

Gesis - Leibniz-Institut für Sozialwissenschaften

Felix Loewe

6. Juni 2013

Inhaltsverzeichnis

1	Einleitung - Vorstellung Gesis	2
2	Vorstellung ALLBUS - Grundkonzeption	2
3	Methodenbericht ALLBUS 2010	2
3.1	Das Frageprogramm des ALLBUS 2010	2
3.1.1	Replikationsfragen	3
3.1.2	Schwerpunktthema: Egozentrierte Netzwerke	3
3.2	Stichprobenverfahren des ALLBUS 2010	4
3.2.1	Die Grundgesamtheit	4
3.2.2	Erste Ziehungsstufe: Auswahl der Gemeinden	5
3.2.3	Zweite Ziehungsstufe: Auswahl der Zielpersonen in den Ge- meinden	6
3.2.4	Die Bildung der Stichprobe aus den gezogenen Personenadressen	8
3.2.5	Gewichtung auf Personenebene	9
3.2.6	Gewichtung auf Haushaltsebene	10
3.2.7	Vergleich von ADM-Stichproben mit Registerstichproben . . .	11
3.2.8	Das ADM-Stichprobenverfahren	11
3.2.9	Unterschiede zwischen Register- und ADM-Stichproben	12
3.2.10	Datenqualität der Stichprobenverfahren	12
4	Schluss	14

1 Einleitung - Vorstellung Gesis

2 Vorstellung ALLBUS - Grundkonzeption

3 Methodenbericht ALLBUS 2010

3.1 Das Frageprogramm des ALLBUS 2010

In der Regel besteht das Frageprogramm des ALLBUS aus einem thematischen Hauptteil, einen Demographie- und einem Replikationsteil. Die thematischen Schwerpunkte werden in regelmäßigen Abständen wiederholt, um einem Hauptaugenmerk des ALLBUS, der Untersuchung des sozialen Wandels, zuträglich zu sein. Da bei jeder ALLBUS Umfrage neben der Wiederholung alter Fragen auch neue Konstrukte dazukommen, ist es von Zeit zu Zeit nötig, eine ALLBUS Umfrage mit dem Schwerpunkt Replikation zu erheben. Dies war bisher anlässlich der Wiedervereinigung 1991, zur Jahrtausendwende 2000 und zum 30-jährigen Jubiläum des ALLBUS 2010 der Fall. Durch die Wiederholung von Frageblöcken sind für Westdeutschland Zeitreihen bis 1980 und im Osten bis 1991 entstanden, die als Grundlage für die Untersuchung von sozialem Wandel in Deutschland dienen können. Durch die getrennten Stichproben in Ost und West kann zudem der Prozess des Zusammenwachsens der beiden deutschen Teilstaaten analysiert werden. Ein besonderes Augenmerk liegt beim ALLBUS 2010 auf der Wiederholung des thematischen Schwerpunktes „egozentrierte Netzwerke“. Die Fragen hierzu nehmen gut 20 Minuten der insgesamt 50 Minuten Befragungszeit ein. Sonstige Replikation und Fragen zur Demographie beanspruchen jeweils 15 Minuten. Im Anschluss an die ALLBUS-Fragen werden die Befragte gebeten, zusätzlich einige Fragen aus der regelmäßigen Umfrage ISSP zu beantworten, die ein internationales Pendant zum ALLBUS in Deutschland darstellt. Auf deren Inhalt soll in dieser Arbeit nicht genauer eingegangen werden.

