern nur das Geburtsjahr verzeichnet ist oder aber ein fiktives Geburtsdatum mit 1. Januar und einem Jahreszusatz eingesetzt wird. Gleichfalls gibt es bei diesem Personenkreis gehäuft Probleme mit der Feststellung der Identität, was zu differierenden Erfassungen der weiteren persönlichen

Merkmale bei den einzelnen Meldevorgängen führt. Daher ist hier auch der Erfolg der maschinellen Klärung von Mehrfachfällen deutlich geringer als im Bundesdurchschnitt.

Im Rahmen der Untersuchungen wurde die Trennschärfe unterschiedlicher Stringvarianten getestet mit der Zielsetzung, aufgedeckte Mehrfachfälle anhand der Meldedaten zu klären. In der Tabelle 9 ist der Wirkungsgrad der Stringvergleiche dargestellt. Um die Effektivität und Genauigkeit bewerten zu können, erfolgte – wie in Tabelle 8 abgebildet – parallel eine primärstatistische Bereinigung des Datenmaterials (=Vergleichsbasis).

noch Tabelle 8: Ergebnis der Mehrfachfallprüfung nach Bundesländern, Gemeindegrößenklassen, Art der Bereinigung und Stichprobenteilzugehörigkeit
Hochgerechnetes Ergebnis der Mehrfachfallprüfung vom 5. Dezember 2001

Land — Gemeindegrößenklasse von ... bis unter ... Einwohnern	Registersätze insgesamt (unabhängig vom Wohnungs- status)	Darunter						
		durch Vergleich mit der zweiten Datenlieferung MR2 maschinell bereinigte Fälle			primärstatistisch bereinigte Fälle		dar.: maschinell bereinigbare Fälle	
		Anzahl		%	Anzahl	%	Anzahl	% ¹⁾
		1	2	3	4	5	6	7
Personen mit Geburtstag am 1. Januar								
Baden-Württemberg	55 056	1 636	3,0	785	1,4	662	84,3	
Bayern	52 266	1 937	3,7	541	1,0	498	92,0	
Berlin	19 182	335	1,8	398	2,1	352	88,4	
Brandenburg	7 894	359	4,6	87	1,1	84	96,6	
Bremen	3 914	112	2,9	89	2,3	77	86,5	
Hamburg	9 840	206	2,1	195	2,0	177	90,8	
Hessen	36 984	1 140	3,1	616	1,7	531	86,2	
Mecklenburg-Vorpommern	5 371	307	5,7	41	0,8	37	90,2	
Niedersachsen	37 343	1 168	3,1	616	1,6	572	92,9	
Nordrhein-Westfalen	96 398	2 346	2,4	1 446	1,5	1 239	85,7	
Rheinland-Pfalz	17 523	480	2,7	219	1,2	193	88,1	
Saarland	4 978	412	8,3	76	1,5	65	85,5	
Sachsen	13 387	554	4,1	121	1,0	113	93,4	
Sachsen-Anhalt	8 119	394	4,8	74	0,9	70	94,6	
Schleswig-Holstein	11 235	621	5,5	186	1,7	167	89,8	
Thüringen	6 923	264	3,8	39	0,6	38	97,4	
Deutschland ...	386 413	12 271	3,2	5 529	1,4	4 875	88,2	
unter 10 000	80 569	2 810	3,5	775	1,0	712	91,9	
10 000 – 50 000	124 759	4 527	3,6	1 645	1,3	1 444	87,8	
50 000 – 800 000	137 816	3 863	2,8	2 260	1,6	1 964	86,9	
800 000 und mehr	43 269	1 071	2,5	849	2,0	755	88,9	
Personen mit unvollständigem Geburtsdatum								
Baden-Württemberg	6 365	415	6,5	149	2,3	119	79,9	
Bayern	6 916	581	8,4	103	1,5	81	78,6	
Berlin	18 854	590	3,2	430	2,3	373	86,7	
Brandenburg	385	93	24,2	/	/	/	/	
Bremen	1 994	103	5,2	53	2,7	44	83,0	
Hamburg	6 538	122	1,9	112	1,7	94	83,9	
Hessen	11 521	450	3,9	285	2,5	226	79,3	
Mecklenburg-Vorpommern	133	14	10,5	–	–	–	–	
Niedersachsen	3 872	200	5,2	125	3,2	109	87,2	
Nordrhein-Westfalen	29 299	1 247	4,3	830	2,8	661	79,6	
Rheinland-Pfalz	2 270	101	4,4	61	2,79	45	73,8	
Saarland	550	33	6,0	19	3,4	16	84,2	
Sachsen	282	32	11,4	/	/	/	/	
Sachsen-Anhalt	66	10	15,2	/	/	/	/	
Schleswig-Holstein	1 377	105	7,6	26	1,8	20	76,9	
Thüringen	139	27	19,4	/	/	/	/	
Deutschland ...	90 561	4 123	4,6	2 214	2,4	1 805	81,5	
unter 10 000	2 673	202	7,6	66	2,5	51	77,3	
10 000 – 50 000	12 581	763	6,1	381	3,0	307	80,6	
50 000 – 800 000	44 767	2 104	4,7	1 148	2,6	926	80,7	
800 000 und mehr	30 540	1 054	3,4	619	2,0	521	84,2	

1) Anteil Spalte 6 an Spalte 4.