3.1.1 Replikationsfragen

Wie oben bereits erwähnt, werden Fragen im ALLBUS regelmäßig wiederholt, mit dem Ziel, eine Datengrundlage für die Erforschung sozialen Wandels zu schaffen. 2010 wurde der Replikation anlässlich des 30-jährigen Jubiläums des ALLBUS besondere Bedeutung beigemessen. Die Auswahl der Replikationsfragen erfolgte dabei in zwei Schritten. Durch eine Vorauswahl wurden 46 Fragen ausgesucht, von denen dann die Nutzer des ALLBUS in einem Onlinevoting 25 Fragen bestimmen konnten, die im ALLBUS 2010 verwendet werden sollten. Die Vorauswahl wurde nach bestimmten Kriterien getroffen. Es sollten möglichst viele Themenbereiche berücksichtigt werden, die schon seit dem Beginn der 80er verwendet werden, um auch langfristigen Wandel abbilden zu können. Auch Fragen vom Beginn der 90-er sind besonders interessant, da sie eine Analyse des Prozesses des Zusammenwachsens von Ost- und Westdeutschland ermöglichen. Zudem wurden Fragen, die besonders häufig von Nutzern der ALLBUS-Daten verwendet oder publiziert wurden, eher berücksichtigt als solche, die nur selten Verwendung fanden. Wie in fast jeder ALLBUS Umfrage kommen Fragen zum politischen Interesse und zur subjektiven Schichteinstufung hinzu

3.1.2 Schwerpunktthema: Egozentrierte Netzwerke

Das Schwerpunktthema wurde bereits 1980, 1990 und 2000 erhoben. 2010 wird es nun umfassender als bisher behandelt. Bei der Untersuchung von sozialen Netzwerken wird der Befragte (ego) in der Regel gebeten, seine wichtigsten sozialen Kontakte (alteri) zu benennen, zu denen dann im folgenden nähere Informationen (Interpretatoren) erfragt werden. 2010 wurden zum einen mehr Interpretatoren der alteri erfasst und zum anderen im Split eine alternativer international besser vergleichbarer Namensgenerator aus den USA verwendet. So wurde die Hälfte der Befragten gebeten, drei Personen außerhalb des Haushaltes zu benennen, mit denen sie am häufigsten privat zusammen ist. Die andere Hälfte sollte diejenigen Personen nennen, mit denen ego in den letzten 6 Monaten über Dinge gesprochen hat, die ihr wichtig waren. Durch die Verwendung beider Fragetechniken wurde sowohl die Fortschreibung der ALLBUS-internen Zeitreihe, als auch die Anknüpfung an internationale Umfragen ermöglicht.

Bei beiden Namensgeneratoren wurden in der Folge die selben Interpretatoren erfasst: Zum einen die zentralen Merkmale der alteri wie zum Beispiel Geschlecht und Alter, sowie der Typ der Beziehung (Partner, Freund, Nachbar, Arbeit, Verwandtschaft). Zum anderen die neuen Interpretatoren zu Quantität und Qualität der Kontakte. So wurde der Migrationshintergrund, die Schulbildung, die wirtschaftliche Lage, die räumliche Entfernung und die Kontakthäufigkeit erfragt. Erstmals wurden auch negative Aspekte der Beziehung zwischen ego und alteri in den Datensatz aufgenommen.

3.2 Stichprobenverfahren des ALLBUS 2010

Das heutige registergestützte Stichprobenverfahren findet beim ALLBUS seit 1992 Verwendung. Davor und aus Kostengründen auch 1998 wurde das sogenannte ADM-Stichprobendesign zu Erhebung der Daten benutzt.

Dieser Abschnitt beleuchtet detailliert das Stichprobenverfahren beim ALLBUS 2010 und stellt seine Vorteile gegenüber dem ADM-Stichprobendesign heraus.

3.2.1 Die Grundgesamtheit

Die Grundgesamtheit des ALLBUS 2010 besteht aus allen in der Bundesrepublik in Privathaushalten wohnenden Personen die vor dem 1.1.1992 geboren wurden, also zu Jahreswechsel 2009/2010 das 18. Lebensjahr vollendet hatten. Sowohl deutsche Staatsbürger als auch Ausländer gehören der Grundgesamtheit an, wobei Personen mit mangelnden Deutschkenntnissen zu den systematischen Ausfällen zählen, da sie nicht interviewt werden können. Da das Auswahlverfahren auf dem Einwohnermelderegister beruht, kommen nur Personen in die Stichprobe, die in Deutschland ihren ersten Wohnsitz gemeldet haben.

3.2.2 Erste Ziehungsstufe: Auswahl der Gemeinden

Beim ALLBUS 2010 wurde zunächst eine Schichtung der Gemeinden nach Bundesländern, Regierungsbezirken, Kreisen und den BIK-Stadtregionen ¹ vorgenommen. In einem zweiten Schritt werden sogenannte Sample-Points auf die Schichten verteilt. Dabei entspricht ein Sample-Point einer festgelegten Anzahl von Adressen. Die Wahrscheinlichkeit, dass ein Sample-Point auf eine Schicht entfällt, ist proportional zum Schichtgewicht. Das Schichtgewicht besteht aus der Anzahl der erwachsenen Einwohner der Gemeinden in einer Schicht, und wird so normiert, dass die Summe aller Schichtgewichte die angestrebte Anzahl an Sample-Points ergibt. Beim ALLBUS 2010 wurden im Westen 111 und im Osten 51 Sample Points vergeben. Das Schichtgewicht stellt dadurch die Anzahl der Gemeinden da, die aus einer Schicht gezogen werden sollen. Durch die Normierung ergeben sich aber Dezimalbrüche als Auswahlwahrscheinlichkeiten für die Zellen. Dadurch wird eine Rundung nötig, die durch ein Computerprogramm mitm folgender Systematik vollzogen wird:

- Rundung mit Wahrscheinlichkeit proportional zu den Nachkommastellen
- wiederholte Anpassung der Zellenbesetzung an die jeweiligen Randsummen
- Begrenzung der Differenz zwischen gebrochener (originaler) Zellenbesetzung und ganzzahliger Zellenbesetzung nach der Allokation auf Werte kleiner 1. (Koch, 1997)

Das Verfahren stellt sicher, dass nach der Auswahl die Verteilung der Gemeinden in der Stichprobe hinsichtlich Bundesland, Regierungsbezirk, Kreis und BIK-Typ mit der Grundgesamtheit praktisch übereinstimmt.(Wasmer et al., 2012) Die Allokation (Verteilung) der Sample-Points erfolgt in zwei Stufen.

Zum Beispiel wurden beim ALLBUS 2010 auf der ersten Stufe der Allokation Schleswig-Hollstein bei einem Schichtgewicht von 4,64 fünf Sample-Points zugewiesen. Dies stand nun beim folgenden Schritt als Eckwert fest. Im zweiten Schritt

¹Zu einer BIK-Region gehören die Gemeinden, deren Pendlerpriorität (Arbeitsort der größten Fraktion der Auspendler) auf eine gemeinsame Kernstadt gerichtet ist, wenn die Pendlerquote mindestens 7% beträgt, also mindestens 7% der Wohnbevölkerung als sozialversicherungspflichtig Beschäftigte in diese Kernstadt einpendeln.

Betrag nun das Schichtgewicht der Zelle „Gemeinde(n) vom BIK-Typ 10 im Kreis 'Pinneberg' “ 0,311. Die kontrollierte Zufallsauswahl sorgte dafür, dass dieser Schicht entweder ein oder kein Sample-Point zugewiesen wurde. (vgl. Abbildung 1 1)

Innerhalb einer Schicht erfolgte nun die Gemeindeauswahl zufällig mit einer Wahrscheinlichkeit proportional zu ihrer Einwohnerzahl.