Um herauszufinden, welche Kombination von Merkmalen ein optimales Ergebnis liefert, wurden alle Stringvarianten getestet und im Anschluss nach hierarchischen Gesichtspunkten gestaffelt. Zuerst wurden alle die Treffer aussortiert, welche die höchste Übereinstimmung hatten (Name, Vorname, Geschlecht, Geburtsort, Geburtsdatum). Durch dieses hierarchische Vorgehen reduzierte sich die Masse kontinuierlich, aber auch die Fehlerwahrscheinlichkeit nahm zu. Bei Bewertung der Zusammenführungsergebnisse wurde klar, dass nur ausgewählte Stringvarianten geeignet sind, Identitäten sicher festzustellen. Das Ergebnis ist in

Tabelle 9 dargestellt. Durch die in der Tabelle aufgeführten Stringvarianten konnten Personenidentitäten festgestellt werden, alle nicht aufgeführten Stringvarianten waren ohne Bedeutung, da sie entweder keine Treffer brachten bzw. Bestandteil eines anderen erfolgreichen Suchstrings waren und man so auf Mehrfachsuchen verzichten kann.

Bei den in der Tabelle 9 dargestellten Suchstrings ist zu unterscheiden zwischen denen „ohne bzw. mit niedriger Fehlerrate“ und denen „mit hoher Fehlerrate“. Im Rahmen weiterer Untersuchungen ist bis zu einem zukünftigen

Tabelle 9: Durch Stringvergleiche maschinell bereinigbare Mehrfachfälle

Hochgerechnetes Ergebnis der Mehrfachfallprüfung vom 5. Dezember 2001

Zusammenführung durch Verknüpfung der Merkmale	Maschinell zusammengeführte Mehrfachfälle	
	Anzahl	%
ohne bzw. mit niedriger Fehlerrate		
Name, Vorname, Geschlecht, Geburtsort, Geburtsdatum	225 609	43,10
Name, Vorname, Geschlecht, Geburtsort	9 699	1,85
Name, Vorname, Geschlecht, Geburtsdatum	231 236	44,17
Name, Geschlecht, Geburtsort, Geburtsdatum	29 393	5,62
Vorname, Geschlecht, Geburtsort, Geburtsdatum	1 455	0,28
Zusammen ...	497 393	95,02
mit hoher Fehlerrate		
Name, Vorname, Geschlecht	25 707	4,91
Name, Geschlecht, Geburtsdatum	358	0,07
Zusammen ...	26 064	4,98
Insgesamt ...	523 456	100

Zensus noch ein Instrumentarium zu entwickeln, mit dessen Hilfe der Geburtsort noch erfolgreicher standardisiert werden kann. Hier ist insbesondere an eine Datenbank gedacht, welche die verschiedenen Schreibweisen und eventuelle frühere Ortsnamen enthält. Bei den Untersuchungen des Einzelmateriale hat sich gezeigt, dass gerade übereinstimmende Geburtsorte in Verbindung mit Namen und Vornamen eine hohe Trefferwahrscheinlichkeit erzielen, insbesondere bei im Ausland geborenen Personen, wo häufig das Geburtsdatum variiert.

Als geeignet für eine Identifizierung von Dubletten mit hinreichender Wahrscheinlichkeit haben sich die Suchstrings aus der Rubrik „ohne bzw. mit niedriger Fehlerrate“ erwiesen. Hier kam es nur vereinzelt zu Fehlzusammenführungen und es konnten insgesamt 95,02% aller echten Dubletten zusammengeführt werden. Die verbleibenden 4,98% echter Dubletten konnten zwar auch maschinell über die Kombinationen von „Name, Vorname, Geschlecht“ oder „Name, Geschlecht, Geburtsdatum“ zusammengeführt werden; gleichzeitig kam es aber zu vielen falschen Dubletten, sodass auf diese maschinellen Zusammenführungen verzichtet werden muss bzw. derart zusammengeführte Dubletten manuell nachbearbeitet werden müssen. Vor dem Hintergrund der bundesweit zu erwartenden rund 26 000 Fälle erscheint dieses Verfahren nicht als zu aufwändig. Eine manuelle Sichtung bringt in der Regel schnell Klarheit über die Echtheit der Dublette bzw. ob eine Befragung notwendig ist.