3.2.3 Zweite Ziehungsstufe: Auswahl der Zielpersonen in den Gemeinden

Nachdem die Gemeinden wie oben beschrieben ausgewählt wurden, werden in einer zweiten Ziehungsstufe für jede Gemeinde pro zugeordnetem Sample-Point 92 Adressen gezogen, um genügend Spielraum für Basisstichprobe, Aufstockungsstichprobe und Ersatzadressen zu haben. Dazu wird von jeder Gemeinde eine numerierte Adressenliste benötigt. Die Zufallsauswahl geschieht dann durch eine Intervallziehung. Die Intervalllänge l entspricht dem abgerundeten Quotienten aus der Anzahl der Einwohner einer Gemeinde und der Anzahl der benötigten Adressen:

$$l = \lfloor \frac{n_{\text{Einwohner}}}{n_{\text{Adressen}}} \rfloor$$

Die Startzahl x_{start} ergibt sich aus der abgerundeten Hälfte der Intervalllänge:

$$x_{\text{start}} = \lfloor l/2 \rfloor$$

Die Adresse mit der Nummer x_{start} wird somit als erstes gezogen. Die zweite Adresse hat die Nummer $x_{\text{start}} + l$ und die folgenden ergeben sich entsprechend durch wiederholte Addition der Intervalllänge l .

Die disproportionale Ziehung in Ost und West und der Zweiteilung Berlins machten es notwendig, auf Ebene der Stadtteile Adressen zu generieren. Hierbei wurden pro Sample-Point zufällig und proportional zur Bevölkerung vier Stadtteile gezogen aus denen jeweils 23 Adressen bestellt wurden.

In größeren Gemeinden wurden, um den Interviewern ihre Arbeit zu erleichtern, ein überhöhter Adresssansatz bestellt - in Gemeinden zwischen 100.000 und 500.000 Einwohnern ein doppelter, und in Gemeinden mit mehr als 500.000 Einwohnern ein

Übersicht 9: Ausschnitt aus dem Schichttableau für den ALLBUS 2010: Normierte Schichtgewichte* für Schleswig-Holstein und ein Beispiel für eine mögliche Verteilung der Sample Points (aus Datenschutzgründen enthält das Beispiel nicht die tatsächlich realisierte Aufteilung der Sample Points)

Kreis	BIK-TYP										Summe
	Einwohner										
	1	2	3	4	5	6	7	8	9	10	
	unter 2.000	2.000 - 5.000	5.000 - 20.000	20.000- 50.000	50.000- 100.000 (Rand)	50.000- 100.000 (Kern)	100.000- 500.000 (Rand)	100.000- 500.000 (Kern)	500.000 und mehr (Rand)	500.000 und mehr (Kern)	
Schleswig-Holstein											
KS Flensburg	---	---	---	---	---	---	0,147	0,402	---	---	0,147
KS Kiel	---	---	---	---	---	---	0,354	---	---	---	0,402
KS Lübeck	---	---	---	---	---	---	0,126	---	---	---	0,354
KS Neumünster	---	---	---	---	---	---	---	---	---	---	0,126
Dithmarschen	0,039	0,030	0,023	0,128	---	---	---	---	---	---	0,220
Herzogtum Lauenburg	0,012	---	0,028	0,041	---	0,022	---	0,085	0,114	---	0,301
Nordfriesland	0,056	0,051	0,076	---	0,049	---	---	---	---	---	0,269
Ostholstein	0,014	0,011	0,115	0,102	---	0,067	---	---	---	---	0,342
Pinneberg	0,005	---	---	---	---	---	---	---	---	---	---
Plön	0,015	0,010	0,038	0,003	---	0,092	0,064	---	---	---	0,492
Rendsburg-Eckernförde	0,037	0,013	---	0,062	0,074	0,168	0,020	---	---	---	0,438
Schleswig-Flensburg	0,036	0,009	0,035	0,000	0,056	0,142	---	---	---	---	0,319
Segeberg	0,013	---	---	0,074	---	0,022	---	---	0,138	0,151	0,418
Steinburg	0,024	0,018	0,025	0,000	0,076	0,054	---	---	0,019	---	0,217
Stormarn	---	---	---	---	---	0,012	---	---	0,213	0,143	0,368
Summe	0,251	0,162	0,341	0,409	0,255	0,196	0,525	1,147	0,631	0,719	4,636

* Gewicht ist proportional zur Zahl der Bevölkerung ab 18 Jahre am Ort der Hauptwohnung am 31.12.2008

1 = Anzahl der Sample Points

Abbildung 1: Ausschnitt aus dem Schichttableau für den ALLBUS 2010 (aus Wasmer et al. (2012))

dreifacher Ansatz. Vor der Ziehung der Personenstichproben wurden hier die Adressen mittels Gemeindeteil, Postleitzahl und Adresse in 8 beziehungsweise 12 räumliche Cluster aufgeteilt und jeweils 4 Cluster pro Sample Point zufällig gezogen. Nur Adressen aus den gezogenen Clustern konnten für die Personenstichprobe gezogen werden.

Letzendlich erhält man für jeden Sample-Point eine Adressliste mit jeweils 92 Adressen.

3.2.4 Die Bildung der Stichprobe aus den gezogenen Personenadressen

Die gezogenen Adressen wurden auf ihre Vollständigkeit überprüft und unbrauchbare Adressen wurden aussortiert. (z.B doppelte Adresse, keine Privatadresse). Aus den restlichen Adressen wurde eine Stichprobe von 80 Adressen gezogen, in der die Personen die Randverteilungen der amtlichen Statistik für die Merkmale Alter und Geschlecht widerspiegeln sollten. Nach dem gleichen Verfahren wurden nochmal 40 Adressen pro Sample-Point ausgewählt, die die Basisstichprobe bilden und in einen weiteren Schritt 12 Adressen, die als QNA-Ersatzadressen² dienen sollten. Die Adressen wurden nach Ortsteil und Postleitzahl sortiert und pro Sample-Point auf 4 Klumpen mit je 10 Einsatz- und 3 Ersatzadressen aufgeteilt. Dadurch wurden für den Fall, dass mehrere Interviewer in einer Gemeinde arbeiteten, die Entfernung zwischen den Zieladressen verringert.

Die restlichen 28 Adressen pro Sample-Point bildeten die Reservestichprobe, aus der im späteren Verlauf der Erhebung die Aufstockungsstichprobe gezogen wurde. Auf Basis der Erfahrungen früherer ALLBUS-Umfragen und dem bisherigen Feldverlauf der Basisstichprobe wurde der Umfang der Aufstockungsstichprobe so bestimmt, dass die angestrebte Fallzahl von 3000 Fällen (2000 West, 1000 Ost) möglichst genau erreicht wurde.

Letzendlich erhielt man insgesamt 2827 verwertbare Interviews, davon 1960 aus dem Westen und 867 aus dem Osten.

²QNA = Qualitätsneutraler Ausfall

3.2.5 Gewichtung auf Personenebene

Wie in jeder ALLBUS-Erhebung seit 1999 wurde Ostdeutschland überrepräsentiert, um eine ausreichend große Fallzahl auch für Analysen einzelner Bevölkerungsgruppen in Ostdeutschland zu erhalten. Bei Analyse der Daten, getrennt nach Ost und West, stellt dies kein Problem dar. Wenn man nun aber mit den erhaltenen Daten eine Auswertung für Gesamtdeutschland durchführen will, ist dadurch eine Gewichtung notwendig, um die disproportionalität der Teilstichproben auszugleichen. Eine Gewichtung verändert allgemein gesprochen die relative Wichtigkeit der Befragten. Ein Gewicht ist ein multiplikativer Faktor, der auf einen oder mehrere Befragte einer Umfrage angewendet werden kann, um die Verteilung der Stichprobe an die wahre Verteilung in der Grundgesamtheit anzugleichen. (Gabler, 1994) Als Referenz für das wahre Verhältnis von erwachsenen Einwohnern von Ost- und Westdeutschland verwendet man Daten des Mikrozensus die möglichst zeitnah zu den Daten des ALLBUS erhoben wurden.