5 Einordnung der Mehrfachfallprüfung in das Verfahren des künftigen Zensus

Die Anordnung von zwei Datenlieferungen hat sich bewährt und sollte auch beim Verfahren des künftigen Zensus Anwendung finden: Die Gemeinden liefern die im Zensusgesetz festgelegten Daten aller zum Zensusstichtag

gemeldeten Personen. Im Abstand von mindestens drei Monaten erfolgt eine zweite Datenlieferung über die zum zweiten Stichtag gemeldeten Personen.

- A. Die erste Datenlieferung stellt den bevölkerungsstatistischen Basiswert zur Ermittlung der amtlichen Einwohnerzahl für Bund, Länder und Gemeinden bereit (Basisbestand I).
- B. Unter Anwendung des bei der Mehrfachfallprüfung entwickelten Instrumentariums erfüllt die zweite Datenlieferung mehrere Zwecke:
 - a) Es werden die Personen festgestellt, die in der ersten und zweiten Datenlieferung mit unverändertem Wohnort enthalten sind, das heißt die Personen, für die sich zwischen den beiden Datenlieferungen keine Änderungen ergaben.
 - b) Es werden die Personen festgestellt,
 - die in der ersten Datenlieferung enthalten und in der zweiten Datenlieferung stichtagsrelevant mit neuem Wohnort gemeldet sind;
 - die in der ersten Datenlieferung enthalten und in der zweiten Datenlieferung nicht mehr enthalten sind (z. B. durch Tod, Fortzug ins Ausland).
 - c) Es werden die Personen festgestellt, die ausschließlich in der zweiten Datenlieferung enthalten sind (z. B. durch Geburt, Zuzug aus dem Ausland).

Die Feststellung der Stichtagsrelevanz in b) und c) erfolgt über das Datum des Beziehens der Wohnung.

- C. Aufgrund der unter B durchgeführten Streichungen und Hinzufügungen der Datenbestände ergeben sich für die Gemeinden die neuen angepassten Basiswerte (Basisbestand II). Wie im empirischen Teil der Ausführungen gezeigt wurde, sind die unsaldierten Summen dieser Anpassungen beträchtlich.
- D. Im Basisbestand II werden die Hauptwohnungs-/Nebenwohnungsverhältnisse auf korrespondierende Einträge weitgehend maschinell überprüft. Unplausibilitäten werden geklärt.
- E. Die erste Datenlieferung wird auf Dubletten geprüft. Die Dubletten gehen in die Befragung mit Ausnahme der Dubletten, die über die aus der zweiten Datenlieferung gewonnenen Erkenntnisse [Schritt B. b)] bzw. über maschinelle Verfahren aufgelöst werden können. Die Befragung der Dubletten stellt Karteileichen und Fehlbestände fest, die aus dem Basisbestand II (Wohnsitz gemäß erster Datenlieferung) herausgelöst bzw. ihm hinzugefügt (Wohnsitz gemäß zweiter Datenlieferung) werden. Aus den Schritten D und E entsteht Basisbestand III.
- F. Der Basisbestand III wird über die Instrumente Haushaltegenerierung und die sich in ihrer Folge ergebenden empirischen Klärungen weiterentwickelt. Es entsteht die amtliche Bevölkerungszahl für Bund, Länder und Gemeinden. [uu](#)

Auszug aus Wirtschaft und Statistik

© Statistisches Bundesamt, Wiesbaden 2004

Für nichtgewerbliche Zwecke sind Vervielfältigung und unentgeltliche Verbreitung, auch auszugsweise, mit Quellenangabe gestattet. Die Verbreitung, auch auszugsweise, über elektronische Systeme/Datenträger bedarf der vorherigen Zustimmung. Alle übrigen Rechte bleiben vorbehalten.

Herausgeber: Statistisches Bundesamt, Wiesbaden

Schriftleitung: Johann Hahlen
Präsident des Statistischen Bundesamtes
Verantwortlich für den Inhalt:
Brigitte Reimann,
65180 Wiesbaden

- Telefon: +49 (0) 6 11/75 20 86
- E-Mail: wirtschaft-und-statistik@destatis.de

Vertriebspartner: SFG Servicecenter Fachverlage
Part of the Elsevier Group
Postfach 43 43
72774 Reutlingen
Telefon: +49 (0) 70 71/93 53 50
Telefax: +49 (0) 70 71/93 53 35
E-Mail: destatis@s-f-g.com

Erscheinungsfolge: monatlich



Allgemeine Informationen über das Statistische Bundesamt und sein Datenangebot erhalten Sie:

- im Internet: www.destatis.de

oder bei unserem Informationsservice
65180 Wiesbaden

- Telefon: +49 (0) 6 11/75 24 05
- Telefax: +49 (0) 6 11/75 33 30
- E-Mail: info@destatis.de