Bei Analyse auf Personenebene ist dazu entsprechend Tabelle 1 die Anzahl der Personen über 18 in West- und Ost-Deutschland zu verwenden.

Tabelle 1: Zahlen für die Ost-West-Gewichtung auf Personenebene: Personen in Privathaushalten über 18 - Mikrozensus 2009 und ALLBUS 2010

Mirkosensus 2009 (in Tausend)			ALLBUS 2010		
West (N_W)	Ost (N_O)	Gesamt (N)	West (n_W)	Ost (n_O)	Gesamt (n)
55412	12367	67779	1960	867	2827

Die Gewichte für Analysen der erwachsenen Bevölkerung in Privathaushalten in ganz Deutschland lassen sich dann wie folgt berechnen:

- Ostdeutschland: $\frac{n}{n_O} \cdot \frac{N_O}{N} = \frac{2827}{867} \cdot \frac{12367}{67779} = 0,59494376$
- Westdeutschland: $\frac{n}{n_W} \cdot \frac{N_W}{N} = \frac{2827}{1960} \cdot \frac{55412}{67779} = 1,17917539$

Die Anzahl der gewichteten Beobachtungen summiert sich dadurch wieder zu n :

$$n_O \cdot \frac{n}{n_O} \cdot \frac{N_O}{N} + n_W \cdot \frac{n}{n_W} \cdot \frac{N_W}{N} = \frac{n(N_O + N_W)}{N} = n \cdot \frac{N}{N} = n$$

Eine entsprechende Gewichtsvariable ist im Datensatz des ALLBUS 2010 enthalten.

3.2.6 Gewichtung auf Haushaltsebene

Der ALLBUS 2010 ist eine Personenstichprobe, was zur Folge hat, dass Personen zwar die gleiche Auswahlwahrscheinlichkeit besitzen, in die Stichprobe zu gelangen, Haushalte jedoch nicht die gleichen Chancen haben. Durch das Design werden Haushalte mit vielen Bewohnern wahrscheinlicher in die Stichprobe aufgenommen als Haushalte mit entsprechend weniger Bewohnern. Um diesen Effekt bei einer Analyse auf Haushaltsebene zu berücksichtigen, empfiehlt es sich eine Transformationsgewichtung vorzunehmen.

Das entsprechende Gewicht beruht auf der „reduzierten Haushaltsgröße“ i . Sie bezeichnet die Anzahl der vor dem 1.1.1992 geborenen Personen im Haushalt, also derjenigen, die zur Grundgesamtheit des ALLBUS 2010 gehören. Zu seiner Berechnung verwendet man den Kehrwert w der entsprechenden Haushaltsgröße i ($w_i = 1/i$). Dadurch resultieren Gewichte kleiner 1, was in der Praxis zu einer ungewünschten Reduzierung der Fallzahl gegenüber den ungewichteten Daten führen würde. n_i bezeichnet im folgenden die Anzahl der Personen die in einem Haushalt mit der reduzierten Haushaltsgröße i wohnen.

$$n_{gew} = \sum_i n_i \cdot w_i$$

Um dies zu verhindern, normiert man das Gewicht indem man durch den mittleren Kehrwert $\bar{w} = 1/n \cdot \sum n_i \cdot w_i$ teilt.

$$w_i^* = \frac{w_i}{\bar{w}} = \frac{n}{\sum_i n_i \cdot w_i} \cdot w_i$$

$$n_{gew}^* = \sum_i n_i \cdot w_i^* = \sum_i n_i \cdot \frac{n \cdot w_i}{\sum_i n_i \cdot w_i} = n \cdot \sum_i \frac{n_i \cdot w_i}{\sum_i n_i \cdot w_i} = n$$

Um Analysen auf Haushaltsebene über Gesamtdeutschland durchzuführen muss zusätzlich analog zu der Gewichtung auf der Ebene der Privatpersonen auch bei den Haushalten mit Hilfe von Tabelle 2 ein Ausgleich der Disproportionalität zwischen Ost und West erfolgen.

Tabelle 2: Zahlen für die Ost-West-Gewichtung auf Haushaltsebene: Privathaushalte mit Personen über 18 - Mikrozensus 2009 und ALLBUS 2010

Mikrozensus 2009 (in Tausend)			ALLBUS 2010		
West (N_W)	Ost (N_O)	Gesamt (N)	West (n_W)	Ost (n_O)	Gesamt (n)
32297	7367	39664	1902	846	2748

- Ostdeutschland: $\frac{n}{n_O} \cdot \frac{N_O}{N} \cdot w_i = \frac{2748}{846} \cdot \frac{7367}{39664} \cdot w_i = 0,60331000 \cdot w_i$
- Westdeutschland: $\frac{n}{n_W} \cdot \frac{N_W}{N} \cdot w_i = \frac{2748}{1902} \cdot \frac{32297}{39664} \cdot w_i = 1,17644571 \cdot w_i$

3.2.7 Vergleich von ADM-Stichproben mit Registerstichproben

Wie oben bereits angedeutet wurden die ALLBUS-Daten bis 1991 mit dem ADM-Stichprobenverfahren erhoben, bevor man 1992 auf das wesentlich teurere und methodisch anspruchsvollere Registerstichproben-Verfahren wechselte, von dem man sich eine höhere Datenqualität erwartete. Im folgenden Abschnitt sollen nach einer kurzen Beschreibung des ADM-Verfahrens die wesentlichen Unterschiede zwischen den Verfahren sowie ihre Vor- und Nachteile vorgestellt werden.

3.2.8 Das ADM-Stichprobenverfahren

Das Stichprobendesign des Arbeitskreises deutscher Marktforschungsinstitute (ADM) findet in den meisten sozialwissenschaftlichen Umfragen in Deutschland Verwendung. (Blohm, 2006) Das Verfahren besteht aus 3 Auswahlstufen. In der ersten Stufe wird eine Stichprobe von Wahlkreisen gezogen. Danach erfolgt eine Auswahl von Haushalten und als letztes die Auswahl einer Zielperson aus diesem Haushalt. Die

Auswahl der Haushalte erfolgt in der Regel durch das Random-Route oder das Adress-Random Verfahren. Bei beiden erhält der Interviewer einen zufälligen Startpunkt, von dem aus er nach festgelegten Begehungsregeln Haushalte auswählt. Beim Random-Route Verfahren geschieht die Auswahl der Zielperson im selben Schritt, während beim Adress-Random aus der gelieferten Adressliste nochmals eine Stichprobe gezogen und dem Interviewer vorgegeben wird.

3.2.9 Unterschiede zwischen Register- und ADM-Stichproben

Die Grundgesamtheit besteht beim ADM-Verfahren aus der wahlberechtigten Bevölkerung in Privathaushalten, während bei einem registergestützten Verfahren alle Einwohner, also Deutsche wie Ausländer, in die Stichprobe gelangen können. Im folgendem können also nur die Daten von erwachsenen Deutschen der Stichproben verglichen werden. Des weiteren hat der Interviewer bei einem ADM-Verfahren mehr Einfluss auf die Auswahl der Haushalte und der Zielpersonen. Es kann nicht ausgeschlossen werden, dass der Interviewer sich trotz fester Regeln bevorzugt leicht zugängliche Haushalte und Personen aussucht, von denen er ein angenehmes Interview erwartet. Auch kann durch die Begehungsregeln nicht jede Eventualität bei der Begehung berücksichtigt werden. Bei Registerstichproben hingegen werden dem Interviewer Adresse und Zielperson vorgegeben, so dass sich keine Spielräume bei deren Auswahl ergeben. In der Theorie ergeben beide Verfahren gleich gute unverzerrte Stichproben. Wie bei Blohm (2006) kann man jedoch vermuten, dass sich dieser Spielraum des Interviewers in der Praxis negativ auf die Qualität der Daten auswirken kann. Demnach ist bei einer Registerstichprobe die höchste Datenqualität und bei einer ADM-Stichprobe mit Random-Route Verfahren die geringste Datenqualität zu erwarten. Eine ADM-Stichprobe mit Adress-Reandom sollte sich demnach dazwischen befinden.

3.2.10 Datenqualität der Stichprobenverfahren

Blohm (2006) empfiehlt die Qualität der Stichproben anhand eines Vergleichs der Nettostichproben mit dem Mikrozensus zu beurteilen. Da die Teilnahme am Mi-

krozensus verpflichtend ist, ist hier kein systematisch Verzerrender Effekt von Non-Response zu erwarten. Mit ihm lassen sich die Verteilungen der Merkmale Geschlecht, Alter, Bildung, Familienstand, Haushaltsgröße und Stellung im Erwerbsleben vergleichen. Als Maßzahl für die Datenqualität wird der Dissimilaritätsindex D verwendet. Die Maßzahl gibt an, wie viel Prozent einer Stichprobe die Kategorien wechseln müssten, damit es zu einer gleichen Verteilung der Kategorien in beiden Stichproben kommt.

$$D = \frac{1}{2} \sum_{i=1}^n \left| \frac{A_i}{A} - \frac{M_i}{M} \right|$$

Hierbei entspricht n der Anzahl der Kategorien, A und M der Anzahl der Personen der jeweiligen Stichproben und A_i beziehungsweise M_i der Anzahl der Personen der Stichprobe A beziehungsweise M in der jeweiligen Kategorie i . Zur Beurteilung der Verfahren berechnet Blohm (2006) für 26 verschiedene Stichproben und die sechs vorhin genannten soziodemographischen Variablen den jeweiligen Dissimilaritätsindex. Bei Registerstichproben ergibt sich für D ein mittlerer Wert von 3,2. Wie vermutet ist der Wert bei Random-Route-Verfahren mit durchschnittlich 5,7 höher und das Verfahren mit Random-Adress bewegt sich mit durchschnittlich 4,7 wie erwartet in der Mitte.

Die durchschnittliche Datenqualität ist bei Registerstichproben scheint also tatsächlich besser zu sein, was die Verwendung des teureren und aufwendigeren Verfahrens beim ALLBUS 2010 bestätigt. Registerstichprobe sind durch Gebühren für die bestellte Adresslisten wesentlich teurer als ADM-Stichproben. Zudem dürfen Daten aus dem Einwohnermelderegister in Deutschland per Gesetz nur herausgegeben werden, wenn die entsprechende Forschung im „öffentlichen Interesse“ liegt. Dadurch scheidet die Methode der Registerstichprobe für die meisten kommerziellen Umfragen aus, weshalb trotz ihrer qualitativen Nachteile die Methode der ADM-Stichprobe in Deutschland wesentlich häufiger verwendet wird.

4 Schluss

Literatur

- Blohm, M. (2006). Datenqualität durch stichprobenverfahren bei der allgemeinen bevölkerungsumfrage der sozialwissenschaften–allbus, *Stichprobenqualität in Bevölkerungsumfragen*. Bonn: Informationszentrum Sozialwissenschaften pp. 37–54.
- Gabler, S. (1994). Ost-west-gewichtung der daten der allbus-baseline-studie 1991 und des allbus 1992, *ZUMA-Nachrichten* **35**: 77–81.
- Koch, A. (1997). Adm-design und einwohnermelderegister-stichprobe, *Stichproben in der Umfragepraxis*, Springer, pp. 99–116.
- Wasmer, M., Scholz, E., Blohm, M., Walter, J. and Jutz, R. (2012). *Konzeption und Durchführung der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften”(ALLBUS) 2010.*, GESIS - Leibniz-Institut für Sozialwissenschaften.